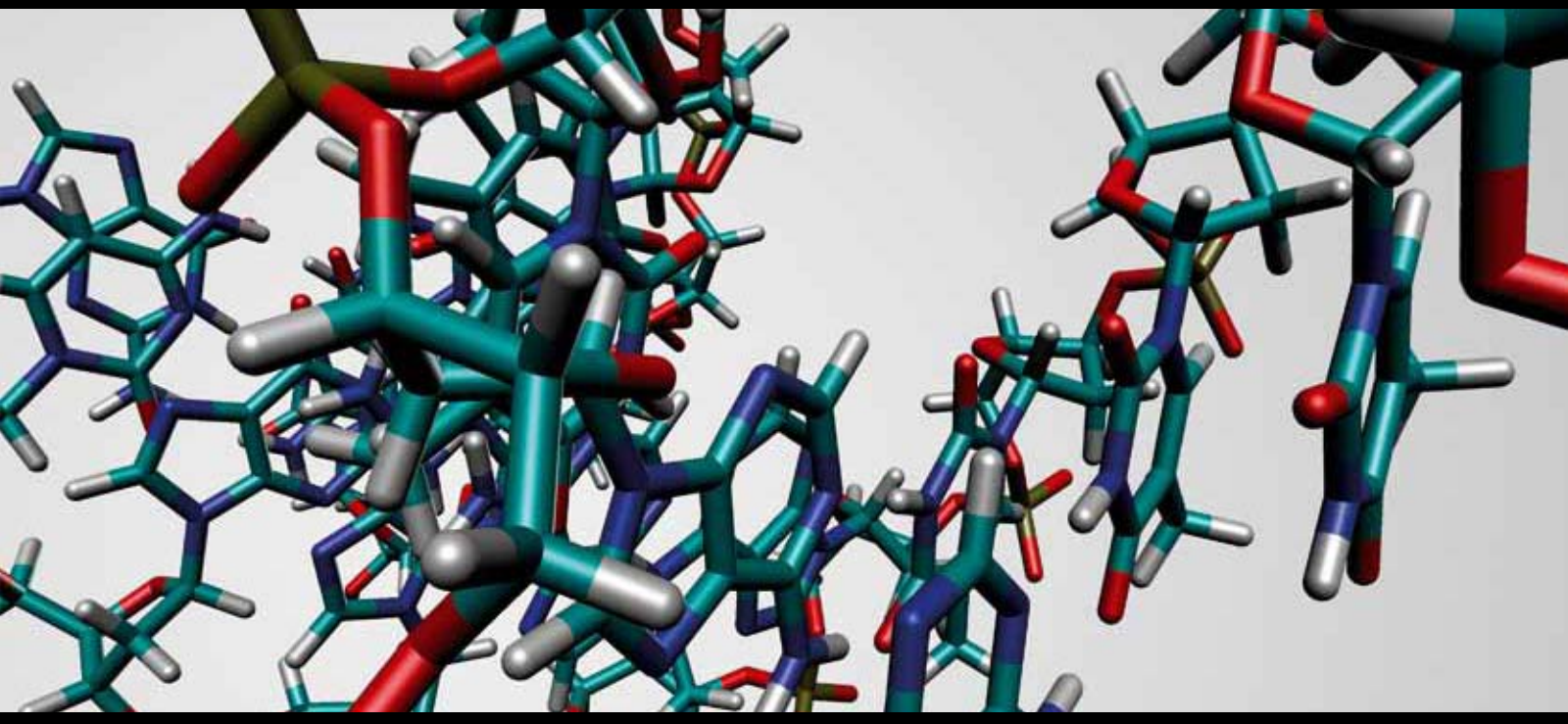
Journal of Nucleic Acids

Artificially Created Nucleic Acids and Peptides/Proteins in Chemical Biology

Guest Editors: Masayasu Kuwahara, Yingfu Li, Eriks Rozners, and Hiroshi Murakami



Artificially Created Nucleic Acids and Peptides/Proteins in Chemical Biology

Journal of Nucleic Acids

Artificially Created Nucleic Acids and Peptides/Proteins in Chemical Biology

Guest Editors: Masayasu Kuwahara, Yingfu Li, Eriks Rozners, and Hiroshi Murakami

Editorial Board

Contents

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2013, Article ID 219263, 2 pages
http://dx.doi.org/10.1155/2013/219263

Editorial

Artificially Created Nucleic Acids and Peptides/Proteins in Chemical Biology

Masayasu Kuwahara,[1] Yingfu Li,[2] Eriks Rozners,[3] and Hiroshi Murakami[4]

[1] *Graduate School of Engineering, Gunma University, Kiryu 376-8515, Japan*
[2] *Departments of Biochemistry and Biomedical Sciences and Chemistry and Chemical Biology, Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, ON, Canada L8S 4K1*
[3] *Department of Chemistry, Binghamton University, State University of New York, Binghamton, NY 13902, USA*
[4] *Graduate School of Arts and Sciences, The University of Tokyo, Komaba, Meguro-ku, Tokyo 153-8902, Japan*

Correspondence should be addressed to Masayasu Kuwahara; mkuwa@gunma-u.ac.jp

Received 27 December 2012; Accepted 27 December 2012

Nucleic acids—DNA and RNA—have been chosen by Mother Nature as the key players for orchestrating the preservation, transfer, and expression of genetic information in all the biological systems on Earth. RNA has also been enlisted to carry out other important cellular functions, such as catalysis and molecular recognition. Within the hands of scientists, the function of nucleic acids has been significantly expanded beyond what is known in nature, and as a result, we are now in the possession of a large array of synthetic, nucleic acid-based catalysts (ribozymes and DNAzymes) and receptors (DNA and RNA aptamers). DNA as a genetic material itself has also been subjected to various chemical modifications in efforts to derive significantly altered or even completely new genetic systems. These systems can be used to create novel peptides and proteins that offer enhanced activities or even completely new properties over their natural protein counterparts. Furthermore, many artificially engineered nucleic acids and proteins have found useful applications as biosensors, diagnostic agents, and therapeutic drugs. This special issue is assembled to reflect recent progress in the important research arena of artificially engineered nucleic acids and proteins.

This issue comprises 10 reviews and 7 research articles that can be grouped into three sections. The first section deals mainly with research on xeno-nucleic acids (XNAs)—nonnatural nucleic acid analogs with significantly altered sugar and/or phosphate backbones. D.-A. Catana et al. provide a review on the use of dinucleotides of dioxaphosphorinane-constrained nucleic acids (CNAs) to tune nucleic acid structures. This is followed by a review by E. Rozners on recent advances in chemical modifications of peptide nucleic acids (PNAs). G. Upert et al. then present a research article on designing cyclic and hairpin PNAs as inhibitors for HIV replication. Z. Wang et al. also present a research article where PNA probes were utilized for live cell imaging of mRNA expression. In their research article, T. Yamamoto et al. examine the gene-silencing effect of bridged/locked nucleic acids (BNA/LNAs). This section is closed out with a research article by S. Saxena et al. in which the molecular crowding effect on the structure and function of RecG (a helicase) was examined.

The second section includes four reviews and three research articles discussing the creation of novel peptides, proteins, transfer RNAs, and peptide mimicries using various selection or screening methods. K. Fukunaga and M. Taki provide a review on the phage display technique with a particular focus on tips for conducting successful phage display experiments. Within the same topic, T. Matsubara reviews the use of phage display for the creation of carbohydrate-mimetic peptides. These are followed by a research article by T. Sumida et al. exploiting the mRNA display technique for the selection of anti-p53 Fab fragments. There are two papers concerning the *in vitro* compartmentalization (IVC) technique, which offers an excellent way to link a genotype to a phenotype in a physically confined environment: the first is a review article by T. Nishikawa et al. on evolving proteins using IVC and the second is a research article by A. Ogawa et al. where IVC

was used to select functional transfer RNAs. T. Kawakami and H. Murakami discuss, in their review article, potential applications of a translation system with a reprogrammed genetic code to prepare a peptide mimetic library. Finally, J. K. Pokorski and D. H. Appella present an on-bead screening approach to create peptide mimicries.

The last section of the issue consists of four review papers on the selection and application of functional nucleic acids. M. McKeague and M. C. DeRosa survey DNA and RNA aptamers derived by SELEX (systematic evolution of ligands via exponential enrichment) for small molecule binding, along with the compilation of nearly 40 improved methodologies of SELEX. K. Tram et al. review the application of fluorescently dressed RNA-cleaving DNAzymes for biosensing. Y. M. Chang et al. discuss the cell-SELEX technique for biomarker discovery. In addition, Y. Kasahara and M. Kuwahara provide a review on SELEX experiments exploring chemically modified nucleic acid libraries.

The ultimate goal in this research arena is to generate scientific knowledge and produce new technologies that will enable the engineering of novel or improved molecular systems to control biological activities at will. Along the way, researchers may help shine light on one of the biggest mysteries surrounding the origin and evolution of life on Earth—the question of why Mother Nature has selected 4 nucleotides to construct nucleic acids and 20 amino acids to construct proteins. We hope this special issue can further spur research efforts in this area.

Acknowledgments

We wish to express our sincere gratitude to all the authors and reviewers for their valuable contributions that have made this special issue successful.

Masayasu Kuwahara
Yingfu Li
Eriks Rozners
Hiroshi Murakami

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2013, Article ID 817350, 7 pages
http://dx.doi.org/10.1155/2013/817350

Review Article

Using Aptamers for Cancer Biomarker Discovery

Yun Min Chang,[1] Michael J. Donovan,[2] and Weihong Tan[1,2]

[1] *Center for Research at Bio/Nano Interface, Department of Chemistry and Department of Physiology and Functional Genomics, Shands Cancer Center, UF Genetics Institute and McKnight Brain Institute, University of Florida, Gainesville, FL, USA*
[2] *Molecular Science and Biomedicine Laboratory, State Key Laboratory of Chemo/Bio-Sensing and Chemometrics, College of Biology, College of Chemistry and Chemical Engineering, Collaborative Innovation Center for Chemistry and Molecular Medicine, Hunan University, Changsha, 410082, China*

Correspondence should be addressed to Weihong Tan; tan@chem.ufl.edu

Received 21 September 2012; Accepted 8 November 2012

Academic Editor: Masayasu Kuwahara

Aptamers are single-stranded synthetic DNA- or RNA-based oligonucleotides that fold into various shapes to bind to a specific target, which includes proteins, metals, and molecules. Aptamers have high affinity and high specificity that are comparable to that of antibodies. They are obtained using iterative method, called (Systematic Evolution of Ligands by Exponential Enrichment) SELEX and cell-based SELEX (cell-SELEX). Aptamers can be paired with recent advances in nanotechnology, microarray, microfluidics, and other technologies for applications in clinical medicine. One particular area that aptamers can shed a light on is biomarker discovery. Biomarkers are important in diagnosis and treatment of cancer. In this paper, we will describe ways in which aptamers can be used to discover biomarkers for cancer diagnosis and therapeutics.

1. Introduction

pproximately 1.5 million Americans were diagnosed with cancer in 2010 [1]. Malignant neoplasm is the second leading cause of death in the world and the leading cause of death in developed nations [2]. Chemotherapy is a common method of treating cancer, but it is largely indiscriminate in that it does not target cancer cells with specificity. Therefore, considerable interest has been shown in developing novel treatments that target only cancer cells, thus avoiding the toxicity of chemotherapy against normal tissues adjacent to the tumor. Such targets can be cancer-specific biomarkers that may be used to assess the changes in expression states of certain proteins or genes within a primary tumor. Since genetic mutations play a key role in modulating the maintenance and progression of cancer cells, fundamental differences in protein levels or gene expression states can be exploited and used for diagnostics and therapies [3]. This paper aims to shed light on the possibility of utilizing aptamers for the discovery of crucial biomarkers for cancers with the goal of improving early-stage diagnosis and therapy.

In recent years, interest has been shown in using aptamers to develop cancer treatments. Currently, AS1411 [4, 5], a potential therapeutic for acute myeloid leukaemia, and NOX-A12 [6, 7], a potential therapeutic for multiple myeloma and non-Hodgkin's lymphoma, aptamers developed by Antisoma and NOXXON, respectively, are in clinical trials [8]. Aptamers are single-stranded oligonucleotides that act like antibodies in recognizing molecular moieties like biomarkers [9]. Because of their ability to fold into secondary or tertiary shapes, aptamers can bind to a wide range of targets, such as metals, proteins, biological cells and tissues, with high specificity [10]. Aptamers generated against various cancer cell lines can be used initially for biomarker discovery and later for diagnostic and therapeutic purposes.

2. Aptamers for Biomarker Discovery

Biomarkers can be expressed in different forms, including, for example, proteins unique to cancer types and subtypes [11]. Proteins are the most useful form of biomarkers since they mirror the genotype and phenotype of a particular disease. Furthermore, since proteins reflect a cell's phenotype, physical alterations of proteins within the cell, as well as overexpression and downregulation of certain proteins, can have profound effects on the cell as a whole. This alteration

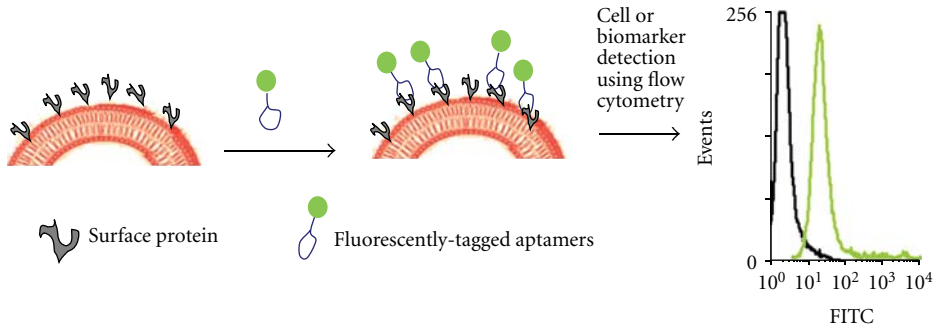


Figure 1: Detection of surface proteins or biomarker in cancer cells using fluorescently tagged aptamers.

in protein composition can be the result of posttranslational modification [12] or mutation at the genetic level, as in the case of cancer [13]. However, it is challenging to develop a biomarker system able to provide accurate evidence that a protein or oncogene is a reflection of a cell's physiological state at some defined stage, essentially because of the robust procedures required to screen for such biomarkers at specific stages of carcinogenesis. These stages of development may not simply lead to overexpression of a single protein but may lead to a change in the ratio of certain proteins. In this case, merely identifying the presence of a protein is not satisfactory. A detection platform that is able to quantify the expression levels of a variety of protein will deliver valuable information. Multiplexed systems that incorporate aptamers for a variety of proteins can be performed, and the aptamers that do bind to protein can be quantified via microarray analysis in order to determine the ratio of specific proteins.

Thus far, the attempts to screen biomarkers have only seen modest results. One traditional method has been the incorporation of antibodies. Yet, antibodies often require a sandwich system in order to detect their target protein [14]. This means that two different types of antibodies must be able to properly identify the target, making large-scale biomarker screening impractical under these conditions. Aptamers, on the other hand, have the ability to overcome such pitfalls, making protein biomarker applications suitable in clinical settings (Figure 1). Aptamers are target specific, able to penetrate cellular membranes, and they can be inexpensively modified and synthesized.

A multitude of applications are available for aptamers, including drug delivery, molecular imaging, and diagnostics. Above all else, however, aptamers for use in biomarker discovery respond to a critical need. Specifically, aptamers can distinguish among thousands of proteins and do so in a short period of time, and they can detect small differences between proteins that are otherwise quite similar in structure, an essential property if proteins are to be differentiated on the cell surface.

(Systematic Evolution of Ligands by Exponential Enrichment) SELEX and cell-based SELEX (cell-SELEX) [15] are two *in vitro* methods used to generate aptamers by iterative positive and negative selection processes that ultimately eliminate non- or weak binding candidate sequences. Aptamers selected in this manner can target overexpressed proteins on the cell surface and even detect small differences among cell-surface proteins. This capability allows aptamers to differentiate unique cellular characteristics, particularly those between cancerous and noncancerous cells based on their apparent biomarker or unique cell-surface homology. Such strategy was used by our group to identify a biomarker for T-cell acute lymphoblastic leukemia (T-ALL), the transmembrane protein tyrosine kinase 7 (PTK7) [16]. By using cell-SELEX, a method akin to SELEX, but one in which live cells are utilized for unbiased selection of aptamers that bind to native forms of the proteins or target, present aptamer sgc8 was selected to bind tightly to PTK7 (K_d = 0.80 ± 0.09 nM) [17] (Figure 2). PTK7 was subsequently identified as a potential biomarker for T-ALL. Furthermore, while sgc8 displayed high selectivity and affinity for its target on most T-ALL and acute myeloid leukemia (AML) cells, as well as some B-cell acute lymphoblastic leukemia (B-ALL) cells, its detection levels in lymphoma and normal human bone marrow cells were not comparable [17] (Table 1). This implies that PTK7 plays a role in the development of most T-ALL and AML cases. Thus, the fundamental differences in expression levels of surface proteins, elucidated with the help of aptamers, play a key role in our ability to differentiate between normal and cancerous cells.

Table 1: Specificity of sgc8 aptamer for T-ALL cancer cells. binding capacity of the aptamer to the cells. 0: <10%; +: 10%–35%; ++: 35%–60%; +++: 60%–85%; ++++: >85%; APL: acute promyelocytic leukemia [17].

Cultured cell lines	sgc8
Molt-4 (T cell ALL)	++++
Sup-T1 (T cell ALL)	++++
Jurkat (T cell ALL)	++++
Sup-B15 (B cell ALL)	+
U266 (B cell myeloma)	0
Toledo (B cell lymphoma)	0
MO2058 (B cell lymphoma)	0
NB-4 (AML, APL)	0

Another example of biomarker discovery using cell-SELEX is the discovery of tenascin-C aptamers using glioblastoma cell line, U251 [18]. This *a priori* approach to

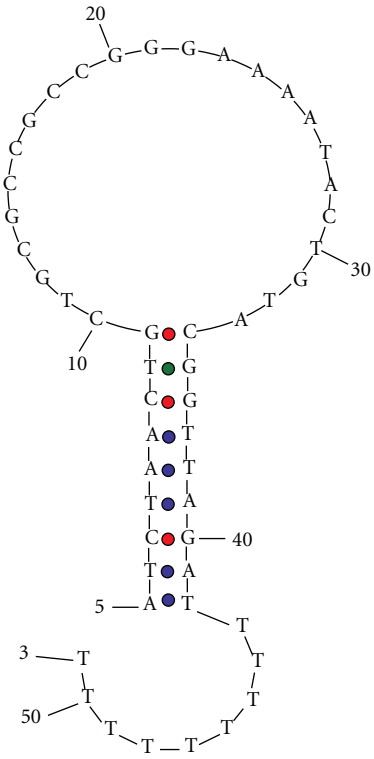


Figure 2: Predicted folded structure of sgc-8 aptamer generated by mfold [17].

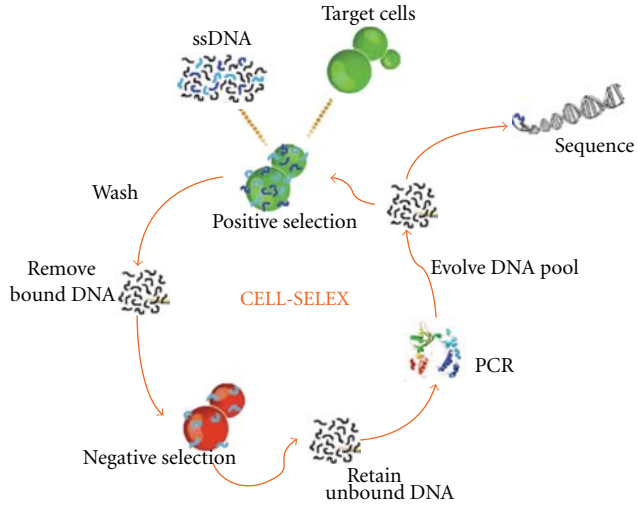


Figure 3: Schematic diagram of cell-SELEX [10].

cell-SELEX helped to reveal biologically pertinent proteins that play a role in cancer that were not previously known. The authors attribute the selection of tenascin-C aptamers using U251 cell lines to the abundance and accessibility of tenascin-C within glioblastoma. This inferred knowledge of overabundance of tenascin in tumor cells was not a coincidence. Tenascin protein was first demonstrated in 1989 as an extracellular matrix protein present in the fetal rat mesenchyme around mammary glands, hair follicles, and teeth, where developmental mesenchyme is present and necessary [19]. Interestingly, adult mammary glands did not contain tenascin, however, when the mammary glands were exposed to carcinogens producing induced tumors, tenascin levels were detectable in fibrous tissues [19]. Upon further inspection, it was shown that tenascin could play a vital role in and serve as a stromal marker for epithelial malignancies. Yoshida et al. report that tenascin is secreted by human lung cancer and glioma cells at a concentration ranging from 1 to 14 μg/mL *in vitro* [20]. Further, patient samples of cerebrospinal fluid with brain tumors showed an increase of tenascin present by >100 ng/mL [20]. Later, immunohistochemical assays revealed that 50 sections of human ovarian tumors showed a greater intensity of staining of tenascin proteins correlating to the high expression of tenascin in cancer cells [20]. Furthermore, malignant and borderline tumors had higher expression levels of tenascin compared to benign tumors [21]. These previous reports of tenascin's role in cancer confirm the aptamer's ability to elucidate how expression levels of certain oncogenic proteins can be used to differentiate between cancerous and noncancerous cells.

Aptamers can not only provide a way of discovering biomarkers that help to differentiate between cancer cells and normal cells, but they can also help to differentiate the stage of carcinogenesis within one subtype of cancer. One study using a DNA-based aptamer with very specific affinity for protein binding revealed 44 different biomarkers able to distinguish among Stage I–III lung cancer [22]. Furthermore, each biomarker correlated with such cofactors as age and frequency of smoking for better treatment outcomes [22]. Under these conditions, early cancer detection is possible, and aptamer-based biomarker discovery with "staging" capability makes it further possible to assess the progression of cancer in terms of the best treatment options.

Unlike SELEX, cell-SELEX uses whole living cells to avoid the disadvantage of selecting aptamers against a nonnative protein conformation (Figure 3). As previously mentioned, this strategy does not require any prior knowledge of the target proteins' conformations or their expression levels within the cell. Since aptamers have the potential of being used in living systems, especially within human biological condition, selecting aptamers using proteins in their native conformation and conditions alleviates the possibility of selecting physiologically nonfunctional aptamers. Furthermore, it bypasses the need for target protein purification, a process that may further disrupt the native conformation of the protein. Cell-SELEX allows all cell-surface molecules or protein to be retained in their native environment, thereby retaining their native folding and physical structures including possible posttranslational modifications throughout the selection process. Thus, aptamers selected with the use of whole live cells will bind to the target with appropriate modifications and conformation on cells. The ability of selecting aptamers in this manner shows great potential in biomedical research and in the development of cell-specific diagnosis and therapeutics in physiological conditions. In addition, using cell-SELEX provides us with the advantage of an unbiased

approach to discovering biomarkers for various diseases, leading to improved cancer diagnostics and therapeutics. We say "unbiased" because prior knowledge of a potential biomarker is not necessary to discover a biomarker using a cell-SELEX approach. "Biased" approach to aptamer selection would mean that a particular biomarker for a certain cancer is already known, but an "unbiased" approach would lend a method of contributing to the discovery of the biomarker itself. This inherent advantage of this approach is the targeting of a disease state, such as cancer, without prior knowledge of the disease's molecular differences. Tumor cell-SELEX provides us with the ability of discovering known and unknown tumor biomarkers [23]. This strategy has the potential of allowing us to discover other uncharacterized biomarkers that illuminate our knowledge on tumor presence, genesis, and progression. It is also likely that discovery of other biologically interesting targets, not previously correlated or connected with cancer, is a possibility. Once such aptamer is selected, it can not only advance diagnosis and treatment of cancer but also help to reveal important biology that defines cancers.

Theoretically, we should be able to use cell-SELEX to profile large numbers of primary tumor cell lines and discover biomarkers for each cancer. Furthermore, such biomarker discoveries can parallel existing cancer genomic data to generate connection between protein biomarkers and genomic biomarkers that differentiate cancerous and non-cancerous cells. This knowledge can provide researchers to develop targeted therapies for cancer that minimize killing of normal cells and physicians a way to treat cancer based on molecular information as opposed to preexisting morphological information of tumors.

3. Imaging Cancer Cells Using Biomarkers and "Membranome"

The term membranome refers to a set of biological membranes that exist in a specific organism. The term was coined by a British biologist, Cavalier-Smith [24], in a way to describe the epigenetics of biological membranes. The term has also been applied to an entire set of proteins [25] or a combination of membrane proteome and lipidome. It is this set of proteins on the cell surface membrane that constitutes its membranome and is a key to understanding biomarker discovery by using aptamers. As previously mentioned, aptamers can be used to detect and differentiate among various cancer cells and their noncancerous subtypes. Such detection is possible because the binding of aptamers occurs mainly on the extracellular domains of cell-surface proteins [26]. This ability of aptamers to bind to such proteins provides a way to identify and discover proteins characteristic of certain cancers. By first characterizing a common protein expressed in a certain cancer, it is possible to image cancer cells by tagging fluorescent dyes to the aptamer's tail. The dyes provide a way to see cancer cells with the naked eye (Figure 1).

Imaging of cancer cells can be enhanced by coupling nanoparticles to aptamers. For example, nanoparticles (NPs) can be bioconjugated with aptamers and used in a colorimetric system where color change occurs upon particle aggregation [10]. In another instance, a multimodal nanoparticle was conjugated with aptamer AS1411 to detect nucleolin, a protein commonly expressed on the membrane of cancer cells, making it possible to image and detect the presence of cancer [9]. By using fluorescence, radioisotope, and MRI imaging techniques, it is possible to detect cancer cells *in vivo* and *in vitro* [9]. Although NPs are toxic *in natura*, by surrounding them with a stable silica shell, or surface modification using aptamers, it might be possible to eliminate the risk of toxicity and conduct clinical studies [9]. This capability also provides a new and improved method to image cancer cells during surgery or develop nanomedicine treatments.

4. Detection of Rare Cancer Cells Using Aptamers

Interestingly, in cancer, circulating tumor cells (CTC) [27, 28] and secreted cancer biomarker, peptide growth factors, cytokines, and hormones [29] can be used as diagnostic markers for cancer diagnosis. Unfortunately, the number of CTCs and other rare secreted protein biomarkers is few. However, we can use aptamers to alleviate detection complications, which can lead to innovations in early cancer detection. This detection method involves testing bodily fluids, such as blood, serum, and sputum, where malignant tumors are present but in very low abundance relative to the large concentration of background cells [10]. To solve this problem, our group has developed aptamer-conjugated nanoparticles (ACNPs) to detect and extract cancer cells from matrices like blood and serum [1], surpassing the capability of antibody-based biomarker discovery. The working principle of ACNPs is conjugating highly selective aptamers to silica-coated magnetic and fluorophore-doped silica nanoparticles to detect and extract target cells from various matrices. Accordingly, ACNPs have shown a theoretical limit of detection (LOD) of 6.6 cells within a complex matrix which is the lowest LOD measured compared to the single-aptamer NPs at 152 cells and original assay LOD of 250 CEM cells [30] (Figure 4). The limit of detection (LOD) was calculated by adding the blank standard deviation multiplied by 3 to the blank signal and determining the LOD from the equation of the line generated by Microsoft Excel. ACNPs are applicable to many types of cancer cells because adaptations, such as chemically modifying aptamers, conjugating different targeted aptamers or changing the type of nanoparticles, can be made for the imaging, detection, and extraction of various cancer cell lines [30]. The versatility and the sensitivity of ACNPs can improve CTC and other rare biomarkers released by cancers detection to change how we diagnose cancer. Such isolation and separation of cancer cells can further be enhanced using aptamer-based fluorescence-activated cell-sorting device as reported by the Famulok group [31]. This method improves cell sorting by reducing false-positive detection of cells that are bound to aptamers due to cellular death or other membrane complications. This approach can be paired with the use of ACNP to accurately detect biomarkers and cancer cells in low concentrations.

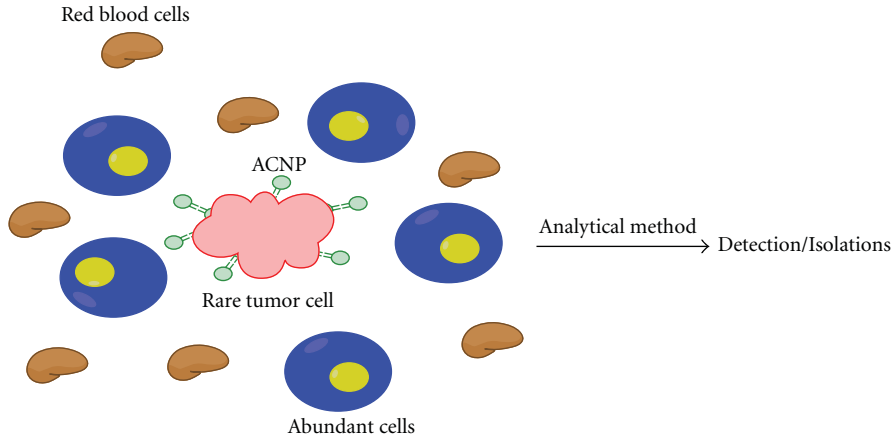


Figure 4: Biomarker concentration within a matrix can be very low. By using aptamer-conjugated nanoparticles, it is possible to detect and discover trace amounts of biomarkers or cells [30].

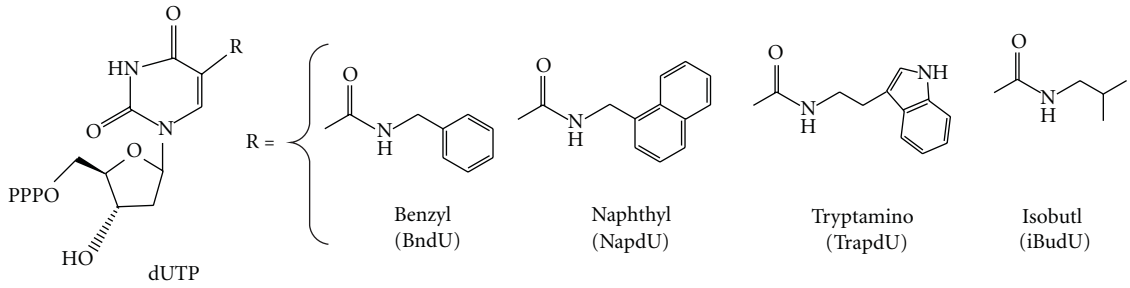


Figure 5: Modifications made for SOMAmer. The nucleotide triphosphate analogs were modified at the 5-UTR position (R) of uridine (dUTP), 5-benzylaminocarbonyl-dU(BndU); 5-naphthylmethylaminocarbonyl-dU (NapdU), 5-tryptaminocarbonyl-dU (TrpdU), and 5-isobutylaminocarbonyl-dU (iBudU) [8].

5. Cell Profiling Using Aptamers and Biomarkers

After cancer cells have been isolated from complex media, aptamer-based cell profiling allows for differentiation between cancer and normal cells based on different protein expression levels. Fluorescent aptamers can be used to profile and study many different types of cancer cells.

The George Church's group at Harvard Medical School [32] has recently developed a logic-gated nanorobot using aptamers for targeted payload delivery. They used aptamer sgc8, which was positively selected using acute lymphoblastic leukemia cells (CCRF-CEM) and negatively selected against a Burkitt's lymphoma cell line (Ramos) [33]. Sgc8 specifically binds PTK7, a receptor present on CCRF-CEM cells, but not Ramos cells. Sgc8-gated nanorobots, which specifically bind and open only in the presence of PTK7 proteins, can recognize the expression level of PTK7 in various cell lines [33]. In another example, Gold et al. have demonstrated the ability of their slow off-rate modified aptamers (SOMAmer) to accurately and specifically recognize cell-surface proteins. Their aptamers modified the uridine 5′-triphosphate (UTP), as illustrated in Figure 5.

Gold et al. further used the modified SOMAmers to develop aptamers for 813 human proteins. Prior to modification, the success rate for protein targeting was <30%. After modifications incorporating the four nucleotides, as shown in Figure 5, the success rate for obtaining an aptamer with a K_d less than ~30 nM was approximately 84% [14]. This is a tremendous increase and one that allows aptamers to be viable components in applications involving biomarkers.

For multiplexed system that Gold et al. incorporated, the reactions took place in a solution, not on a surface. This allowed for the advantageous use of kinetics. In solutions where aptamers with a higher binding affinity are present, they will act as a competitive inhibitor against aptamers with lower binding affinities. Aptamers that did not bind to any proteins were washed away while the ones that did bind to proteins present were kept to be analyzed. A microarray of complementary strands to the aptamers was prepared. This microarray was able to quantify the aptamers present after washing, effectively giving a quantity for the protein levels present. This multiplexed system not only was able to analyze a large array of aptamers but also able to collect data to determine the significance of certain ratios of proteins present.

6. Aptamers for Known Biomarkers

Validating biomarkers is a challenging exercise; accordingly, few *bona fide* proteins have been identified. However, one glycoprotein has been thoroughly studied and generally

agreed to be a biomarker for prostate cancer. This 33-34 kDa glycoprotein is termed prostate specific antigen (PSA). Serum PSA is released into the blood stream after being produced by the prostate epithelium. By using the SELEX method and using a unique genetic algorithm, Ikebukuro et al. were able to derive an aptamer sequenced for PSA. The genetic algorithm was incorporated after traditional SELEX methods. The aptamers obtained had a K_d in the tens of nanomolar range. The group was also able to sense PSA concentrations between 40 and 100 nM when incorporating their aptamers [33]. Such aptamer construct can be used for biosensing.

The finding by Ikebukuro et al. also raises an important point as to the proximity of biomarkers. Antigens, which are present in the blood, may signal a certain disease is present, but do not give evidence to its exact location. For biomarkers that are closely positioned to the cancerous cells, aptamers may be used in conjunction with drugs in order to develop a theranostic platform. Yet, if the biomarker is not in direct proximity to the cancer cell, an additional form of testing will need to be administered to determine where the unhealthy cells are. It is clear that not all biomarkers are the same in terms of their applications. For biomarkers such as PSA, their use is strictly diagnostic. For biomarkers that are on the unhealthy cells surface, they can serve in diagnostics and also be used as a target for therapy.

7. Conclusions

As indicated in this paper, aptamers have shown promise in cancer studies, clinical diagnostics and therapeutic applications. Importantly, aptamers can be used for both clinical purposes, as well as biomarker discovery. However, to further advance the use of aptamers in the laboratory and clinical settings, the number of aptamers for specific targets must be increased. Also, there should be a push to select for aptamers that bind to cancer-specific or other disease-specific proteins that are outlined by cancer genomic studies and resources such as (Catalog of Somatic Mutations in Cancer) COSMIC [34, 35], OncoMap [36], and (The Cancer Genome Atlas) TCGA [37–42]. Using the cell-SELEX technique, previously unknown overexpressed proteins on the surface membrane of cancer cells can be identified as potential markers of carcinogenesis. At the same time, however, cell-SELEX relies on saturation levels of surface proteins, and it can be difficult to discover rare biomarkers that have low abundance on the cell membrane. Fortunately, negative selection for surface proteins known to be overabundant within the cell can circumvent this problem. Thus, aptamers can provide a systematic and accurate way to discover biomarkers for cancer, as well as other diseases, and a theranostic platform for practical clinical applications, particularly early-stage cancer detection.

Acknowledgements

This work is supported by grants awarded by the National Institutes of Health (GM066137, GM079359 and CA133086), by the National Key Scientific Program of China (2011CB911000), NSFC (Grant 21221003) and China National Instrumentation Program 2011YQ03012412.

References

[1] C. D. Medley, S. Bamrungsap, W. Tan, and J. E. Smith, "Aptamer-conjugated nanoparticles for cancer cell detection," *Analytical Chemistry*, vol. 83, no. 3, pp. 727–734, 2011.

[2] J. Temsamani and P. Vidal, "The use of cell-penetrating peptides for drug delivery," *Drug Discovery Today*, vol. 9, no. 23, pp. 1012–1019, 2004.

[3] W. C. Hahn, C. M. Counter, A. S. Lundberg, R. L. Beijersbergen, M. W. Brooks, and R. A. Weinberg, "Creation of human tumour cells with defined genetic elements," *Nature*, vol. 400, no. 6743, pp. 464–468, 1999.

[4] P. J. Bates, D. A. Laber, D. M. Miller, S. D. Thomas, and J. O. Trent, "Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer," *Experimental and Molecular Pathology*, vol. 86, no. 3, pp. 151–164, 2009.

[5] E. Merit Reyes-Reyes, Y. Teng, and P. J. Bates, "A new paradigm for aptamer therapeutic AS1411 action: uptake by macropinocytosis and its stimulation by a nucleolin-dependent mechanism," *Cancer Research*, vol. 70, no. 21, pp. 8617–8629, 2010.

[6] S. G. Sayyed, H. Hägele, O. P. Kulkarni et al., "Podocytes produce homeostatic chemokine stromal cell-derived factor-1/CXCL12, which contributes to glomerulosclerosis, podocyte loss and albuminuria in a mouse model of type 2 diabetes," *Diabetologia*, vol. 52, no. 11, pp. 2445–2454, 2009.

[7] M. N. Darisipudi, O. P. Kulkarni, S. G. Sayyed et al., "Dual blockade of the homeostatic chemokine CXCL12 and the pro-inflammatory chemokine CCL2 has additive protective effects on diabetic kidney disease," *The American Journal of Pathology*, vol. 179, no. 1, pp. 116–124, 2011.

[8] A. D. Keefe, S. Pai, and A. Ellington, "Aptamers as therapeutics," *Nature Reviews Drug Discovery*, vol. 9, no. 7, pp. 537–550, 2010.

[9] D. Hwang, H. Ko, J. Lee et al., "A nucelolin-targeted multimodal nanoparticle imaging probe for tracking cancer cells using an aptamer," *Journal of Nuclear Medicine*, vol. 51, pp. 98–105, 2010.

[10] X. Fang and W. Tan, "Aptamers generated from cell-SELEX for molecular medicine: a chemical biology approach," *Accounts of Chemical Research*, vol. 43, no. 1, pp. 48–57, 2010.

[11] N. L. Henry and D. F. Hayes, "Cancer biomarkers," *Molecular Oncology*, vol. 6, no. 2, pp. 140–146, 2012.

[12] M. Mann and O. N. Jensen, "Proteomic analysis of post-translational modifications," *Nature Biotechnology*, vol. 21, no. 3, pp. 255–261, 2003.

[13] W. Hancock, A. Apffel, J. Chakel et al., "Integrated genomic/proteomic analysis," *Analytical Chemistry*, vol. 71, no. 21, pp. 742A–748A, 1999.

[14] L. Gold, D. Ayers, J. Bertino et al., "Aptamer-based multiplexed proteomic technology for biomarker discovery," *PLoS ONE*, vol. 5, no. 11, artcile e15004, 2010.

[15] K. Sefah, D. Shangguan, X. Xiong, M. B. O'Donoghue, and W. Tan, "Development of DNA aptamers using Cell-SELEX," *Nature Protocols*, vol. 5, no. 6, pp. 1169–1185, 2010.

[16] D. Shangguan, Z. Cao, L. Meng et al., "Cell-specific aptamer probes for membrane protein elucidation in cancer cells," *Journal of Proteome Research*, vol. 7, no. 5, pp. 2133–2139, 2008.

[17] D. Shangguan, Y. Li, Z. Tang et al., "Aptamers evolved from live cells as effective molecular probes for cancer study," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 32, pp. 11838–11843, 2006.

[18] D. A. Daniels, H. Chen, B. J. Hicke, K. M. Swiderek, and L. Gold, "A tenascin-C aptamer identified by tumor cell SELEX:

Systematic evolution of ligands by exponential enrichment," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 26, pp. 15416–15421, 2003.

[19] R. Chiquet-Ehrismann, E. J. Mackie, C. A. Pearson, and T. Sakakura, "Tenascin: an extracellular matrix protein involved in tissue interactions during fetal development and oncogenesis," *Cell*, vol. 47, no. 1, pp. 131–139, 1986.

[20] J. Yoshida, T. Wakabayashi, S. Okamoto et al., "Tenascin in cerebrospinal fluid is a useful biomarker for the diagnosis of brain tumour," *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 57, no. 10, pp. 1212–1215, 1994.

[21] K. E. Wilson, S. P. Langdon, A. M. Lessells, and W. R. Miller, "Expression of the extracellular matrix protein tenascin in malignant and benign ovarian tumours," *British Journal of Cancer*, vol. 74, no. 7, pp. 999–1004, 1996.

[22] R. M. Ostroff, W. L. Bigbee, W. Franklin et al., "Unlocking biomarker discovery: large scale application of aptamer proteomic technology for early detection of lung cancer," *PLoS ONE*, vol. 5, no. 12, article e15003, 2010.

[23] T. Kunii, S. I. Ogura, M. Mie, and E. Kobatake, "Selection of DNA aptamers recognizing small cell lung cancer using living cell-SELEX," *Analyst*, vol. 136, no. 7, pp. 1310–1312, 2011.

[24] T. Cavalier-Smith, "Membranome and membrane heredity in development and evolution," in *Organelles, Genomes and Eukaryote Phylogeny an Evolutionary Synthesis in the Age of Genomics*, R. P. Hirt and D. S. Horner, Eds., 2012.

[25] D. Ghosh, R. C. Beavis, and J. A. Wilkins, "The identification and characterization of membranome components," *Journal of Proteome Research*, vol. 7, no. 4, pp. 1572–1583, 2008.

[26] P. Dua, S. Kim, and D. K. Lee, "Nucleic acid aptamers targeting cell-surface proteins," *Methods*, vol. 54, no. 2, pp. 215–225, 2011.

[27] F. Tanaka, K. Yoneda, N. Kondo et al., "Circulating tumor cell as a diagnostic marker in primary lung cancer," *Clinical Cancer Research*, vol. 15, no. 22, pp. 6980–6986, 2009.

[28] P. Cen, X. Ni, J. Yang, D. Y. Graham, and M. Li, "Circulating tumor cells in the diagnosis and management of pancreatic cancer," *Biochimica et Biophysica Acta*, vol. 1826, no. 2, pp. 350–356, 2012.

[29] J. B. Welsh, L. M. Sapinoso, S. G. Kern et al., "Large-scale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 6, pp. 3410–3415, 2003.

[30] C. D. Medley, S. Bamrungsap, W. Tan, and J. E. Smith, "Aptamer-conjugated nanoparticles for cancer cell detection," *Analytical Chemistry*, vol. 83, no. 3, pp. 727–734, 2011.

[31] G. Mayer, M. S. L. Ahmed, A. Dolf, E. Endl, P. A. Knolle, and M. Famulok, "Fluorescence-activated cell sorting for aptamer SELEX with cell mixtures," *Nature Protocols*, vol. 5, no. 12, pp. 1993–2004, 2010.

[32] S. M. Douglas, I. Bachelet, and G. M. Church, "A logic-gated nanorobot for transport of molecular payload," *Science*, vol. 355, no. 6070, pp. 831–834, 2012.

[33] D. Sangguan, Y. Li, Z. Tang et al., "Aptamers evolved from live cells as effective molecular probes for cancer study," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 32, pp. 11838–11843, 2006.

[34] S. Bamford, E. Dawson, S. Forbes et al., "The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website," *British Journal of Cancer*, vol. 91, no. 2, pp. 355–358, 2004.

[35] http://www.sanger.ac.uk/genetics/CGP/cosmic/.

[36] L. E. MacConaill, C. D. Campbell, S. M. Kehoe et al., "Profiling critical cancer gene mutations in clinical tumor samples," *PLoS ONE*, vol. 4, no. 11, article e7887, 2009.

[37] P. S. Hammerman, D. N. Hayes, M. D. Wilkerson et al., "Comprehensive genomic characterization of squamous cell lung cancers," *Nature*, vol. 489, pp. 519–525, 2012.

[38] T. Cancer and G. Atlas, "Comprehensive molecular characterization of human colon and rectal cancer," *Nature*, vol. 487, pp. 330–337, 2012.

[39] R. McLendon, A. Friedman, D. Bigner et al., "Comprehensive genomic characterization defines human glioblastoma genes and core pathways," *Nature*, vol. 455, no. 7216, pp. 1061–1068, 2008.

[40] D. C. Koboldt, R. S. Fulton, M. D. McLellan et al., "Comprehensive molecular portraits of human breast tumours," *Nature*, pp. 1–10, 2012.

[41] T. C. Larman, S. R. DePalma, A. G. Hadjipanayisa et al., "Spectrum of somatic mitochondrial mutations in five cancers," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 14087–14091, 2012.

[42] http://cancergenome.nih.gov/abouttcga/policies/publication-guidelines.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 958683, 8 pages
doi:10.1155/2012/958683

Review Article

Lighting Up RNA-Cleaving DNAzymes for Biosensing

Kha Tram, Pushpinder Kanda, and Yingfu Li

Departments of Biochemistry and Biomedical Sciences and Chemistry and Chemical Biology, Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, 1280 Main Street West, Hamilton, ON, Canada L8S 4K1

Correspondence should be addressed to Yingfu Li, liying@mcmaster.ca

Received 7 August 2012; Accepted 3 October 2012

Academic Editor: Masayasu Kuwahara

The development of the *in vitro* selection technique has allowed the isolation of functional nucleic acids, including catalytic DNA molecules (DNAzymes), from random-sequence pools. The first-ever catalytic DNA obtained by this technique in 1994 is a DNAzyme that cleaves RNA. Since then, many other RNase-like DNAzymes have been reported from multiple *in vitro* selection studies. The discovery of various RNase DNAzymes has in turn stimulated the exploration of these enzymatic species for innovative applications in many different areas of research, including therapeutics, biosensing, and DNA nanotechnology. One particular research topic that has received considerable attention for the past decade is the development of RNase DNAzymes into fluorescent reporters for biosensing applications. This paper provides a concise survey of the most significant achievements within this research topic.

1. Introduction

A biosensor is an analytical device composed of two key components: a molecular recognition element (MRE) that seeks a target of interest for binding and a signal transducer that works to translate the target-MRE interaction into a detectable signal. Proteins, particularly antibodies, receptors, and enzymes, have been the traditional choice of MREs in the design of biosensors for many decades. However, nucleic acids that possess a defined function, such as ligand binding and/or catalysis, have emerged as very attractive MREs over the past 20 years [1–3]. These "functional nucleic acids (FNAs)" include DNA and RNA aptamers, ribozymes (RNA-based enzymes), DNAzymes (DNA-based enzymes), and aptazymes (ribozyme-aptamer or DNAzyme-aptamer conjugates in which the catalytic activity of the enzyme domain is regulated by the ligand binding to the aptamer domain). FNAs, and particularly DNA aptamers and DNAzymes, are inherently more stable than proteins, resulting in more robust biosensors that can function for an increased period of time. FNAs can be created to recognize a broad range of targets by a simple test-tube evolution technique known as SELEX (Systematic Evolution of Ligands by EXponential Enrichment) or *in vitro* selection [4–6], a process that is short, does not require the use of animals or cells, and has little restriction on the nature of targets and the choice of experimental conditions. FNAs can be chemically synthesized at a relatively low cost with excellent batch-to-batch consistency. They can be facilely immobilized onto a solid matrix. They are easy to modify with sensing probes to allow the detection by many different methods. Binding of targets to an FNA can be coupled to novel amplification techniques to achieve a high level of signal amplification. Several reviews have been published in recent years that have comprehensively discussed the exploration of FNAs for biosensing applications [1–3]. This review is intended to focus on RNA-cleaving DNAzymes, denoted RNase DNAzymes, as a specific class of FNAs and their use in the development of biosensors that utilize fluorescence as the signal output. Special attention will be given to the most recent developments in this area.

2. Deriving RNA-Cleaving DNAzymes by *In Vitro* Selection

Compared to DNA, RNA has significantly reduced chemical stability [7], due to the inherent transesterification reaction between a phosphodiester linkage and the nearby 2′-hydroxyl group (Figure 1). In addition, RNA and DNA are

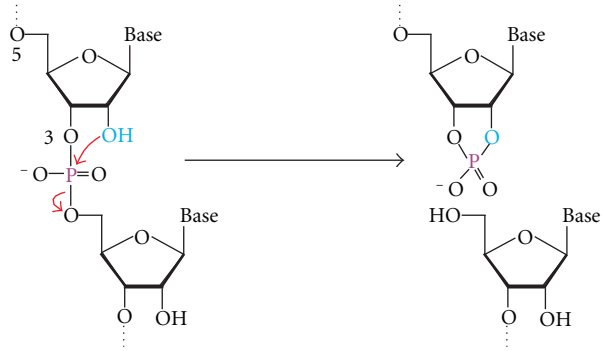


Figure 1: The chemical reaction catalyzed by RNase DNAzymes. These DNAzymes cleave a phosphodiester bond using the 2′-hydroxyl (light blue) as the nucleophile.

two polymers made of almost identical building blocks, making them prone to interacting through Watson-Crick base pairing. From these two viewpoints, it is not very difficult to make a DNA catalyst that binds an RNA substrate and activates a suitable 2′-hydroxyl group to render its attack on a nearby phosphodiester [8–10].

All DNAzymes known to date, including RNase DNAzymes, have been obtained from a large, random-sequence DNA library *via in vitro* selection experiments [4–6]. A typical DNAzyme selection (Figure 2(a)) uses a DNA library containing up to 10^{16} different single-stranded DNA molecules. Such a library is chemically synthesized and contains a central random domain flanked by two short constant regions as primer binding sites for polymerase chain reaction (PCR). The library is subjected to a function-based selection step that separates active sequences from inactive ones. The surviving molecules are amplified by PCR and the amplified DNA is used for the next round of selection. The separation-amplification cycle (selection cycle) is repeated for a number of times until the pool exhibits a desirable catalytic activity. This is followed by cloning and sequencing experiments to obtain individual DNAzymes for characterization and application.

A function-based selection step that can be used to isolate RNase DNAzymes is illustrated in Figure 2(b). The RNA substrate is linked to a DNA pool to isolate *cis*-acting DNAzymes (DNAzymes that cleave an attached RNA sequence). The cleavage makes the active constructs smaller, which travel faster on gel electrophoresis. They are purified, amplified by PCR, and used for the next selection round. This process is repeated until a robust RNA-cleaving activity is observed. Initially, the selected DNAzymes are *cis*-acting, as covalent attachment of the substrate offers an easy way to select for an enzymatic activity; however, *cis*-acting DNAzymes can be made into *trans*-acting DNAzymes (true enzymes) simply by separating the substrate sequence from the DNAzyme.

The first DNAzyme selection, conducted in Gerald Joyce's laboratory in 1994, led to the isolation of an RNase DNAzyme that cleaves a single RNA linkage embedded in a DNA sequence in the presence of Pb^{2+} [11]. The DNAzyme exhibits a cleavage rate of $1\,min^{-1}$, translating into a rate enhancement of 10^5-fold over the spontaneous cleavage of RNA. Since then, a plethora of *in vitro* selection experiments have been performed to derive diverse RNase DNAzymes with widely different characteristics. The list includes DNAzymes utilizing Mg^{2+} [12], Ca^{2+} [13], Zn^{2+} [14], Mn^{2+} [15], and UO_2^{2+} [16] as metal-ion cofactors, a cofactorless DNAzyme [17] and DNAzymes that use the amino acid histidine as the cofactor [18].

The Joyce Laboratory also made an effort of selecting DNAzymes that can cleave an all-RNA substrate under physiological conditions, which led to the discovery of two DNAzymes known as 10–23 and 8–17 (Figure 3) [19]. 10–23 has the ability to cleave any purine-pyrimidine junction within an RNA chain as long as its two-binding arms are properly engineered so that they can form stable duplexes with nucleotides flanking the cleavage junction. 10–23 has been used to target many messenger and viral RNAs for therapeutic applications [8, 9]. The other DNAzyme isolated from this selection, 8–17, has also been isolated in several other *in vitro* selection experiments [13, 14, 20, 21]. Sequence variants of 8–17 are capable of using different metal-ion cofactors, such as Mg^{2+} [19], Ca^{2+} [13], Mn^{2+} [20], and Zn^{2+} [16] as well as cleaving fourteen of the possible sixteen dinucleotide junctions of RNA [20]. 8–17 has been widely studied as biosensors for the detection of metal ions and other ligands [10]; some examples will be discussed below.

3. Dressing Existing RNase DNAzymes with a Fluorescence-Signaling Module

Fluorescence is a powerful technique that has found widespread applications in bioanalysis [22]. Its popularity comes from high detection sensitivity, minimal invasiveness, capability for real-time detection, and the availability of many different fluorophores and related quenchers. The isolation of many RNase DNAzymes has created an opportunity for developing fluorescent biosensors based on such enzymes.

The most popular way of making a fluorescent sensor out of an RNase DNAzyme is to report the DNAzyme action through the use of a fluorophore/quencher (F/Q) pair. This can be conveniently achieved by a judicious arrangement of F and Q so that they are in close proximity prior to catalysis but separate from each other after catalysis.

Three types of F/Q arrangements have been commonly adopted: F and Q can be placed on the bases flanking the cleavage site (Figure 4(a)) [23], at the end of a duplex formed between a DNAzyme and its substrate (Figure 4(b)) [24], or at the end of a hairpin-shaped substrate (Figure 4(c)) [25].

4. Isolating New RNase DNAzymes That Cleave a Fluorogenic Substrate

In addition to converting existing RNase DNAzymes into fluorescent reporters, novel RNase DNAzymes can also be directly isolated from random-sequence DNA libraries to cleave a substrate predeposited with a F/Q pair. To

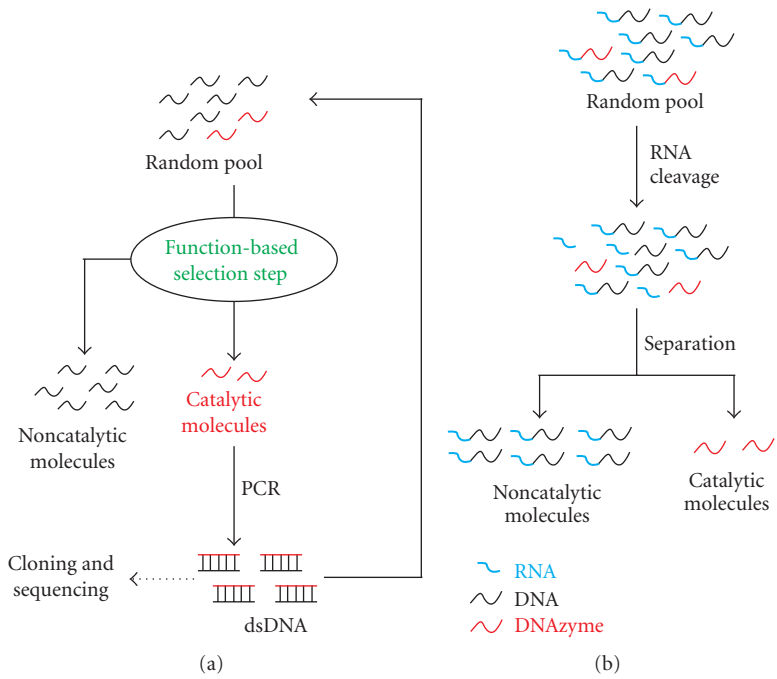


Figure 2: Deriving RNase DNAzymes by *in vitro* selection. (a) Schematic of isolating DNAzymes from a DNA pool. (b) Gel electrophoresis-based method for selecting RNase DNAzymes. For clarity, only the function-based selection step is shown.

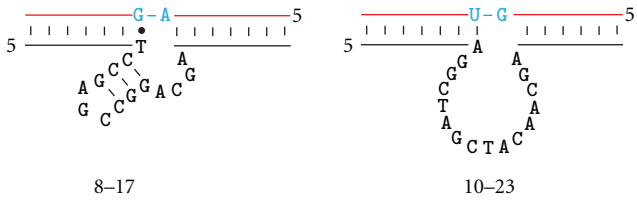


Figure 3: Secondary structures of two best studied RNase DNAzymes: 8–17 and 10–23. The RNA substrate strands are shown in red and the reaction centers are shown in blue. Both 10–23 and 8–17 bind their RNA substrates by two-binding arms made of short Watson-Crick duplexes, which are simplified as line drawings.

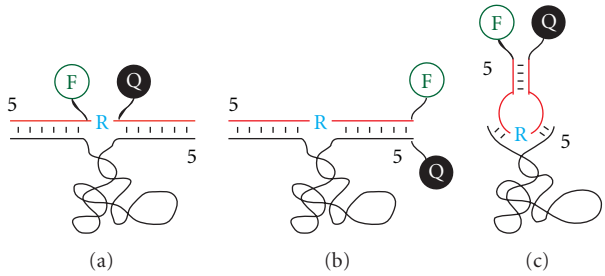


Figure 4: Three common arrangements of fluorophore (F) and quencher (Q) in the design of fluorescent RNase DNAzyme-based biosensors. F and Q are appended to the nucleobases that flank the cleavage site (a), placed at the end of duplex formed between the substrate and DNAzyme strands (b), or located at the end of the hairpin-shaped substrate strand (c). DNAzyme and substrate strands are shown in black and red, respectively. The cleavage site is marked by a blue "R".

achieve high levels of fluorescence signal by the dequenching mechanism, F and Q need to be placed as close to the cleavage site as possible. However, doing so can significantly affect the catalytic performance of an existing RNase DNAzyme because the bulky fluorophores and quenchers may prevent the DNAzyme/substrate from achieving optimal structural folding. Isolating RNase DNAzymes directly from a random-sequence library to cleave a premodified substrate would overcome this issue [23]. Our laboratory has carried out several *in vitro* selection experiments to isolate such DNAzymes [26–30]. These DNAzymes cleave a chimeric DNA/RNA substrate containing a single ribonucleotide as the cleavage site that is located between a nucleotide modified with a fluorophore and a nucleotide modified with a quencher (Figure 5(a)).

Several fluorogenic RNase DNAzymes have been obtained and their catalytic and signaling properties are characterized [31–34]. Most of these DNAzymes are robust catalysts, exhibiting a catalytic rate at or above 1 min^{-1}. They can achieve 10- to 30-fold signal enhancements. Some of these DNAzymes have very interesting secondary structures, such as 3-way junction [28, 30], 4-way junction [32], and 5-way junction [29]. The secondary structures of two representative signaling DNAzymes are provided in Figure 5(b).

5. Using Fluorescent RNase DNAzymes as Metal Ion Sensors

Certain metal ions, such as Pb^{2+} and Hg^{2+}, impose a great threat to human health, and biosensors that have the ability

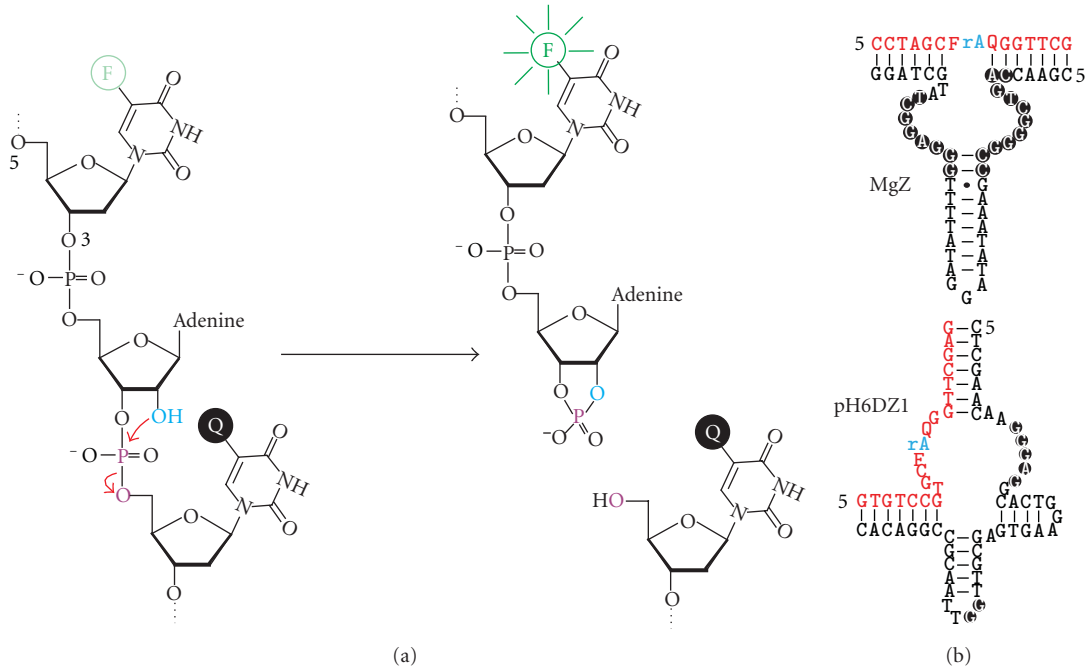


Figure 5: Fluorogenic DNAzymes obtained from random-sequence DNA libraries. (a) The chemical transformation catalyzed by these DNAzymes. These DNAzymes cleave a phosphodiester bond using a nearby 2′-hydroxyl (light blue) as the nucleophile. The special nucleic acid substrate to be cleaved has two features: (1) it contains a single ribonucleotide (riboA) as the cleavage site embedded in a DNA sequence, and (2) the cleavage site is immediately sandwiched between two DNA nucleotides modified with a fluorophore (F; specifically fluorescein) and a quencher (Q; specifically DABCYL). (b) Secondary structures of two representative signaling DNAzymes. rA: adenine ribonucleotide (the cleavage site; shown in blue); F, fluorescein-modified dT; Q, DABCYL-modified dT. Conserved nucleotides are shown in black circles. The substrate strands are shown in red.

to sensitively and selectively detect such metal ions are highly coveted. DNAzymes typically need divalent metal ions as cofactors to perform efficient catalysis and thus, the catalytic activity of a properly engineered metallo DNAzyme can be exploited to detect metal ions. Significant work has been carried out, mostly in the laboratory of Yi Lu, to develop RNase DNAzymes into sensors for toxic metal ions. Here we will discuss a few examples where the detection was done using fluorescence as the signal output.

The first-ever fluorescent DNAzyme-based metal-ion sensor, reported by the Lu group, is for the detection of Pb^{2+} [24]. The DNAzyme employed is a variant of 8–17 that has the robust RNA cleavage activity in the presence of Pb^{2+}. The DNAzyme was converted into a fluorescent sensor using the labeling scheme shown in Figure 4(b). This sensor has a detection limit of 10 nM for Pb^{2+} and exhibits an 80-fold selectivity for Pb^{2+} over other metal ions. The same group recently reported another fluorescent Pb^{2+} sensor using the RNase DNAzyme originally selected by the Joyce group in the presence of Pb^{2+} [11]. This sensor was found to be ~40,000 times more specific than the 8–17 based Pb^{2+} sensor [35].

A fluorescent sensor for $UO_2{}^{2+}$ was also reported by the Lu group, using an RNase DNAzyme selected for $UO_2{}^{2+}$ [16]. This sensor is able to detect $UO_2{}^{2+}$ down to 45 pM and exhibits a selectivity of more than one million-fold over other metal ions, making it the best $UO_2{}^{2+}$ sensor known to date.

The same group has also engineered a DNAzyme sensor for Hg^{2+} using an altered version of the uranium-responsive DNAzyme [36]. They replaced a short duplex motif of the DNAzyme with a few T-T mismatches, a DNA element known to bind Hg^{2+}. This simple manipulation leads to the deactivation of the DNAzyme. However, when Hg^{2+} is present, the strong binding of Hg^{2+} to the DNA motif reactivates the DNAzyme. This Hg^{2+} sensor exhibits a detection limit of 4 nM, which is below the threshold for toxicity set by the US Environmental Protection Agency.

6. Constructing Fluorescent RNase Aptazymes by Rational Design

To detect molecules other than metal ions, DNAzymes need to carry an aptamer domain as the recognition element. The unified aptamer-DNAzyme systems are often referred to as "allosteric DNAzymes" or simply "aptazymes". Rational design can be used to engineer RNase aptazymes. This can be done by judiciously integrating an aptamer with an RNase DNAzyme to create an allosteric DNAzyme whose catalytic activity is regulated by the binding of the target analyte to the aptamer domain. A great amount of research work has been carried out to devise various strategies to engineer RNA or DNA aptazymes for biosensing purposes [37–40]. Here we will discuss two strategies published in our own group on linking an ATP-binding DNA aptamer to fluorescently dressed RNase DNAzymes.

The first strategy, illustrated in Figure 6(a), is modeled after the structure-switching aptamer design principle [41]. The strategy uses three separate oligonucleotides: an ATP-binding DNA aptamer linked to an RNase DNAzyme, a regulatory oligonucleotide that binds part of the DNAzyme sequence and part of the aptamer domain, and a fluorogenic substrate [42]. Without ATP, the regulatory oligonucleotide prevents the substrate from binding to the DNAzyme; the binding of ATP to the aptamer results in the release of the regulatory oligonucleotide. The freed DNAzyme can now bind and cleave its fluorogenic RNA substrate, leading to the increase of fluorescence. Two DNAzymes were used for this demonstration: 8–17 and pH7DZ1.

The second strategy is based on an intramolecular structure-switching mechanism using the same ATP-binding DNA aptamer and pH6DZ1 as the DNAzyme (Figure 6(b)) [43]. The aptamer sequence is inserted into a hairpin motif of pH6DZ1; the sequence of the aptamer is slightly altered so that part of the aptamer can form an intramolecular stem with several catalytically essential nucleotides of pH6DZ1 in the absence of ATP. Upon addition of ATP, the aptamer domain switches from the duplex structure to the complex structure with ATP, which frees up the sequestered nucleotides to create the active structure of the DNAzyme to cleave its fluorogenic substrate.

7. Isolating Fluorescent RNase Aptazymes *via In Vitro* Selection

Although rational design can be used to engineer RNase aptazymes from existing RNase DNAzymes and aptamers, *in vitro* selection can be explored to isolate novel RNase aptazymes from random-sequence pools. Such an approach relies on two selection steps to tune the dependence of the catalytic activity of the DNAzyme on the target of interest: a negative selection step to remove sequences that are catalytic in the absence of target or in the presence of undesired targets; a positive selection step with the target of interest to enrich sequences that are target-specific.

Our group has used this approach to derive fluorescent RNase DNAzymes that are able to detect bacteria [44, 45]. The goal of the work was to determine whether it was feasible to develop a fluorescent RNase DNAzyme sensor that can detect *Escherichia coli* (*E. coli*), representing the bacteria of interest but does not recognize other bacteria represented by *Bacillus subtilis* (*B. subtilis*). A library of DNA molecules containing a fluorogenic chimeric DNA/RNA substrate was first subjected to a negative selection step with the crude extracellular mixture (CEM) from *B. subtilis*. The cleaved sequences were discarded, and the uncleaved ones were recovered and subjected to a positive selection step with the CEM from *E. coli*. The cleaved sequences from this positive selection step were amplified by PCR and used as the pool for the next round of negative/positive selection. A DNAzyme, named RFD-EC1, was obtained after 20 selection cycles. RFD-EC1 was found to be highly specific for *E. coli* and did not show any activity in the presence of a host of other Gram-negative and Gram-positive bacteria. With the addition of a culturing step, RFD-EC1 is able to detect a single seeding *E. coli* cell. Using the CEM from *E. coli* (which represents a complex mixture of small molecules and proteins) as the target of interest (rather than using a defined biomarker for *E. coli*) allows the selection process to choose a target from the mixture that not only bind strongly to the aptazyme but is also absent from the control bacteria. The key advantage of this method is that it avoids laborious steps to identify and purify a suitable target that is unique only to the target bacteria.

8. Using RNase DNAzymes to Achieve Signal Amplification

RNase DNAzymes can be neatly incorporated into signal amplification strategies to achieve highly sensitive detection. Here we will discuss three examples.

The first study used the aforementioned Pb^{2+} dependent RNase DNAzyme and an RNA-containing molecular beacon (MB) (as the substrate) to achieve signal amplification. The MB-type substrate provides highly efficient fluorescence quenching; the multiple turnover ability of the DNAzyme—cleaving one MB molecule after another—means that a large signal can be generated from a low concentration of the RNase DNAzyme. The outcome is significantly increased detection sensitivity [46].

The second study cleverly integrated T4 DNA ligase, an RNA-cleaving DNAzyme and again an RNA-containing molecular beacon (MB) to achieve the detection of ATP. Since ATP is a required cofactor for T4 DNA ligase, the presence of ATP in a test sample leads to the activation of the ligase's DNA ligation ability. The activated ligase joins two pieces of DNA to make a functional RNase DNAzyme. The assembled DNAzyme then cleaves the RNA linkage placed within the loop region of the MB to generate a fluorescence signal (Figure 7(a)) [25].

The third study employed a histidine-dependent RNA-cleaving DNAzyme, an endonuclease, and an MB to create an amplified sensor for the detection of histidine. When histidine is present, the DNAzyme is activated, leading to the cleavage of an RNA-containing substrate that is designed to bind strongly with the DNAzyme. The cleavage event produces two DNA fragments that can no longer hold onto the DNAzyme strongly. One released fragment goes on to hybridize with the MB creating the double-stranded recognition site for a specific endonuclease, which carries out the cleavage of the MB for fluorescence signal generation (Figure 7(b)) [47].

The last two biosensing systems discussed above are encoded with two layers of signal amplification capabilities due to the multiple-turnover nature of both the DNAzyme and the protein enzyme.

9. Conclusions

DNA is best known as the hereditary material for the storage of genetic information in living organisms. However, a great amount of work has been done in the past 20

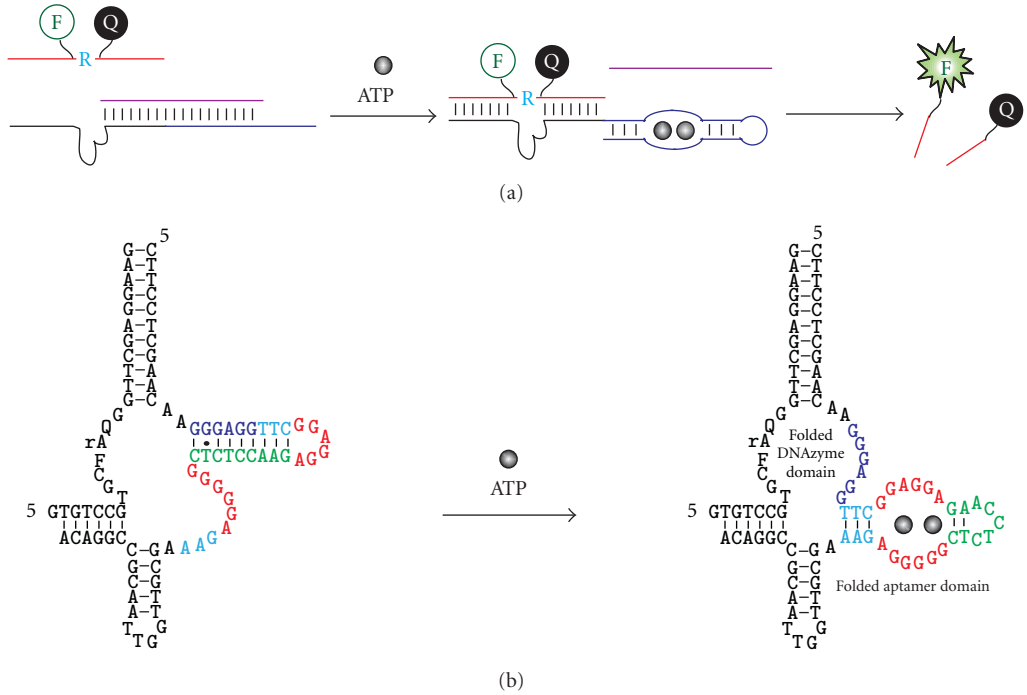


Figure 6: Engineering fluorescent RNase aptazymes by rational design. (a) The use of a regulatory oligonucleotide. The DNAzyme (black line) is joined to an ATP-binding DNA aptamer (blue line) and the combined sequence can form a duplex structure with a regulatory oligonucleotide (purple strand), which prevents the DNAzyme domain from binding the fluorogenic substrate. Binding of ATP to the aptamer domain causes the release of the regulatory oligonucleotide, freeing up the DNAzyme for binding and cleaving the substrate. (b) An ATP-responsive DNAzyme with internal regulatory nucleotides. The DNAzyme is designed from pH6DZ1 and the same ATP-binding aptamer. Nucleotides in the sensor sequence are color-coded for easy tracking of their relative positions in the two alternative structures. In the absence of ATP, the aptamer is designed to form an inhibitory duplex with the catalytic core; in the presence of ATP, the aptamer pulls away from the DNAzyme, which activates the DNAzyme leading to the generation of a fluorescent signal.

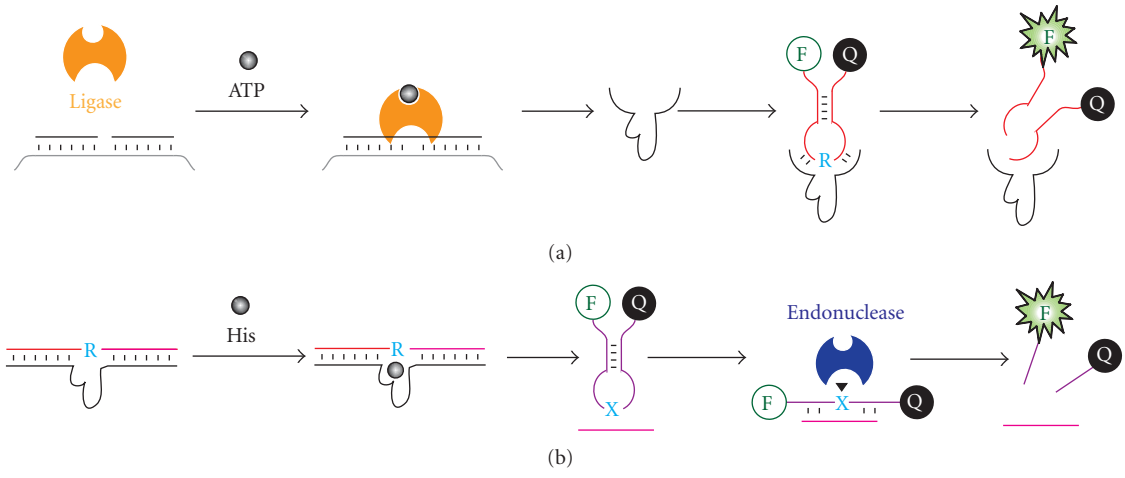


Figure 7: RNase DNAzyme for signal amplification. (a) Ligase/DNAzyme/MB sensor for ATP detection. ATP activates T4 DNA ligase to assemble a whole DNAzyme from two shorter DNA molecules; the DNAzyme cleaves a ribonucleotide (blue "R") containing molecular beacon (MB) to generate a fluorescence signal. (b) DNAzyme/MB/endonuclease sensors for histidine. Histidine (His) activates the DNAzyme; the cleavage of the substrate by the DNAzyme produces a small DNA fragment (shown as the pink line) that bind an MB to assemble the recognition site (blue "X") for a nicking endonuclease, which cleaves the MB to generate a fluorescence signal.

years that have convincingly shown that DNA is also a versatile polymer from which receptors (DNA aptamers) and catalysts (DNAzymes) can be derived. Since the discovery of the first-ever DNAzyme by the Joyce group in 1994 [11], tremendous progress has been made in DNAzyme research in several aspects, including isolation of many DNAzymes that can collectively catalyze more than a dozen different chemical reactions, demonstration that DNAzymes can achieve large rate enhancements, and application of DNAzymes as chemical and biological tools [9]. One of the highly studied classes of DNAzymes is RNase DNAzyme. This review has sampled some key studies where RNase DNAzymes have been examined for fluorescence-based biosensing applications.

Several conclusions can be drawn. First, RNase DNAzymes can be made to detect a very broad range of targets, from simple chemical species, such as toxic metal ions and small biological cofactors, to complex biological samples, such as the CEM of *E. coli*. We expect that the list of analytes will expand when this field continues to develop.

Second, both rational design and *in vitro* selection approaches can be applied to design fluorogenic RNase DNAzymes to detect a specific analyte. Rational design is usually adopted if there is a known aptamer for the analyte. Since aptamers that bind a variety of targets do exist, rational design can still be a fruitful way of turning these aptamers into fluorescent RNase DNAzyme sensors. However, for analytes without a preisolated aptamer, the selection approach becomes an excellent choice. This point has been effectively demonstrated by the Lu group through the creation of a novel uranium sensor [16], as well as by our own group through the development of a novel *E. coli* sensor [44, 45]. We expect that similar approaches will be explored in the future for creating sensors for many other targets.

Third, fluorogenic RNase DNAzyme systems offer a unique way of achieving signal amplification, as illustrated by the two examples discussed earlier where amplified ATP and histidine sensors were engineered. Both examples elegantly use DNA manipulating enzymes (DNA ligase and nicking endonuclease) to produce an RNase DNAzyme (in the case of DNA ligase) or act on the product of an RNase DNAzyme to amplify the signal. Such couplings are made possible simply because DNAzymes are DNA molecules at the first place and can be easily teamed up with DNA manipulating enzymes to do unique things.

The real-time detection capability and excellent chemical stability, combined with the power of *in vitro* selection, will continue to make fluorogenic RNase DNAzyme-based sensors an excellent option for many biosensing applications to come.

References

[1] N. K. Navani and Y. Li, "Nucleic acid aptamers and enzymes as sensors," *Current Opinion in Chemical Biology*, vol. 10, no. 3, pp. 272–281, 2006.

[2] W. Mok and Y. Li, "Recent progress in nucleic acid aptamer-based biosensors and bioassays," *Sensors*, vol. 8, no. 11, pp. 7050–7084, 2008.

[3] J. Liu, Z. Cao, and Y. Lu, "Functional nucleic acid sensors," *Chemical Reviews*, vol. 109, no. 5, pp. 1948–1998, 2009.

[4] D. L. Robertson and G. F. Joyce, "Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA," *Nature*, vol. 344, no. 6265, pp. 467–468, 1990.

[5] C. Tuerk and L. Gold, "Systemic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase," *Science*, vol. 249, no. 4968, pp. 505–510, 1990.

[6] A. D. Ellington and J. W. Szostak, "*In vitro* selection of RNA molecules that bind specific ligands," *Nature*, vol. 346, no. 6287, pp. 818–822, 1990.

[7] Y. Li and R. R. Breaker, "Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2′-hydroxyl group," *Journal of the American Chemical Society*, vol. 121, no. 23, pp. 5364–5372, 1999.

[8] S. K. Silverman, "*In vitro* selection, characterization, and application of deoxyribozymes that cleave RNA," *Nucleic Acids Research*, vol. 33, no. 19, pp. 6151–6163, 2005.

[9] K. Schlosser and Y. Li, "Biologically inspired synthetic enzymes made from DNA," *Chemistry and Biology*, vol. 16, no. 3, pp. 311–322, 2009.

[10] K. Schlosser and Y. Li, "A versatile endoribonuclease mimic made of DNA: characteristics and applications of the 8-17 RNA-cleaving DNAzyme," *ChemBioChem*, vol. 11, no. 7, pp. 866–879, 2010.

[11] R. R. Breaker and G. F. Joyce, "A DNA enzyme that cleaves RNA," *Chemistry and Biology*, vol. 1, no. 4, pp. 223–229, 1994.

[12] R. R. Breaker and G. F. Joyce, "A DNA enzyme with Mg^{2+}-dependent RNA phosphoesterase activity," *Chemistry and Biology*, vol. 2, no. 10, pp. 655–660, 1995.

[13] D. Faulhammer and M. Famulok, "The Ca^{2+} ion as a cofactor for a novel RNA-cleaving deoxyribozyme," *Angewandte Chemie—International Edition*, vol. 35, no. 23-24, pp. 2837–2841, 1997.

[14] J. Li, W. Zheng, A. H. Kwon, and Y. Lu, "*In vitro* selection and characterization of a highly efficient Zn(II)-dependent RNA-cleaving deoxyribozyme," *Nucleic Acids Research*, vol. 28, no. 2, pp. 481–488, 2000.

[15] J. C. F. Lam, J. B. Withers, and Y. Li, "A complex RNA-cleaving DNAzyme that can efficiently cleave a pyrimidine-pyrimidine junction," *Journal of Molecular Biology*, vol. 400, no. 4, pp. 689–701, 2010.

[16] J. Liu, A. K. Brown, X. Meng et al., "A catalytic beacon sensor for uranium with parts-per-trillion sensitivity and millionfold selectivity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 7, pp. 2056–2061, 2007.

[17] C. R. Geyer and D. Sen, "Evidence for the metal-cofactor independence of an RNA phosphodiester-cleaving DNA enzyme," *Chemistry and Biology*, vol. 4, no. 8, pp. 579–593, 1997.

[18] A. Roth and R. R. Breaker, "An amino acid as a cofactor for a catalytic polynucleotide," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 11, pp. 6027–6031, 1998.

[19] S. W. Santoro and G. F. Joyce, "A general purpose RNA-cleaving DNA enzyme," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 9, pp. 4262–4266, 1997.

[20] R. P. G. Cruz, J. B. Withers, and Y. Li, "Dinucleotide junction cleavage versatility of 8–17 deoxyribozyme," *Chemistry and Biology*, vol. 11, no. 1, pp. 57–67, 2004.

[21] K. Schlosser and Y. Li, "Tracing sequence diversity change of RNA-cleaving deoxyribozymes under increasing selection

pressure during *in vitro* selection," *Biochemistry*, vol. 43, no. 30, pp. 9695–9707, 2004.

[22] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Kluwer Academic/Plenum Publishers, New York, NY, USA, 1999.

[23] W. Chiuman and Y. Li, "Efficient signaling platforms built from a small catalytic DNA and doubly labeled fluorogenic substrates," *Nucleic Acids Research*, vol. 35, no. 2, pp. 401–405, 2007.

[24] J. Li and Y. Lu, "A highly sensitive and selective catalytic DNA biosensor for lead ions," *Journal of the American Chemical Society*, vol. 122, no. 42, pp. 10466–10467, 2000.

[25] L. M. Lu, X. B. Zhang, R. M. Kong, B. Yang, and W. Tan, "A ligation-triggered DNAzyme cascade for amplified fluorescence detection of biological small molecules with zero-background signal," *Journal of the American Chemical Society*, vol. 133, no. 30, pp. 11686–11691, 2011.

[26] S. H. J. Mei, Z. Liu, J. D. Brennan, and Y. Li, "An efficient RNA-cleaving DNA enzyme that synchronizes catalysis with fluorescence signaling," *Journal of the American Chemical Society*, vol. 125, no. 2, pp. 412–420, 2003.

[27] Z. Liu, S. H. J. Mei, J. D. Brennan, and Y. Li, "Assemblage of signaling DNA enzymes with intriguing metal-ion specificities and pH dependences," *Journal of the American Chemical Society*, vol. 125, no. 25, pp. 7539–7545, 2003.

[28] W. Chiuman and Y. Li, "Revitalization of six abandoned catalytic DNA species reveals a common three-way junction framework and diverse catalytic cores," *Journal of Molecular Biology*, vol. 357, no. 3, pp. 748–754, 2006.

[29] W. Chiuman and Y. Li, "Evolution of high-branching deoxyribozymes from a catalytic DNA with a three-way junction," *Chemistry and Biology*, vol. 13, no. 10, pp. 1061–1069, 2006.

[30] W. Chiuman and Y. Li, "Simple fluorescent sensors engineered with catalytic DNA "MgZ" based on a non-classic allosteric design," *PLoS ONE*, vol. 2, no. 11, Article ID e1224, 2007.

[31] S. A. Kandadai and Y. Li, "Characterization of a catalytically efficient acidic RNA-cleaving deoxyribozyme," *Nucleic Acids Research*, vol. 33, no. 22, pp. 7164–7175, 2005.

[32] Y. Shen, J. D. Brennan, and Y. Li, "Characterizing the secondary structure and identifying functionally essential nucleotides of pH6DZ1, a fluorescence-signaling and RNA-cleaving deoxyribozyme," *Biochemistry*, vol. 44, no. 36, pp. 12066–12076, 2005.

[33] M. M. Ali, S. A. Kandadai, and Y. Li, "Characterization of pH3DZ1—an RNA-cleaving deoxyribozyme with optimal activity at pH 3," *Canadian Journal of Chemistry*, vol. 85, no. 4, pp. 261–273, 2007.

[34] S. A. Kandadai, W. W. K. Mok, M. M. Ali, and Y. Li, "Characterization of an RNA-cleaving deoxyribozyme with optimal activity at pH 5," *Biochemistry*, vol. 48, no. 31, pp. 7383–7391, 2009.

[35] T. Lan, K. Furuya, and Y. Lu, "A highly selective lead sensor based on a classic lead DNAzyme," *Chemical Communications*, vol. 46, no. 22, pp. 3896–3898, 2010.

[36] J. Liu and Y. Lu, "Rational design of "turn-on" allosteric DNAzyme catalytic beacons for aqueous mercury ions with ultrahigh sensitivity and selectivity," *Angewandte Chemie—International Edition*, vol. 46, no. 40, pp. 7587–7590, 2007.

[37] J. Tang and R. R. Breaker, "Rational design of allosteric ribozymes," *Chemistry and Biology*, vol. 4, no. 6, pp. 453–459, 1997.

[38] G. A. Soukup and R. R. Breaker, "Engineering precision RNA molecular switches," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 7, pp. 3584–3589, 1999.

[39] D. Y. Wang, B. H. Y. Lai, and D. Sen, "A general strategy for effector-mediated control of RNA-cleaving ribozymes and DNA enzymes," *Journal of Molecular Biology*, vol. 318, no. 1, pp. 33–43, 2002.

[40] M. Levy and A. D. Ellington, "ATP-dependent allosteric DNA enzymes," *Chemistry and Biology*, vol. 9, no. 4, pp. 417–426, 2002.

[41] R. Nutiu and Y. Li, "Structure-switching signaling aptamers," *Journal of the American Chemical Society*, vol. 125, no. 16, pp. 4771–4778, 2003.

[42] J. C. Achenbach, R. Nutiu, and Y. Li, "Structure-switching allosteric deoxyribozymes," *Analytica Chimica Acta*, vol. 534, no. 1, pp. 41–51, 2005.

[43] Y. Shen, W. Chiuman, J. D. Brennan, and Y. Li, "Catalysis and rational engineering of trans-acting pH6DZ1, an RNA-cleaving and fluorescence-signaling deoxyribozyme with a four-way junction structure," *ChemBioChem*, vol. 7, no. 9, pp. 1343–1348, 2006.

[44] M. M. Ali, S. D. Aguirre, H. Lazim, and Y. Li, "Fluorogenic DNAzyme probes as bacterial indicators," *Angewandte Chemie—International Edition*, vol. 50, no. 16, pp. 3751–3754, 2011.

[45] S. D. Aguirre, M. M. Ali, S. D. Aguirre, P. Kanda, and Y. Li, "Detection of bacteria using fluorogenic DNAzymes," *The Journal of Visualized Experiments*, no. 63, article 3961, 2012.

[46] X. B. Zhang, Z. Wang, H. Xing, Y. Xiang, and Y. Lu, "Catalytic and molecular beacons for amplified detection of metal ions and organic molecules with high sensitivity," *Analytical Chemistry*, vol. 82, no. 12, pp. 5005–5011, 2010.

[47] R. Kong, X. Zhang, Z. Chen et al., "Unimolecular catalytic DNA biosensor for amplified detection of L-histidine via an enzymatic recycling cleavage strategy," *Analytical Chemistry*, vol. 83, no. 20, pp. 7603–7607, 2011.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 215876, 17 pages
doi:10.1155/2012/215876

Review Article

Dioxaphosphorinane-Constrained Nucleic Acid Dinucleotides as Tools for Structural Tuning of Nucleic Acids

Dan-Andrei Catana,[1] Brice-Loïc Renard,[1] Marie Maturano,[1] Corinne Payrastre,[1] Nathalie Tarrat,[2,3] and Jean-Marc Escudier[1]

[1] *Laboratoire de Synthèse et Physicochimie de Molécules d'Intérêt Biologique, CNRS UMR 5068, Université Paul Sabatier, 31062 Toulouse, France*
[2] *Université de Toulouse, INSA, UPS, INP, LISBP, 135 Avenue de Rangueil, 31077 Toulouse, France*
[3] *INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, 31400 Toulouse, France*

Correspondence should be addressed to Jean-Marc Escudier, escudier@chimie.ups-tlse.fr

Received 1 June 2012; Accepted 2 August 2012

Academic Editor: Eriks Rozners

We describe a rational approach devoted to modulate the sugar-phosphate backbone geometry of nucleic acids. Constraints were generated by connecting one oxygen of the phosphate group to a carbon of the sugar moiety. The so-called dioxaphosphorinane rings were introduced at key positions along the sugar-phosphate backbone allowing the control of the six-torsion angles α to ζ defining the polymer structure. The syntheses of all the members of the D-CNA family are described, and we emphasize the effect on secondary structure stabilization of a couple of diastereoisomers of α,β-D-CNA exhibiting wether B-type canonical values or not.

1. Introduction

It is now clear that nucleic acids play several different roles in the living cell from genetic code storage to the catalysis of chemical reactions in ribosome. All of these particular behaviours are associated with various and very often transient structures of these polymers. The most prevalent secondary structure of nucleic acids is the double helix that can adopt either A- or B-type depending on the hydration level and/or the 2′-deoxyribosyl or ribosyl nature of the hybridized strands. While the backbone organization of double-stranded DNA and RNA is normally quite regular, there are many other secondary and tertiary structures that DNA and RNA molecules can adopt *in vivo* [1]. It is also well established that these disparate structures, which are predisposed to promote a significant local conformational heterogeneity in the sugar-phosphate backbone, play a crucial role in the fundamental biological processes where protein-nucleic acid interactions, folding, or catalytic activity are involved [2]. As a consequence nucleic acids can fold into biologically relevant distinct structures such as bulges, hairpin loops, *U*-turns, adenosine platforms, branched junctions, or quadruplexes (Figure 1). As proposed by few studies, the sugar/phosphate backbone of these unusual motifs exhibit a variety of conformations, which markedly differ from the regular conformational states of duplex DNA and RNA molecules [3–8]. However, the intrinsic role impar-ed to the phosphate diester backbone in respect with bases sequence in stamping these structures is still not properly defined.

The determination of the precise biological role played by nonstandard helical conformations during the biochemically important processes (e.g., protein-DNA complexation, DNA processing, and DNA packaging) is also an area of intense study [9, 10]. An important study based on an analysis of available high-resolution crystallographic data and molecular simulation techniques has shown that, in contrast to free B-DNA structures, protein-bound B-DNA oligomers regularly involve noncanonical backbone geometries [11].

These unusual backbone states are believed to contribute to the specific recognition of DNA by proteins in assisting, at some stages, the fine structural adjustments that are required

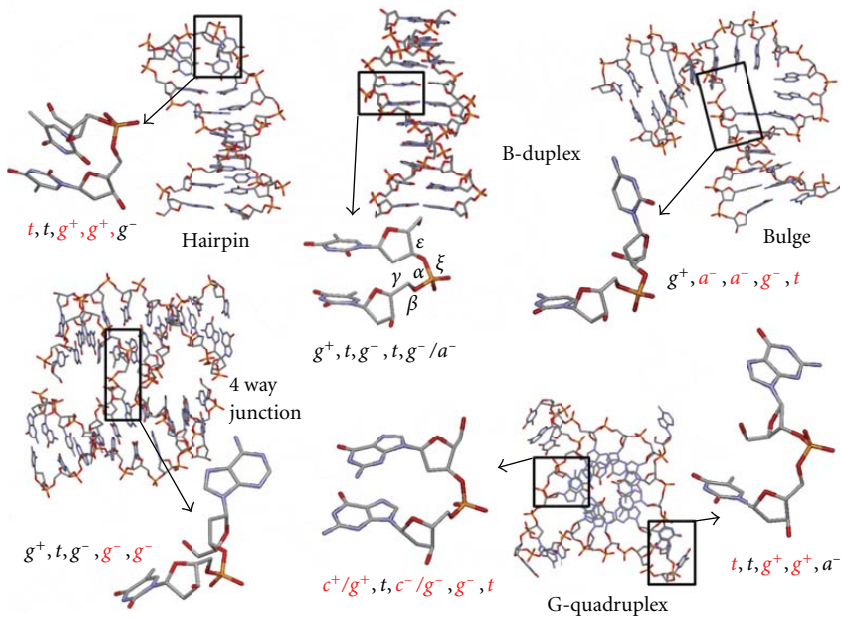


Figure 1: Examples of DNA secondary structures and associated backbone-torsion angles $\gamma/\beta/\alpha/\varepsilon/\zeta$ of representative dinucleotide units. The following 6-fold staggered pattern of the torsional angles is used: *cis* = 0 ± 30° (c), *gauche*(+) = 60 ± 30° (g^+), *anticlinal*(+) = 120 ± 30° (a^+), *trans* = 180 ± 30° (*t*), *anticlinal*(−) = 240 ± 30° (a^-), and *gauche*(−) = 300 ± 30° (g^-). The notation g^-/a^- is used to designate a torsion angle on the border of *gauche*(−) and *anticlinal*(−). PDB ID: Hairpin, 1ii1; duplex, 436d; Bulge, 1jrv; 4-Way junction, 1zez, and G-quad, 1kf1.

between DNA and proteins to form stable complexes. There are many examples in which DNA/protein complex formation results in DNA bending without disruption of the Watson-Crick base pairing [12–15]. Whereas this bending can be essential for complexes formation, it is generally sequence specific but with a strong impact on the sugar/phosphate backbone and can reach up to 90°. Unfortunately, experimental studies which aimed at determining the structural and functional implications of such helical deformations are somewhat complicated by the intrinsically transient nature of the corresponding backbone states. Stable structural analogues of these distorted backbone geometries would be very useful in the elucidation of the role that helical deformations play in nucleic acid interactions with proteins.

Mainly driven by the need of antisens research, most of the conformationally restricted oligonucleotides have been designed to enhance duplex formation ability and stability. Therefore, many efforts have been devoted to the synthesis of analogues with sugar-puckering conformational restriction of the North type [16–18, 18, 19]. To our knowledge, less attention has been paid to the design of conformationally restricted nucleosides with the aim of mimicking nucleic acid secondary structures containing non-Watson-Crick pairs or unpaired nucleotides. We are interested in the development of conformationally constrained dinucleotide building units in which the backbone torsional angles α–ζ can have predefined values that differ significantly from the typical values observed in DNA and RNA duplexes. In that context, the present paper will describe the last proposals and recent advances towards the introduction of conformational constraints into nucleotides by means of cyclic-phosphate structures.

The introduction of constraint on the sugar-phosphate backbone by connecting a phosphate to a base, sugar moiety, or another phosphate of the same strand gave new opportunity to provide conformationally constrained nucleic acids mimics. The pioneering work of the Sekine's group in the late 90′s illustrated this approach. They were interested in developing mimics of the *U*-turn structure [20]. This sharply bent conformation has been commonly found in the anticodon loop of tRNAs and later discovered at the active site of hammerhead ribozymes. Therefore, they focused on the preparation of two cyclic diuridylates (compounds **I** and **II**, Figure 2), in which the two nucleosides moieties were connected either by an amide group or by a carbamate function for **I**, or by introducing a bridge between the 5′-phosphate group and the 5-*C* position of the uracil moiety for **II** [21]. When incorporated within oligonucleotides, these modified nucleotides **I** were both able to induce a severe bent into the oligomer, whereas the cyclouridilic derivatives of type **II** could either allow the formation of the duplex with the R_P configured phosphotriester moiety or be a good motif to mimic the *U*-turn structure with the S_P configured rigid-cyclouridilic acid derivative [22, 23].

Later on, the Poul Nielsen's group showed that the ring-closing metathesis (RCM) reaction was a suitable methodology towards the synthesis of conformationally restricted dinucleotide structures (compounds **III** to **VII**) in order to preorganize a single-stranded nucleic acid and to either form stabilized duplexes or to induce stabilization in other secondary structures [24–29]. The approach is based on the synthesis of dinucleotide units (or trinucleotide units) with a phosphotriester linkage constructed by RCM between an

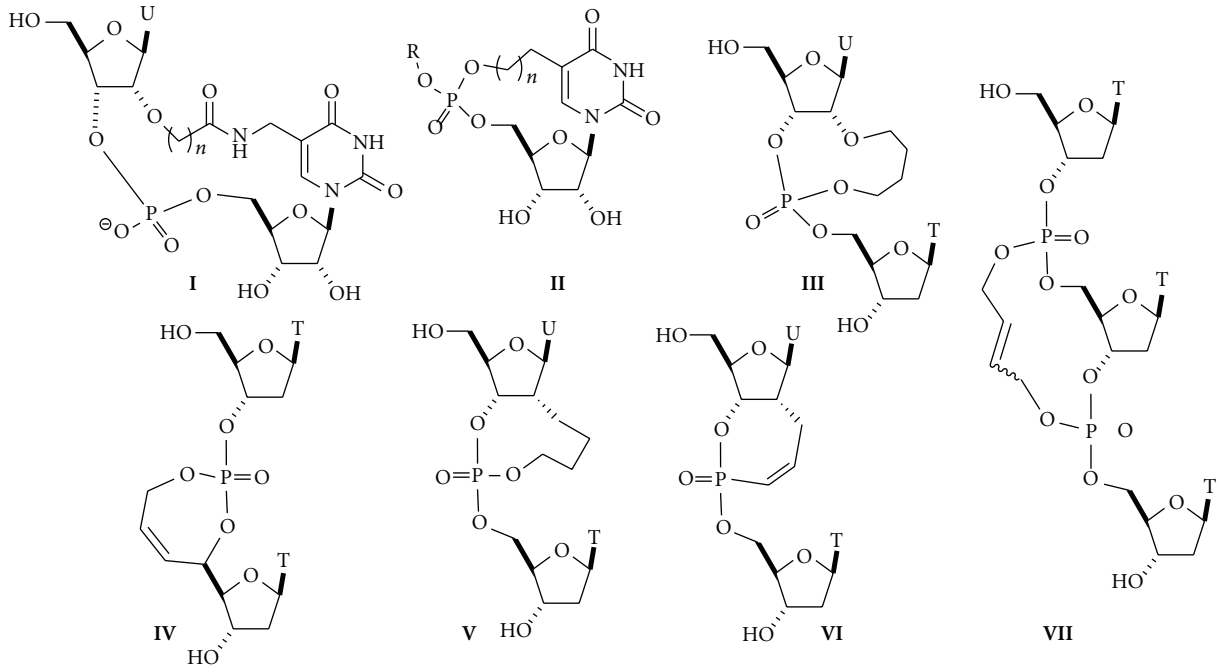


Figure 2: Selected macrocyclic-constrained nucleotides.

allyl-protected phosphate and another double-bond introduced at the appropriate location on the nucleoside either on the sugar or on the base moiety. Whereas all the constrained dinucleotide structures evaluated in duplex context showed destabilizing behavior, the R_P isomer of **V** provided the first example of stabilized three-way junction, in particular when the hairpin moiety was composed of ribonucleotides with an increase stability of +2.2°C rising to 2.7°C with the addition of Mg^{2+}.

The cyclic structures proposed there to modulate the sugar/phosphate backbone were composed of the smallest of a seven-membered cyclic phosphotriester to a very large macrocycle (up to eighteen members) and therefore exhibited rather flexible and undefined structures. In order to have a more rationalized approach to the design of covalently constrained nucleic acids (CNA) with specific canonical or noncanonical backbone conformations, we have developed dimeric building units in which two or three backbone torsion angles $\alpha - \zeta$ are part of a well-defined six-membered ring structure (Figure 3).

The so-called D-CNAs are dinucleotides, in which a set of backbone torsion angles α–ζ is stereocontrolled to canonical or noncanonical values by a 1,3,2-dioxaphosphorinane ring structure. For a given dinucleotide step, there are fourteen possible [*β-D-deoxyribo*]-configured D-CNA stereoisomers which formally result from the introduction of a methylene or ethylene linker between a nonbridging phosphate oxygen and the 2′/4′-carbons (methylene linker) or the 3′/5′-carbons (ethylene linker) of the sugar moiety.

Herein, we disclose the synthesis of each member of the D-CNA family, discuss their structural parameters which were established by means of X-ray diffraction analysis or NMR, and finally emphasize on the behaviour of *α,β*-D-CNA within duplex or hairpin secondary structure.

2. Synthesis of *α,β*-Constrained Nucleic Acids Dinucleotides (*α,β*-D-CNA and *α,β*-P-CNA)

Our retrosynthetic analysis for the synthesis of *α,β*-D-CNA dinucleotides is based on the very simple strategy that consists of using both steric and anomeric effects to stereocontrol the cyclization reaction of a dinucleotide precursor, in which the phosphate oxyanions can attack an activated carbon atom. The preparations of the *α,β*-D-CNA dithymidine diastereoisomers are disclosed in Scheme 1 [30, 31].

The key compounds of these pathways are the diastereopure 5′(*S*) and 5′(*R*)-*C*-hydroxyethyl-substituted nucleosides **3** and **7**, respectively. The former was obtained after reduction of the ester moiety of the product **2** of a diastereoselective Mukayama's reaction catalysed by $BiCl_3/ZnI_2$ on the aldehyde **1** [32–34]. The starting aldehyde **1** was prepared by a Pfitzner-Moffatt oxidation procedure of the primary hydroxyl function of the thymidine after a classical three-step protection/deprotection sequence [35, 36]. The 5′(*R*) isomer **7** was generated from **1** through a Sakuraï's allylation with a *ω*-subsituted-allyltrimethylsilane [37, 38] followed by a three-step oxidative cleavage protocol of the double bond of the 5′-*C*-hydroxypentenyl-thymidine **6** isolated by silica gel chromatography from its diastereoisomer. Selective tosylation of the primary-hydroxyl function was achieved in good yield by reaction with tosyl chloride in the presence of pyridine [39] to provide the corresponding 5′-*C*-tosyloxyethylthymidines that were coupled with the

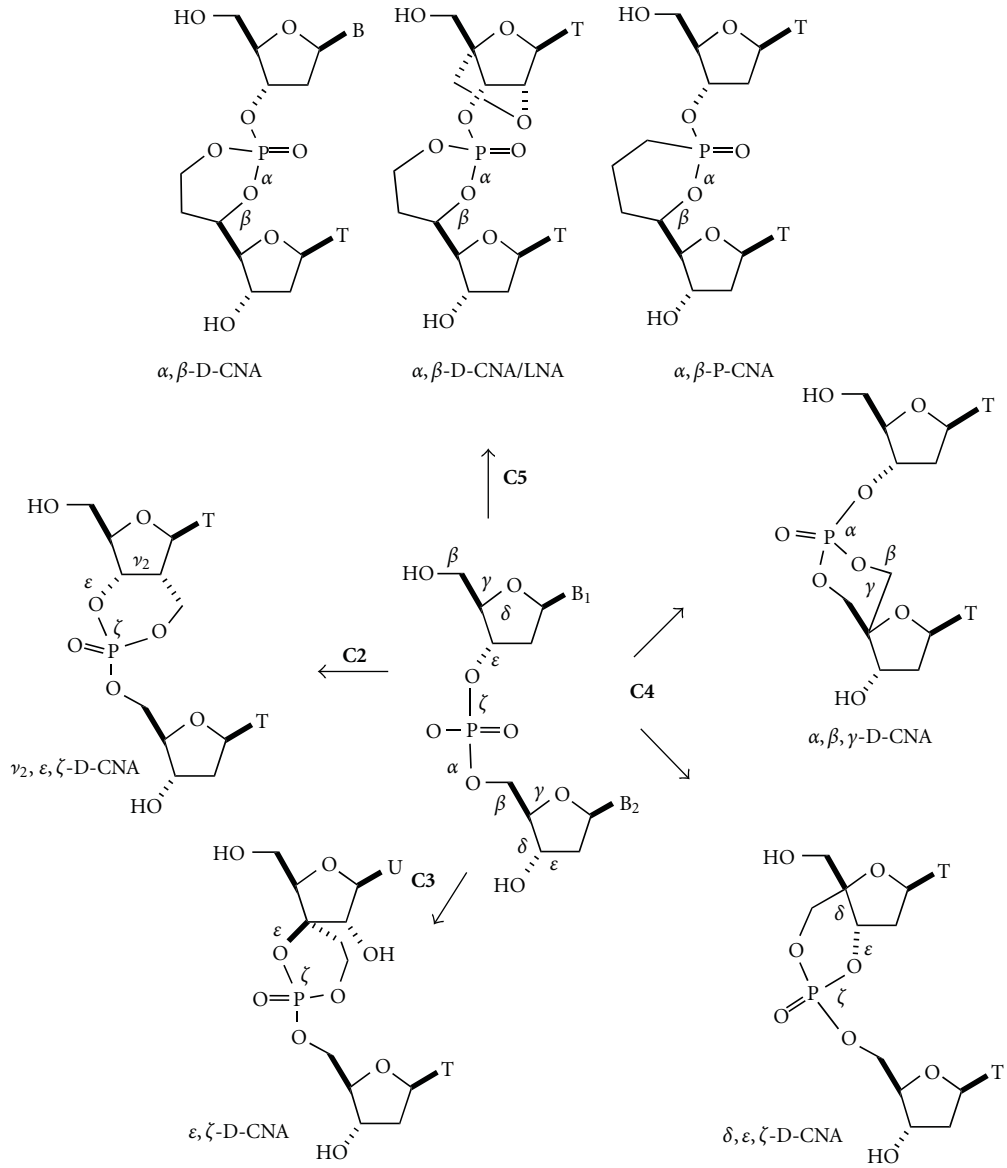


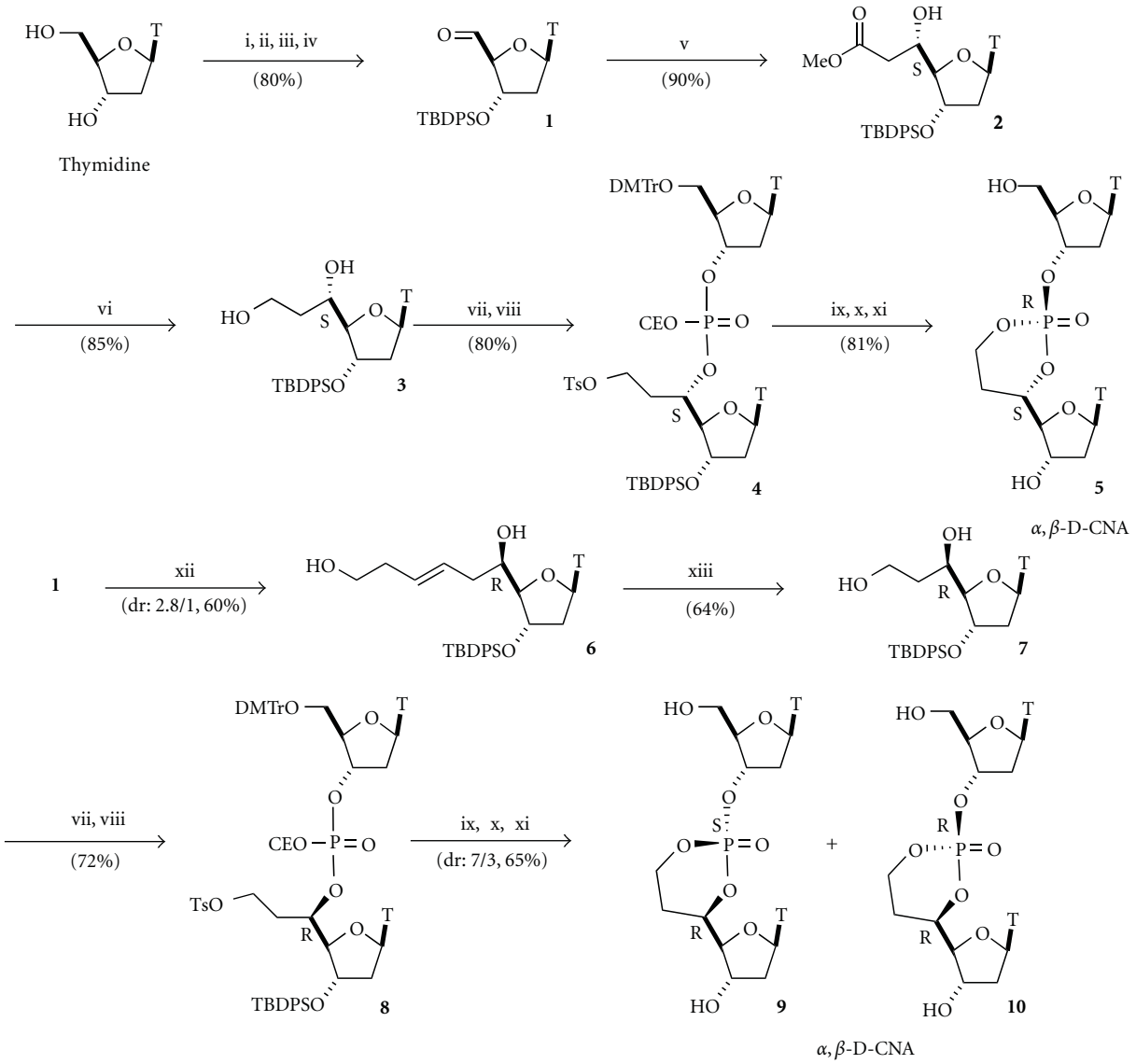
Figure 3: D-CNA dinucleotides building blocks for sugar/phosphate torsion angles control.

commercially available thymidine phosphoramidite under a standard phosphoramidite procedure [40] to give the acyclic dinucleotides **4** and **8** after the oxidation step, respectively.

The cyclization reaction for the formation of the dioxaphosphorinane structure occurred by the treatment of **4** or **8** in basic medium to generate the phosphate anion that can displace the tosylate group. Surprisingly, the (S_C, R_P) isomer **5** of α,β-D-CNA was exclusively obtained from **4** whereas a lower stereoselectivity of 7/3 was observed for the formation of the (R_C, S_P) **9** and (R_C, R_P) **10** α,β-D-CNA from the 5′(*R*)-C precursor **8**. After deprotection of the hydroxyl functions both, α,β-D-CNA were structurally characterized and revealed that the major isomers were those with the dioxaphosphorinane ring in the chair conformation (See Section 6).

Following the same chemical synthesis pathway, α,β-D-CNA analogues have been prepared by introducing a LNA-modified nucleoside during the phosphoramidite coupling to lead to LNA/α,β-D-CNA [41] (Figure 4), while changing the oxidation procedure from water/iodine to sulfur provided after cyclisation Thio-dioxa- and oxo-oxathiaphosphorinane structures (thio-α,β-D-CNA) [42]. Finally, starting from uridine or 2′-OMe-uridine, *ribo*-α,β-D-CNA could be achieved with the same diastereoselectivity outcome during the cyclisation process [43].

Therefore, the high diastereoselectivity observed for the formation of **5** led us to develop a strategy to synthesize the missing isomer in order to complete the set of CNA structural element. We turned our interest to phosphonate analogues of the D-CNA,, in which the dioxaphosphorinane ring



Scheme 1: Synthesis of α,β-Dioxaphosphorinane-Constrained Nucleic Acid dinucleotides (α,β-D-CNA). *Reagents and conditions:* (i) TBDMSiCl, pyr, 95%; (ii) TBDPSiCl, imid, DMF, 97%; (iii) PTSA, MeOH, 95%; (iv) DCC, $Cl_2HCCOOH$, DMSO then oxalic acid, 90%; (v) cat $BiCl_3$, ZnI_2, methyl acetate silylketene, DCM, 90%; (vi) $NaBH_4$, EtOH, 85%; (vii) TsCl, $CHCl_3$, pyr, 90; (viii) 5′-*O*-DMT-3′-*O*-diisopropylamino-cyanoethoxyphosphite thymidine, 1*H*-tetrazole, acetonitrile, then I_2/H_2O, 89%; (ix) Et_3N, DMF, 90°C, 95%; (x) 3% TFA/DCM, 95%; (xi) TBAF, THF, 90%; (xii) (a) ω-*t*-butyldimethyl-silyloxy-allyltrimethylsilane, $Et_2O{:}BF_3$, DCM, (b) $TiCl_4$, DCM, 60%; (xiii) (a) OsO_4 cat, *N*-methylmorpholine *N*-oxide, H_2O, (b) $NaIO_4$, MeOH, (c) $NaBH_4$, EtOH. 64%.

would be replaced by a cyclic phosphonate called phostone providing Phostone-Constrained Nucleic Acids building blocks (P-CNA) [44]. We speculated that an intramolecular Arbuzov reaction, performed on the phosphite dinucleotide intermediate **14** similar to that prepared for the synthesis of D-CAN, would be suitable to reach this target (Scheme 2).

Starting from the thymidine aldehyde **1**, allylation under Hosomi and Sakurai's condition [45] gave pure 5′(*S*)-*C*-allylthymidine **11** that underwent a selective hydroboration/oxidation of the double-bond after protection of the secondary hydroxyl function to provide **12**. The required 5′-*C*-tosyloxypropylthymidine **13** was reached by tosylation of the primary hydroxyl function and removal of the trimethylsilyl protective group. The key phosphite intermediate **14** resulted from the standard coupling of **13** with the commercially cyanoethyl protected thymidine phosphoramidite using usual tetrazole activation and without oxidation step. In optimized Arbuzov reaction conditions (micro-waves irradiation and addition of LiBr), the cyanoethyl group was eliminated after the attack of the phosphorus on the activated carbon, leading to the formation of a 2/1 diastereoisomeric mixture of phostones. The removal of the 5′ and 3′-hydroxyl function protective groups and silica gel chromatography led to the isolation of the P-CNA **15** and **16**. The (S_C, S_P)

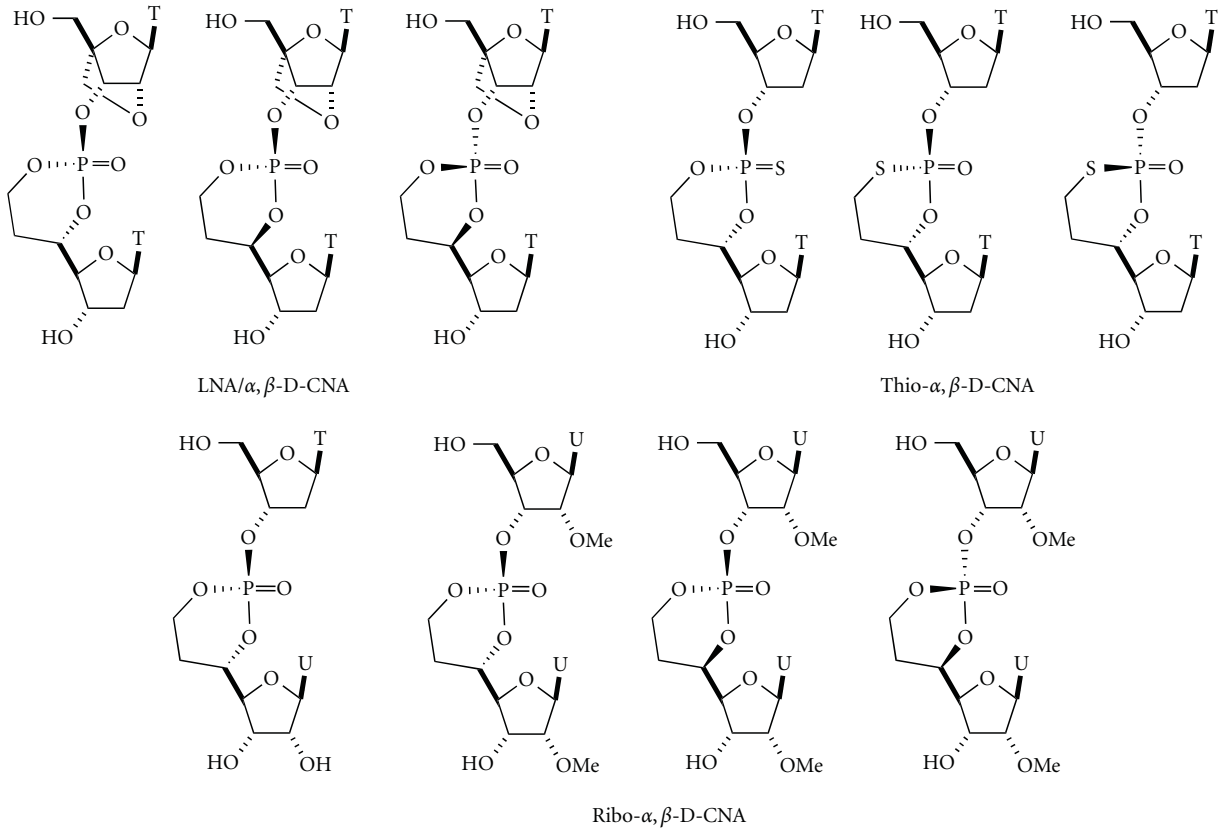


Figure 4: LNA-, thio-, and *ribo*-α,β-D-CAN.

α,β-P-CNA isomer **16** was isolated as the minor isomer and was corresponding to the structural analogue of the missing α,β-D-CNA.

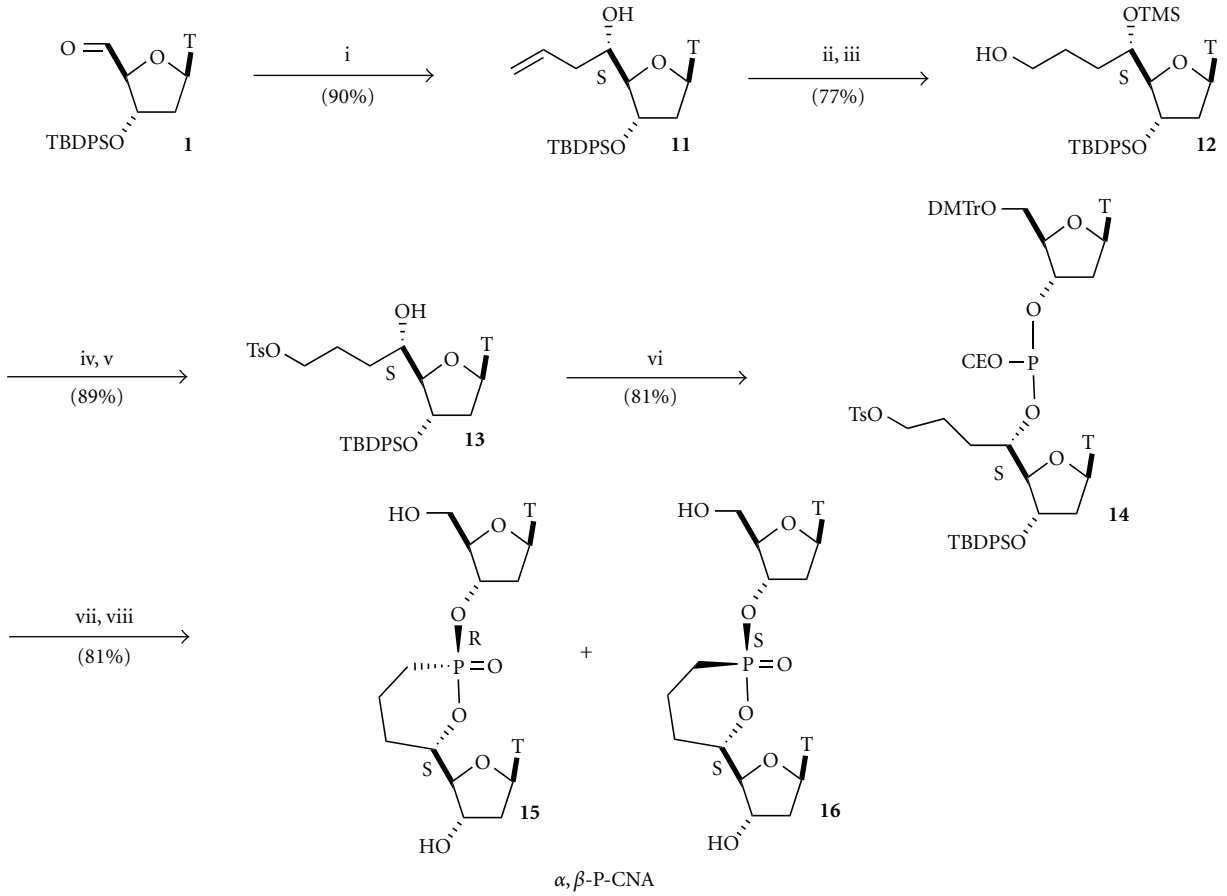
3. Synthesis of α,β,γ-Constrained Nucleic Acids Dinucleotides (α,β,γ-D-CNA) and δ,ε,ζ-Constrained Nucleic Acids Dinucleotides (δ,ε,ζ-D-CNA)

These two representatives of the D-CNA family originate from the connection of the phosphate to the 4′-*C*-carbon atom either of the downstream sugar moiety for the α,β,γ-D-CNA or of the upstream sugar moiety in the case of the δ,ε,ζ-D-CNA (Figure 3) [46]. Therefore, their synthesis started from a common intermediate 4′-*C*-hydroxymethylthymidine **17** obtained by a treatment of the thymidine aldehyde **1** under Cannizzaro's conditions (Schemes 3 and 4) [47].

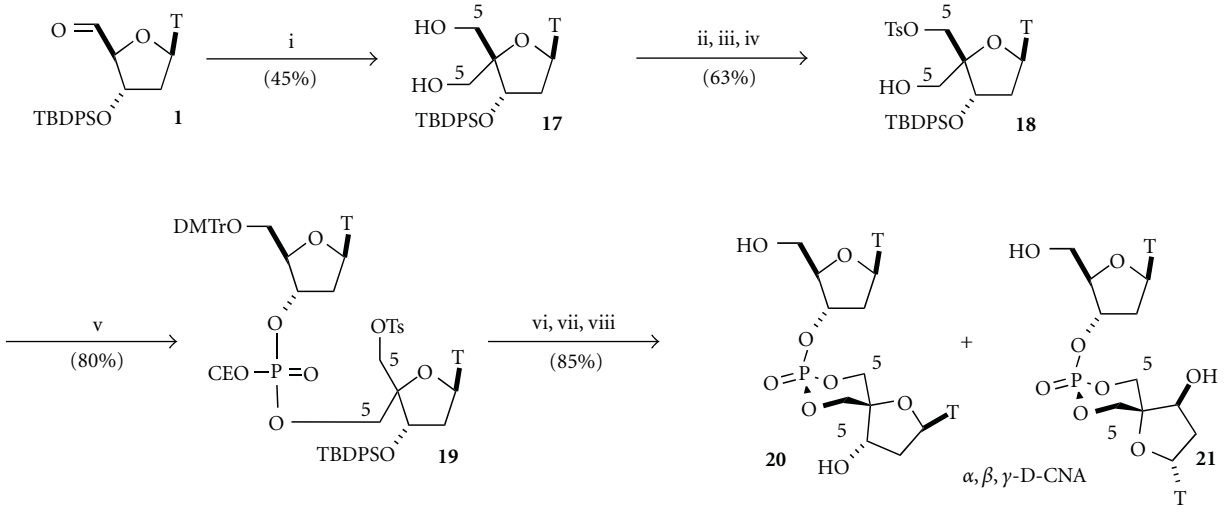
In the case of α,β,γ-D-CNA thymidine dinucleotides, the dioxaphosphorinane ring structure was formed as previously described for α,β-D-CNA by displacement of a tosyl group by a phosphate anion generated by the removal of a phosphate cyanoethyl protective group in basic medium (Scheme 3). The acyclic precursor involved is the dithymidine **19** prepared by coupling 5′-*O*-tosyl-4′-*C*-hydroxymethylthymidine **18** with the commercially available thymidine phosphoramidite using standard phosphoramidite technology. A three-step procedure involving first a selective protection of the 5′′-hydroxyl function of **17**, followed by tosylation of the residual primary 5′-hydroxyl function, and finally removal under acidic conditions of the dimetoxytrityl group furnished the required 5′-*O*-tosyl-4′-*C*-hydroxymethyl thymidine **18**. The removal of the cyanoethyl group from **19** by treatment with potassium carbonate in dimethylformamide generated the phosphate anion which by heating at 90°C provided the formation of two *cis*- and *trans*- isomers of protected α,β,γ-D-CNA in a 2/1 ratio in favor of the *cis*. After the removal of the 5′- and 3′-protective groups, the α,β,γ-D-CNA *cis* **20** could be separated from the *trans* isomer **21** and characterized.

A similar approach towards the synthesis of δ,ε,ζ-D-CNA, that is, introduction of a tosyl group on the 5′′-hydroxyl function, phosphoramidite coupling, and nucleophilic attack of the phosphate has been investigated, but it turned out to be troublesome and the desired 5′′-*O*-activated nucleoside could only be obtained in very poor yield. Therefore, we choose to use the well-known phosphotriester methodology that allows the formation of a phosphoester from a phosphate with an alcohol in the presence of an activator such as 1-(mesitylene-2-sulfonyl)-3-nitro-1,2,4-triazole (MSNT) [48].

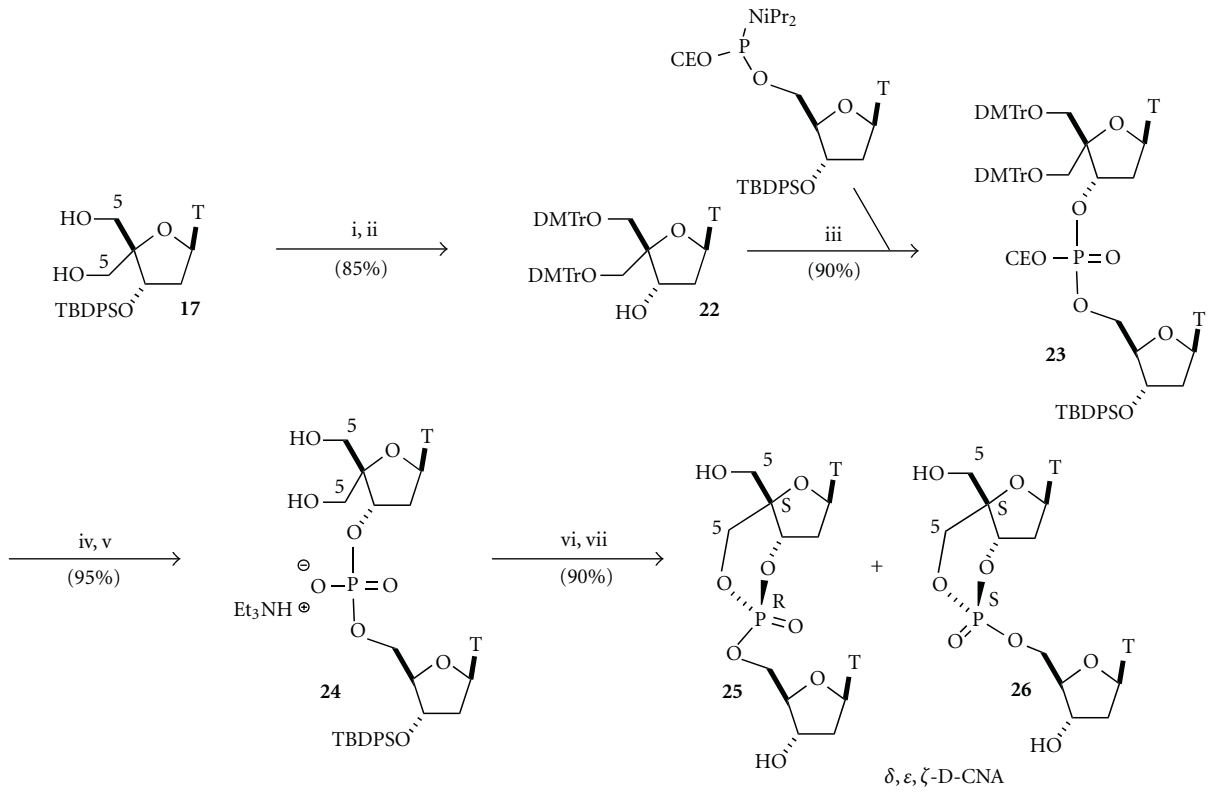
Starting from the diol **17**, both primary hydroxyl functions were protected as dimethoxytrityl ether and the 3′-*O*-silyl protective group was removed by treatment with



Scheme 2: Synthesis of α,β-Phostone-Constrained Nucleic Acid dinucleotides (α,β-P-CNA). *Reagents and conditions:* (i) allyltrimethylsilane, $BF_3 : Et_2O$, 90%, (ii) Me_3SiCl, pyr, rt, 3 h, 95%, (iii) $BH_3 : Me_2S$, THF, rt, 2 h then $NaOH/H_2O_2$, rt, 0.5 h, 90%, (iv) TsCl, $CHCl_3$, pyr, rt, 16 h, 85%, (v) PTSA, MeOH, rt, 1 h, 90%, (vi) thymidine phosphoramidite,1H-tetrazole, rt, 81%, (vii) LiBr, acetonitrile, 90°C, MW 3 h, 85%, and (viii) 3%TFA, CH_2Cl_2, then TBAF, THF, rt, 1 h, 90%.



Scheme 3: Synthesis of α,β,γ-Dioxaphosphorinane-Constrained Nucleic Acid dinucleotides (α,β,γ-D-CNA). *Reagents and conditions:* (i) HCOH, 2N NaOH, dioxane then $NaBH_4$, 45%, (ii) DMTrCl, pyr, 70%, (iii) TsCl, $CHCl_3$, Pyr, 90%, (iv) 3% TFA, DCM, quant. (v) 5′-*O*-DMT-3′-*O*-diisopropylamino-cyanoethoxyphosphite thymidine, 1*H*-tetrazole, acetonitrile, then I_2/H_2O, 80%, (vi) K_2CO_3, DMF, 90°C, 100%, (vii) TBAF, THF, 85%, and (viii) 3% TFA/DCM, quant.



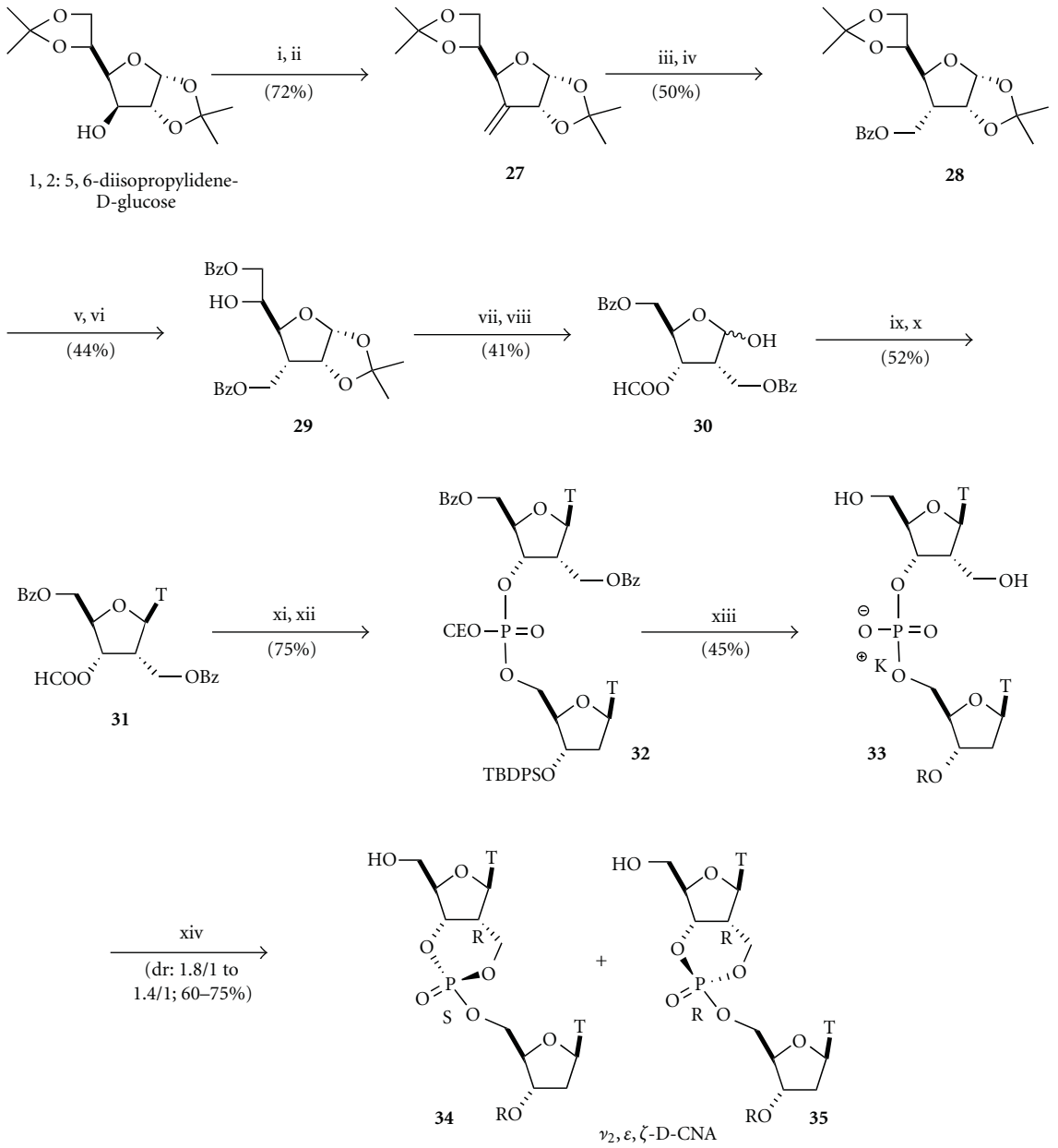
Scheme 4: Synthesis of δ, ε, ζ-Dioxaphosphorinane-Constrained Nucleic Acid dinucleotides (δ,ε,ζ-D-CNA). *Reagents and conditions:* (i) DMTrCl, DMAP, pyr, 90%, (ii) TBAF, THF, 95% and (iii) 5′-thymidine phosphoramidites, 1*H*-tetrazole, acetonitrile, then I_2/H_2O, 90%, (iv) 3% TFA, DCM, 95%, (v) Et_3N, acetonitrile, 60°C, quant, and (vi) 1-(mesitylene-2-sulfonyl)-3-nitro-1,2,4-triazole (MSNT), pyr, 80°C, quant, (vii) TBAF, THF, 90%.

fluoride ion to produce the nucleoside **22** (Scheme 4). A phosphoramidite coupling with a 5′-*O*-phosphoramidite-thymidine gave dinucleotide **23** that was consecutively treated in acidic medium to remove the dimethoxytrityl protective groups and in basic medium with triethylamine to eliminate the cyanoethyl phosphate protective group. The key phosphodiester **24** was then available to undergo the cyclisation process according to the phosphotriester methodology. Even if two primary hydroxyl functions were present, only the 5′′-hydroxyl reacted under the MSNT catalyst to form the dioxaphosphorinane ring. This high regioselectivity was unfortunately combined with no diastereoselectivity in the neither ring formation nor the formation of a 1/1 mixture of ($S_{C4'}$, R_P) and ($S_{C4'}$, S_P) diastereoisomers **25** and **26**, respectively. The poor diastereoselectivity could be explained by the fact that due to the fused sugar ring none of these compounds feature a chair conformation of the dioxaphosphorinane structure, which is indicative that there is not a more favorable intermediate during the cyclisation process.

4. Synthesis of ν_2,ε,ζ-Constrained Nucleic Acids Dinucleotides (ν_2,ε,ζ-D-CNA)

The synthesis of ν_2,ε,ζ-D-CNA implied the connection of the phosphate to the 2′-*C*-carbon of the sugar moiety through a methylene link [49]. To achieve this goal, instead of starting from a nucleoside precursor, we choose to reproduce a protocol previously described by Marquez and Coll. that used the commercially available 1,2 : 5,6-diisopropylidene-*D-glucose* and through an elegant rearrangement gave the pivotal protected 2-deoxy-*C*-(hydroxymethyl)-*D-ribo*furanose **30** (Scheme 5) [50]. Then a Vorbrüggen et al.'s procedure [51] could install the thymine base and a phosphoramidite coupling would provide the dinucleotide that could undergo the dioxaphosphorinane ring formation, here again by the phosphotriester method leading to the target ν_2,ε,ζ-D-CNA.

The secondary 3-hydroxyl function of 1,2 : 5,6-diisopropylidene-*D-glucose* was oxidised to ketone to be substrate for a Wittig homologation with methyltriphenylphosphonium on the 3-*C* position providing the sugar **27** with an exocyclic double bond. The hydroxymethyl function at 3-*C* was generated by a hydroboration/oxidation that occurred from the top-face of the sugar resulting in the formation of the required *R*-configured 3-carbon. Benzoylation of the resulting hydroxyl function provided the fully protected 3-deoxy-3-hydroxymethyl-*D-allose* **28**. Acidic hydrolysis of the 5,6-isopropylidene followed by a tricky selective benzoylation of the primary hydroxyl function led to **29** with the unprotected 5-secondary alcohol. Acetolysis of the 1,2-isopropylidene gave the 6-*O*-benzoyl-3-deoxy-3-benzoyloxymethyl-*D-allose* that was subsequently treated with sodium periodate to cleave the diol system. After rearrangement, the 2-deoxy-2-benzoyloxymethyl-*D-ribose*



Scheme 5: Synthesis of ν_2,ε,ζ-Dioxaphosphorinane-Constrained Nucleic Acid dinucleotides (ν_2,ε,ζ-D-CNA). *Reagents and conditions*: (i) PDC, Ac_2O, DCM, 80°C, 87%; (ii) nBuLi, Ph_3PMeBr, THF, 83%; (iii) BH_3Me_2S, THF, 0°C then H_2O_2 25%, NaOH 2 N, H_2O/THF (1/1), 81%; (iv) BzCl, pyr, DMAP, rt, 61%; (v) HCl 0.06 N, MeOH, 55°C, 2 h, 81%; (vi) BzCl, pyr, 0°C, 2 h, 54%; (vii) AcOH 80%, 80°C, 46%; (viii) KIO_4, EtOH/H_2O (1/1), rt, 18 h, 90%; (ix) Ac_2O, pyr, 94%; (x) TMSOTf 0.55 M in toluene, bis(trimethylsilyl)thymine, acetonitrile reflux, 56%; (xi) NH_4OH 28%, rt, quant; (xii) 3′-*O*-*t*-butyldiphenylsilylthymidine-5′-*O*-phosphoramidite, 1*H*-tetrazole, acetonitrile, then collidine, I_2/H_2O, 75%; (xiii) K_2CO_3, MeOH/H_2O (4/1), rt, 6 h, 45%; (xiv) 1-(mesitylene-2-sulfonyl)-3-nitro-1,2,4-triazole (MSNT), pyr, rt, 60–75%.

analogue has been isolated as a mixture of anomers. After protection of the anomeric position with an acetate function, thymine was introduced by a Vorbrüggen's procedure and the thymidine analogue **31** was obtained in a 1/9 ratio of α/β anomers. Removal of the residual-formyl group by aqueous ammonia gave the suitable nucleoside for a phosphoramidite coupling with a 5′-*O*-phosphoramidite-thymidine ending in the formation of the acyclic dinucleotide **32**. Potassium carbonate treatment, to remove the base labile benzoyl and cyanoethyl protective groups proceeded with a concomitant loss of the *t*-butyldiphenylsilyl group and dinucleotide **33** was obtained as a 1/1 mixture of fully deprotected, and 3′-*O*-silylated dinucleotide. These dinucleotides were separated and submitted to the cyclisation activated by 1-(mesitylene-2-sulfonyl)-3-nitro-1,2,4-triazole (MSNT) to furnish ($R_{C2'}$, S_P) and ($R_{C2'}$, R_P) ν_2,ε,ζ-D-CNA **34** and **35**, respectively.

While the ring formation occurred with a 1/1.8 ratio in the case of the partially protected dinucleotide, the diastereoselectivity was lowered to 1/1.4 for the fully deprotected dinucleotide.

5. Synthesis of *xylo*-ε,ζ-Constrained Nucleic Acids Dinucleotides (*xylo*-ε,ζ-D-CNA)

The restrains on only the torsional angles ε and ζ requires the formation of a spiro connection between the sugar and the dioxaphosphorinane rings by introduction of an ethylene linker between the 3′-*C*-carbon atom of the sugar moiety and the phosphate (Figure 3) [52]. To date, on the four possible stereoisomers, we have reported the synthesis of the *xylo*-configured D-CNAs because they represent a class of distorted structures directly available from commercially uridine (Scheme 6). A similar approach to that proposed for α,β-D-CNA has been followed for the preparation of the *xylo*-ε,ζ-D-CAN, that is, aldol condensation to introduce the ethylene link on the 3′-*C* and activation through a tosylation to form the dioxaphosphorinane ring after phosphoramidite coupling.

Uridine was selectively protected on the 5′-*O* and 2′-*O* by *t*-butyldimethylsilyl group following the Ogilvie's procedure [53] before being oxidized the with Dess-Martin periodinane [54] to give the keto-uridine **36**. A stereoselective Mukaïyama's addition of the *t*-butyldimethylsilyl-methylketene acetal occurred on the *Re* face of the carbonyl as determined by NOE experiments on the adduct **37**. Reduction of the ester function turned to be rather difficult using $NaBH_4$ and the solution came from DIBAH; however; in a modest yield. The primary hydroxyl function was then selectively tosylated to provide the 3′-tosyloxyethyl-*xylo*-uridine **38** suitable to be engaged in the phosphoramidite coupling with the 5′-*O*-phosphoramidite-thymidine. The acyclic 3′-*C* -tosyloxyethyluridine/thymidine dinucleotide **39** was then submitted to basic treatment at room temperature, and the generated phosphate anion cleanly displaced the tosyl group to form a 1/1 diastereoisomeric mixture of protected ($S_{C3'}$, R_P) and ($S_{C3'}$, S_P) *xylo*-ε,ζ-D-CNA **40** and **41**, respectively, which have been separated on reverse phase HPLC after deprotection. Whereas a relative instability of the phosphotriester could be expected due to the presence of the secondary hydroxyl function, the spiro structure with an "*S*" configuration of 3′-*C* fixed their relative positions away to the necessary "on line" conformation avoiding any transesterification process [55, 56].

6. Structural Assignment

The determination of the values of the constrained torsional angles within D-CNA structures relied on the establishment of the geometry of the dioxaphosphorinane ring whether in chair conformation or not. Some of D-CNAs were crystallized and solid phase structures were determined by X-ray diffraction analysis for ($R_{C5'}$, S_P) α,β-D-CNA TT (compound **9**, Scheme 1), ($S_{C5'}$, R_P) α,β-D-CNA TU (Figure 4), and ($S_{C4'}$, S_P) δ,ε,ζ-D-CNA TT (compound **26**, Scheme 4).

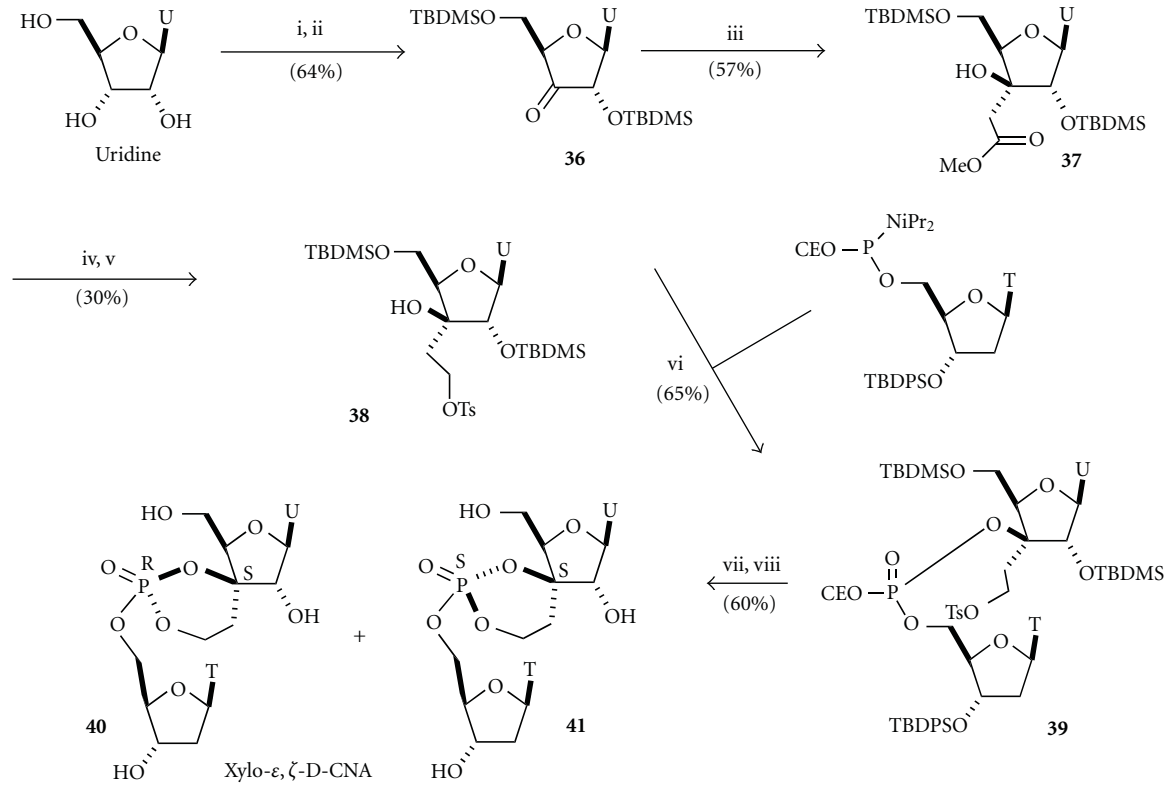
Table 1: Summary of the backbone torsion angles derived from the canonical B_I-, A-DNA duplex structures and of the synthesized D-CNA[a].

Name	Isomer	Torsion angles					
		α	β	γ	δ	ε	ζ
B_I-type		g^-	t	g^+	a^+/t	t	g^-/a^-
A-type		g^-	t	g^+	g^+/a^+	a^-/t	g^-
α,β-D-CNA	($R_{C5'}$, S_P)	g^-	t				
	($S_{C5'}$, R_P)	g^+	t				
	($R_{C5'}$, R_P)	a^+	a^-/t				
	($S_{C5'}$, S_P)	a^-	a^+/t				
α,β-P-CNA	($S_{C5'}$, S_P)	g^+	t				
	($R_{C5'}$, R_P)	a^-/t	t				
LNA/α,β-D-CNA	($R_{C5'}$, S_P)	g^-	t		g^+/a^+		
	($S_{C5'}$, R_P)	g^+	t		g^+/a^+		
	($R_{C5'}$, R_P)	a^+	a^-/t		g^+/a^+		
α,β,γ-D-CNA	*cis*	g^-	g^-	g^-			
	trans	g^+	c/g^+	g^-/a^-			
δ,ε,ζ-D-CNA	($S_{C4'}$, R_P)				a^+/t	g^-	g^-
	($S_{C4'}$, S_P)				a^+/t	g^-	a^+/t
ν_2,ε,ζ-D-CNA	($R_{C2'}$, S_P)				a^+	g^+	g^-
	($R_{C2'}$, R_P)				a^+	g^+	t
ε,ζ-D-CNA	($S_{C3'}$, R_P)				c	t	g^+
	($S_{C3'}$, S_P)				c	t	g^-

[a]The following 6-fold staggered pattern of the torsional angles is used: *cis*: 0 ± 30° (c), *gauche* (+) = 60 ± 30° (g^+), *anticlinal* (+): 120 ± 30° (a^+), *trans*: 180 ± 30° (t), *anticlinal* (−) = 240 ± 30° (a^-), and *gauche* (−) = 300 ± 30° (g^-). The notation a^+/t is used to designate a torsion angle on the border of *anticlinal* (+) and *trans*.

Moreover, NMR analysis of the H/H and H/P coupling constants of the protons involved in the dioxaphosphorinane ring or in the sugar moieties either corroborated the results of the X-ray analysis or allowed for the establishment of the rings conformations. Interestingly, $^3J_{H/P}$ coupling constants between relevant protons within the dioxaphosphorinane ring and the phosphorous gave important information because they exhibit specific values dependant on the relative axial or equatorial position of the proton within the six membered ring, that is, $^3J_{Hax/P} < ca \cdot 3\,Hz$ and $^3J_{Heq/P} < ca \cdot 20\,Hz$, respectively [57]. Therefore, a careful examination of these data allowed for the determination of the dioxaphosphorinane ring conformation, whereas $^3J_{H/H}$ coupling constants gave also information on the sugar puckering. The conformational ranges of the constrained torsional angles within D-CNA determined by these methods are summarized in Table 1. Torsional angles' values depicted in A- or B-type duplex are given as reference and are considered as canonical values for the regular double-helix structure [58].

Among all the sets of constrained torsional angles, the values exhibited for α and β by the ($R_{C5'}$, S_P) isomer of



Scheme 6: Synthesis of *xylo*-ε,ζ-Dioxaphosphorinane-Constrained Nucleic Acid dinucleotides (*xylo*-ε,ζ-D-CNA). *Reagents and conditions:* (i) TBDMSCl, $AgNO_3$, Pyr, THF, 74%, (ii) Dess-Martin periodinane, Pyr, DCM, rt, 87%, (iii) *t*-butyldimethylsilyl-methyl-ketene, $BF_3 \cdot Et_2O$, DCM, −78°C, 57%, (iv) DIBAH, DCM, 0°C, 2 h then rt 12 h, 46%, (v) TsCl, Pyr, $CHCl_3$, 64%, (vi) 3′-O-*t*-Butyldiphenylsilylthymidine-5′-*O*-phosphoramidite (2 eq), tetrazole, acetonitrile then collidine, I_2/H_2O, 65%, (vii) K_2CO_3, DMF, rt., 81%, and (viii) TBAF (3.3 eq), THF, rt, 74%.

α,β-D-CNA (or its analogue LNA/α,β-D-CNA) are identical to those observed for the A- or B- type duplex. In contrast, and as expected by the proposed approach, all the others constrained dinucleotides feature-torsional angle' values greatly differ from the canonical ones. Therefore, these members offer an extraordinary diversity in the relative spatial arrangement of the bases moieties allowing the description of an unusual local shape of nucleic acids. In order to illustrate this point, Figure 5 shows a superimposition of ($R_{C5'}$, S_P)-, ($S_{C5'}$, R_P) α,β-D-CNA and *cis*-, *trans*-α,β,γ-D-CNA featuring a (g^-, t), (g^+, t), (g^-, g^-), and (g^+, c/g^+) set of value for α and β, respectively. Whereas ($R_{C5'}$, S_P) α,β-D-CNA analogue stands for a good mimic of B-type dinucleotide with the thymine bases mostly stacked, it is nicely illustrated that the two bases can be oriented in rather different planes in the others D-CNA dinucleotides.

The sugar puckering of each nucleosides within D-CNA was estimated by the empirical equation of Altona-Sundaralingan using the $^3J_{H/H}$ coupling constants: C2′-*endo* (%) = $[J_{1'/2}/(J'_{1'/2} + J'_{3'/4})] \times 100$ [59]. Due to an increase of the electronegativity of the 3′ oxygen by introduction of a neutral internucleotidic linkage, the sugar pucker of the upstream nucleoside is favored in its C2′-*endo* conformation (South) [60]. The determination of the impact that a neutral phosphotriester linkage would display in the conformational North/South equilibrium is particularly important as it is well recognized that this conformational state is of major importance for the DNA duplex formation ability.

The examination of the relevant coupling constants showed that in all cases for D-CNA built with 2′-deoxyribose, the sugar puckering of the 5′-upstream or 3′-downstream nucleoside were in the C2′-*endo* conformation. However, for α,β-, α,β,γ-, and ν_2,ε,ζ-D-CNA the 5′-upstream nucleoside sugar puckering equilibrium was strongly displaced toward the C2′-*endo* conformation (South) compared to natural 2′-deoxyribose units [61]. Interestingly, in the cases of *ribo*-α,β-D-CNA (Figure 4), even the 2′-OMe-*ribose* unit was pushed into the C2′-endo conformation upon the influence of the neutralized internucleotidic linkage. Only the *xylose*-configured sugar within *xylo*-ε,ζ-D-CNA adopted a North conformation (C2′-*exo*).

Dioxaphosphorinane-modified sugar/phosphate backbone of dinucleotide could therefore represent a promising methodology to provide alternative backbone conformations. It is likely that D-CNA within DNA or RNA oligomers would be able to modulate the shape and the folding with significant-conformational distortion of secondary nucleic acid structures.

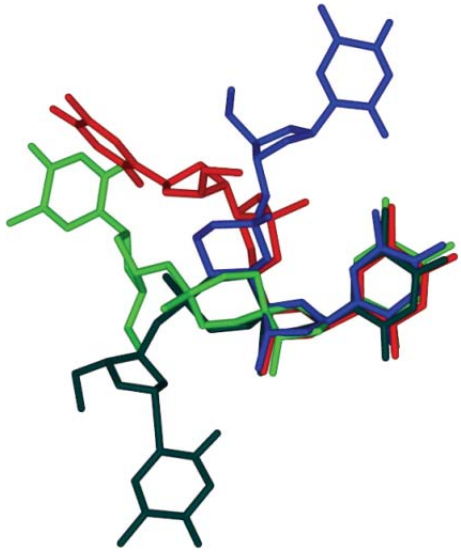


Figure 5: superimposition of minimized structures of α,β-D-CNA (g^-, t) blue; α,β-D-CNA (g^+, t) red; α,β,γ-D-CNA (g^-, g^-, g^-) green; α,β,γ-D-CNA (g^+, c/g^+, g^-/a^-) dark green.

7. Survey of α,β-D-CNA Dinucleotides Behavior within Oligonucleotides

We focused our interest on the study of the impact of the restraint on one specific torsional angle, α, through the behavior within oligodeoxynucleotide (ODN) of a couple of α,β-D-CNA diastereoisomers featuring either canonical or noncanonical α/β combination [62]. As shown previously (Scheme 1 and Table 1), ($S_{C5'}$, R_P) and ($R_{C5'}$, S_P) α,β-D-CNA derivatives **5** and **9** can be easily prepared and their structural assignment showed that the ($R_{C5'}$, S_P) α,β-D-CNA **9** exhibited a canonical value set (*gauche*(−), *trans*) for α and β, whereas its diastereoisomer ($S_{C5'}$, R_P) α,β-D-CNA **5** differed only on the α value which was changed to the *gauche*(+) conformation while maintaining β in the *trans* configuration. Therefore, we dispose of a unique couple of modified nucleotides that will give us new insight on the impact of backbone preorganization either in the B-type duplex geometry or with a strong torsional stress applied on α corresponding to that observed in DNA/protein complex or in unpaired secondary structures such as hairpin or bulges.

A molecular dynamic simulation has been run on dA_{10}/dT_{10} duplex whether modified or not with one TT step constrained with ($R_{C5'}$, S_P) or ($S_{C5'}$, R_P) α,β-D-CNA denoted as **ODNref**, **ODNgm**, and **ODNgp**, respectively (Figure 6) [63]. This study gave us two main results: compared to unmodified duplex the structure seems to accommodate the canonical restraint on α with a straightness of the double-helix whereas the *gauche*(+) conformation induced a bend without loss of the Watson-Crick base pairing. Therefore, these observations let us speculate that controlling the torsion of an ODN into its B-type canonical form should enhance the duplex formation ability, whereas displacing it to around +70° might result in the formation of localized distortion able to stabilize unpaired conformations.

Interestingly, analysis of the atomic fluctuations derived from these simulations indicated that in **ODNgm** all of these fluctuations were diminished in both strands which could be indicative of a potential duplex stabilisation, whereas in **ODNgp** they were unchanged compared to those observed in **ODNref**.

Therefore, we investigated the behavior of α,β-D-CNA within ODNs by thermal denaturation studies by means of UV experiments. Selected results are reported in Table 2 for ($R_{C5'}$, S_P) α,β-D-CNA (denoted to as **CNAgm**), in Table 3 for a comparative study between ($R_{C5'}$, S_P) and ($S_{C5'}$, R_P) α,β-D-CAN, and in Table 4 for hairpin structures stabilisation by ($S_{C5'}$, R_P) α,β-D-CNA (denoted to as **CNAgp**).

All the ODNs containing D-CNA were obtained by automated synthesis according to the phosphoramidite methodology. The phosphoramidite building blocks of ($R_{C5'}$, S_P) and ($S_{C5'}$, R_P) α,β-D-CNA were synthesized by conventional method and their incorporation within ODNs occurred similarly to standard phosphoramidite with no

Table 2: Sequences and melting temperatures of **CNAgm** containing duplexes.

Entry	Sequence[a]	T_m[b] (°C)	ΔT_m (°C)
1	5′-GCGTTTTTTGCT-3′ 3′-CGCAAAAAACGA-5′	49.0	—
2	5′-GCGTT***TxT***TTGCT-3′ 3′-CGCAAAAAACGA-5′	55.0	+6.0
3	5′-GCGTT***TxTTxT***GCT-3′ 3′-CGCAAAAAACGA-5′	59.0	+10.0
4	5′-GCG***TxTTxTTxT***GCT-3′ 3′-CGCAAAAAACGA-5′	64.0	+15.0
5	5′-GCAAAAACTTGC-3′ 3′-CGTTTTTGAACG-5′	48.0	—
6	5′-GCAAAAAC***TxT***GC-3′ 3′-CGTTTTTGAACG-5′	53.2	+5.2
7	5′-GCAAAAACTTGC-3′ 3′-CGTTT***TxT***GAACG-5′	52.2	+4.2
8	5′-GCAAAAACTTGC-3′ 3′-CGTT***TxT***TGAACG-5′	54.3	+6.3
9	5′-GCAAAAACTTGC-3′ 3′-CGT***TxT***TTGAACG-5′	53.4	+5.4
10	5′-GCAAAAACTTGC-3′ 3′-CG***TxT***TTTGAACG-5′	52.4	+4.4
11	5′-GCAAAAAC***TxT***GC-3′ 3′-CGTTT***TxT***GAACG-5′	55.9	+7.9
12	5′-GCAAAAAC***TxT***GC-3′ 3′-CGTT***TxT***TGAACG-5′	56.4	+8.4
13	5′-GCAAAAAC***TxT***GC-3′ 3′-CGT***TxT***TTGAACG-5′	57.6	+9.6
14	5′-GCAAAAAC***TxT***GC-3′ 3′-CG***TxT***TTTGAACG-5′	57.8	+9.8

[a] ***TxT***: ($R_{C5'}$, S_P) α,β-D-CNA TT (**CNAgm**) within the strand. [b]UV melting experiments were carried out in sodium phosphate buffer (10 mM, pH 7.0) containing NaCl (100 mM) and EDTA (1 mM).

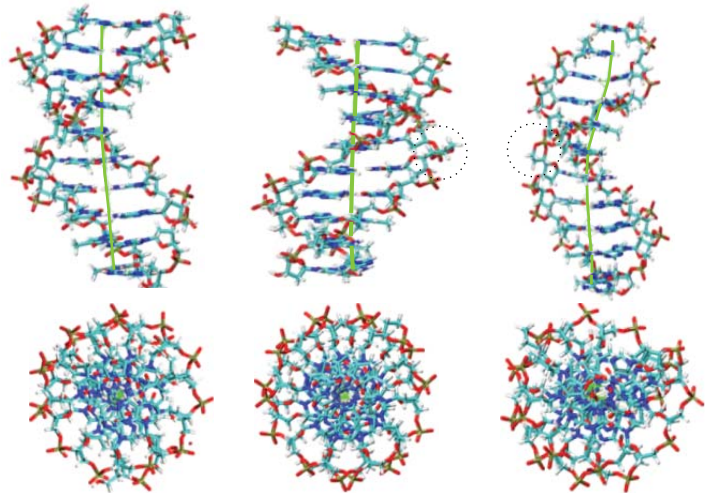


Figure 6: Side and top views of B-type duplex produced by molecular dynamics simulations, the double-helix axes are outlined in green. (a) Unmodified dA_{10}/dT_{10} **ODNref**, (b) $dA_{10}/dT_4\boldsymbol{TxT}T_4$ **ODNgm** with $\boldsymbol{TxT} = (R_{C5'}, S_P)$ α,β-D-CNA TT, or (c) **ODNgp** with $\boldsymbol{TxT} = (S_{C5'}, R_P)$ α,β-D-CNA TT.

change in automated synthesis protocols but with a smooth deprotection in ammonia at room temperature.

The introduction of a canonical constraint within ODN resulted in a remarkable stabilizing effect on duplex formed with DNA counterparts ($\Delta Tm = +5.0 \pm 1°C$/mod, Table 2, entries 2, 6–10) [64]. These increases in Tm values are insensitive to salt concentration suggesting that the effects observed were primarily conformational rather than electrostatic. Thus, **CNAgm** represents a rare example of constrained nucleotide that significantly increases the hybridizing properties of ODNs without forcing the sugar pucker into the C3′-*endo* conformation, demonstrating that the preorganization concept can also be successfully applied to other torsional angles than those involved in the sugar moiety puckering. The ability of **CNAgm** to adapt to the B-conformation of the double-helix is outlined by its additive stabilizing effect (entries 2–4) when included in the same strand and also when the two strands are modified with one **CNAgm**, with a maximum effect when constraints were close to the 3′-ends preventing end frying (entries 11–14).

It is noteworthy that a rather moderate effect was observed with RNA counterparts ($\Delta Tm = +0.9$ to $+3.0°C$/mod), which could originate from the reluctance of the upstream-furanose unit of the α,β-D-CNA to undergo a significant conformational change from 2′-*endo* to 3′-*endo* in the hybrid duplex DNA/RNA due to the loss of the internucleotidic negative charge.

On the other hand, when we prepared sequences either modified by **CNAgm** or by **CNAgp** we were able to have insight on the cost in terms of thermal stability of a dramatic change of restraint on α from *gauche*(−) to *gauche*(+). As expected, incorporation of **CNAgp** featuring noncanonical (*gauche*(+), *trans*) α/β combination resulted in an important loss in duplex stability ($-4.2°$ to $-13.6°C$/mod, Table 3) depending on the sequence length and composition. Short decamer and rather unstable oligothymidilate exhibited the higher destabilized level (entries 1–3) while increasing the size to 18-mer (entries 4–10) and 24-mer (entries 11, 12) modulated the impact of the *gauche*(+) restraint around $-5 \pm 1°C$/mod and also minimized the positive effect on duplex formation ability of **CNAgm** from $+5 \pm 1°C$/mod to $+1 \pm 1°C$/mod. We showed that regardless of the type of restriction applied to ODN a high level of sequence discrimination was maintained as natural duplexes do.

Interestingly, exceptions in the destabilization effect of **CNAgp** (entries 6 and 13) and in the positive impact of **CNAgm** (entry 13) appeared. However, the first sequence is hemipalindromic and the second is fully self-complementary. Therefore, they can exist either as duplex or as hairpin structure. A further experiment showed that the observed transitions were indeed a combination of melting temperature from equilibrium made of high hairpin transition and lower duplex melting. In the case of the Drew-Dickerson sequence (entry 13), **CNAgm** was able to displace the equilibrium in favor of the duplex, whereas **CNAgp** displaced it to the hairpin structure because constraint was imposed within the loop.

In order to emphasize the effect of stabilization of unpaired region of secondary nucleic acids structure by **CNAgp**, we engaged the synthesis of modified T_4 loop within hairpin that could differ in their stem composition and especially in the AT or CG loop-closing base pair (Table 4) [65].

In a hairpin, which is a single-stranded structure, the sugar/phosphate backbone orientation is reversed by means of the loop moiety. The necessary torsional stress is not spread throughout all the loop constituents but ensured by a sharp-turn position called the "turning phosphate" that displays a *gauche*(+) transition of α torsional angle [66]. Therefore, it was tempting to speculate that **CNAgp** could

Table 3: Comparison of melting temperatures of **CNAgm** or **CNAgp** containing duplexes.

Entry	Sequence[a]	T_m[b] °C (ΔT_m °C)		
		$x = PO_2^-$	$x =$ **CNAgm**	$x =$ **CNAgp**
1	dA_{14}/dT_4***TxT***T_4	24.2	30.1 (+5.9)	10.5 (−13.7)
2	dA_{14}/dT_6***TxT***T_6	35.0	38.1 (+3.1)	24.0 (−11.0)
3	5′-GCGC***TxT***GCCG-3′ 3′-CGCGAACGGC-5′	53.0	58.0 (+5.0)	44.0 (−9.0)
4	5′-ACAT***TxT***GAAATGCAAATG-3′ 3′-TGTAAACTTTACGTTTAC-5′	49.0	52.0 (+3.0)	44.1 (−4.9)
5	5′-ACATTTGAAATGCAAATG-3′ 3′-TGTAAACTTTACG***TxT***TAC-5′	49.0	54.9 (+5.9)	43.3 (−5.7)
6	5′-ACATTTGAAATGCAAATG-3′ 3′-TGTAAAC***TxT***TACGTTTAC-5′	49.0	54.5 (+5.5)	49.0 (0.0)
7	5′-CTCATGAATATGCAAATC-3′ 3′-GAGTACTTATACG***TxT***TAG-5′	52.4	55.0 (+2.6)	47.2 (−5.2)
8	5′-CTCATGAATATGCAAATC-3′ 3′-GAGTAC***TxT***ATACGTTTAG-5′	52.4	56.5 (+4.1)	47.6 (−4.8)
9	5′-TGCTCAGTAAATAATGCA-3′ 3′-ACGAGTCATTTA***TxT***ACGT-5′	55.6	59.0 (+3.4)	49.3 (−6.3)
10	5′-TGCTCAGTAAATAATGCA-3′ 3′-ACGAGTCA***TxT***TATTACGT-5′	55.6	58.8 (+3.2)	49.6 (−6.0)
11	5′-ATCTCATTTGAAATGCAAATGGAA-3′ 3′-TAGAGTAAACTTTACG***TxT***TACCTT-5′	57.9	59.3 (+1.4)	53.5 (−4.4)
12	5′-TGTCTCATGAATATGCAAATCACA-3′ 3′-ACAGAGTACTTATACG***TxT***TAGTGT-5′	59.9	61.3 (+1.4)	55.7 (−4.2)
13	5′-CGCGAA***TxT***CGCG-3′ 3′-GCGC***TxT***AAGCGC-5′	62.0[c]	58.0 (−4.0)[c]	27.0 (−35.0) 82.0 (+20.0)

[a] ***TxT***: modified TT within the strand. [b]UV melting experiments were carried out in sodium phosphate buffer (10 mM, pH 7.0) containing NaCl (100 mM) and EDTA (1 mM). [c]Only one transition observed.

play the role of a preorganized "turning phosphate" and as a consequence could induce hairpin stabilization [67]. **CNAgp** was installed in all the possible positions within the loop, and the thermal stabilities were evaluated by UV melting curves analysis. In the case of T_4-looped hairpin structures with AT closing base pair, the central position was best suited for **CNAgp** (Table 4, entry 3 versus 2, 4, and 5) with a maximum in ΔTm of +3.0°C. However, if a constraint in the middle of the loop helped the hairpin folding, when installed at the 3′-end of the loop, **CNAgp** strongly was destabilized by −7.0°C (entry 5). Remarkably, with CG closing base pair, **CNAgp** behaves as a stabilizing analogue in any position within the loop (entries 6–9). Indeed, circular dichroism experiments showed that when **CNAgp** was placed in the middle of the loop (entry 8), the stem structure was not altered, whereas when located in 3′-end region a stem rearrangement occurred that could participate to the stabilization enhancement observed (entry 9). Similar results were depicted for T_5-looped hairpin structures with ΔTm up to +5.0°C [68].

Eventually, we showed that the two diastereoisomers of α,β-D-CNA featuring a fixed torsional angle alpha either in the B-type canonical value *gauche*(−) or in atypical *gauche*(+) conformation are powerful building blocks allowing high level of duplex or hairpin stabilization as expected according to the preorganization concept. This is evidence that controlling the sugar/phosphate backbone not only in terms of sugar puckering is a promising approach toward the control of nucleic acids secondary structures.

8. Conclusions

The development of nucleotides analogues, for the purpose of mimicking nucleic acids secondary structures, started with the pioneering work of Sekine with his approach toward *U*-turn loop, and then the design of conformationally constrained nucleotides grew up through the Nielsen's ring-closing metathesis pathway. Finally, the introduction of dioxaphosphorinane element at key positions along the sugar/phosphate backbone proved to be a rational concept to gain control on torsional angle sets. We have synthesized most of the possible members of the D-CNA family; all this structural units provide control on α to ζ torsional angle associated with a broad range of backbone conformations. Interestingly, the dioxaphosphorinane ring structures within D-CNAs are reasonably stable towards the oligonucleotide synthesis conditions according to a special care during the final deprotection step and are especially inert towards enzymatic degradation such as snake venon phosphodiesterase [69] as expected for phosphotriesters [70]. As a consequence, they are potential elements for the elaboration of synthetic nucleic acids with programmable folding and stability either

Table 4: Thermal melting temperatures (°C) of **CNAgp** within hairpin structures.

Entry	Sequence (5′-3′)[a]	T_m (°C)[b]	ΔT_m (°C)
1	ATCCTA*TTTT*TAGGAT	52.0	—
2	ATCCTA***TxT****TT*TAGGAT	53.0	+1.0
3	ATCCTA*T****TxT****T*TAGGAT	55.0	+3.0
4	ATCCTA*TT* ***TxT***TAGGAT	50.0	−2.0
5	ATCCTA*TTT****TxT***AGGAT	45.0	−7.0
6	AGGATCC*TTTT*GGATCCT	74.0	—
7	AGGATCC***TxT****TT*GGATCCT	75.3	+1.3
8	AGGATCC*T****TxT****T*GGATCCT	77.5	+3.5
9	AGGATCC*TT****TxT***GGATCCT	76.7	+2.7

[a]***TxT*** denotes an ($S_{C5'}$, R_P) α,β-D-CNA-modified TT dinucleotide (**CNAgp**) and italic character denotes loop nucleotide. Underlined bases indicate the loop-closing base pair. [b]Melting temperatures (T_m values) were measured as the maximum of the first derivate of the UV melting curve (OD_{260} versus temperature, 20–90°C, 0.5°C/min) which was recorded at concentration around 5 μM in sodium phosphate buffer (10 mM, pH 7.0) containing NaCl (100 mM) and EDTA (1 mM).

in the double-helix or in the unpaired secondary structures. As a proof of concept, we demonstrated that relying on the restrain applied within oligodeoxynucleotides by means of α,β-D-CNA, high level of duplex formation ability or hairpin stabilization could be achieved. Therefore, at least with these leading components of the family, torsional stress applied to the sugar/phosphate backbone could be used in probing the necessary flexibility, or in contrary rigidity, of the nucleic acids architecture during the interaction with ligands or biomacromolecules. However, D-CNAs can be seen as a new alphabet for the conception of shape-defined nucleic acids, and if ten years ago C. Leumann [71] concluded a review by "*a large field that has not yet been tapped is the use of conformationally constrained nucleosides for the stabilization of secondary structural elements as, for example hairpin loops and bulges,*" there is still a long way before being able to properly address the use of each member of the family and to understand or predict the behavior of D-CNA within nucleic acids. Nevertheless, this new kind of nucleotide analogues could be the basis for the development of synthetic oligonucleotides for the modulation of protein/nucleic acids complex formation [72].

Acknowledgment

The authors are grateful to the Agence Nationale de la Recherche for giving a grant to B. L. Renard and for financial support to the CNA project (ANR-11-BS07-012-01).

References

[1] P. Belmont, J. F. Constant, and M. Demeunynck, "Nucleic acid conformation diversity: from structure to function and regulation," *Chemical Society Reviews*, vol. 30, no. 1, pp. 70–81, 2001.

[2] A. A. Antson, "Single stranded RNA binding proteins," *Current Opinion in Structural Biology*, vol. 10, no. 1, pp. 87–94, 2000.

[3] E. Ennifar, M. Yusupov, P. Walter et al., "The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges," *Structure*, vol. 7, no. 11, pp. 1439–1449, 1999.

[4] V. Tereshko, S. T. Wallace, N. Usman, F. E. Wincott, and M. Egli, "X-ray crystallographic observation of "*in-line*" and "*adjacent*" conformations in a bulged self-cleaving RNA/DNA hybrid," *RNA*, vol. 7, no. 3, pp. 405–420, 2001.

[5] L.-W. Hung, E. L. Holbrook, and S. R. Holbrook, "The crystal structure of the Rev binding element of HIV-1 reveals novel base pairing and conformational variability," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 10, pp. 5107–5112, 2000.

[6] W. G. Scott, J. T. Finch, and A. Klug, "The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage," *Cell*, vol. 81, no. 7, pp. 991–1002, 1995.

[7] J. A. Ippolito and T. A. Steitz, "A 1.3-Å resolution crystal structure of the HIV-1 trans-activation response region RNA stem reveals a metal ion-dependent bulge conformation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 17, pp. 9819–9824, 1998.

[8] J. H. Cate, A. R. Gooding, E. Podell et al., "RNA tertiary structure mediation by adenosine platforms," *Science*, vol. 273, no. 5282, pp. 1696–1699, 1996.

[9] X.-J. Lu, Z. Shakked, and W. K. Olson, "*A*-form conformational motifs in ligand-bound DNA structures," *Journal of Molecular Biology*, vol. 300, no. 4, pp. 819–840, 2000.

[10] W. K. Olson, A. A. Gorin, X.-J. Lu, L. M. Hock, and V. B. Zhurkin, "DNA sequence-dependent deformability deduced from protein-DNA crystal complexes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 19, pp. 11163–11168, 1998.

[11] P. Várnai, D. Djuranovic, R. Lavery, and B. Hartmann, "α/γ transitions in the B-DNA backbone," *Nucleic Acids Research*, vol. 30, no. 24, pp. 5398–5406, 2002.

[12] F. Guo, D. N. Gopaul, and G. D. Van Duyne, "Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 13, pp. 7143–7148, 1999.

[13] M. Lewis, G. Chang, N. C. Horton et al., "Crystal structure of the lactose operon repressor and its complexes with DNA and inducer," *Science*, vol. 271, no. 5253, pp. 1247–1254, 1996.

[14] L. Tora and H. Timmers, "The TATA box regulates TATA-binding protein (TBP) dynamics in vivo," *Trends in Biochemical Sciences*, vol. 35, no. 6, pp. 309–314, 2010.

[15] K. Furuita, S. Murata, J. G. Jee, S. Ichikawa, A. Matsuda, and C. Kojima, "Structural feature of bent DNA recognized by HMGB1," *Journal of the American Chemical Society*, vol. 133, no. 15, pp. 5788–5790, 2011.

[16] A. J. A. Cobb, "Recent highlights in modified oligonucleotide chemistry," *Organic and Biomolecular Chemistry*, vol. 5, no. 20, pp. 3260–3275, 2007.

[17] N. M. Bell and J. Micklefield, "Chemical modification of oligonucleotides for therapeutic, bioanalytical and other applications," *ChemBioChem*, vol. 10, no. 17, pp. 2691–2703, 2009.

[18] C. Zhou and J. Chattopadhyaya, "The synthesis of therapeutic locked nucleos(t)ides," *Current Opinion in Drug Discovery and Development*, vol. 12, no. 6, pp. 876–898, 2009.

[19] J. Lebreton, J.-M. Escudier, L. Arzel, and C. Len, "Synthesis of bicyclonucleosides having a C-C bridge," *Chemical Reviews*, vol. 110, no. 6, pp. 3371–3418, 2010.

[20] K. Seio, T. Wada, K. Sakamoto, S. Yokoyama, and M. Sekine, "Chemical synthesis and properties of conformationally fixed

diuridine monophosphates as building blocks of the RNA turn motif," *Journal of Organic Chemistry*, vol. 63, no. 5, pp. 1429–1443, 1998.

[21] K. Seio, T. Wada, K. Sakamoto, S. Yokoyama, and M. Sekine, "Chemical synthesis and conformational properties of a new cyclouridylic acid having an ethylene bridge between the uracil 5-position and 5′-phosphate group," *Journal of Organic Chemistry*, vol. 61, no. 4, pp. 1500–1504, 1996.

[22] M. Sekine, O. Kurasawa, K.-I. Shohda, K. Seio, and T. Wada, "Synthesis and properties of oligonucleotides having a phosphorus chiral center by incorporation of conformationally rigid 5′-cyclouridylic acid derivatives," *Journal of Organic Chemistry*, vol. 65, no. 20, pp. 6515–6524, 2000.

[23] K. Seio, T. Wada, and M. Sekine, "Synthesis and properties of oligothymidylates incorporating an artificial bend motif," *Helvetica Chimica Acta*, vol. 83, no. 1, pp. 162–180, 2000.

[24] A. M. Sørensen and P. Nielsen, "Synthesis of conformationally restricted dinucleotides by ring-closing metathesis," *Organic Letters*, vol. 2, no. 26, pp. 4217–4219, 2000.

[25] A. M. Sørensen, K. E. Nielsen, B. Vogg, J. P. Jacobsen, and P. Nielsen, "Synthesis and NMR-studies of dinucleotides with conformationally restricted cyclic phosphotriester linkages," *Tetrahedron*, vol. 57, no. 51, pp. 10191–10201, 2001.

[26] P. Børsting and P. Nielsen, "Tandem ring-closing metathesis and hydrogenation towards cyclic dinucleotides," *Chemical Communications*, no. 18, pp. 2140–2141, 2002.

[27] P. Børsting, A. M. Sørensen, and P. Nielsen, "A ring-closing metathesis strategy towards conformationally restricted di- and trinucleotides," *Synthesis*, no. 6, pp. 797–801, 2002.

[28] P. Bersting, K. E. Nielsen, and P. Nielsenl, "Stabilisation of a nucleic acid three-way junction by an oligonucleotide containing a single 2′-C to 3′-O-phosphate butylene linkage prepared by a tandem RCM-hydrogenation method," *Organic and Biomolecular Chemistry*, vol. 3, no. 11, pp. 2183–2190, 2005.

[29] P. Børsting, M. S. Christensen, S. I. Steffansen, and P. Nielsen, "Synthesis of dinucleotides with 2′-C to phosphate connections by ring-closing metathesis," *Tetrahedron*, vol. 62, no. 6, pp. 1139–1149, 2006.

[30] I. Le Clézio, J.-M. Escudier, and A. Vigroux, "Diastereoselective synthesis of a conformationally restricted dinucleotide with predefined α and β torsional angles for the construction of α,β-constrained nucleic acids (α,β-CNA)," *Organic Letters*, vol. 5, no. 2, pp. 161–164, 2003.

[31] C. Dupouy, I. Le Clézio, P. Lavedan, H. Gornitzka, J.-M. Escudier, and A. Vigroux, "Diastereoselective synthesis of conformationally restricted dinucleotides featuring canonical and noncanonical α/β torsion angle combinations (α,β-D-CNA)," *European Journal of Organic Chemistry*, no. 24, pp. 5515–5525, 2006.

[32] T. Mukaiyama, "Titanium tetrachloride in organic synthesis," *Angewandte Chemie—International Edition*, vol. 16, no. 12, pp. 817–826, 1977.

[33] C. Le Roux, H. Gaspard-Iloughmane, J. Dubac, J. Jaud, and P. Vignaux, "New effective catalysts for Mukaiyama-Aldol and -Michael reactions: $BiCl_3$-metallic iodide systems," *Journal of Organic Chemistry*, vol. 58, no. 7, pp. 1835–1839, 1993.

[34] J.-M. Escudier, I. Tworkowski, L. Bouziani, and L. Gorrichon, "Synthèse stéréosélective de thymidine substituée en C-5′," *Tetrahedron Letters*, vol. 37, no. 27, pp. 4689–4692, 1996.

[35] K. E. Pfitzner and J. G. Moffatt, "Sulfoxide-carbodiimide reactions. I. A facile oxidation of alcohols," *Journal of the American Chemical Society*, vol. 87, no. 24, pp. 5661–5670, 1965.

[36] G. H. Jones, M. Taniguchi, D. Tegg, and J. G. Moffatt, "4′-Substituted nucleosides. 5. Hydroxymethylation of nucleoside 5′-aldehydes," *Journal of Organic Chemistry*, vol. 44, no. 8, pp. 1309–1317, 1979.

[37] V. Banuls and J.-M. Escudier, "Allylsilanes in the preparation of 5′-C-hydroxy or bromo alkylthymidines," *Tetrahedron*, vol. 55, no. 18, pp. 5831–5838, 1999.

[38] V. Banuls, J.-M. Escudier, C. Zedde, C. Claparols, B. Donnadieu, and H. Plaisancié, "Stereoselective synthesis of (5′S)-5′-C-(5-bromo-2-penten-1-yl)-2′-deoxyribofuranosyl thymine, a new convertible nucleoside," *European Journal of Organic Chemistry*, no. 24, pp. 4693–4700, 2001.

[39] G. W. Kabalka, M. Varma, R. S. Varma, P. C. Srivastava, and F. F. Knapp, "Tosylation of alcohols," *Journal of Organic Chemistry*, vol. 51, no. 12, pp. 2386–2388, 1986.

[40] M. H. Caruthers, "Chemical synthesis of DNA and DNA analogues," *Accounts of Chemical Research*, vol. 24, no. 9, pp. 278–284, 1991.

[41] C. Dupouy, P. Lavedan, and J.-M. Escudier, "Synthesis and structure of dinucleotides featuring canonical and non-canonical A-type duplex α, β and δ torsion angle combinations (LNA/α,β-D-CNA)," *European Journal of Organic Chemistry*, no. 31, pp. 5256–5264, 2007.

[42] I. Le Clézio, A. Vigroux, and J.-M. Escudier, "Diastereoselective and regioselective synthesis of conformationally restricted thio-dioxa- and oxo-oxathiaphosphorinane dinucleotides featuring noncanonical α/β torsion angle combinations (α,β-CNAs)," *European Journal of Organic Chemistry*, no. 12, pp. 1935–1941, 2007.

[43] M. Maturano, D.-A. Catana, P. Lavedan et al., "Synthesis and structural study of *ribo*-dioxaphosphorinane-constrained nucleic acid dinucleotides (ribo-α,β-D-CNA)," *European Journal of Organic Chemistry*, no. 4, pp. 721–730, 2012.

[44] D.-A. Catana, M. Maturano, C. Payrastre, P. Lavedan, N. Tarrat, and J.-M. Escudier, "Synthesis of phostone-constrained nucleic acid (P-CNA) dinucleotides through intramolecular Arbuzov's reaction," *European Journal of Organic Chemistry*, no. 34, pp. 6857–6863, 2011.

[45] A. Hosomi and H. Sakurai, "Syntheses of γ,δ-unsaturated alcohols from allylsilanes and carbonyl compounds in the presence of titanium tetrachloride," *Tetrahedron Letters*, vol. 17, no. 16, pp. 1295–1298, 1976.

[46] I. Le Clézio, H. Gornitzka, J.-M. Escudier, and A. Vigroux, "Constrained nucleic acids (CNA). Part 2. Synthesis of conformationally restricted dinucleotide units featuring non-canonical $\alpha/\beta/\gamma$ or $\delta/\varepsilon/\zeta$ torsion angle combinations," *Journal of Organic Chemistry*, vol. 70, no. 5, pp. 1620–1629, 2005.

[47] R. D. Youssefyeh, J. P. H. Verheyden, and J. G. Moffatt, "4′-Substituted nucleosides. 4. Synthesis of some 4′-hydroxymethyl nucleosides," *Journal of Organic Chemistry*, vol. 44, no. 8, pp. 1301–1309, 1979.

[48] S. S. Jones, B. Rayner, C. B. Reese, A. Ubasawa, and M. Ubasawa, "Synthesis of the 3′-terminal decaribonucleoside nonaphosphate of yeast alanine transfer ribonucleic acid," *Tetrahedron*, vol. 36, no. 20-21, pp. 3075–3085, 1980.

[49] C. Dupouy, P. Lavedan, and J.-M. Escudier, "Synthesis and structure of dinucleotides with S-type sugar puckering and noncanonical ε and ζ torsion angle combination (ν_2, ε,ζ-D-CNA)," *European Journal of Organic Chemistry*, no. 7, pp. 1285–1294, 2008.

[50] C. K. H. Tseng, V. E. Marquez, G. W. A. Milne et al., "A ring-enlarged oxetanocin A analogue as an inhibitor of HIV infectivity," *Journal of Medicinal Chemistry*, vol. 34, no. 1, pp. 343–349, 1991.

[51] H. Vorbrüggen, K. Krolikiewiecz, and B. Bennua, "Nucleoside syntheses, XXII1) Nucleoside synthesis with trimethylsilyl

triflate and perchlorate as catalysts," *Chemische Berichte*, vol. 114, no. 4, pp. 1234–1255, 1981.

[52] C. Dupouy, P. Lavedan, and J. M. Escudier, "Synthesis of spiro ε,ζ-D-CNA in xylo configuration featuring noncanonical δ/ε/ζ torsion angle combination," *Tetrahedron*, vol. 63, no. 46, pp. 11235–11243, 2007.

[53] G. H. Hakimelahi, Z. A. Proba, and K. K. Ogilvie, "High yield selective 3′-silylation of ribonucleosides," *Tetrahedron Letters*, vol. 22, no. 52, pp. 5243–5246, 1981.

[54] D. B. Dess and J. C. Martin, "A useful 12-i-5 triacetoxyperiodinane (the dess-martin periodinane) for the selective oxidation of primary or secondary alcohols and a variety of related 12-i-5 species," *Journal of the American Chemical Society*, vol. 113, no. 19, pp. 7277–7287, 1991.

[55] R. A. Torres and T. C. Bruice, "The mechanism of phosphodiester hydrolysis: near in-line attack conformations in the hammerhead ribozyme," *Journal of the American Chemical Society*, vol. 122, no. 5, pp. 781–791, 2000.

[56] H. Lönnberg, "Cleavage of RNA phosphodiester bonds by small molecular entities: a mechanistic insight," *Organic and Biomolecular Chemistry*, vol. 9, no. 6, pp. 1687–1703, 2011.

[57] D. G. Gorenstein, R. Rowell, and J. Findlay, "Stereoelectronic effects in the reactions of epimeric 2-aryloxy-2-oxy-1,3,2-dioxaphosphorinanes and oxazaphosphorinanes," *Journal of the American Chemical Society*, vol. 102, no. 15, pp. 5077–5081, 1980.

[58] B. Schneider, S. Neidle, and H. M. Berman, "Conformations of the sugar-phosphate backbone in helical DNA crystal structures," *Biopolymers*, vol. 42, no. 1, pp. 113–124, 1997.

[59] C. Altona and M. Sundaralingam, "Gonformational analysis of the sugar ring in nucleosides and nucleotides. Improved method for the interpretation of proton magnetic resonance coupling constants," *Journal of the American Chemical Society*, vol. 95, no. 7, pp. 2333–2344, 1973.

[60] M. Ikehara, "2′-substituted 2′-deoxypurinenucleotides their conformation and properties," *HeteroCycles*, vol. 21, no. 1, pp. 75–90, 1984.

[61] D. J. Wood, K. K. Ogilvie, and F. E. Hruska, "Proton magnetic resonance studies of dinucleotide conformation in aqueous solution. 2′-deoxythymidylyl-(3′,5′)-2′-deoxy-3′-thymidylate, d(TpTp)," *Canadian Journal of Chemistry*, vol. 53, no. 18, pp. 2781–2790, 1975.

[62] C. Dupouy, N. Iché-Tarrat, M. P. Durrieu, F. Rodriguez, J. M. Escudier, and A. Vigroux, "Watson-Crick base-pairing properties of nucleic acid analogues with stereocontrolled α and β torsion angles (α,β-D-CNAs)," *Angewandte Chemie—International Edition*, vol. 45, no. 22, pp. 3623–3627, 2006.

[63] C. Dupouy, N. Iché-Tarrat, M. P. Durrieu, A. Vigroux, and J. M. Escudier, "α,β-D-CNA induced rigidity within oligonucleotides," *Organic and Biomolecular Chemistry*, vol. 6, no. 16, pp. 2849–2851, 2008.

[64] A. Boissonnet, C. Dupouy, P. Millard, M. P. Durrieu, N. Tarrat, and J. M. Escudier, "α,β-D-CNA featuring canonical and noncanonical α/β torsional angles behaviours within oligonucleotides," *New Journal of Chemistry*, vol. 35, no. 7, pp. 1528–1533, 2011.

[65] C. Dupouy, P. Millard, A. Boissonnet, and J. M. Escudier, "α,β-D-CNA preorganization of unpaired loop moiety stabilizes DNA hairpin," *Chemical Communications*, vol. 46, no. 28, pp. 5142–5144, 2010.

[66] C. W. Hilbers, H. A. Heus, M. J. P. Van Dongen, and S. S. Wijmenga, "The hairpin elements of nucleic acid structure: DNA and RNA folding," in *Nucleic Acids and Molecular Biology*, F. Eckstein and D. M. J. Lilley, Eds., vol. 8, p. 56, Springer, Berlin, Germany, 1994.

[67] I. Le Clézio, C. Dupouy, P. Lavedan, and J.-M. Escudier, "Synthesis and structure of an α,β-D-CNA featuring a noncanonical α/β torsion angle combination within a tetranucleotide," *European Journal of Organic Chemistry*, no. 23, pp. 3894–3900, 2007.

[68] Unpublished results.

[69] P. S. Miller, K. N. Fang, N. S. Kondo, and P. O. P. Ts'o, "Syntheses and properties of adenine and thymine nucleoside alkyl phosphotriesters, the neutral analogs of dinucleoside monophosphates," *Journal of the American Chemical Society*, vol. 93, no. 24, pp. 6657–6665, 1971.

[70] C. Zhou, O. Plashkevych, and J. Chattopadhyaya, "Double sugar and phosphate backbone-constrained nucleotides: synthesis, structure, stability, and their incorporation into oligodeoxynucleotides," *Journal of Organic Chemistry*, vol. 74, no. 9, pp. 3248–3265, 2009.

[71] C. J. Leumann, "DNA analogues: from supramolecular principles to biological properties," *Bioorganic and Medicinal Chemistry*, vol. 10, no. 4, pp. 841–854, 2002.

[72] O. Martínez, V. Ecochard, S. Mahéo et al., "α,β-D-constrained nucleic acids are strong terminators of thermostable DNA polymerases in polymerase chain reaction," *PloS ONE*, vol. 6, no. 10, Article ID e25510, 2011.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 748913, 20 pages
doi:10.1155/2012/748913

Review Article

Challenges and Opportunities for Small Molecule Aptamer Development

Maureen McKeague and Maria C. DeRosa

Department of Chemistry, Carleton University, 1125 Colonel By Drive, Ottawa, ON, Canada K1S 5B6

Correspondence should be addressed to Maria C. DeRosa, maria_derosa@carleton.ca

Received 30 July 2012; Accepted 8 September 2012

Academic Editor: Masayasu Kuwahara

Aptamers are single-stranded oligonucleotides that bind to targets with high affinity and selectivity. Their use as molecular recognition elements has emerged as a viable approach for biosensing, diagnostics, and therapeutics. Despite this potential, relatively few aptamers exist that bind to small molecules. Small molecules are important targets for investigation due to their diverse biological functions as well as their clinical and commercial uses. Novel, effective molecular recognition probes for these compounds are therefore of great interest. This paper will highlight the technical challenges of aptamer development for small molecule targets, as well as the opportunities that exist for their application in biosensing and chemical biology.

1. Aptamers as Molecular Recognition Elements

Historically, nucleic acids were associated with the storage and genetic coding of information and have long been thought to be less complex than proteins [1]. However, like proteins, nucleic acids are able to fold into intricate tertiary structures that have the potential to perform a variety of functions including gene-regulation, catalytic activity and ligand-binding [2]. Interest in these so-called "functional" nucleic acids was prompted by the ever-increasing number of discoveries of non-coding ribonucleic acids (RNAs) displaying catalytic or binding properties [2].

Two decades ago, several researchers revolutionized molecular recognition by developing synthetic RNA motifs that bound specifically to molecular targets [3–5]. These RNA structures, called aptamers, were selected using an *in vitro* selection procedure called systematic evolution of ligands by exponential enrichment (SELEX) [3]. Like antibodies, these synthetically derived molecular recognition probes were found to be selective and able to bind to their targets with high affinity.

Currently, there is a growing need for rapid, robust, and inexpensive methods for sensing and diagnostic purposes [6]. As molecular recognition is the cornerstone of sensing, there has been increased focus on the development of new molecular recognition probes for sensing applications [7]. While antibodies have long been considered to be the standard in molecular recognition and the use of antibodies as recognition probes predates the 1950s, the relatively new technology of aptamers offers several advantages [8]. Firstly, the *in vitro* aptamer selection process allows a greater control over aptamer binding conditions. Nonphysiological salt concentrations, temperatures and pH can be used in successful selections [9]. Due to the robustness of the phosphodiester backbone, aptamers can exhibit an improved stability over their protein-based antibody counterparts. In particular, aptamers can be reversibly denatured by changing the surrounding conditions. For example, a change in pH, temperature, ionic strength, or use of denaturants irreversibly denatures antibodies, while aptamers simply unfold. The aptamer structure can then regain functionality upon return of the original binding conditions [6]. Due to the nucleic acid nature of aptamers, they bind to complementary nucleic acids as well as their targets, which can be exploited in sensing schemes or as "antidotes" *in vivo*. Aptamer generation has been achieved for a wide variety of targets including small molecules [10], proteins [11], viruses [12], and whole cells [13]. Unlike antibodies, aptamers can also

be generated for targets that are toxic as well as for targets that do not elicit an immune response *in vivo* [8]. Once selected, aptamers are manufactured using well-established automated chemical solid-phase synthesis [14, 15]. The accuracy and reproducibility of this procedure allows for a relative ease in producing aptamers at large scales, with very little batch-to-batch variation in activity [16]. Additionally, aptamer sequences can be modified with reporter molecules throughout this solid-phase synthesis; this allows for labeling at judiciously chosen nucleotide positions to minimize any effect on the functionality of the aptamer [17, 18].

Aptamers also offer advantages over other synthetically created molecular recognition systems such as molecular imprinted polymers (MIPs). While MIP synthesis can be simple and cheap, and the resulting MIPs are unaffected by changes in heat and pH [19], MIPs typically display high cross-reactivity [20] and are not particularly amenable to chemical modifications.

Of course, aptamers are not without their disadvantages. Unlike antibodies or MIPS, their tertiary structure is highly dependent on solution conditions, and they are easily degraded in blood. Furthermore, antibodies have a significantly higher chemical diversity with 20 amino acids. However, some of these problems can be addressed, for example, through chemical modifications to increase nuclease resistance or increase the diversity of the nucleic acid pools.

1.1. Systematic Evolution of Ligands by Exponential Enrichment (SELEX). The concept of *in vitro* evolution was first reported in the 1960s with the observation that, in a cell-free system, the RNA genome of the Qβ bacteriophage could be evolved during replication to form RNAs that were more efficiently copied by the viral replicase [21]. Later, they were able to evolve sequences for other traits such as resistance to ethidium bromide [22]. Despite the importance of these early discoveries, however, the true potential of *in vitro* evolution was not realized until several decades later, after the introduction of modern biotechnological advances such as the invention of polymerase chain reaction (PCR), the isolation of reverse transcriptase and the ability to generate long oligonucleotides containing random nucleotide regions using solid-phase synthesis. Equipped with these modern techniques, in 1990, three separate groups reported *in vitro* selection and evolution of functional nucleic acids [23]. Tuerk and Gold [3] used the term SELEX for their process of selecting RNA ligands against T4 DNA polymerase; Ellington and Szostak [4] performed *in vitro* selection to select RNA ligands (for which they coined the term "aptamers") against various organic dyes; Robertson and Joyce [5] evolved the *Tetrahymena* self-splicing intron to carry out a DNA cleavage reaction.

Since its invention, several researchers have performed SELEX to isolate nucleic acids with a wide variety of functions. While several modifications of the procedure have been made by various groups, the general SELEX process remains the same (Figure 1 shows the process for DNA aptamers). Typically, SELEX begins with an initial library (often referred to as "a pool") of random nucleic acid sequences (either RNA or DNA depending on the nature

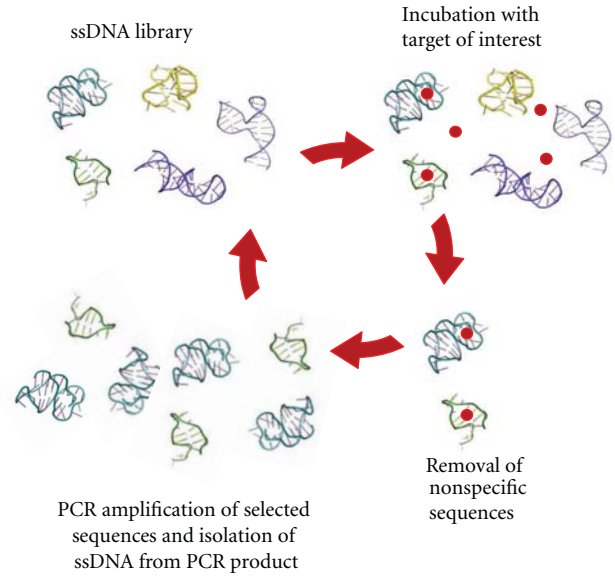


Figure 1: The systematic evolution of ligands by exponential enrichment process (SELEX). Beginning with a large library of DNA, iterative cycles of target incubation, library partitioning, and amplification are performed to select aptamers.

of the research). SELEX libraries ideally consist of 30–80 random nucleotide positions flanked by primer-binding sites necessary for PCR amplification [24]. The library is then incubated with the target of interest and several washing steps are employed to remove nonfunctional sequences. For small molecule targets, the target is usually immobilized onto a solid-support matrix to permit partitioning of binding and nonbinding sequences [25]. The next steps in SELEX include the elution of the binding sequences from the target and the polymerase chain reaction (PCR) amplification (reverse transcription PCR for the RNA aptamers) of those binding sequences to yield an enriched library for subsequent, more stringent, selection rounds [10]. As the interactions that lead to molecular recognition between the binding sequences and the target are noncovalent in nature, mild conditions can be used to separate the two species. Elution using heat, high concentrations of the target molecule, or chaotropic agents, such as urea, can be performed. The strength of the molecular interactions within the target-aptamer complex will dictate the conditions required for elution [7].

Once separated from the target, the few binding DNA sequences are amplified by PCR to yield a practical amount of sample to continue the process. As PCR generates double-stranded DNA and aptamers are single stranded, the DNA aptamer sequence is separated from its complement using one of a number of techniques, such as gel electrophoresis or using an agarose resin [26, 27]. Single-stranded RNA aptamers are generated from the double-stranded DNA PCR products by *in vitro* transcription [10], thus no further processing is required before reintroduction of an enriched RNA pool into the next SELEX round.

SELEX progress can be monitored by modifying the aptamer strand with a traceable label, to determine when

more stringent conditions should be applied [28]. The enriched library generated from a round of selection is subjected to further selection rounds that serve to increase the affinity of the library for the target molecule (positive selections). After several rounds, the enriched library is cloned, sequenced and characterized to isolate aptamers with the desired properties. Once these sequences are elucidated, solid-phase chemical synthesis is used to reproducibly synthesize aptamers in large quantities.

1.2. Adaptations to the SELEX Process. An enormous advantage of SELEX is its flexibility with respect to methodology, binding conditions and library design. The first SELEX modifications introduced the inclusion of negative or counter selection steps to eliminate sequences displaying affinity for either the solid-support matrix or compounds sharing structural similarity to the target. The majority of more recent SELEX modifications involve changing the stringency, the platform on which selection is performed or the type of target [29, 30]. The library used in the selection can also be modified to include fixed regions of known functionality or increase the diversity of structures available for selection either through initial pool design [28, 31, 32], or by inclusion of mutagenic PCR to alter the pool from round to round [33]. The SELEX process has also been automated by several groups [34–36]. Table 1 lists several modifications to the original SELEX process. Regardless of whether the listed SELEX modifications involved changes to target immobilization, nucleic acid library, selection stringency, amplification, or monitoring of the enrichment, the goal of these changes was to either generate improved aptamers or to simplify the SELEX procedure.

2. SELEX Targets

As can be noted from Figure 2, less than a quarter of existing aptamers have been generated for small molecule targets. With the success of the first *in vitro* selection experiments to small organic dyes [4], much of the original SELEX focus was on developing aptamers for small molecules. However, once it was found that aptamers could be easily selected for proteins and cells, new aptamers for small molecules became less prevalent. These larger targets, containing more functional groups and structural motifs, give a higher probability of finding sequences that can interact with the target via hydrogen bonds, electrostatic interactions, and hydrophobic interactions [6].

2.1. Small Molecule-Binding Aptamers. Despite this trend towards larger targets, there are many compelling reasons for pursuing the identification of new small molecule-binding aptamers. Small molecules play key roles in many biological processes due to their ability to diffuse across cell membranes [37]. These targets may be harmful, such as toxins and carcinogens, or beneficial, such as drugs or nutrients. In cells, small molecules serve as cell signaling molecules, pigments, or as part of defense mechanisms [38]. In molecular biology, they can be used as antibiotics or other important drugs

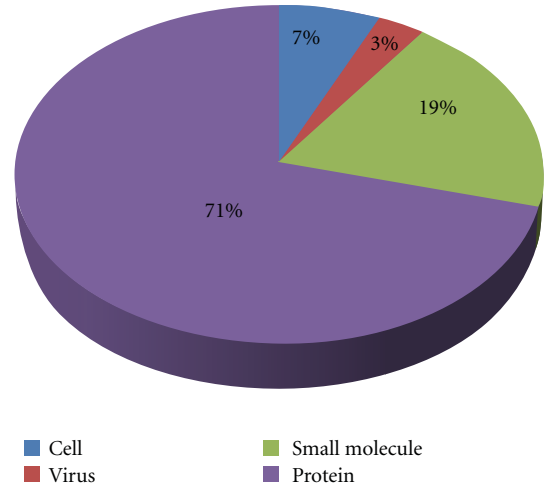


Figure 2: Breakdown, by target type, of aptamers selected between 1990 and 2011. This list was generated using the Aptamer Base [41] http://aptamerbase.semanticscience.org/ (accessed July 9, 2012) (accessed July 9, 2012).

[39]. In the food industry, small molecules are important for energy storage or can act as pesticides [40]. Aptamers for small molecules may be applied to a wide variety of applications in medicine, agriculture, and environmental analysis. Tables 2 and 3 list the small molecule targets for which DNA and RNA aptamers, respectively, have been characterized.

2.2. Conceptual Challenges for Small Molecule Aptamers. Although they represent a minor proportion of all aptamers, small molecule-binding aptamers are among the most successful and widely studied aptamers in the literature. For example, the ATP aptamer is second only to the thrombin aptamer in terms of the number of publications using the sequence in an aptamer-based assay, sensor, or biosensor in the last ten years. The cocaine and theophylline aptamers are the fifth and seventh most frequently used aptamers for biosensing, respectively [42]. It has already been described that aptamers are ideal molecular recognition probes for small molecules [8, 43], based on their ability to achieve a remarkably high degree of selectivity. The first example of this unparalleled selectivity was observed in 1994, when the selected RNA aptamer for theophylline displayed a 10,000 times weaker binding affinity to caffeine, a xanthine that differs by a single methyl group. This selectivity was found to be a 10-fold improvement on the selectivity for the antibodies for these targets [44]. Several groups have also exploited the ability of aptamers to distinguish between small molecule enantiomers [45]. Initially, several RNA aptamers displayed partial discrimination between various L and D amino acids [46, 47]. Then, in 1996, Geiger et al. [48] reported the selection of RNA aptamers that bound to L-arginine with high affinity and enantioselectivity. More recently, enantioselective DNA aptamers have been selected for the small molecule drug (R)-thalidomide [49] and separate aptamers have been identified for (S) and (R)-ibuprofen [50].

One possible explanation for the scarcity of new small molecule aptamers is the impression that aptamers cannot

Table 1: A list of modifications to the SELEX process and their descriptions.

Method	Description	Reference
Atomic force microscopy (AFM)-SELEX	AFM-SELEX uses a dynamic atomic force microscopy tip to pick up and visualize aptamer-target complexes. This SELEX requires only one round of selection.	[99]
Automated SELEX	This SELEX uses automated systems for the procedure to reduce the time and labour required.	[34]
Blended SELEX	In this technique, a lead chemical compound is attached covalently or non-covalently to a nucleic acid library. Each nucleic acid conjugate in the starting library is a variant of the chemical compound moiety and allows up to 10^{15} variants of the small molecule to be screened for the most active of these composite assemblies.	[100]
Cell-SELEX	Cell-SELEX generates aptamers that can bind specifically to a cell of interest. Commonly, a cancer cell line is used as the target to generate aptamers that can differentiate that cell from other cancers or normal cells.	[101]
Capillary electrophoresis (CE)-SELEX	The separation of bound and nonbound oligonucleotides is performed using capillary electrophoresis.	[102]
Chimeric SELEX	Chimeric SELEX uses two or more different oligonucleotide libraries for production of chimeric aptamers with more than one wanted feature or function. Each of the parent libraries will be selected first to a distinct feature; the resulting aptamers are then fused together.	[103]
Conditional SELEX	This SELEX uses regulator molecules during the selection, thus, allowing aptamer binding to the target to be regulated.	[104]
Counter selection/ subtractive SELEX	This technique employs additional rounds of SELEX to remove sequences that bind to similar target structures.	[44]
Covalent/ Crosslinking SELEX	This process is used to select aptamers that contain reactive groups which are capable of covalent linking to a target protein.	[105]
Deconvolution SELEX	Deconvolution SELEX is used to generate aptamers for complex targets. Typically selection is performed on mixtures (or a cell). Once aptamers have been generated, a second part of SELEX involves discriminating which aptamers bind to which parts of the complex mixture.	[106]
Electrophoretic mobility shift assay (EMSA)-SELEX	The partitioning step of SELEX occurs through the use of electrophoretic mobility shift assay (EMSA) at every round.	[107]
Expression cassette SELEX	This is a special form of blended SELEX that involves transcription factors and optimizes aptamer activity for gene therapy applications.	[108]
Fluorescence-activated cell sorting (FACS) SELEX	This SELEX makes use of fluorescence-activated cell sorting to differentiate and separate aptamer-bound cells.	[59]
FluMag SELEX	Here the library is modified with fluorescein instead of radiolabels for quantification purposes. Additionally, the target is immobilized to magnetic beads instead of agarose.	[109]
Genomic SELEX	The SELEX library is constructed from an organism's genome and target proteins and metabolites from the same organism are used to elucidate meaningful interactions.	[110]
In vivo SELEX	*In vivo* SELEX uses transient transfection in an iterative procedure in cultured vertebrate cells to select for RNA-processing signals.	[111]
Indirect SELEX	The target used in the selection is not the aptamer binder; however, it becomes required for aptamer binding to the new target.	[112]
Mod-SELEX	Mod-SELEX uses a library of oligonucleotides with chemical substitutions that result in nuclease-resistant aptamers.	[113]
Multivalent aptamer isolation (MAI) SELEX	This process is used to generate aptamer pairs for a given target.	[114]
Microfluidics SELEX	This SELEX uses microfluidic technologies, creating an automatic, and miniature SELEX platform for fast aptamer screening.	[115, 116]
Monolex	Monolex involves a single affinity chromatography step, followed by physical segmentation of the affinity material, to obtain the highest affinity aptamers.	[117]

Table 1: Continued.

Method	Description	Reference
Multiplexed massively parallel SELEX	This allows analysis of large numbers of transcription factors in parallel through the use of affinity-tagged proteins, bar-coded selection oligonucleotides, and multiplexed sequencing.	[118]
Multi-stage SELEX	Multistage SELEX is a modified version of chimeric selex. Here, the fused aptamer components then go through an additional selection with all the targets.	[119]
Negative selection	An additional step, performed typically at the beginning of selection, removes sequences that have an affinity for the selection matrix.	[48]
Next generation SELEX	This SELEX uses designed oligonucleotide libraries that tile through a pre-mRNA sequence. The pool is then partitioned into bound and unbound fractions, which are quantified by a two-color microarray.	[120]
Non-SELEX (NCEEM)	This process involves repetitive steps of partitioning with no amplification steps.	[121]
Photo SELEX	Aptamers bearing photo-reactive groups that can photo cross-link to a target and/or photo activate a target molecule are used.	[122]
Primer-free SELEX	This SELEX involves removal of the primer-annealing sequences from the library prior to selection, preventing unwanted primer-based secondary structures.	[123]
Serial analysis of gene expression (SAGE) or high-throughput SELEX	SAGE SELEX links oligomers from SELEX with longer DNA molecules that can be efficiently sequenced.	[124]
Spiegelmer technology	The aptamer selection is performed with the natural D-nucleic acids but on the opposite enantiomer of the chiral target molecule. After sequencing, the aptamers are synthesized as L-isomers for binding to the desired enantiomer of the target.	[125]
Slow off-rate modified aptamers (SOMAmer)	The selection is performed with oligonucleotide libraries that are uniformly functionalized at the 5′-position resulting in high-quality aptamers.	[28]
Tailored SELEX	This is an integrated method to identify aptamers with only 10 fixed nucleotides through ligation and removal of primer binding sites within the SELEX process.	[126]
Target expressed on cell surface (TECS) SELEX	Recombinant proteins on the cell surface are used directly as the selection target.	[127]
Tissue-SELEX	This method is for generating aptamers capable of binding to tissue targets.	[106]
Toggle-SELEX	The selection is performed on different targets in alternating rounds.	[128]
Yeast Genetic SELEX	This method optimizes *in vitro* selected aptamers by creating a library of degenerate aptamers and performing a secondary selection *in vivo* using a yeast three (one)-hybrid system.	[129]

bind these smaller targets with the high affinity required for most sensing applications. Work by Carothers et al. attempted to determine the effect of target structure and size on binding affinity [51]. Using aptamers for 6 small molecule targets from the literature, as well as aptamers obtained from his own selections for two other small molecules, Carothers determined that the target molecular weight was proportional to the resulting aptamer affinity (larger targets resulted in lower K_d values). This finding was consistent with the findings of other studies between affinity and target mass [52]. However, the target theophylline which has a very small mass (180 g/mol) did not follow this general trend. It was therefore concluded that targets with fewer rotatable bonds, and therefore fewer degrees of freedom, can result in improved aptamer affinity. Nevertheless, while many aptamers that bind to small molecule targets display affinities in the low to mid micromolar range, there are several aptamers that have recently been isolated with K_d values in the low nanomolar range (e.g., BPA [53] and oxytetracycline [54]). Furthermore, riboswitches, which are widely considered as containing "natural aptamers," bind exclusively to small molecules and ions, and several of these display remarkably strong binding. For example, the guanine riboswitch has a K_d of 5 nM [55] and the thiamine pyrophosphate-sensing riboswitch has an affinity in the picomolar range [56]. The glycine riboswitch is particularly noteworthy for its ability to selectively bind to one of the smallest target of any natural or artificial aptamer [57].

Table 2: A listing of DNA aptamers reported in the open literature* (up until July 2012) that have been confirmed to bind to small molecule targets. The dissociation constant (K_d), a measure of binding affinity, is included as well as the year of aptamer development.

Target	Binding affinity (K_d)	Year	Reference
Reactive green 19	33 μM	1992	[130]
Adenosine monophosphate and adenosine triphosphate	6 μM	1995	[131]
L-arginine	2.5 mM	1995	[132]
L-argininamide	0.25 mM	1995	[132]
Anionic porphyrins	0.4–4.9 μM	1996	[88]
Sulforhodamine B	190 nM	1998	[61]
Cellobiose	600 nM	1998	[133]
7,8-dihydro-8-hydroxy-2′-deoxyguanosine	270 nM	1998	[134]
Cholic acid	5–67.5 μM	2000	[135]
Hematoporphyrin	1.6 μM	2000	[136]
L-tyrosinamide	4.5 μM	2001	[137]
Sialyllactose	4.9 μM	2004	[138]
Ethanolamine	6–19 nM	2005	[139]
(R)-thalidomide	1 μM	2007	[49]
Hoechst derivative 7e	878 nM	2007	[140]
17β-estradiol	0.13 μM	2007	[141]
Lys-Arg-Azobenzene-Arg	0.33 μM	2007	[142]
Tetracycline	64 nM	2008	[143]
L and D arginine	580–810 μM	2008	[144]
Daunomycin	10 nM	2008	[145]
Oxytetracycline	10 nM	2008	[54]
Ochratoxin A	200 nM	2008	[26]
Dopamine	700 nM	2009	[146]
8-hydroxy-2′-deoxyguanosine	100 nM	2009	[147]
Diclofenac	42.7–166.34 nM	2009	[148]
(S) and (R)-ibuprofen	1.5–5.2 μM	2010	[50]
Adenosine triphosphate	3.7 μM	2010	[31]
Fumonisin B_1 (FB_1)	100 nM	2010	[7]
Acetamiprid	4.98 μM	2011	[149]
Kanamycin	78.8 nM	2011	[150]
L-tryptophan	1.757 μM	2011	[151]
Bisphenol A	8.3 nM	2011	[53]
Ochratoxin A	96–293 nM	2011	[152]
Phenylphosphonic dichloride	>50 μM	2011	[153]
Organophosphorus pesticides (phorate, profenofos, isocarbophos and omethoate)	0.8–2.5 μM	2012	[154]
Polychlorinated biphenyls (PCB77)	4.02, 8.32 μM	2012	[155]
Polychlorinated biphenyls (PCB72 and PCB106)	60–100 nM	2012	[156]
Ampicillin	9.4–13.4 nM	2012	[157]

*Only aptamer sequences that have experimentally determined K_d values were included in this table.

Nature's effectiveness at developing small molecule aptamers should provide an indication that there is considerable untapped potential in this field.

In the early 1990s, in an effort to promote the power of SELEX, numerous papers and reviews boasted that *in vitro* selection is facile, inexpensive, and fail-safe, which may have contributed to little interest in publications for new selections. On the contrary, SELEX can be very laborious and it has been estimated that less than 30% of selections result in aptamers [28]. Additionally, patents for virtually every application of aptamers have placed a stranglehold on aptamer innovation [58]. As a result, very few research groups have chosen to invest the time and expense to develop aptamers for new small molecule targets, especially considering the unique technical challenges that arise when selecting for small molecule binding aptamers, as is discussed in the next section.

3. Technical Challenges for Small Molecule Aptamers

3.1. Target Immobilization. The separation of target-bound sequences from those with no affinity for the target is a critical step in the SELEX process. For protein targets,

Table 3: A listing of RNA aptamers reported in the open literature* (up until July 2012) that have been confirmed to bind to small molecule targets. The dissociation constant (K_d), a measure of binding affinity, is included as well as the year of aptamer development.

Target	Binding affinity (K_d)	Year	Reference
Organic dyes	100–600 μM	1990	[158]
D-tryptophan	18 μM	1992	[46]
L-valine	2.9 mM	1994	[159]
Theophylline	100 nM	1994	[44]
Cyanocobalamin	88 nM	1994	[160]
L-citrulline	62–68 μM	1994	[47]
Flavin mononucleotide	0.5 μM	1994	[161]
Flavin adenine dinucleotide	137–273 μM	1994	[161]
Kanamycin A	$\leq$300 nM	1995	[162]
Neomycin	100 nM	1995	[163]
Tobramycin	2–3 nM	1995	[164]
Lividomycin	$\leq$300 nM	1995	[162]
Nicotinamide adenine dinucleotide	2.5 μM	1995	[165]
Riboflavin	1–5 μM	1995	[165]
Biotin	5 μM	1995	[166]
L-arginine	330 nM	1996	[48]
Dopamine	2.8 μM	1997	[167]
7-methyl-guanosine	5 μM	1997	[168]
CCdApPuro	10 nM	1997	[169]
Chloramphenicol	25–65 μM	1997	[170]
Viomycin	11–21 μM	1997	[171]
Sulforhodamine B	310 nM	1998	[172]
Streptomycin	1–10 μM	1998	[173]
L-isoleucine	200–500 μM	1998	[174]
7,8-dihydro-8-hydroxy-2′-deoxyguanosine (8-oxodG)	0.27–2.8 μM	1998	[134]
Xanthine	3.3 μM	1998	[175]
Guanine	1.3 μM	1998	[175]
Malachite green	$\leq$1 μM	1999	[176]
Phosphatidylcholine	$\geq$100 μM	1999	[177]
Cyclic adenosine monophosphate	10 μM	2000	[178]
Adenosine triphosphate	127–223 μM	2000	[179]
L-tyrosine	35 μM	2000	[180]
S-adenosyl homocysteine	0.2–0.8 μM	2000	[181]
Neomycin	1.8 μM	2000	[182]
Moenomycin A	300–400 nM	2001	[183]
Sialyl Lewis X	0.085–10 nM	2001	[184]
Tetracycline	1 μM	2001	[185]
Kanamycin B	180 nM	2001	[186]
Adenine	10 μM	2002	[187]
Flavin adenine dinucleotide	50 μM	2002	[188]
L-isoleucine	1–7 mM	2003	[189]
Adenosine triphosphate	2 μM	2003	[190]
Morpholine-based GTP analog	20, 33 μM	2003	[191]
4,4′-methylenedianiline	0.45–15 μM	2004	[192]
Tobramycin	30–100 nM	2004	[34]
Kanamycin	10–30 nM	2004	[34]
Adenosine triphosphate	5 μM	2004	[193]
Isoleucine	0.9 mM	2005	[194]
L-histidine	8–54 μM	2005	[195]
Codeine	2.5–4 μM	2006	[33]

Table 3: Continued.

Target	Binding affinity (K_d)	Year	Reference
Mesomesoprotoporphyrin IX	188–445 nM	2006	[87]
Thyroxine	50 μM	2007	[196]
Tobramycin	16 μM	2007	[197]
10-carboxy-2,7-di-t-butyl-trans-12c,12d-dimethyl-12c,12d-dihydrobenzo[e]pyrene	2.7 μM	2007	[198]
Dimethylindole red	87 nM	2008	[199]
Cyanine 3 dye	60 μM	2010	[200]
Aniline-substituted sulforhodamine analogue	3.5 μM	2010	[201]
Atrazine	2 μM	2010	[202]
Sphingosylphosphorylcholine	20–250 nM	2010	[203]
Black hole quencher	4.7 μM	2011	[204]
4-dimethylaminobenzylidene imidazolinone	464 nM	2011	[205]
Glutathione	41.8, 48.9 nM	2011	[206]
Heteroaryldihydropyrimidine	50 nM	2011	[207]

*Only aptamer sequences that have experimentally determined K_d values were included in this table.

partitioning can be achieved using a matrix that selectively adsorbs the target and any interacting aptamer sequences. For example, nitrocellulose filters are a cheap and convenient matrix for this purpose due to their nucleic acid permeability and their ability to retain proteins by hydrophobic adsorption. With cell targets, partitioning can be accomplished by centrifugation, fluorescence activated cell sorting (FACS), [59] or by gentle washing of adherent cells [13]. In the case of both these target types, the selection can be accomplished without chemical modification of the target; this is ideal since it increases the likelihood of finding aptamers capable of binding the molecule in its unaltered form. This is typically not possible with small molecule aptamer selections. Thus, the primary complication arises from the need to immobilize the target to a solid support matrix, for example, magnetic beads, acrylic beads, agarose/sepharose, to facilitate the partitioning process. Early small molecule aptamers were selected for targets for which premade agarose material was commercially available [10]. In the absence of commercially available material, there is a wide array of conjugation chemistries that are available for preparing these materials for SELEX experiments. However, these are all dependent on the presence of certain functional groups that allow for coupling, which are not always present on the desired target. For cases where conjugation is possible, the proportionally large amount of column material, in comparison to the target, that is presented to the nucleic acid pool during each SELEX round can result in high nonspecific binding of the library. As chemical modification of the target is required to facilitate column immobilization, the library is exposed to chemically-modified target rather than the desired, unmodified target molecule, increasing the likelihood of selecting sequences that display binding properties towards the matrix and/or the linker arm. Despite negative selection steps, carry-over of such sequences is difficult to avoid [60]. Many aptamer applications, particularly those *in vivo*, require the selected sequences to bind the target free in solution. Therefore, any aptamer affinity derived from partial binding to the matrix or from chemical modifications will reduce the functionality of the aptamer in the intended applications. For example, the published rhodamine aptamer displays a weaker binding to the target rhodamine when in solution compared to when it is immobilized on the matrix used in the selection [61].

3.2. Measurement of Binding Affinity (K_d). While new methods for the determination of binding affinity are constantly being developed, this is often the limiting factor in the rapid development and testing of aptamers. This is particularly true for small molecule binding aptamers. To measure K_d, a constant concentration of either the aptamer or target is titrated with an increasing concentration of the other component to yield a binding isotherm. A list of common methods for determining aptamer binding affinity can be found in the Table 4. A brief evaluation of their applicability to small molecule binding aptamers is also provided below.

As can be seen in Table 4, relatively few of these common K_d methods are effective for measuring aptamer binding to small molecules. Separation-based techniques are among the most common approaches for determining binding affinity, and many of these are more challenging for small molecule targets than for proteins. In particular, separation-based methods that rely on a dramatic change in the size of the aptamer-target complex upon target binding are of limited use when the target is much smaller than the aptamer. Other methods require that the target has some intrinsic fluorescence/absorbance, which is often not the case. For targets lacking these properties, an alternative is to label the target, which can affect the chemical properties of the small molecule and interfere with aptamer binding. Surface mass-sensitive detection methods such as QCM and SPR are typically limited to large targets such as proteins [62]. These approaches generally require one binding partner to be tethered to the surface. In cases where the aptamer is surface-bound, the sensitivity of the technique may be compromised by the small overall mass change caused by small molecule binding. As an alternative, the target could be attached to the surface, but once again this chemical modification of the target can negatively impact binding affinity. Other

Table 4: Methods for determining aptamer binding affinity.

Method	Description of method and applicability to small molecules	Sample reference
	Spectroscopy-based methods	
Fluorescence intensity	The fluorescence of the aptamer or target may be quenched or increased upon binding. This method requires a fluorescent small molecule target or requires labelling of the target.	[208]
Fluorescence polarization	A fluorophore is excited with polarized light and, due to rotational diffusion, the size of the fluorophore will dictate the proportion of polarized light that is emitted. This method requires a fluorescent small molecule target or target labelling. It can be used with a fluorescently tagged aptamer, however, the method is less sensitive as the overall change in mass upon binding a small molecule will be less dramatic.	[209]
UV-vis absorption	This method requires a change in intensity or wavelength of absorption in either the aptamer or target's UV-vis spectrum. In some cases, melting studies can be used to determine K_d.	[210]
Circular dichroism (CD)	CD refers to the differential absorption of left and right circularly polarized light. Upon aptamer binding to the target, the CD spectra may change but a significant difference in conformation upon target binding is required for this method to have good sensitivity.	[211]
Nuclear magnetic spectroscopy (NMR)	By comparing the heteronuclear single quantum coherence spectroscopy (HSQC) of individual amide protons in the free and bound aptamer, it is possible to observe changes in the chemical shifts of the peaks. This method requires conformation changes in the aptamer for good sensitivity.	[212]
	Mass-sensitive surface-based measurements	
Surface plasmon resonance (SPR)	Either the target or aptamer can be coupled to a chip; by flowing various concentrations of the nontethered ligand, changes in refractive index can be measured as the aptamer-target complex forms. If the small molecule target is immobilized, its ability to bind to the aptamer may be compromised. Immobilization of the aptamer, however, leads to a less sensitive measurement as the smaller target will cause less of a change at the surface.	[33]
Quartz crystal microbalance (QCM)	This method uses piezoelectric crystals to correlate the mass accumulated (target binding) on the surface with a decrease of the resonance frequency of the quartz crystal. Once again, small molecule target immobilization could affect binding affinity. Immobilization of the aptamer leads to a less sensitive measurement because there will less of a mass change upon target binding.	[213]
	Separation-based methods	
High-performance liquid chromatography (HPLC)	Zone separations of the free aptamer, target, and aptamer-target complex can be used to assess the equilibrium distribution of these components. This method is particularly difficult with small molecule targets as they have less of an effect on the separation of aptamer-target complex from the free aptamer.	[214]
Capillary electrophoresis (CE), kinetic capillary electrophoresis (KCE), affinity probe capillary electrophoresis (APCE)	This method is similar to HPLC except that it using an electric field to separate the components of the mixture by size and charge. Small molecule targets can be a challenge, typically requiring labeling of the small molecule although label-free KCE UV has recently been described. Once again, separation of the aptamer-target complex from the free aptamer can be more difficult in the case of small molecule targets.	[64, 215, 216]
Microfree-flow electrophoresis (μ-FFE)	This technique separates aptamer and aptamer-target complex based on their electrophoretic mobilities. Sample is continuously streamed into a planar flow channel while an electric field is applied perpendicularly to the direction of flow, deflecting analyte streams as they travel through the flow channel according to their mobility. Once again, this method is less effective with small molecule targets.	[217]
Equilibrium dialysis	Equilibrium dialysis allows the aptamer, target and the complex to equilibrate in a two compartment cell separated by a semipermeable membrane that allows only the smallest component to pass through. This method can be hampered by nonspecific adsorption of small molecule targets to the membrane.	[26]

Table 4: Continued.

Method	Description of method and applicability to small molecules	Sample reference
Ultrafiltration/nitrocellulose filtration	This method is similar to dialysis. The aptamer and target are incubated to allow binding. The fraction of the smallest unbound component is forced through a filter and measured. Once again, nonspecific adsorption to the membrane can cause this method to be unreliable.	[218]
Affinity chromatography	Either the target or aptamer is covalently immobilized to a solid-phase support. The other component is incubated with the support and the amount of binding is calculated. As with other methods, chemical modification of the target or the aptamer to allow for immobilization can affect binding.	[7]
Electrophoretic mobility shift assay (EMSA)	The presence of the target will cause an increase in molecular weight of the aptamer-target complex, resulting in a change in electrophoretic mobility and a gel shift. This approach is not effective with small molecule targets unless a significant conformational change is observed upon binding.	[219]
Optical thermophoresis	Based on the directed movement of molecules along temperature gradients, the thermophoresis of an aptamer typically differs significantly from that of an aptamer-target complex because of changes in size, charge, or solvation energy. This method requires fluorescent labelling which could affect binding. Also, it could be less sensitive for small molecule aptamers due to the smaller change in mass upon target binding.	[220]
Other methods		
Isothermal titration calorimetry (ITC)	This method allows simultaneous determination of K_d, stoichiometry, and thermodynamic properties. It relies on the fact that formation of the aptamer-target complex is an exothermic process. Effective for small molecule aptamers, particularly if a large conformational change occurs upon target binding.	[221]
High-throughput affinity quantitative PCR binding assay	With this method, an aptamer duplex is incubated with the target. The concentration of aptamer released by this binding event is then measured using real time PCR.	[222]
In-line probing	Spontaneous cleavage of the RNA backbone is affected by local structural characteristics, which in turn are impacted by target binding. Can be effective for small molecule aptamers but requires conformational changes upon target binding and is only applicable to RNA aptamers.	[223]
Footprinting assays	This method determines the region of aptamer sequence where target binding occurs by exploiting that the target may protect the aptamer from enzymatic cleavage/chemical reactions. Footprinting assays are easier with larger targets or require conformational changes with target binding.	[224, 225]

methods that detect a change in aptamer conformation upon binding to target can be applicable to small molecule aptamers, but measurable conformation change is not a universal property of all aptamers [63]. Some recent reports of approaches for determining aptamer binding affinity have recognized the unique challenges for small molecule aptamers and attempted to address them using more innovative approaches such as automated microchip electrophoresis and atomic force spectroscopy, although no technique can be considered generally applicable to small molecule aptamers at this stage [64, 65].

4. Opportunities for Small Molecule Aptamers

While aptamer technology has existed for over two decades, the challenges imposed on the development of aptamers for small molecules has resulted in very few novel aptamers that can bind to practical small molecular targets. Nevertheless, there are many opportunities for the innovative application of small molecule binding aptamers in biosensing and chemical biology. A diverse range of natural and synthetic compounds fall under the designation of small molecules, including organic compounds, amino acids, steroids, carbohydrates, and nucleotides. These molecules play a variety of beneficial roles; they are therapeutics, dyes, cofactors, metabolites, and neurotransmitters. Unfortunately, they may also be harmful substances, such as pollutants, food adulterants, carcinogens, and drugs of abuse. There are many reasons why effective tools for the detection of small molecules are needed now more than ever before. These include the growing recognition of the role of small molecules in biological systems, the extensive application of synthetic small molecules as drugs, and the increasing need to monitor contamination our environment and food supply. Many examples of aptamer-based sensing approaches have been designed for the detection of small molecules,

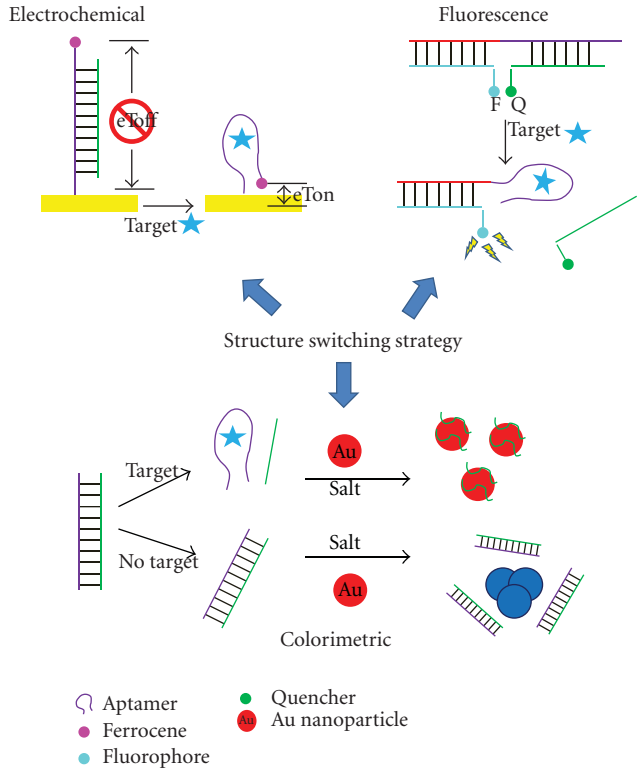


FIGURE 3: Illustration of sample electrochemical, fluorescence, and colorimetric assays using the structure switching strategy and small molecule-binding aptamers.

however, most are proof-of-concept systems using the ATP, cocaine, or theophylline aptamer. Electrochemical, fluorescence, colorimetric, and other approaches that have been employed for aptamer-based sensing have been highlighted in several recent reviews [66–69]. However, the general applicability of many of these biosensors to small molecule targets, other than the three main targets mentioned, has yet to be confirmed. As shown in Tables 2 and 3, however, there are many small molecule aptamer sequences that are available for biosensor development. Several of these targets are relevant to applications in environmental monitoring, agriculture, food safety, and medicine [70]; this is a clear motivation to move aptamer-based biosensing past the few proof-of-concept systems and to validate aptamer-based sensing approaches for real-world bioanalytical applications.

One notable approach to small-molecule biosensing has been the "structure-switching" strategy, described by Li and colleagues, which relates to the unique ability of nucleic aptamers to bind to both their cognate target and to a complementary sequence [71]. This ability to switch from a nucleic acid duplex to an aptamer-target complex, and its concomitant structural changes, have been shown to be a generally applicable method for converting a recognition event into detectable signal. As such, it has been applied to several fluorescence, electrochemical, or colorimetric (nanoparticle-based) assays (see Figure 3). This strategy is particularly appealing given that a large conformational change upon target binding is virtually guaranteed, thus many of the typical limitations encountered in small molecule detection are avoided. Both DNA and RNA aptamers can be employed in this approach [72] and it has been used in solution and on surfaces [73]. Although many reports use this methodology for the sensing of adenosine or ATP [71], this method has been applied to detect several other small molecule targets including theophylline [72], cocaine [74], histidine [75], OTA [76], L-argininamide [77], tyrosinamide [78], GTP [79], and arginine [80].

Recent reviews have examined how aptamers can be combined with other functional moieties without affecting their ability to recognize and bind to their cognate target [81, 82]. This property can allow aptamers to serve as regulatory elements for nucleic acid enzymes (either natural ribozymes or synthetic DNAzymes) or other actuator parts, allowing control of a variety of functions such as gene-expression regulation [83]. Small molecule-binding sequences are particularly convenient for the preparation of these chimeras. This is perhaps not surprising considering that natural riboswitches, which are known to be key regulators of several biosynthetic pathways, contain an aptamer domain that serves as a high affinity sensor for a specific small molecule. In one of the original examples of these chimeric systems, one of the stem-loop sections of the widely studied hammerhead ribozyme was replaced with the ATP-binding aptamer. As a result, ATP binding was required for activation [84]. A similar approach can be applied to the development of allosteric aptamers for sensing. Conjugation of the malachite green aptamer to the flavin mononucleotide (FMN) aptamer created a FMN sensor where binding of malachite green and the concomitant increase in fluorescence could only be achieved after FMN bound to its aptamer domain [85]. ATP and theophylline sensors were also made using the same approach. This strategy has been suggested for the fluorescence detection of cellular metabolites; a combination of the endogenously expressed aptamer conjugate and cellular dyes could enable intracellular detection. Thus, the continued development of biologically relevant small molecule aptamers for these chimeric systems is important in order to enable further advances in areas such as *in vivo* imaging and synthetic biology.

Small molecule-binding-aptamers can also serve as important tools for elucidating the role that biologically-important small molecules play in modulating critical cellular regulatory circuits. Intramers, aptamers specific for intracellular target proteins, have already been shown to be useful tools for probing important protein-based networks *in vivo* [86]. By perturbing the intracellular pools of physiologically important small molecules, small-molecule-binding aptamers also have the potential to improve our understanding of biological systems. This strategy has been explored by Marletta using intracellular expression of a heme-binding RNA aptamer to predictably modulate *E. coli* heme biosynthesis as a model for a product feedback inhibited system. This group demonstrated that *in vitro* selected, heme-binding RNA aptamers could specifically sequester intracellular heme when expressed *in vivo* and perturb the heme-mediated inhibition of the heme biosynthetic pathway in a measurable way [87]. Using heme-binding DNA

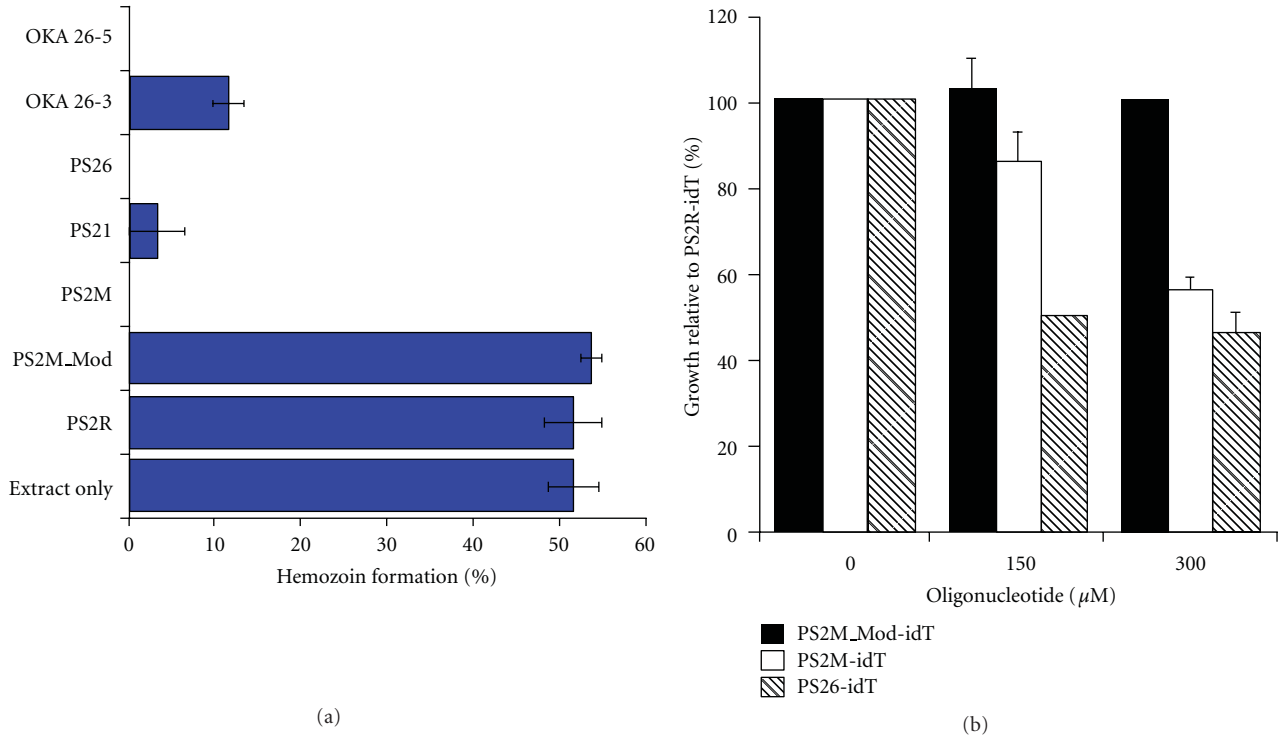


Figure 4: (a) Confirmation of aptamer-mediated inhibition of hemozoin formation within parasite lysates. Heme-binding DNA aptamers (OKA 26-5 and 26-3; PS26, PS21, and PS2M) inhibit hemozoin formation but control oligonucleotides (PS2M_Mod and PS2R) have no effect. (b) The growth of parasites incubated in red blood cells that had been preloaded with nuclease resistant DNA aptamers (PS2M_idT and PS26-idT) is significantly inhibited in comparison to those exposed to red blood cells loaded with control oligonucleotides (PS2M_Mod-idT). Used with permission from PNAS.

aptamers [88], Marletta and colleagues also demonstrated that aptamers can successfully control an essential, *P. falciparum*-specific metabolic pathway [89]. In this study, the pathway of interest was the parasite's heme detoxification pathway, a key target for malaria controls. Hemoglobin ingested by the parasite is degraded in vacuolar structures at low pH (between 4.5 and 5.5), releasing free heme. This free heme is cytotoxic, thus, heme detoxification by polymerization into hemozoin is a critical process for plasmodia survival. The heme-binding DNA aptamers were shown to interfere with hemozoin formation in two *in vitro* assays, one using a model lipid-catalyzed system and the other using parasite-derived lysates containing the native hemozoin formation components. IC_{50} values for inhibition by the aptamers were comparable to that of chloroquine, a known inhibitor of hemozoin production. Additionally, when preloaded into red blood cells, nuclease resistant heme-binding aptamers induced parasite toxicity in a manner consistent with inhibiting the hemozoin production process in early stage parasites (see Figure 4). These examples show the potential utility of small molecule-binding aptamers as new chemical tools for probing biophysical processes as well as their potential use as leads in antimalarial drug development.

Recently, the *in vivo* utility of a dopamine-binding DNA aptamer as a tool to investigate neurobiological processes was demonstrated in a preclinical animal model of schizophrenia [90]. In a similar fashion as the heme example, the role of the dopamine aptamer was to sequester excess dopamine in a particular brain region (the nucleus accumbens) and to monitor whether an abnormal behavior could be reversed as a result. The drug MK-801, an N-Methyl-D-aspartate (NMDA) receptor antagonist, has been used to model the cognitive dysfunction observed in individuals with schizophrenia. In this schizophrenia model, rats are trained to press a bar for chocolate pellet rewards. After 5 days of training, the rats undergo an "extinction" test, where no chocolate is presented upon lever pressing. While normal rats quickly realize this and stop pressing the lever, rats under the influence of MK-801 show a cognitive defect known as perseveration, meaning that they continued to press the lever at a high rate and are unable to inhibit this behavioral tendency. This defect has been linked to high levels of dopamine in the nucleus accumbens region of the brain. The ability of a dopamine-binding DNA aptamer to reverse these MK-801-induced cognitive deficits when injected directly into the nucleus accumbens was determined. Figure 5(a) shows that injection of the dopamine aptamer (filled triangles) reversed the MK-801-induced elevation in lever pressing to levels as seen in rats not treated with MK-801 (empty squares, X with dashed line). Injection of buffer (empty diamonds) or of a random oligonucleotide (dashed line) had no effect on moderating the pressing behavior. Thus, it appears that the aptamer was successful in sequestering the excess dopamine within the nucleus accumbens, resulting

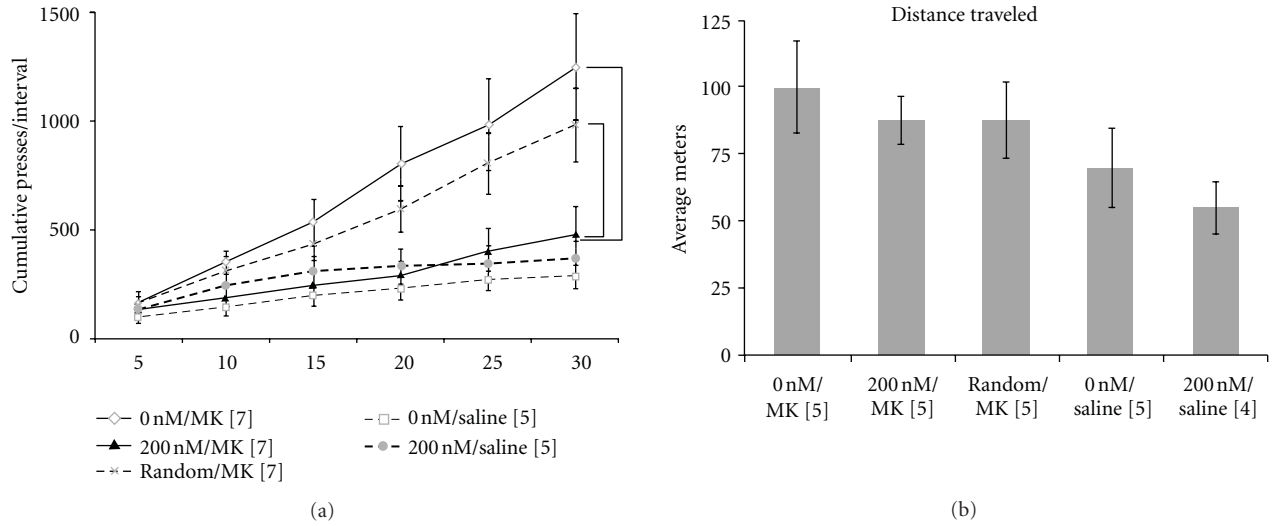


Figure 5: (a) Pretreatment of animals with the dopamine aptamer reversed the effects of MK-801 administration. Animals given MK-801 (empty diamonds; 0 nM/MK) show higher cumulative presses in this behavioural test in comparison to animals not given this drug (empty squares, dashed line; 0 nM/Saline). The group receiving aptamer pretreatment (filled triangles; 200 nM/MK), however, showed similar levels of cumulative presses as those that were not given any MK-801. A random oligonucleotide pretreatment, however, had no dampening effect on the number of presses (X with dashed line; Random/MK). (b) Aptamer pretreatment (200 nM/MK) did not significantly affect locomotor activity as measured by distance traveled in an elevated cross maze. Used with permission from PLoS One.

in "normal" behavior. Interestingly, it was also shown that the aptamer treatment did not impair locomotor activity in the animals (Figure 5(b)). The near-selective effect of the aptamer on reversing cognitive deficits without drastic negative motor consequences lends support for the use of DNA aptamers in the further study of preclinical animal models of mental health disease and as possible drug leads. New neurotransmitter-binding aptamers, as well as strategies for the delivery of aptamers across the blood-brain-barrier, will be required to realize the full potential of aptamers in this regard.

5. Conclusions and Outlook

The selection and application of small molecule-binding aptamers come with a unique set of challenges that have hampered their research and commercialization. Nevertheless, efforts to expand the suite of aptamers for pertinent small molecules need to continue. A major bottleneck in small molecule aptamer development and application occurs at the point of K_d determination. Approaches for binding affinity measurement that are more generally applicable to small molecule targets are required. Additionally, there is a negative perception that there are published aptamers that have little affinity to their targets [91]. Because of these issues associated with K_d measurements for small molecules, several complementary methods may be required in order to achieve a true sense of the aptamer binding affinity for its target. Ideally, these methods should allow aptamer-target binding to occur free in solution, to remove any contribution from matrix binding. A new aptamer database is available (http://aptamerbase.semanticscience.org/) containing valuable information about the experimental conditions under which the aptamers were selected and their binding affinity quantified [41]. Efforts to mine this data to better tailor SELEX and binding affinity experiments for small molecules is currently underway. Aptamer-based assays also need to begin to move away from the proof-of-concept targets and exploit the largely untapped resource of existing small molecule-binding aptamers. Signs of this shift to more relevant aptamer-based assays can be seen with the large number of biosensors developed using the ochratoxin A (OTA) aptamer. This small molecule mycotoxin contaminates a wide variety of food commodities such as cereals and wine. Several recent reports have used OTA-binding sequences [26] to develop affinity clean-up columns [92, 93] and biosensors [94–98] for evaluation under actual food testing conditions. Applications of small molecule-binding aptamers in other burgeoning areas, such as metabolomics, drug discovery, and synthetic biology, could also soon see dramatic growth. Continued effort in the development of aptamers for important small molecules is required in order for this field to realize its full potential.

Acknowledgments

The authors thank Alexander Wahba for assistance and the Natural Sciences and Engineering Council of Canada (NSERC), the Canadian Foundation for Innovation (CFI), the Ontario Research Fund (ORF), and an Ontario Early Researcher Award for funding.

References

[1] K. E. Deigan and A. R. Ferré-D'Amaré, "Riboswitches: discovery of drugs that target bacterial gene-regulatory RNAs,"

Accounts of Chemical Research, vol. 44, no. 12, pp. 1329–1338, 2011.

[2] J. E. Weigand and B. Suess, "Aptamers and riboswitches: perspectives in biotechnology," *Applied Microbiology and Biotechnology*, vol. 85, no. 2, pp. 229–236, 2009.

[3] C. Tuerk and L. Gold, "Systemic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase," *Science*, vol. 249, no. 4968, pp. 505–510, 1990.

[4] A. D. Ellington and J. W. Szostak, "In vitro selection of RNA molecules that bind specific ligands," *Nature*, vol. 346, no. 6287, pp. 818–822, 1990.

[5] D. L. Robertson and G. F. Joyce, "Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA," *Nature*, vol. 344, no. 6265, pp. 467–468, 1990.

[6] M. Mascini, *Aptamers in Bioanalysis*, John Wiley & Sons, Hoboken, NJ, USA, 2009.

[7] M. McKeague, C. R. Bradley, A. De Girolamo, A. Visconti, J. David Miller, and M. C. DeRosa, "Screening and initial binding assessment of fumonisin B1 aptamers," *International Journal of Molecular Sciences*, vol. 11, no. 12, pp. 4864–4881, 2010.

[8] S. D. Jayasena, "Aptamers: an emerging class of molecules that rival antibodies in diagnostics," *Clinical Chemistry*, vol. 45, no. 9, pp. 1628–1650, 1999.

[9] S. M. Nimjee, C. P. Rusconi, and B. A. Sullenger, "Aptamers: an emerging class of therapeutics," *Annual Review of Medicine*, vol. 56, pp. 555–583, 2005.

[10] S. Jhaveri and A. Ellington, "In vitro selection of RNA aptamers to a small molecule target," *Current Protocols in Nucleic Acid Chemistry*, Chapter 9, Unit 9.5, 2002.

[11] V. Bardoczy and T. Meszaros, "Aptamer selection for macromolecular (Protein) and for small molecule targets," in *Proceedings of the Periodica Polytechnica Abstracts of PhD Conference*, 2006.

[12] Z. Balogh, G. Lautner, V. Bardóczy, B. Komorowska, R. E. Gyurcsányi, and T. Mészáros, "Selection and versatile application of virus-specific aptamers," *FASEB Journal*, vol. 24, no. 11, pp. 4187–4195, 2010.

[13] K. Sefah, D. Shangguan, X. Xiong, M. B. O'Donoghue, and W. Tan, "Development of DNA aptamers using Cell-SELEX," *Nature protocols*, vol. 5, no. 6, pp. 1169–1185, 2010.

[14] S. L. Beaucage and M. H. Caruthers, "Deoxynucleoside phosphoramidites-A new class of key intermediates for deoxypolynucleotide synthesis," *Tetrahedron Letters*, vol. 22, no. 20, pp. 1859–1862, 1981.

[15] S. L. Beaucage and R. P. Iyer, "Advances in the synthesis of oligonucleotides by the phosphoramidite approach," *Tetrahedron*, vol. 48, no. 12, pp. 2223–2311, 1992.

[16] A. D. Keefe, S. Pai, and A. Ellington, "Aptamers as therapeutics," *Nature Reviews Drug Discovery*, vol. 9, no. 7, pp. 537–550, 2010.

[17] S. L. Beaucage and R. P. Iyer, "The synthesis of modified oligonucleotides by the phosporamidite approach and their applications," *Tetrahedron*, vol. 49, no. 28, pp. 6123–6194, 1993.

[18] S. L. Beaucage and R. P. Iyer, "The functionalization of oligonucleotides via phosphoramidite derivatives," *Tetrahedron*, vol. 49, no. 10, pp. 1925–1963, 1993.

[19] B. T. S. Bui and K. Haupt, "Molecularly imprinted polymers: synthetic receptors in bioanalysis," *Analytical and Bioanalytical Chemistry*, vol. 398, no. 6, pp. 2481–2492, 2010.

[20] Z. X. Xu, H. J. Gao, L. M. Zhang, X. Q. Chen, and X. G. Qiao, "The biomimetic immunoassay based on molecularly imprinted polymer: a comprehensive review of recent progress and future prospects," *Journal of Food Science*, vol. 76, no. 2, pp. R69–R75, 2011.

[21] D. R. Mills, R. L. Peterson, and S. Spiegelman, "An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 58, no. 1, pp. 217–224, 1967.

[22] R. Saffhill, H. Schneider-Bernloehr, L. E. Orgel, and S. Spiegelman, "In vitro selection of bacteriophage Qβ ribonucleic acid variants resistant to ethidium bromide," *Journal of Molecular Biology*, vol. 51, no. 3, pp. 531–539, 1970.

[23] D. S. Wilson and J. W. Szostak, "In vitro selection of functional nucleic acids," *Annual Review of Biochemistry*, vol. 68, pp. 611–647, 1999.

[24] S. Silverman and S. K, "Artificial functional nucleic acids: aptamers, ribozymes, and deoxyribozymes identified by in vitro selection," *Functional Nucleic Acids For Analytical Applications*, vol. 1, pp. 47–108, 2009.

[25] B. Vant-Hull, L. Gold, and D. A. Zichi, "Theoretical principles of in vitro selection using combinatorial nucleic acid libraries," *Current Protocols in Nucleic acid Chemistry*, Chapter 9, Unit 9.1, 2000.

[26] J. A. Cruz-Aguado and G. Penner, "Determination of ochratoxin A with a DNA aptamer," *Journal of Agricultural and Food Chemistry*, vol. 56, no. 22, pp. 10456–10461, 2008.

[27] M. Svobodova, A. Pinto, P. Nadal, and C. K. OSullivan, "Comparison of different methods for generation of single-stranded DNA for SELEX processes," *Analytical and Bioanalytical Chemistry*, vol. 404, no. 3, pp. 835–842, 2012.

[28] L. Gold, D. Ayers, J. Bertino et al., "Aptamer-based multiplexed proteomic technology for biomarker discovery," *PLoS One*, vol. 5, no. 12, Article ID 15004, 2010.

[29] R. Stoltenburg, C. Reinemann, and B. Strehlitz, "SELEX-A (r)evolutionary method to generate high-affinity nucleic acid ligands," *Biomolecular Engineering*, vol. 24, no. 4, pp. 381–403, 2007.

[30] J. H. Davis and J. W. Szostak, "Isolation of high-affinity GTP aptamers from partially structured RNA libraries," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 18, pp. 11616–11621, 2002.

[31] X. Luo, M. Mckeague, S. Pitre et al., "Computational approaches toward the design of pools for the in vitro selection of complex aptamers," *RNA*, vol. 16, no. 11, pp. 2252–2262, 2010.

[32] K. M. Ruff, T. M. Snyder, and D. R. Liu, "Enhanced functional potential of nucleic acid aptamer libraries patterned to increase secondary structure," *Journal of the American Chemical Society*, vol. 132, no. 27, pp. 9453–9464, 2010.

[33] M. N. Win, J. S. Klein, and C. D. Smolke, "Codeine-binding RNA aptamers and rapid determination of their binding constants using a direct coupling surface plasmon resonance assay," *Nucleic Acids Research*, vol. 34, no. 19, pp. 5670–5682, 2006.

[34] J. C. Cox, P. Rudolph, and A. D. Ellington, "Automated RNA selection," *Biotechnology Progress*, vol. 14, no. 6, pp. 845–850, 1998.

[35] P. W. Goertz, J. C. Cox, and A. D. Ellington, "Automated selection of aminoglycoside aptamers," *Journal of the Association for Laboratory Automation*, vol. 9, no. 3, pp. 150–154, 2004.

[36] A. Wochner, B. Cech, M. Menger, V. A. Erdmann, and J. Glökler, "Semi-automated selection of DNA aptamers using magnetic particle handling," *BioTechniques*, vol. 43, no. 3, pp. 344–353, 2007.

[37] M. J. Cho and R. Juliano, "Macromolecular versus small-molecule therapeutics: drug discovery, development and clinical considerations," *Trends in Biotechnology*, vol. 14, no. 5, pp. 153–158, 1996.

[38] M. L. Ashour and M. Wink, "Genus Bupleurum: a review of its phytochemistry, pharmacology and modes of action," *Journal of Pharmacy and Pharmacology*, vol. 63, no. 3, pp. 305–321, 2011.

[39] T. Roemer, J. Davies, G. Giaever, and C. Nislow, "Bugs, drugs and chemical genomics," *Nature Chemical Biology*, vol. 8, no. 1, pp. 46–56, 2012.

[40] T. A. Walsh, "The emerging field of chemical genetics: potential applications for pesticide discovery," *Pest Management Science*, vol. 63, no. 12, pp. 1165–1171, 2007.

[41] J. Cruz-Toledo, M. McKeague, X. Zhang et al., "Aptamer base: a collaborative knowledge base to describe aptamers and SELEX experiments," *Database*, vol. 2012, Article ID bas006, 2012.

[42] M. Mascini, I. Palchetti, and S. Tombelli, "Nucleic acid and peptide aptamers: fundamentals and bioanalytical aspects," *Angewandte Chemie*, vol. 51, pp. 1316–1332, 2012.

[43] F. Michael, "Oligonucleotide aptamers that recognize small molecules ," *Current Opinion in Structural Biology*, vol. 9, pp. 324–329, 1999.

[44] R. D. Jenison, S. C. Gill, A. Pardi, and B. Polisky, "High-resolution molecular discrimination by RNA," *Science*, vol. 263, no. 5152, pp. 1425–1429, 1994.

[45] M. Michaud, E. Jourdan, A. Villet, A. Ravel, C. Grosset, and E. Peyrin, "A DNA aptamer as a new target-specific chiral selector for HPLC," *Journal of the American Chemical Society*, vol. 125, no. 28, pp. 8672–8679, 2003.

[46] M. Famulok and J. W. Szostak, "Stereospecific recognition of tryptophan agarose by in vitro selected RNA," *Journal of the American Chemical Society*, vol. 114, pp. 3990–3991, 1992.

[47] M. Famulok, "Molecular recognition of amino acids by RNA-aptamers: an L-citrulline binding RNA motif and its evolution into an L-arginine binder," *Journal of the American Chemical Society*, vol. 116, no. 5, pp. 1698–1706, 1994.

[48] A. Geiger, P. Burgstaller, H. Von der Eltz, A. Roeder, and M. Famulok, "RNA aptamers that bind L-arginine with sub-micromolar dissociation constants and high enantioselectivity," *Nucleic Acids Research*, vol. 24, no. 6, pp. 1029–1036, 1996.

[49] A. Shoji, M. Kuwahara, H. Ozaki, and H. Sawai, "Modified DNA aptamer that binds the (R)-isomer of a thalidomide derivative with high enantioselectivity," *Journal of the American Chemical Society*, vol. 129, no. 5, pp. 1456–1464, 2007.

[50] Y. S. Kim, C. J. Hyun, I. A. Kim, and M. B. Gu, "Isolation and characterization of enantioselective DNA aptamers for ibuprofen," *Bioorganic and Medicinal Chemistry*, vol. 18, no. 10, pp. 3467–3473, 2010.

[51] J. M. Carothers, J. A. Goler, Y. Kapoor, L. Lara, and J. D. Keasling, "Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity," *Nucleic Acids Research*, vol. 38, no. 8, Article ID gkq082, pp. 2736–2747, 2010.

[52] P. Pfeffer and H. Gohlke, "DrugScoreRNA—knowledge-based scoring function to predict RNA—Ligand interactions," *Journal of Chemical Information and Modeling*, vol. 47, no. 5, pp. 1868–1876, 2007.

[53] M. Jo, J. Y. Ahn, J. Lee et al., "Development of single-stranded DNA aptamers for specific bisphenol a detection," *Oligonucleotides*, vol. 21, no. 2, pp. 85–91, 2011.

[54] J. H. Niazi, S. J. Lee, Y. S. Kim, and M. B. Gu, *Bioorg. Med. Chem*, vol. 16, pp. 1254–1261, 2008.

[55] M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, and R. R. Breaker, "Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria," *Cell*, vol. 113, no. 5, pp. 577–586, 2003.

[56] R. Welz and R. R. Breaker, "Ligand binding and gene control characteristics of tandem riboswitches in Bacillus anthracis," *RNA*, vol. 13, no. 4, pp. 573–582, 2007.

[57] M. Kwon and S. A. Strobel, "Chemical basis of glycine riboswitch cooperativity," *RNA*, vol. 14, no. 1, pp. 25–34, 2008.

[58] S. Missailidis and A. Hardy, "Aptamers as inhibitors of target proteins," *Expert Opinion on Therapeutic Patents*, vol. 19, no. 8, pp. 1073–1082, 2009.

[59] G. Mayer, M. S. L. Ahmed, A. Dolf, E. Endl, P. A. Knolle, and M. Famulok, "Fluorescence-activated cell sorting for aptamer SELEX with cell mixtures," *Nature Protocols*, vol. 5, no. 12, pp. 1993–2004, 2010.

[60] D. J. Schneider, R. Vanderslice, and L. Gold, "Flow cell SELEX," US Patent 5,861,254, 1999.

[61] C. Wilson and J. W. Szostak, "Isolation of a fluorophore-specific DNA aptamer with weak redox activity," *Chemistry and Biology*, vol. 5, no. 11, pp. 609–617, 1998.

[62] C. Yao, Y. Qi, Y. Zhao, Y. Xiang, Q. Chen, and W. Fu, "Aptamer-based piezoelectric quartz crystal microbalance biosensor array for the quantification of IgE," *Biosensors and Bioelectronics*, vol. 24, no. 8, pp. 2499–2503, 2009.

[63] R. J. White, A. A. Rowe, and K. W. Plaxco, "Re-engineering aptamers to support reagentless, self-reporting electrochemical sensors," *Analyst*, vol. 135, no. 3, pp. 589–594, 2010.

[64] J. Hu and C. Easley, "A simple and rapid approach for measurement of dissociation constants of DNA aptamers against proteins and small molecules *via* automated microchip electrophoresis," *Analyst*, vol. 136, no. 17, pp. 3461–3468, 2011.

[65] T. H. Nguyen, L. J. Steinbock, H. J. Butt, M. Helm, and R. Berger, "Measuring single small molecule binding via rupture forces of a split aptamer," *Journal of the American Chemical Society*, vol. 133, no. 7, pp. 2025–2027, 2011.

[66] P. S. Lau and Y. Li, "Functional nucleic acids as molecular recognition elements for small organic and biological molecules," *Current Organic Chemistr*, vol. 15, no. 4, pp. 557–575, 2011.

[67] A. K. H. Cheng, D. Sen, and H. Z. Yu, "Design and testing of aptamer-based electrochemical biosensors for proteins and small molecules," *Bioelectrochemistry*, vol. 77, no. 1, pp. 1–12, 2009.

[68] R. E. Wang, Y. Zhang, J. Cai, W. Cai, and T. Gao, "Aptamer-based fluorescent biosensors," *Current Medicinal Chemistry*, vol. 18, pp. 4175–4184, 2011.

[69] E. J. Cho, J. W. Lee, and A. D. Ellington, "Applications of aptamers as sensors," *Annual Review of Analytical Chemistry* , pp. 241–264, 2009.

[70] M. McKeague, A. Giamberardino, and M. C. DeRosa, "Advances in aptamer-based biosensors for food safety," in *Environmental Biosensors*, V. Somerset, Ed., pp. 17–42, InTech, 2011.

[71] R. Nutiu and Y. Li, "Structure-switching signaling aptamers," *Journal of the American Chemical Society*, vol. 125, no. 16, pp. 4771–4778, 2003.

[72] P. S. Lau, B. K. Coombes, and Y. Li, "A General approach to the construction of structure-switching reporters from RNA aptamers," *Angewandte Chemie International*, vol. 49, pp. 7938–7942, 2010.

[73] C. Carrasquilla, P. S. Lau, Y. Li, and J. D. Brennan, "Stabilizing structure-switching signaling RNA aptamers by entrapment in sol-gel derived materials for solid-phase assay," *Journal of the American Chemical Society*, vol. 134, pp. 10998–11005, 2012.

[74] D. Zheng, R. Zou, and X. Lou, "free fluorescent detection of ions, proteins, and small molecules using structure-switching aptamers, SYBR gold, and exonuclease," *Analytical Chemistry*, vol. 84, pp. 3554–3560, 2012.

[75] J. Liang, Z. Chen, L. Guo, and L. Li, "Electrochemical sensing of L-histidine based on structure-switching DNAzymes and gold nanoparticle-graphene nanosheet composites ," *Chemical Communications* , vol. 47, pp. 5476–5478, 2011.

[76] J. Chen, Z. Fang, J. Liu, and L. Zeng, "A simple and rapid biosensor for ochratoxin A based on a structure-switching signaling aptamer," *Food Control*, vol. 25, pp. 555–560, 2012.

[77] X. Hun and Z. Wang, "L-Argininamide biosensor based on S1 nuclease hydrolysis signal amplification," *Microchimica Acta*, vol. 176, pp. 209–216, 2012.

[78] Z. Zhu, T. Schmidt, M. Mahrous et al., "Optimization of the structure-switching aptamer-based fluorescence polarization assay for the sensitive tyrosinamide sensing," *Analytica Chimica Acta*, vol. 707, pp. 191–196, 2011.

[79] R. Nutiu and Y. Li, "In vitro selection of structure-switching signaling aptamers," *Angewandte Chemie* , vol. 44, pp. 1061–1065, 2005.

[80] E. L. Null and Y. Lu, "Rapid determination of enantiomeric ratio using fluorescent DNA or RNA aptamers," *Analyst*, vol. 135, no. 2, pp. 419–422, 2010.

[81] M. Famulok, J. S. Hartig, and G. Mayer, "Functional aptamers and aptazymes in biotechnology, diagnostics, and therapy," *Chemical Reviews*, vol. 107, no. 9, pp. 3715–3743, 2007.

[82] J. L. Vinkenborg, N. Karnowski, and M. Famulok, "Aptamers for allosteric regulation," *Nature Chemical Biology*, vol. 7, no. 8, pp. 519–527, 2011.

[83] M. N. Win, J. C. Liang, and C. D. Smolke, "Frameworks for Programming Biological Function through RNA Parts and Devices," *Chemistry and Biology*, vol. 16, no. 3, pp. 298–310, 2009.

[84] J. Tang and R. R. Breaker, "Rational design of allosteric ribozymes," *Chemistry and Biology*, vol. 4, no. 6, pp. 453–459, 1997.

[85] M. N. Stojanovic and D. M. Kolpashchikov, "Modular aptameric sensors," *Journal of the American Chemical Society*, vol. 126, no. 30, pp. 9266–9270, 2004.

[86] M. Famulok, M. Blind, and G. Mayer, "Intramers as promising new tools in functional proteomics," *Chemistry and Biology*, vol. 8, no. 10, pp. 931–939, 2001.

[87] J. C. Niles and M. A. Marletta, "Utilizing RNA aptamers to probe a physiologically important heme-regulated cellular network," *ACS Chemical Biology*, vol. 1, no. 8, pp. 515–524, 2006.

[88] Y. Li, C. R. Geyer, and D. Sen, "Recognition of anionic porphyrins by DNA aptamers," *Biochemistry*, vol. 35, no. 21, pp. 6911–6922, 1996.

[89] J. C. Niles, J. L. DeRisi, and M. A. Marletta, "Inhibiting Plasmodium falciparum growth and heme detoxification pathway using heme-binding DNA aptamers," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 32, pp. 13266–13271, 2009.

[90] M. R. Holahan, D. Madularu, E. M. McConnell, R. Walsh, and M. C. DeRosa, "Intra-accumbens injection of a dopamine aptamer abates MK-801-induced cognitive dysfunction in a model of schizophrenia," *PLoS One*, vol. 6, no. 7, Article ID e22239, 2011.

[91] G. Penner, IVD Technology, 2012.

[92] A. De Girolamo, M. McKeague, J. D. Miller, M. C. DeRosa, and A. Visconti, "Determination of ochratoxin A in wheat after clean-up through a DNA aptamer-based solid phase extraction column," *Food Chemistry*, vol. 127, no. 3, pp. 1378–1384, 2011.

[93] A. De Girolamo, L. Le, G. Penner, R. Schena, and A. Visconti, "Analytical performances of a DNA-ligand system using time-resolved fluorescence for the determination of ochratoxin A in wheat," *Analytical and Bioanalytical Chemistry*, vol. 403, pp. 2627–2634, 2012.

[94] C. Yang, V. Lates, B. Prieto-Simón, J. Marty, and X. Yang, "Aptamer-DNAzyme hairpins for biosensing of Ochratoxin A ," *Biosensors and Bioelectronics*, vol. 32, pp. 208–212, 2012.

[95] L. Bonel, J. C. Vidal, P. Duato, and J. R. Castillo, "An electrochemical competitive biosensor for ochratoxin A based on a DNA biotinylated aptamer," *Biosensors and Bioelectronics*, vol. 26, no. 7, pp. 3254–3259, 2011.

[96] Z. Guo, J. Ren, J. Wang, and E. Wang, "Single-walled carbon nanotubes based quenching of free FAM-aptamer for selective determination of ochratoxin A," *Talanta*, vol. 85, no. 5, pp. 2517–2521, 2011.

[97] L. Wang, W. Ma, W. Chen et al., "An aptamer-based chromatographic strip assay for sensitive toxin semi-quantitative detection," *Biosensors and Bioelectronics*, vol. 26, no. 6, pp. 3059–3062, 2011.

[98] H. Kuang, W. Chen, D. Xu et al., "Fabricated aptamer-based electrochemical 'signal-off' sensor of ochratoxin A," *Biosensors and Bioelectronics*, vol. 26, no. 2, pp. 710–716, 2010.

[99] Y. Miyachi, N. Shimizu, C. Ogino, and A. Kondo, "Selection of DNA aptamers using atomic force microscopy," *Nucleic Acids Research*, vol. 38, no. 4, article e21, 2010.

[100] D. Smith, G. P. Kirschenheuter, J. Charlton, D. M. Guidot, and J. E. Repine, "In vitro selection of RNA-based irreversible inhibitors of human neutrophil elastase," *Chemistry and Biology*, vol. 2, no. 11, pp. 741–750, 1995.

[101] Y. Kim, C. Liu, and W. Tan, "Aptamers generated by Cell SELEX for biomarker discovery," *Biomarkers in Medicine*, vol. 3, no. 2, pp. 193–202, 2009.

[102] S. D. Mendonsa and M. T. Bowser, "In vitro selection of high-affinity DNA ligands for human IgE using capillary electrophoresis," *Analytical Chemistry*, vol. 76, no. 18, pp. 5387–5392, 2004.

[103] D. H. Burke and J. H. Willis, "Recombination, RNA evolution, and bifunctional RNA molecules isolated through chimeric SELEX," *RNA*, vol. 4, no. 9, pp. 1165–1175, 1998.

[104] J. D. Smith and L. Gold, "Conditional-selex," US Patent 6706482, 2004.

[105] K. B. Jensen, B. L. Atkinson, M. C. Willis, T. H. Koch, and L. Gold, "Using in vitro selection to direct the covalent attachment of human immunodeficiency virus type 1 Rev protein to high-affinity RNA ligands," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 26, pp. 12220–12224, 1995.

[106] K. N. Morris, K. B. Jensen, C. M. Julin, M. Weil, and L. Gold, "High affinity ligands from in vitro selection: complex targets," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 6, pp. 2902–2907, 1998.

[107] R. Y. L. Tsai and R. R. Reed, "Identification of DNA recognition sequences and protein interaction domains of the multiple-Zn-finger protein Roaz," *Molecular and Cellular Biology*, vol. 18, no. 11, pp. 6447–6456, 1998.

[108] R. E. Martell, J. R. Nevins, and B. A. Sullenger, "Optimizing aptamer activity for gene therapy applications using expression cassette SELEX," *Molecular Therapy*, vol. 6, no. 1, pp. 30–34, 2002.

[109] R. Stoltenburg, C. Reinemann, and B. Strehlitz, "FluMag-SELEX as an advantageous method for DNA aptamer selection," *Analytical and Bioanalytical Chemistry*, vol. 383, no. 1, pp. 83–91, 2005.

[110] M. Dobbelstein and T. Shenk, "In vitro selection of RNA ligands for the ribosomal L22 protein associated with Epstein-Barr virus-expressed RNA by using randomized and cDNA-derived RNA libraries," *Journal of Virology*, vol. 69, no. 12, pp. 8027–8034, 1995.

[111] L. R. Coulter, M. A. Landree, and T. A. Cooper, "Identification of a new class of exonic splicing enhancers by in vivo selection," *Molecular and Cellular Biology*, vol. 17, no. 4, pp. 2143–2150, 1997.

[112] J. Kawakami, H. Imanaka, Y. Yokota, and N. Sugimoto, "In vitro selection of aptamers that act with Zn^{2+}," *Journal of Inorganic Biochemistry*, vol. 82, no. 1–4, pp. 197–206, 2000.

[113] A. D. Keefe and S. T. Cload, "SELEX with modified nucleotides," *Current Opinion in Chemical Biology*, vol. 12, no. 4, pp. 448–456, 2008.

[114] Q. Gong, J. Wang, K. M. Ahmad et al., "Selection strategy to generate aptamer pairs that bind to distinct sites on protein targets," *Analytical Chemistry*, vol. 84, no. 12, pp. 5365–5371, 2012.

[115] C. J. Huang, H. I. Lin, S. C. Shiesh, and G. B. Lee, "Integrated microfluidic system for rapid screening of CRP aptamers utilizing systematic evolution of ligands by exponential enrichment (SELEX)," *Biosensors and Bioelectronics*, vol. 25, no. 7, pp. 1761–1766, 2010.

[116] X. Lou, J. Qian, Y. Xiao et al., "Micromagnetic selection of aptamers in microfluidic channels," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 9, pp. 2989–2994, 2009.

[117] A. Nitsche, A. Kurth, A. Dunkhorst et al., "One-step selection of Vaccinia virus-binding DNA aptamers by MonoLEX," *BMC Biotechnology*, vol. 7, article no. 48, 2007.

[118] A. Jolma, T. Kivioja, J. Toivonen et al., "Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities," *Genome Research*, vol. 20, no. 6, pp. 861–873, 2010.

[119] L. Wu and J. F. Curran, "An allosteric synthetic DNA," *Nucleic Acids Research*, vol. 27, no. 6, pp. 1512–1516, 1999.

[120] D. C. Reid, B. L. Chang, S. I. Gunderson, L. Alpert, W. A. Thompson, and W. G. Fairbrother, "Next-generation SELEX identifies sequence and structural determinants of splicing factor binding in human pre-mRNA sequence," *RNA*, vol. 15, no. 12, pp. 2385–2397, 2009.

[121] M. Berezovski, M. Musheev, A. Drabovich, and S. N. Krylov, "Non-SELEX selection of aptamers," *Journal of the American Chemical Society*, vol. 128, no. 5, pp. 1410–1411, 2006.

[122] E. N. Brody, M. C. Willis, J. D. Smith, S. Jayasena, D. Zichi, and L. Gold, "The use of aptamers in large arrays for molecular diagnostics," *Molecular Diagnosis*, vol. 4, no. 4, pp. 381–388, 1999.

[123] J. D. Wen and D. M. Gray, "Selection of genomic sequences that bind tightly to Ff gene 5 protein: primer-free genomic SELEX," *Nucleic acids research*, vol. 32, no. 22, article e182, 2004.

[124] E. Roulet, S. Busso, A. A. Camargo, A. J. G. Simpson, N. Mermod, and P. Bucher, "High-throughput SELEX-SAGE method for quantitative modeling of transcription-factor binding sites," *Nature Biotechnology*, vol. 20, no. 8, pp. 831–835, 2002.

[125] S. Klußmann, A. Nolte, R. Bald, V. A. Erdmann, and J. P. Fürste, "Mirror-image RNA that binds D-adenosine," *Nature Biotechnology*, vol. 14, no. 9, pp. 1112–1115, 1996.

[126] A. Vater, F. Jarosch, K. Buchner, and S. Klussmann, "Short bioactive Spiegelmers to migraine-associated calcitonin gene-related peptide rapidly identified by a novel approach: tailored-SELEX," *Nucleic acids research*, vol. 31, no. 21, article 130, 2003.

[127] S. P. Ohuchi, T. Ohtsu, and Y. Nakamura, "Selection of RNA aptamers against recombinant transforming growth factor-β type III receptor displayed on cell surface," *Biochimie*, vol. 88, no. 7, pp. 897–904, 2006.

[128] R. White, C. Rusconi, E. Scardino et al., "Generation of species cross-reactive aptamers using "toggle" SELEX," *Molecular Therapy*, vol. 4, no. 6, pp. 567–573, 2001.

[129] L. A. Cassiday and L. J. Maher III, "Yeast genetic selections to optimize RNA decoys for transcription factor NF-κB," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 7, pp. 3930–3935, 2003.

[130] A. D. Ellington and J. W. Szostak, "Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures," *Nature*, vol. 355, no. 6363, pp. 850–852, 1992.

[131] D. E. Huizenga and J. W. Szostak, "A DNA aptamer that binds adenosine and ATP," *Biochemistry*, vol. 34, no. 2, pp. 656–665, 1995.

[132] K. Harada and A. D. Frankel, "Identification of two novel arginine binding DNAs," *EMBO Journal*, vol. 14, no. 23, pp. 5798–5811, 1995.

[133] Q. Yang, I. J. Goldstein, H.-Y. Mei, and D. R. Engelke, "DNA ligands that bind tightly and selectively to cellobiose," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 10, pp. 5462–5467, 1998.

[134] S. M. Rink, J. C. Shen, and L. A. Loeb, "Creation of RNA molecules that recognize the oxidative lesion 7,8-dihydro-8-hydroxy-2′-deoxyguanosine (8-oxodG) in DNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 20, pp. 11619–11624, 1998.

[135] T. Kato, T. Takemura, K. Yano, K. Ikebukuro, and I. Karube, "In vitro selection of DNA aptamers which bind to cholic acid," *Biochimica et Biophysica Acta. Gene Structure and Expression*, vol. 1493, no. 1-2, pp. 12–18, 2000.

[136] A. Okazawa, H. Maeda, E. Fukusaki, Y. Katakura, and A. Kobayashi, "In vitro selection of hematoporphyrin binding DNA aptamers," *Bioorganic and Medicinal Chemistry Letters*, vol. 10, no. 23, pp. 2653–2656, 2000.

[137] E. Vianini, M. Palumbo, and B. Gatto, "In vitro selection of DNA aptamers that bind L-tyrosinamide," *Bioorganic and Medicinal Chemistry*, vol. 9, no. 10, pp. 2543–2548, 2001.

[138] M. M. Masud, M. Kuwahara, H. Ozaki, and H. Sawai, "Sialyllactose-binding modified DNA aptamer bearing additional functionality by SELEX," *Bioorganic and Medicinal Chemistry*, vol. 12, no. 5, pp. 1111–1120, 2004.

[139] D. Mann, C. Reinemann, R. Stoltenburg, and B. Strehlitz, "In vitro selection of DNA aptamers binding ethanolamine," *Biochemical and Biophysical Research Communications*, vol. 338, no. 4, pp. 1928–1934, 2005.

[140] S. Sando, A. Narita, and Y. Aoyama, "Light-up Hoechst-DNA aptamer pair: generation of an aptamer-selective fluorophore from a conventional DNA-staining dye," *ChemBioChem*, vol. 8, no. 15, pp. 1795–1803, 2007.

[141] Y. S. Kim, H. S. Jung, T. Matsuura, H. Y. Lee, T. Kawai, and M. B. Gu, "Electrochemical detection of 17β-estradiol using DNA aptamer immobilized gold electrode chip," *Biosensors and Bioelectronics*, vol. 22, no. 11, pp. 2525–2531, 2007.

[142] G. Hayashi, M. Hagihara, C. Dohno, and K. Nakatani, "Photoregulation of a peptide-RNA interaction on a gold surface," *Journal of the American Chemical Society*, vol. 129, no. 28, pp. 8678–8679, 2007.

[143] J. H. Niazi, S. J. Lee, and M. B. Gu, "Single-stranded DNA aptamers specific for antibiotics tetracyclines," *Bioorganic and Medicinal Chemistry*, vol. 16, no. 15, pp. 7245–7253, 2008.

[144] K. Ohsawa, T. Kasamatsu, J. I. Nagashima et al., "Arginine-modified DNA aptamers that show enantioselective recognition of the dicarboxylic acid moiety of glutamic acid," *Analytical Sciences*, vol. 24, no. 1, pp. 167–172, 2008.

[145] A. Wochner, M. Menger, D. Orgel et al., "A DNA aptamer with high affinity and specificity for therapeutic anthracyclines," *Analytical Biochemistry*, vol. 373, no. 1, pp. 34–42, 2008.

[146] R. Walsh and M. C. DeRosa, "Retention of function in the DNA homolog of the RNA dopamine aptamer," *Biochemical and Biophysical Research Communications*, vol. 388, no. 4, pp. 732–735, 2009.

[147] Y. Miyachi, N. Shimizu, C. Ogino, H. Fukuda, and A. Kondo, "Selection of a DNA aptamer that binds 8-OHdG using GMP-agarose," *Bioorganic and Medicinal Chemistry Letters*, vol. 19, no. 13, pp. 3619–3622, 2009.

[148] C. B. Joeng, J. H. Niazi, S. J. Lee, and M. B. Gu, "ssDNA aptamers that recognize diclofenac and 2-anilinophenylacetic acid," *Bioorganic and Medicinal Chemistry*, vol. 17, no. 15, pp. 5380–5387, 2009.

[149] J. He, Y. Liu, M. Fan, and X. Liu, "Isolation and identification of the DNA aptamer target to acetamiprid," *Journal of Agricultural and Food Chemistry*, vol. 59, no. 5, pp. 1582–1586, 2011.

[150] K. M. Song, M. Cho, H. Jo et al., "Gold nanoparticle-based colorimetric detection of kanamycin using a DNA aptamer," *Analytical Biochemistry*, vol. 415, no. 2, pp. 175–181, 2011.

[151] X. Yang, T. Bing, H. Mei, C. Fang, Z. Cao, and D. Shangguan, "Characterization and application of a DNA aptamer binding to l-tryptophan," *Analyst*, vol. 136, no. 3, pp. 577–585, 2011.

[152] L. Barthelmebs, J. Jonca, A. Hayat, B. Prieto-Simon, and J. L. Marty, "Enzyme-Linked Aptamer Assays (ELAAs), based on a competition format for a rapid and sensitive detection of Ochratoxin A in wine," *Food Control*, vol. 22, no. 5, pp. 737–743, 2011.

[153] A. Renaud De La Faverie, F. Hamon, C. Di Primo et al., "Nucleic acids targeted to drugs: SELEX against a quadruplex ligand," *Biochimie*, vol. 93, no. 8, pp. 1357–1367, 2011.

[154] L. Wang, X. Liu, Q. Zhang et al., "Selection of DNA aptamers that bind to four organophosphorus pesticides," *Biotechnology Letters*, vol. 34, no. 5, pp. 869–874, 2012.

[155] S. Xu, H. Yuan, S. Chen, A. Xu, J. Wang, and L. Wu, "Selection of DNA aptamers against polychlorinated biphenyls as potential biorecognition elements for environmental analysis," *Analytical Biochemistry*, vol. 423, no. 2, pp. 195–201, 2012.

[156] J. Mehta, E. Rouah-Martin, B. Van Dorst et al., "Selection and characterization of PCB-binding DNA aptamers," *Analytical Chemistry*, vol. 84, no. 3, pp. 1669–1676, 2012.

[157] K.-M. Song, E. Jeong, W. Jeon, M. Cho, and C. Ban, "Aptasensor for ampicillin using gold nanoparticle based dual fluorescence-colorimetric methods," *Analytical and Bioanalytical Chemistry*, vol. 402, no. 6, pp. 2153–2161, 2012.

[158] J. W. Szostak, "Enzymatic activity of the conserved core of a group I self- splicing intron," *Nature*, vol. 322, no. 6074, pp. 83–86, 1986.

[159] I. Majerfeld and M. Yarus, "An RNA pocket for an aliphatic hydrophobe," *Nature Structural Biology*, vol. 1, no. 5, pp. 287–292, 1994.

[160] J. R. Lorsch and J. W. Szostak, "In vitro selection of RNA aptamers specific for cyanocobalamin," *Biochemistry*, vol. 33, no. 4, pp. 973–982, 1994.

[161] P. Burgstaller and M. Famulok, "Isolation of RNA aptamers for biological cofactors by in vitro selection," *Angewandte Chemie*, vol. 33, no. 10, pp. 1084–1087, 1994.

[162] S. M. Lato, A. R. Boles, and A. D. Ellington, "In vitro selection of RNA lectins: using combinatorial chemistry to interpret ribozyme evolution," *Chemistry and Biology*, vol. 2, no. 5, pp. 291–303, 1995.

[163] M. G. Wallis, U. Von Ahsen, R. Schroeder, and M. Famulok, "A novel RNA motif for neomycin recognition," *Chemistry and Biology*, vol. 2, no. 8, pp. 543–552, 1995.

[164] Y. Wang and R. R. Rando, "Specific binding of aminoglycoside antibiotics to RNA," *Chemistry and Biology*, vol. 2, pp. 281–290, 1995.

[165] C. T. Lauhon and J. W. Szostak, "RNA aptamers that bind flavin and nicotinamide redox cofactors," *Journal of the American Chemical Society*, vol. 117, no. 4, pp. 1246–1257, 1995.

[166] C. Wilson, J. Nix, and J. Szostak, "Functional requirements for specific ligand recognition by a biotin-binding rna pseudoknot," *Biochemistry*, vol. 37, no. 41, pp. 14410–14419, 1998.

[167] C. Mannironi, A. Di Nardo, P. Fruscoloni, and G. P. Tocchini-Valentini, "In vitro selection of dopamine RNA ligands," *Biochemistry*, vol. 36, no. 32, pp. 9726–9734, 1997.

[168] A. A. Haller and P. Sarnow, "In vitro selection of a 7-methylguanosine binding RNA that inhibits translation of capped mRNA molecules," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 16, pp. 8521–8526, 1997.

[169] M. Welch, I. Majerfeld, and M. Yarus, "23S rRNA similarity from selection for peptidyl transferase mimicry," *Biochemistry*, vol. 36, no. 22, pp. 6614–6623, 1997.

[170] D. H. Burke, D. C. Hoffman, A. Brown, M. Hansen, A. Pardi, and L. Gold, "RNA aptamers to the peptidyl transferase inhibitor chloramphenicol," *Chemistry and Biology*, vol. 4, no. 11, pp. 833–843, 1997.

[171] M. G. Wallis, B. Streicher, H. Wank et al., "In vitro selection of a viomycin-binding RNA pseudoknot," *Chemistry and Biology*, vol. 4, no. 5, pp. 357–366, 1997.

[172] L. A. Holeman, S. L. Robinson, J. W. Szostak, and C. Wilson, "Isolation and characterization of fluorophore-binding RNA aptamers," *Folding and Design*, vol. 3, no. 6, pp. 423–431, 1998.

[173] S. T. Wallace and R. Schroede, "In vitro selection and characterization of streptomycin-binding RNAs: recognition discrimination between antibiotics," *RNA*, vol. 4, no. 1, pp. 112–123, 1998.

[174] I. Majerfeld and M. Yarus, "Isoleucine:RNA sites with associated coding sequences," *RNA*, vol. 4, no. 4, pp. 471–478, 1998.

[175] D. Kiga, Y. Futamura, K. Sakamoto, and S. Yokoyama, "An RNA aptamer to the xanthine/guanine base with a distinctive mode of purine recognition," *Nucleic Acids Research*, vol. 26, no. 7, pp. 1755–1760, 1998.

[176] D. Grate and C. Wilson, "Laser-mediated, site-specific inactivation of RNA transcripts," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 11, pp. 6131–6136, 1999.

[177] A. Khvorova, Y. G. Kwak, M. Tamkun, I. Majerfeld, and M. Yarus, "RNAs that bind and change the permeability of phospholipid membranes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 19, pp. 10649–10654, 1999.

[178] M. Koizumi and R. R. Breaker, "Molecular recognition of cAMP by an RNA aptamer," *Biochemistry*, vol. 39, no. 30, pp. 8983–8992, 2000.

[179] S. Jhaveri, M. Rajendran, and A. D. Ellington, "In vitro selection of signaling aptamers," *Nature Biotechnology*, vol. 18, no. 12, pp. 1293–1297, 2000.

[180] C. Mannironi, C. Scerch, P. Fruscoloni, and G. P. Tocchini-Valentini, "Molecular recognition of amino acids by RNA aptamers: the evolution into an L-tyrosine binder of a dopamine-binding RNA motif," *RNA*, vol. 6, no. 4, pp. 520–527, 2000.

[181] K. Gebhardt, A. Shokraei, E. Babaie, and B. H. Lindqvist, "RNA aptamers to S-adenosylhomocysteine: kinetic properties, divalent cation dependency, and comparison with anti-S-adenosylhomocysteine antibody," *Biochemistry*, vol. 39, no. 24, pp. 7255–7265, 2000.

[182] J. A. Cowan, T. Ohyama, D. Wang, and K. Natarajan, "Recognition of a cognate RNA aptamer by neomycin B: quantitative evaluation of hydrogen bonding and electrostatic interactions," *Nucleic Acids Research*, vol. 28, no. 15, pp. 2935–2942, 2000.

[183] H. Schürer, K. Stembera, D. Knoll et al., "Aptamers that bind to the antibiotic moenomycin A," *Bioorganic and Medicinal Chemistry*, vol. 9, no. 10, pp. 2557–2563, 2001.

[184] S. Jeong, T.-Y. Eom, S.-J. Kim, S.-W. Lee, and J. Yu, "In vitro selection of the RNA Aptamer against the Sialyl Lewis X and its inhibition of the cell adhesion," *Biochemical and Biophysical Research Communications*, vol. 281, no. 1, pp. 237–243, 2001.

[185] C. Berens, A. Thain, and R. Schroeder, "A tetracycline-binding RNA aptamer," *Bioorganic and Medicinal Chemistry*, vol. 9, no. 10, pp. 2549–2556, 2001.

[186] M. Kwon, S. M. Chun, S. Jeong, and J. Yu, "In vitro selection of RNA against kanamycin B," *Molecules and Cells*, vol. 11, no. 3, pp. 303–311, 2001.

[187] M. Meli, J. Vergne, J.-L. Décout, and M.-C. Maurel, "Adenine-aptamer complexes. A bipartite RNA site that binds the adenine nucleic base," *Journal of Biological Chemistry*, vol. 277, no. 3, pp. 2104–2111, 2002.

[188] M. Roychowdhury-Saha, S. M. Lato, E. D. Shank, and D. H. Burke, "Flavin recognition by an RNA aptamer targeted toward FAD," *Biochemistry*, vol. 41, no. 8, pp. 2492–2499, 2002.

[189] C. Lozupone, S. Changayil, I. Majerfeld, and M. Yarus, "Selection of the simplest RNA that binds isoleucine," *RNA*, vol. 9, no. 11, pp. 1315–1322, 2003.

[190] N. K. Vaish, R. Larralde, A. W. Fraley, J. W. Szostak, and L. W. McLaughlin, "A novel, modification-dependent ATP-binding aptamer selected from an RNA library incorporating a cationic functionality," *Biochemistry*, vol. 42, no. 29, pp. 8842–8851, 2003.

[191] Z. Huang and J. W. Szostak, "Evolution of aptamers with a new specificity and new secondary structures from an ATP aptamer," *RNA*, vol. 9, no. 12, pp. 1456–1463, 2003.

[192] U. Brockstedt, A. Uzarowska, A. Montpetit, W. Pfau, and D. Labuda, "In vitro evolution of RNA aptamers recognizing carcinogenic aromatic amines," *Biochemical and Biophysical Research Communications*, vol. 313, no. 4, pp. 1004–1008, 2004.

[193] P. L. Sazani, R. Larralde, and J. W. Szostak, "A small aptamer with strong and specific recognition of the triphosphate of ATP," *Journal of the American Chemical Society*, vol. 126, no. 27, pp. 8370–8371, 2004.

[194] M. Legiewicz and M. Yarus, "A more complex isoleucine aptamer with a cognate triplet," *Journal of Biological Chemistry*, vol. 280, no. 20, pp. 19815–19822, 2005.

[195] I. Majerfeld, D. Puthenvedu, and M. Yarus, "RNA affinity for molecular L-histidine; genetic code origins," *Journal of Molecular Evolution*, vol. 61, no. 2, pp. 226–235, 2005.

[196] D. Lévesque, J. D. Beaudoin, S. Roy, and J. P. Perreault, "In vitro selection and characterization of RNA aptamers binding thyroxine hormone," *Biochemical Journal*, vol. 403, no. 1, pp. 129–138, 2007.

[197] D. P. Morse, "Direct selection of RNA beacon aptamers," *Biochemical and Biophysical Research Communications*, vol. 359, no. 1, pp. 94–101, 2007.

[198] H. W. Lee, S. G. Robinson, S. Bandyopadhyay, R. H. Mitchell, and D. Sen, "Reversible photo-regulation of a hammerhead ribozyme using a diffusible effector," *Journal of Molecular Biology*, vol. 371, no. 5, pp. 1163–1173, 2007.

[199] T. P. Constantin, G. L. Silva, K. L. Robertson et al., "Synthesis of new fluorogenic cyanine dyes and incorporation into RNA fluoromodules," *Organic Letters*, vol. 10, no. 8, pp. 1561–1564, 2008.

[200] K. Endo and Y. Nakamura, "A binary Cy3 aptamer probe composed of folded modules," *Analytical Biochemistry*, vol. 400, pp. 103–109, 2010.

[201] J. Lee, K. H. Lee, J. Jeon, A. Dragulescu-Andrasi, F. Xiao, and J. Rao, "Combining SELEX screening and rational design to develop light-up fluorophore-RNA aptamer pairs for RNA tagging," *ACS Chemical Biology*, vol. 5, no. 11, pp. 1065–1074, 2010.

[202] J. Sinha, S. J. Reyes, and J. P. Gallivan, "Reprogramming bacteria to seek and destroy an herbicide," *Nature Chemical Biology*, vol. 6, no. 6, pp. 464–470, 2010.

[203] K. Horii, K. Omi, Y. Yoshida et al., "Development of a sphingosylphosphorylcholine detection system using RNA aptamers," *Molecules*, vol. 15, no. 8, pp. 5742–5755, 2010.

[204] A. Murata, S. I. Sato, Y. Kawazoe, and M. Uesugi, "Small-molecule fluorescent probes for specific RNA targets," *Chemical Communications*, vol. 47, no. 16, pp. 4712–4714, 2011.

[205] J. S. Paige, K. Y. Wu, and S. R. Jaffrey, "RNA mimics of green fluorescent protein," *Science*, vol. 333, no. 6042, pp. 642–646, 2011.

[206] J. Bala, A. Bhaskar, A. Varshney, A. K. Singh, S. Dey, and P. Yadava, "In vitro selected RNA aptamer recognizing glutathione induces ROS-mediated apoptosis in the human breast cancer cell line MCF 7," *RNA Biology*, vol. 8, no. 1, pp. 101–111, 2011.

[207] J. L. Lau, M. M. Baksh, J. D. Fiedler et al., "Evolution and protein packaging of small-molecule RNA aptamers," *ACS Nano*, vol. 5, pp. 7722–7729, 2011.

[208] J. Flinders, S. C. DeFina, D. M. Brackett, C. Baugh, C. Wilson, and T. Dieckmann, "Recognition of planar and nonplanar ligands in the malachite green—RNA aptamer complex," *ChemBioChem*, vol. 5, no. 1, pp. 62–72, 2004.

[209] J. A. Cruz-Aguado and G. Penner, "Fluorescence polarization based displacement assay for the determination of small

molecules with aptamers," *Analytical Chemistry*, vol. 80, no. 22, pp. 8853–8855, 2008.

[210] A. Guedin, L. Lacroix, and J. L. Mergny, "Thermal melting studies of ligand DNA interactions," *Methods in Molecular Biology*, vol. 613, pp. 25–35, 2010.

[211] P. Lin, R. Chen, C. Lee, Y. Chang, C. Chen, and W. Chen, "Studies of the binding mechanism between aptamers and thrombin by circular dichroism, surface plasmon resonance and isothermal titration calorimetry," *Colloids and Surfaces B*, vol. 88, pp. 552–558, 2011.

[212] J. H. Lee, M. D. Canny, A. De Erkenez et al., "A therapeutic aptamer inhibits angiogenesis by specifically targeting the heparin binding domain of VEGF165," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 52, pp. 18902–18907, 2005.

[213] Y. Sultan, R. Walsh, C. Monreal, and M. C. DeRosa, "Preparation of functional aptamer films using layer-by-layer self-assembly," *Biomacromolecules*, vol. 10, no. 5, pp. 1149–1154, 2009.

[214] Q. Deng, I. German, D. Buchanan, and R. T. Kennedy, "Retention and separation of adenosine and analogues by affinity chromatography with an aptamer stationary phase," *Analytical Chemistry*, vol. 73, no. 22, pp. 5415–5421, 2001.

[215] A. P. Drabovich, M. Berezovski, V. Okhonin, and S. N. Krylov, "Selection of smart aptamers by methods of kinetic capillary electrophoresis," *Analytical Chemistry*, vol. 78, no. 9, pp. 3171–3178, 2006.

[216] J. Bao, S. M. Krylova, O. Reinstein, P. E. Johnson, and S. N. Krylov, "Label-free solution-based kinetic study of aptamer-small molecule interactions by kinetic capillary electrophoresis with UV detection revealing how kinetics control equilibrium," *Analytical Chemistry*, vol. 83, pp. 8387–8390, 2011.

[217] R. T. Turgeon, B. R. Fonslow, M. Jing, and M. T. Bowser, "Measuring aptamer equilibria using gradient micro free flow electrophoresis," *Analytical Chemistry*, vol. 82, no. 9, pp. 3636–3641, 2010.

[218] B. Hall, S. Arshad, K. Seo et al., "In vitro selection of RNA aptamers to a protein target by filter immobilization," *Current Protocols in Molecular Biology*, no. 88, pp. 24.3.1–24.3.27, 2009.

[219] C. Gaillard and F. Strauss, "DNA loops and semicatenated DNA junctions," *BMC Biochemistry*, vol. 1, article no. 1, pp. 1–7, 2000.

[220] P. Baaske, C. J. Wienken, P. Reineck, S. Duhr, and D. Braun, "Optical thermophoresis for quantifying the buffer dependence of aptamer binding," *Angewandte Chemie*, vol. 49, no. 12, pp. 2238–2241, 2010.

[221] A. S. R. Potty, K. Kourentzi, H. Fang et al., "Biophysical characterization of DNA aptamer interactions with vascular endothelial growth factor," *Biopolymers*, vol. 91, no. 2, pp. 145–156, 2009.

[222] S. S. Oh, K. Plakos, X. Lou, Y. Xiao, and H. T. Soh, "In vitro selection of structure-switching, self-reporting aptamers," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 32, pp. 14053–14058, 2010.

[223] E. E. Regulski and R. R. Breaker, "In-line probing analysis of riboswitches," *Methods in Molecular Biology*, vol. 419, pp. 53–67, 2008.

[224] W. Yoshida, K. Sode, and K. Ikebukuro, "Homogeneous DNA sensing using enzyme-inhibiting DNA aptamers," *Biochemical and Biophysical Research Communications*, vol. 348, no. 1, pp. 245–252, 2006.

[225] S. A. McManus and Y. Li, "Multiple occurrences of an efficient self-phosphorylating deoxyribozyme motif," *Biochemistry*, vol. 46, no. 8, pp. 2198–2204, 2007.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 713510, 15 pages
doi:10.1155/2012/713510

Review Article

Genetically Encoded Libraries of Nonstandard Peptides

Takashi Kawakami and Hiroshi Murakami

Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Meguro-ku, Tokyo 153-8902, Japan

Correspondence should be addressed to Hiroshi Murakami, murah@bio.c.u-tokyo.ac.jp

Received 1 June 2012; Accepted 12 August 2012

Academic Editor: Masayasu Kuwahara

The presence of a nonproteinogenic moiety in a nonstandard peptide often improves the biological properties of the peptide. Nonstandard peptide libraries are therefore used to obtain valuable molecules for biological, therapeutic, and diagnostic applications. Highly diverse non-standard peptide libraries can be generated by chemically or enzymatically modifying standard peptide libraries synthesized by the ribosomal machinery, using posttranslational modifications. Alternatively, strategies for encoding non-proteinogenic amino acids into the genetic code have been developed for the direct ribosomal synthesis of non-standard peptide libraries. In the strategies for genetic code expansion, non-proteinogenic amino acids are assigned to the nonsense codons or 4-base codons in order to add these amino acids to the universal genetic code. In contrast, in the strategies for genetic code reprogramming, some proteinogenic amino acids are erased from the genetic code and non-proteinogenic amino acids are reassigned to the blank codons. Here, we discuss the generation of genetically encoded non-standard peptide libraries using these strategies and also review recent applications of these libraries to the selection of functional non-standard peptides.

1. Introduction

Nonstandard peptides, also known as unnatural peptides or peptidomimetics, are peptide-based small molecules containing a moiety that does not exist in standard (i.e., natural) peptides composed of only 20 proteinogenic amino acids. The nonproteinogenic moiety in nonstandard peptides, such as a nonproteinogenic side chain, a modified backbone, or a macrocyclized backbone, often contributes to improving the peptide's cell permeability, stability against peptidases, and conformational rigidity, thereby affording specific high affinity toward its target molecule [1–5]. Naturally occurring nonribosomal peptides (e.g., immunosuppressant cyclosporine A) are representative of nonstandard peptides, and given the success of nonribosomal peptides as therapeutics, the development of methods to construct highly diverse drug-like nonstandard peptide libraries is important for the discovery of novel drug candidates. Chemical synthesis can generate highly modified drug-like nonstandard peptide libraries, but the size of these libraries is relatively small (with a diversity of up to 10^6 unique compounds). In contrast, by using genotype-phenotype linking technology, ribosomal synthesis can generate genetically encoded peptide libraries with extremely high diversity (up to 10^{13} compounds) [6–9]. However, ribosomally synthesized peptide libraries are typically composed of just the 20 proteinogenic amino acids. This had led to the development of various strategies to generate highly diverse nonstandard peptide libraries; these strategies integrate biology-based methods to construct highly diverse libraries and chemistry-based methods to construct highly modified drug-like libraries.

One approach to generate genetically encoded nonstandard peptide libraries involves the posttranslational modification of ribosomally synthesized peptide libraries. Examples include using chemical or enzymatic reactions for the posttranslational macrocyclization of standard peptide libraries. In the chemical cyclization approach, Heinis and coworkers used 1, 3, 5-tris(bromomethyl)benzene to react with three cysteine residues in peptide libraries displayed on phages, and the resulting bicyclic peptide libraries were used to obtain bicyclic peptide inhibitors against several enzymes [10–12]. In the enzymatic cyclization approach, Bosma et al. reported the bacterial display of peptide libraries cyclized with a thioether-bridge that was posttranslationally added using enzymes, and demonstrated thioether-bridged peptide selection against a model protein (streptavidin) [13].

Another example of the posttranslational conversion of standard peptide libraries to nonstandard ones is the modification of peptides with a moiety that selectively interacts with a target protein. Li and Roberts chemically modified a peptide library by reacting the sulfhydryl group of cysteine residues in the library with a bromoacetamide derivative of penicillin, and the penicillin-modified peptide library was subjected to mRNA display selection against penicillin binding protein 2a [14]. An excellent review discussing these posttranslational strategies has recently been published elsewhere [15].

Another approach to generate genetically encoded nonstandard peptide libraries involves the engineering of a ribosomal translation system, that is, cotranslational incorporation of nonproteinogenic amino acids into ribosomally synthesized polypeptides (Figure 1). This approach is further divided into two strategies: genetic code expansion and genetic code reprogramming. In this paper, we discuss both strategies and their application for the generation of genetically encoded nonstandard peptide libraries as well as the use of these libraries for the selection/evolution of functional nonstandard peptides.

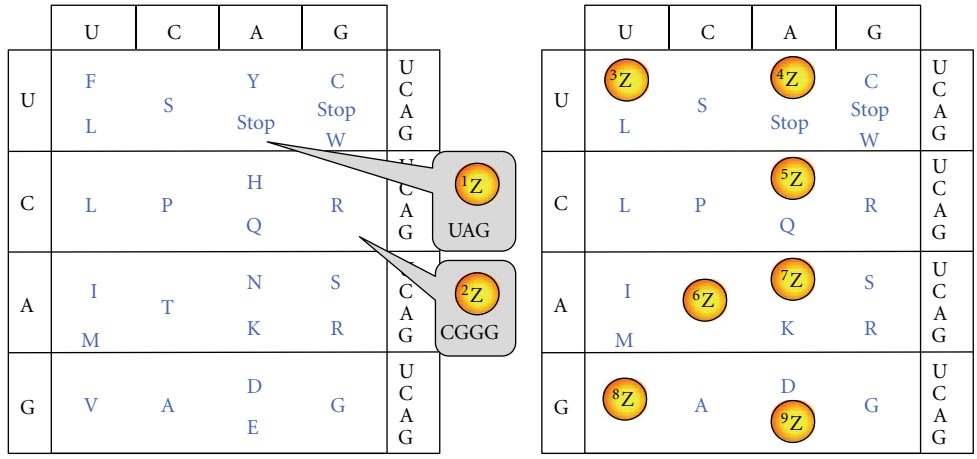
2. Genetic Code Expansion

The genetic code expansion strategy (Figure 1(a)) has been used mainly for the study of proteins rather than short peptides because it allows the site-specific incorporation of a 21st amino acid in addition to the 20 proteinogenic amino acids. In 1989, the amber suppression method, the first example of a genetic code expansion strategy, was independently reported by Schultz's group and Chamberlin's group (Figure 1(a)) [16, 17]. In these reports, amber suppressor tRNA(CUA) carrying a nonproteinogenic amino acid was synthesized by chemoenzymatic acylation, which was developed by Hecht's group [18], and the resulting aminoacyl-tRNA (aa-tRNA) was added to an *in vitro* translation system (conventional crude lysate system). Accordingly, a polypeptide containing the nonproteinogenic amino acid was synthesized from an mRNA bearing a site-specifically mutated amber UAG codon to incorporate the nonproteinogenic amino acid. Later, Chamberlin and coworkers also reported that the other nonsense codons (opal UGA codon and ochre UAA codon) could be suppressed effectively by the corresponding nonproteinogenic aa-tRNA in rabbit reticulocyte lysate [19]. The extension of this nonsense suppression method to use a four-base codon was reported by Sisido's group. In this approach, a nonproteinogenic aa-tRNA bearing an ACCU four-base anticodon was used to incorporate a nonproteinogenic amino acid at an AGGU four-base codon [20]. They also expanded this approach to other four-base [21–23] and even five-base codons [24] and showed the simultaneous incorporation of multiple nonproteinogenic amino acids into proteins [25–28]. Later, the genetic code expansion strategy was further extended from *in vitro* to *in vivo*. In 1995, Dougherty, Lester, and coworkers reported *in vivo* nonsense suppression, where UAG-containing mRNA and suppressor tRNA, which was chemically acylated with a nonproteinogenic amino acid, were injected into a *Xenopus* oocyte to incorporate the nonproteinogenic amino acid into an ion channel expressed on the surface of the oocyte [29, 30]. Further progress came with the development of orthogonal tRNA(CUA) and aminoacyl-tRNA synthetase (aaRS) pairs, which produce an orthogonal tRNA(CUA) carrying a nonproteinogenic amino acid *in vivo* under multiple turnover conditions. This technology allows us to not only synthesize a protein carrying a nonproteinogenic amino acid at a specific position in various types of cells, but also obtain a protein containing a nonproteinogenic amino acid in a larger quantity than by *in vitro* genetic code expansion [31–34]. The use of various orthogonal tRNA/aaRS pairs has allowed for the synthesis of proteins carrying various artificial functional groups, such as a biochemical group (e.g., sulfate, acetate, or methylate) [35–38], fluorescent probe [39–43], photo-cross-linker [44–49], photo-caged group [50–56], and bioorthogonal reactive group [57–63], for the study of protein structure and function. Similarly to *in vitro* genetic code expansion, *in vivo* amber suppression using orthogonal tRNA/aaRS pairs was extended to ochre and opal codon suppression [64–66] and four-base codon suppression to incorporate two nonproteinogenic amino acids simultaneously into proteins *in vivo* [67, 68].

3. Selection of Functional Nonstandard Peptides Generated by Genetic Code Expansion

The *in vivo* genetic code expansion strategy has recently been used to generate nonstandard peptide libraries. Using orthogonal aaRS/tRNA(CUA) pairs, Young et al. generated a backbone-cyclized peptide library containing a nonproteinogenic amino acid and performed in-cell selection of a nonstandard peptide inhibitor from the library (Figure 2(a)) [69]. The orthogonal aaRS/tRNA(CUA) pairs were expressed in *Escherichia coli* (*E. coli*) in order to assign L-*p*-benzoylphenylalanine to the amber codon [46]. Additionally, the split intein catalyzed ligation of proteins and peptides (SICROPPS) method, which was developed by Benkovic's group [70, 71], was used to cyclize the backbone of a ribosomally synthesized peptide library containing L-*p*-benzoylphenylalanine. The activity-based in-cell selection of human immunodeficiency virus (HIV) protease inhibitors from the cyclic peptide library was performed by linking the protease inhibitory activity of a cyclic peptide to cell viability of the host *E. coli*. The most abundant peptide obtained after two rounds of selection inhibited HIV protease activity at a low micromolar concentration (IC_{50} = 0.96 μM). Interestingly, the L-*p*-benzoylphenylalanine residue of the peptide formed a covalent Schiff-base adduct with the ε-amino group of Lys14 of the HIV protease.

Although this study using *in vivo* selection succeeded in selecting new functional nonstandard peptides, the size of the library was limited by the transformation efficiency of *E. coli*. In contrast, *in vitro* display methods, such as ribosome display [7] and mRNA display (or *in vitro* virus)



(a) Genetic code expansion (amber codon, 4-base codon) (b) Genetic code reprogramming (blank codon)

Figure 1: Generation of genetically encoded libraries of nonstandard peptides. (a) Genetic code expansion. Nonproteinogenic amino acids are assigned to the amber codon and four-base codons. (b) Reprogramming the genetic code. Nonproteinogenic amino acids are reassigned to blank codons generated by reconstructing a cell-free translation system with a reduced number of amino acids and protein factors. ^{n}Z (n = 1–9) represents one of the nonproteinogenic amino acids.

[8, 9], can generate highly diverse libraries (up to 10^{13} compounds) because library construction does not involve cell transformation. Moreover, *in vitro* genetic code expansion has an advantage over *in vivo* expansion because membrane-impermeable, metabolically unstable or cytotoxic nonproteinogenic amino acids can be used as the building blocks for genetically encoded nonstandard peptides. The first proof of concept study of an *in vitro* display selection from a library containing a nonproteinogenic amino acid was reported by Li et al. (Figure 2(b)) [72, 73]. A nonstandard peptide library displayed on mRNA was prepared by using a lysate translation system containing amber suppressor tRNA(CUA) precharged with biocytin and a puromycin-modified mRNA library containing one NNS (S = C or G) codon. The authors demonstrated that mRNA containing the UAG codon (i.e., encoding a biocytin-containing peptide) was selectively recovered from the prepared library by pulldown using streptavidin beads. The same group combined a similar genetic code expansion strategy with a posttranslational chemical cyclization method to generate a nonstandard cyclic peptide library [74]. A peptide library containing *N*-methyl-L-phenylalanine assigned to the amber codon was displayed on mRNA and chemically converted to a cyclic peptide library by linking the *N*-terminal α-amino group to the ε-amino group of a lysine residue using disuccinimidyl glutarate. This library was used to select Gαi1 binding peptides, and the best binding cyclic peptide had very high affinity against Gαi1 (2.1 nM). Although this cyclic peptide also showed higher proteolytic stability than its linear counterpart, the *in vitro*-selected peptides in this study unfortunately did not contain the nonproteinogenic amino acid. Muranaka et al. also reported the selection of a streptavidin-binding peptide from nonstandard peptide libraries prepared by combining the amber and four-base codon suppression approaches (Figure 2(c)) [75]. Messenger RNA libraries composed of nine NNK (K = U or G) codons and one VGGU (V = C, A, or G) four-base codon were used to construct nonstandard peptide libraries in the mRNA display format, and a streptavidin-binding non-standard peptide was successfully selected. The selected peptide contained only one nonproteinogenic amino acid (L-*p*-benzoylphenylalanine) residue (Figure 2(c)), even though multiple nonproteinogenic amino acids could have appeared in the sequence of the used libraries.

4. Limitation of Genetic Code Expansion for the Preparation of Nonstandard Peptide Libraries

In principle, highly modified non-standard peptides containing multiple nonproteinogenic amino acids can be produced by genetic code expansion, but this strategy actually imposes considerable restrictions on the synthesis of such peptides. In this strategy, the incorporation of nonproteinogenic amino acids competes with other normal events during translation. For example, an amber suppressor nonproteinogenic aa-tRNA(CUA) competes with endogenous release factor 1 (RF-1), which results in the production of truncated as well as full-length peptides from one mRNA template [16, 17]. Similarly, a 4-base suppressor nonproteinogenic aa-tRNA(ACCU) competes with endogenous Arg-tRNA^{Arg}(CCU) for decoding the AGG(U) codon, which results in the production of a mixture of in-frame and out-of-frame products from one mRNA template [20]. This feature makes it difficult to simultaneously incorporate multiple nonproteinogenic amino acids into a peptide since the yield of a full-length (in-frame) peptide decreases exponentially as the number of nonproteinogenic amino acids increases. Moreover, in these genetic code expansion strategies, an mRNA containing the nonsense or 4-base codon does not encode one peptide but a mixture of multiple peptide

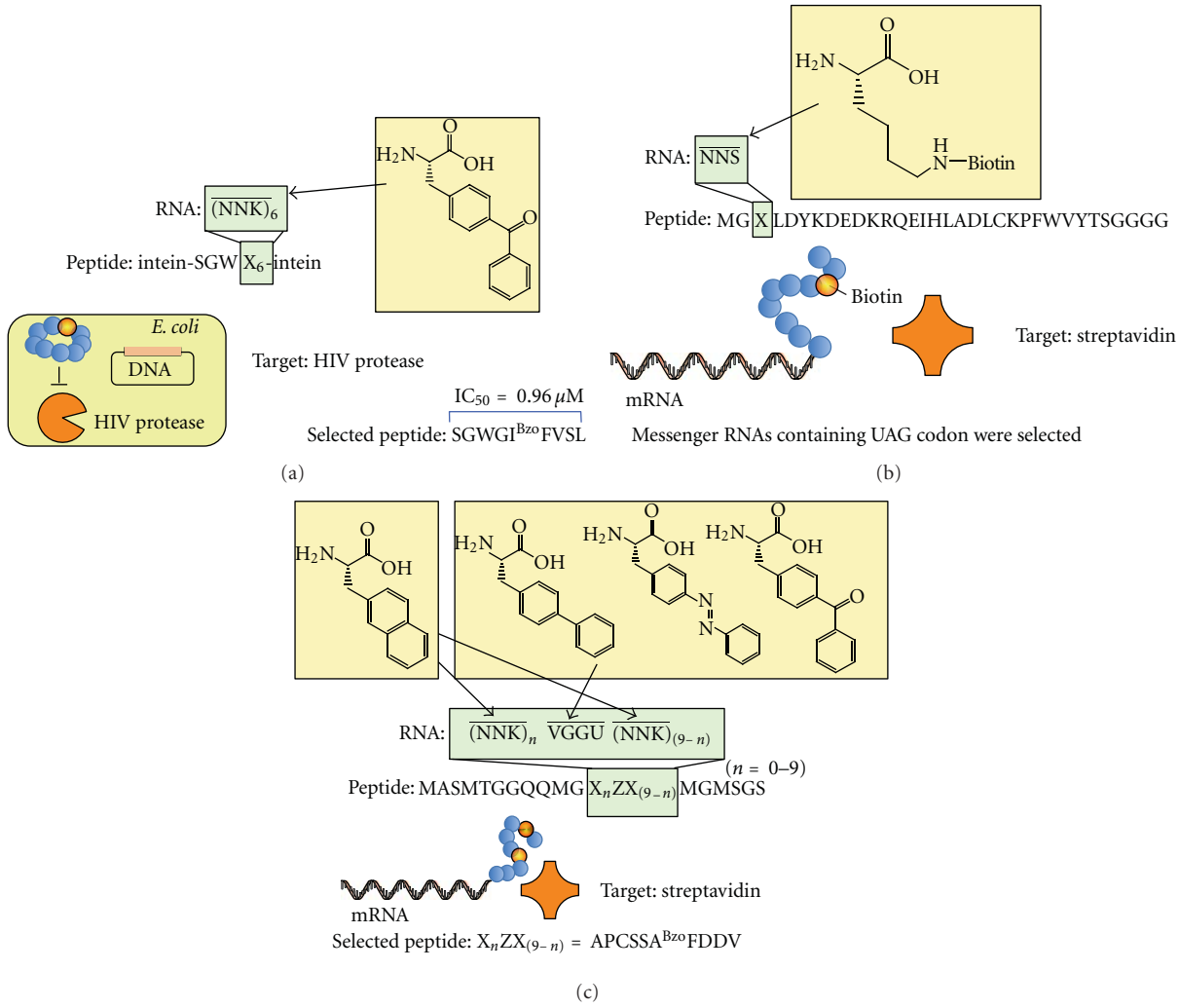


Figure 2: Selection of nonstandard peptides from libraries generated by genetic code expansion. (a) Selection of backbone-cyclized L-*p*-benzoylphenylalanine-containing peptide inhibitors. Orthogonal tRNA(CUA)/aaRS pairs were expressed in *E. coli* in order to assign L-*p*-benzoylphenylalanine to the amber codon. SICROPPS was used to cyclize the peptides. (b) Proof-of-concept study of mRNA display with a nonproteinogenic amino acid. Biocytin was assigned to the amber codon by adding biocytin-tRNA(CUA) to a rabbit reticulocyte lysate. The peptides were expressed in mRNA display format and pulled down using streptavidin beads. (c) Selection of streptavidin binding nonstandard peptides. Four nonproteinogenic amino acids were assigned to the amber codon and four-base codons (UAG, L-2-naphthylalanine; AGGU, L-*p*-biphenylalanine; CGGU, L-*p*-benzoylphenylalanine; GGGU, L-*p*-phenylazophenylalanine). All corresponding nonproteinogenic aa-tRNAs were added to an *E. coli* cell-free translation system. The peptide libraries were expressed in mRNA display format and used for *in vitro* selection against streptavidin. X represents one of the proteinogenic or nonproteinogenic amino acids, and Z represents one of the nonproteinogenic amino acids.

products. This could be a more critical problem for peptide selection because this phenomenon would make it difficult to identify an active species from selected cDNA sequences. Recently, as one of the approaches used to decrease these competing translation events, Chin and coworkers evolved an orthogonal mutant ribosome that is thought to have a decreased functional interaction with RF-1 and demonstrated the increased incorporation efficiency of a nonproteinogenic amino acid on the amber codon [76]. However, since the orthogonal mutant ribosome was still recognized by RF-1 to some extent, a mixture of full-length and truncated proteins was expressed from one UAG-containing mRNA. Similarly, Liu and coworkers prepared a ribosome with the C-terminal domain of ribosomal protein L11 to decrease RF-1-mediated translation termination; however, the complete reassignment of the amber codon was not shown [77]. Conversely, Sakamoto's group [78–80] and Wang's group [81, 82] independently reported using a different approach to knock out RF-1 from *E. coli* strains, and both groups demonstrated complete reassignment of the amber UAG codon from translation termination to nonproteinogenic amino acid incorporation *in vivo*. In Wang's approach, the RF-1 knockout strains had a substantially decreased growth rate, most likely because of the failure of translation termination at amber UAG codons in endogenous mRNAs that are essential for the growth of *E. coli*. The combination

of this RF-1 knockout approach with the recently reported genome-wide codon replacement method to replace all 314 endogenous UAG codons to UAA stop codons in *E. coli* [83] might solve this problem. Nevertheless, these RF-1 knockout strategies enabled the complete reassignment of the amber codon to a nonproteinogenic amino acid *in vivo*; however, they have only been applied to the amber codon thus far.

The difficulty of complete codon reassignment *in vivo* clearly comes from the fact that the translation events that compete with the incorporation of nonproteinogenic amino acids are required to maintain the life of the host cells. Conversely, engineering an *in vitro* translation system is not as complicated as for an *in vivo* translation system; even so, the simultaneous complete reassignment of multiple codons to different nonproteinogenic amino acids was not achieved until 2003. For example, Roberts and coworkers used a lysate translation system pretreated with aminoethanol-Sepharose to remove endogenous tRNAs, which compete with the incorporation of a nonproteinogenic amino acid at sense codons. However, a peptide synthesized by using this tRNA-depleted system still consisted of a mixture of proteinogenic and nonproteinogenic amino acids [2]. Other approaches to decrease competitive translation events against nonproteinogenic amino acid incorporation *in vitro* include heat inactivation of RF-1 [84], aptamer-mediated RF-1 inhibition [85], and antibody-mediated RF-1 inhibition [86], yet none of these resulted in complete codon reassignment. Another interesting approach is the pretreatment of a lysate translation system with a phenylalanyl-tRNA synthetase (PheRS) inhibitor, and nearly complete reassignment of the Phe sense codon to naphthylalanine was demonstrated, although its general application to multiple codons has yet to be reported [87].

5. Genetic Code Reprogramming

The first demonstration of the simultaneous complete reassignment of multiple codons to different nonproteinogenic amino acids (i.e., genetic code reprogramming) was reported by Forster et al. in 2003 (Figure 3(a)) [88]. The key feature of genetic code reprogramming is the use of a reconstituted translation system, instead of crude lysate translation system, to eliminate translation events that compete with the incorporation of no-proteinogenic amino acids. In the report by Forster et al., purified ribosomes, initiation factors, and elongation factors were mixed with fMet-tRNA^{ini}, Glu-tRNA^{Glu}, and three kinds of chemoenzymatically synthesized nonproteinogenic aa-tRNA which were assigned in the reprogrammed genetic code shown in Figure 3(a), and a non-standard peptide containing three different nonproteinogenic amino acids was synthesized. An interesting feature of their aaRS-free reconstituted translation system is that since orthogonality of the tRNA against aaRS is not required, in principle, tRNA with any body sequence and anticodon can be used to incorporate amino acids [89]. On the other hand, in their genetic code reprogramming system using chemoenzymatic tRNA-acylation, the maximum number of nonproteinogenic amino acids that were simultaneously reassigned in the reprogrammed genetic code was not more than three [88, 90–92], probably because the authors used a technically demanding chemoenzymatic acylation method to prepare the aa-tRNAs [18].

To achieve genetic code reprogramming by a more simple tRNA-acylation method, Szostak's group took advantage of the substrate tolerance of wild-type aaRSs, namely, that "wild-type" aaRSs can charge some proteinogenic amino acid analogs [93]. This feature has been used mainly to incorporate a proteinogenic amino acid analog into proteins *in vivo* (in an auxotrophic strain in many cases) in a residue-specific manner to produce heterogeneous proteins containing proteinogenic and nonproteinogenic amino acids as engineered protein material [94–97]. In the first report of genetic code reprogramming by aaRS-catalyzed acylation, Josephson et al. described a landmark example [93] in which nonproteinogenic amino acids were misacylated onto endogenous tRNAs by wild-type aaRSs in a reconstituted translation system [98] lacking the corresponding proteinogenic amino acids, and 10 different nonproteinogenic amino acids were incorporated simultaneously into a peptide as assigned in the reprogrammed genetic code (Figure 3(b)). Importantly, they also demonstrated that the same non-standard peptide could be displayed on its encoding mRNA via a puromycin linker to show the compatibility of genetic code reprogramming with mRNA display. Later, Szostak's group found that over 90 nonproteinogenic amino acids could be charged onto tRNAs by wild-type and mutant aaRSs and showed that over 50 nonproteinogenic amino acids could be ribosomally incorporated into peptides [99, 100]. They also synthesized a non-standard peptide containing 13 different nonproteinogenic amino acids. In terms of the number of different nonproteinogenic amino acids that can be incorporated simultaneously into ribosomal peptides, this non-standard peptide still holds the record. The advantage of Szostak's reprogramming method is that (1) the misacylated tRNAs can be generated *in situ* under multiple turnover conditions and (2) non-standard peptides can be expressed by merely adding nonproteinogenic amino acids to the translation system from which the corresponding proteinogenic amino acids are withdrawn. On the other hand, the number of available nonproteinogenic amino acids is limited since aaRSs can generally mischarge only those that are structurally similar to proteinogenic amino acids, and it is not possible to introduce multiple analogs of the same amino acid type at the same time (e.g., two proline analogs). Moreover, since proteinogenic amino acids are more acceptable to both aaRS acylation and ribosomal translation, even a tiny amount of proteinogenic amino acid contamination causes the production of undesired product that consists of a mixture of proteinogenic and nonproteinogenic amino acids [93].

As another approach to achieve genetic code reprogramming by a more general tRNA-acylation method, Murakami et al. reported a highly flexible-ribozyme- (flexizyme-) based tRNA-acylation system [101]. In the first demonstration of genetic code reprogramming using flexizymes, three sense codons were reassigned to three nonproteinogenic amino acids charged onto orthogonal tRNAs (Figure 3(c)).

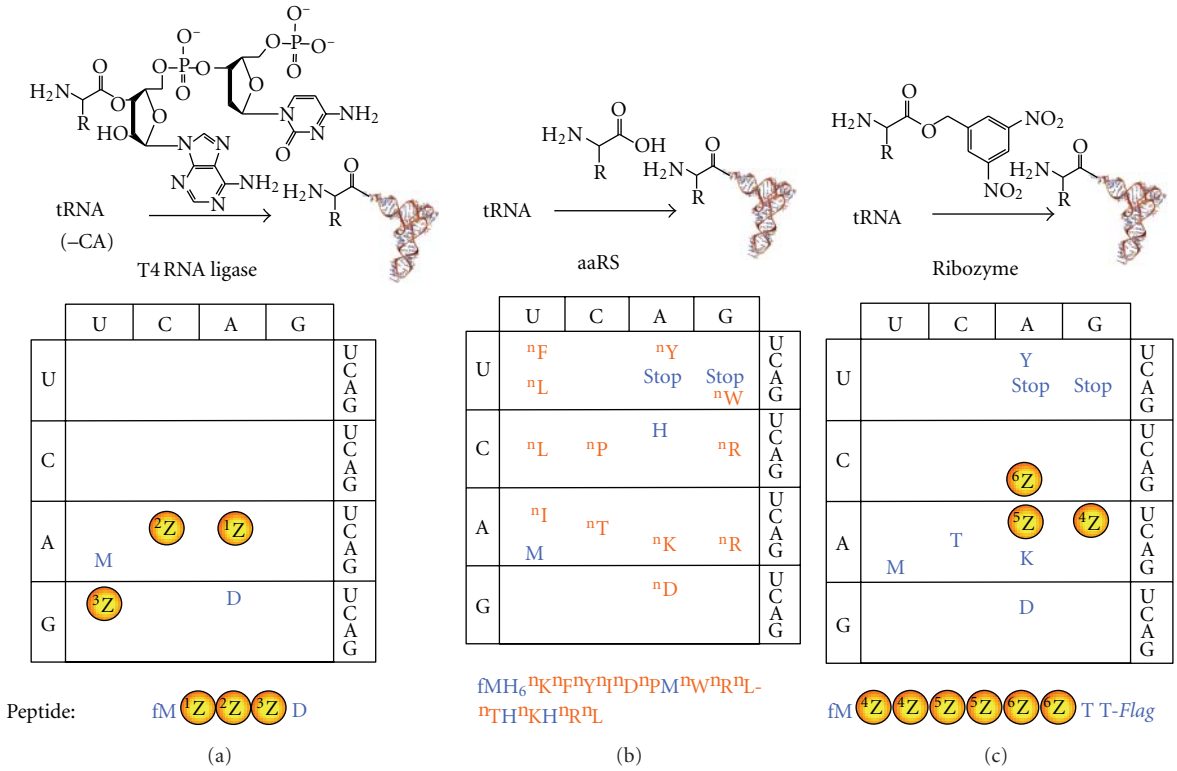


Figure 3: Comparison of the aminoacylation methods used for genetic code reprogramming. (a) Chemoenzymatic acylation method. The dinucleotide carrying the nonproteinogenic amino acid (pdCpA-aa) was enzymatically ligated to tRNA lacking 3′-nucleotides (pCpA). Aminoacyl-tRNAs prepared by the chemoenzymatic acylation method were added to an aaRS-free reconstituted translation system to synthesize a non-standard peptide containing three nonproteinogenic residues (^{1}Z, L-propargylglycine; ^{2}Z, L-*O*-methylserine; ^{3}Z, L-allylglycine). (b) aaRS tRNA-misacylation method. Aminoacyl-tRNAs carrying nonproteinogenic amino acids were enzymatically generated in a reconstituted translation system. A non-standard peptide containing 13 nonproteinogenic residues was synthesized (^{N}X, an analog of proteinogenic amino acid X). (c) Flexizyme-based acylation method. An activated amino acid was mixed with tRNAs and flexizyme to synthesize aminoacyl-tRNA. Aminoacyl-tRNAs prepared by the flexizyme-based method were added to a reconstituted translation system to synthesize a non-standard peptide containing six nonproteinogenic residues (^{4}Z, L-acetyllysine; ^{5}Z, L-citrulline; ^{6}Z, L-*p*-iodophenylalanine).

According to the reprogrammed genetic code, a 17 mer non-standard peptide possessing six nonproteinogenic amino acids was synthesized in a genetically encoded manner.

The prototype [102] of the flexizyme used in genetic code reprogramming, which was evolved from a random RNA pool, has expanded its acceptability of tRNAs [103, 104] and amino acids for aminoacylation and turned into highly flexible ribozymes for tRNA aminoacylation [101, 105, 106]. Because the flexizymes recognize their cognate aromatic group, they are able to charge a wide variety of not only L-amino acids with nonproteinogenic side chains [3, 107–111], but also amino acids with an altered backbone such as *N*-alkyl amino acids [112–114], *N*-acyl amino acids [115], D-amino acids [116], *β*-amino acids, oligopeptides [117, 118], and even hydroxy acids [119, 120]. Moreover, as the preparation of aa-tRNAs by flexizymes is simple and straightforward, it should facilitate the parallel preparation of various aa-tRNAs in a high-throughput manner, which enables us to carry out genetic code reprogramming conveniently. In fact, more than 160 kinds of (amino) acids were charged onto tRNAs using the flexizyme system in the past few years. Since flexizymes also recognize conserved 3′-terminal three bases (CCA-3′) of tRNA, they can aminoacylate virtually any tRNA [103, 104]. Recently, Cornish and coworkers took advantage of this tRNA flexibility of flexizymes to aminoacylate base-modified tRNAs with their noncognate amino acids and used the prepared aa-tRNAs to investigate differences in ribosome recognition between the amino acid charged onto cognate and noncognate tRNAs [121].

Bearing the advantages and disadvantages of each genetic code reprogramming method in mind, scientists in this research field moved onto the next step to select functional non-standard peptides from genetically encoded non-standard peptide libraries.

6. Selection of Functional Nonstandard Peptides Generated by Genetic Code Reprogramming

The first model study of the *in vitro* display selection of a non-standard peptide prepared with a reconstituted translation system was reported by Forster et al. [122]. Five kinds of biotinylated peptides bearing different lengths of polypeptide

spacer were synthesized in a ribosome display format, and biotinylated peptides with a spacer long enough to exit from the ribosome tunnel were selectively pulled down using streptavidin beads. However, in this paper, a diverse non-standard peptide library was not constructed and a novel functional non-standard peptide was not obtained. Very recently, several papers have been published on the selection of functional non-standard peptides by combining genetic code reprogramming with mRNA display selection. In this section, we review the preparation of non-standard libraries and the results of the selection in these reports.

In 2011, Suga's group published an important study in which highly diverse non-standard peptide libraries containing multiple different nonproteinogenic amino acids were constructed in an mRNA display format, and a novel non-standard peptide was selected from the libraries [5, 123]. Previously, the same group used the flexizyme system to examine the synthesis of various non-standard peptides in a ribosomal translation system [124] and reported the following findings: (1) an *N*-chloroacetyl-amino acid residue and cysteine residue on the same peptide were spontaneously reacted *in situ* using a reconstituted translation system to give a thioether-cyclized peptide without an intermolecular side reaction between the *N*-terminal chloroacetyl group on the peptide and the sulfhydryl group of the other translation component such as a cysteine monomer and DTT [3, 108, 112, 113, 115, 125]; (2) *N*-methyl amino acids with an aromatic side chain or noncharged and nonbulky side chains are efficiently incorporated into peptides by the ribosome, and multiple *N*-methyl amino acids can be incorporated simultaneously into a peptide to give various sequences of thioether-cyclized *N*-methyl-peptides [112]; and (3) the translation initiation apparatus accepts D-amino acids with hydrophobic side chain as relatively good initiators, and pre-*N*-acylation of D-aa-tRNA dramatically increases the efficiency of translation initiation [116]. On the basis of these studies, Yamagishi et al. generated non-standard peptide libraries in which *N*-chloroacetyl-D-tryptophan was reassigned to the AUG start codon and *N*-methyl-phenylalanine, *N*-methyl-serine, *N*-methyl-glycine, and *N*-methyl-alanine were individually reassigned to the UUU, CUU, AUU, and GCU codons, respectively (Figure 4(a)) [5]. The thioether-cyclized *N*-methyl-peptide libraries containing 8–15 random residues were expressed in the mRNA display format by using a translation system reconstituted from a limited set of proteinogenic amino acids and their cognate aaRSs. After reverse transcription, the non-standard peptide libraries were used without any purification for *in vitro* selection against site-specifically biotinylated E6-associated protein (E6AP) together with negative selection using E6AP-free beads. Significantly, unlike the previous report of *in vitro* selection from cyclic peptide library containing an *N*-methyl amino acid with genetic code expansion where the selected peptides did not contain the *N*-methyl amino acid [74], all of the peptides obtained after this *in vitro* selection contained multiple different *N*-methyl amino acids inside the macrocyclized peptide ring. The most abundant cyclic *N*-methyl-peptide among the sequenced clones was chemically synthesized and further characterized. The cyclic *N*-methyl-peptide exhibited high affinity for E6AP (0.6 nM), and despite the fact that the peptide sequence was obtained by affinity selection, it inhibited self-ubiquitination of E6AP *in vitro*. More importantly, removing all of the *N*-methyl groups from the backbone of the cyclic *N*-methyl-peptide dramatically decreased its affinity for E6AP, suggesting that such a cyclic *N*-methyl-peptide is nearly impossible to obtain using conventional approaches, such as the combination of selection from standard peptide libraries and following *N*-methyl modification of the selected peptide. Also, the cyclic *N*-methyl peptide showed 200-times higher affinity than its linear counterpart, which demonstrates that macrocyclization of the *N*-methyl-peptide is important for its binding to E6AP. Moreover, it was shown that both the *N*-methylated backbone and macrocyclization of the cyclic *N*-methyl peptide also contributed to the stability of the cyclic *N*-methyl-peptide against peptidases in human plasma.

The same research group also obtained protein kinase Akt2 inhibitors by mRNA display selection from thioether-cyclized peptide libraries generated by using the flexizyme system (Figure 4(b)) [126]. Peptide libraries containing 4–12 random residues cyclized with either L or D isomers of *N*-chloroacetyl-tyrosine reassigned to the initiation AUG codon were expressed and used for selection against Akt2 immobilized on Ni-NTA beads together with negative selection against Akt2-free beads to remove His-tagged translation components and undesired bead binders. After six rounds of selection, the selected cyclic L- and D-peptides converged to completely distinct sequences, indicating that one configurational difference in the macrocyclized peptides may provide different sequence spaces. In this study, even though enzymatically inactive Akt2 was used for selection, some of the selected peptides showed potent inhibitory activity against active Akt2. On the other hand, despite the fact that the recovery rate after the six rounds of selection between the enriched L- and D-peptide display libraries was at the same level, the most abundant L- and D-peptides showed a different level of inhibitory activity. In addition, the frequency of the selected peptides did not correlate with the inhibitory potency of the selected peptides, indicating the difficulty in identifying the best inhibitor by using only binding-based selection. Even so, the authors successfully obtained potent Akt2 inhibitors by screening chemically synthesized multiple peptides discovered by *in vitro* selection. Interestingly, *in vitro* inhibition assays of the selected L-peptides revealed not only high kinase family specificity but also high Akt isoform specificity, for example, Akt2 (IC_{50} = 110 nM) over Akt1 ($IC_{50} \geq 25{,}000$ nM) and Akt3 (IC_{50} = 4,200 nM) for Pakti-L1 peptide.

To obtain strong inhibitors by using binding-based *in vitro* selection, Morimoto et al. performed mRNA display selection from cyclic peptide libraries containing *N*-6-trifluoroacetyl-L-lysine (^{CF3}K), a nonproteinogenic amino acid inhibitor of its target enzyme (Figure 4(c)) [127]. In this paper, in addition to each enantiomer of *N*-chloroacetyl-tyrosine at the initiation AUG codon, ^{CF3}K was simultaneously reassigned to the elongation AUG codon. Thioether-cyclized ^{CF3}K-peptide libraries containing 7–11 random residues were prepared and used to select against

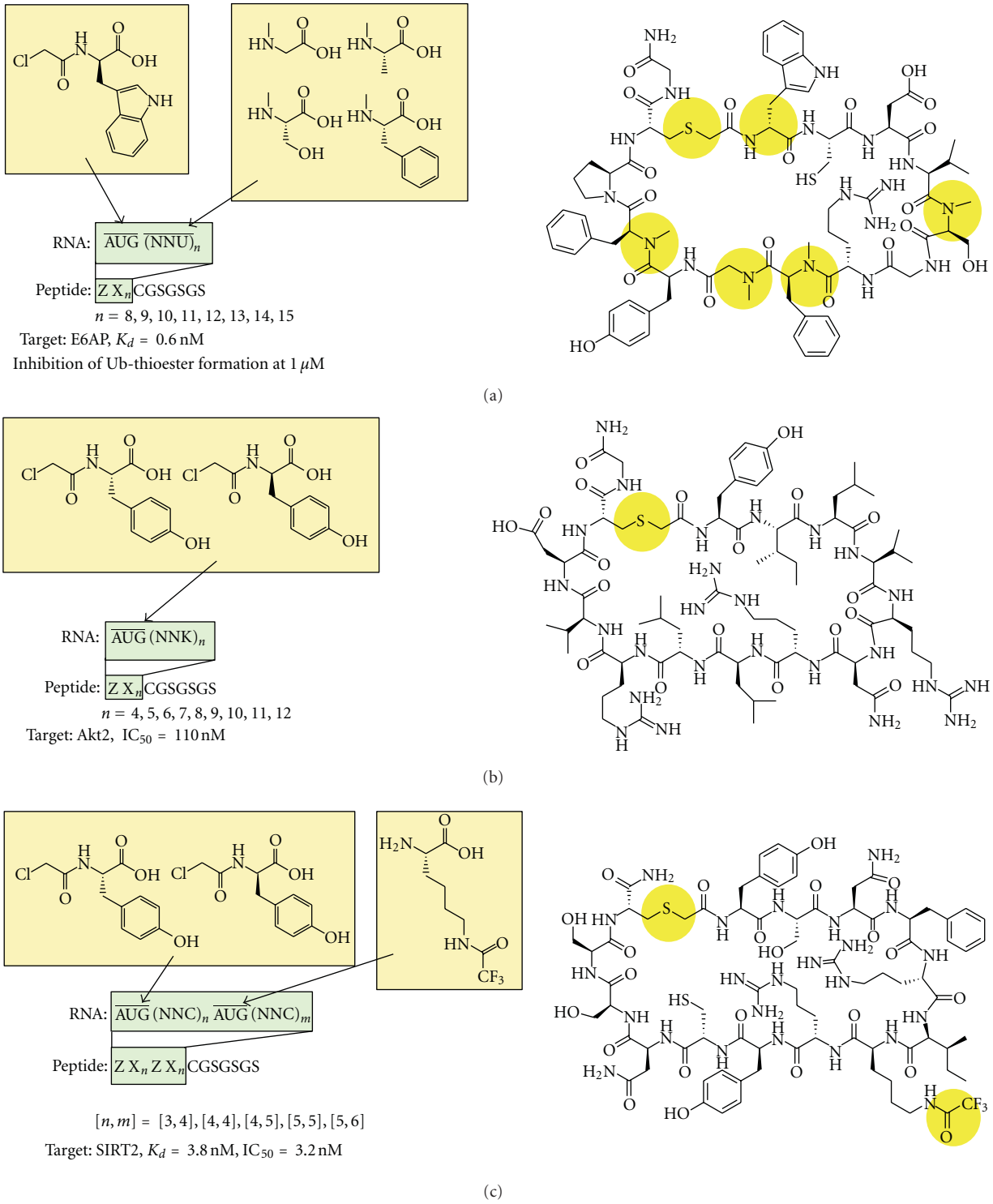


Figure 4: Selection of nonstandard peptides from libraries generated by flexizyme-based genetic code reprogramming. The flexizyme system was used to acylate tRNAs with nonproteinogenic amino acids. The nonstandard peptide libraries were expressed in a reconstituted cell-free translation system in the mRNA display format. (a)Selection of cyclic *N*-methyl-peptides against E6AP. *N*-Chloroacetyl-D-tryptophan was reassigned to AUG, and four *N*-methyl-L-amino acids were independently reassigned to one of the NNU codons. (b)Selection of thioether-cyclized peptides against Akt2. Each enantiomer of *N*-chloroacetyltyrosine was reassigned to the initiation AUG codon. (c)Selection of cyclic peptides containing *N*-6-trifluoroacetyl-L lysine against sirtuin 2 (SIRT2). Each enantiomer of *N*-chloroacetyl-tyrosine was reassigned to the initiation AUG codon and *N*-6-trifluoroacetyl-L-lysine (^{CF3}K) was reassigned to the elongation AUG codon.

NAD-dependent deacetylase sirtuin-2 (SIRT2) immobilized on Ni-NTA beads. After six or seven rounds of selection, although various sequences were seen at the randomized positions, the R-I/V-^{CF3}K-RY sequence was preferred in the vicinity of the *N*-6-trifluoroacetyl-L-lysine residue. After confirming the inhibition of *in vitro*-selected cyclic ^{CF3}K-peptides expressed in the translation system, *in vitro* binding and inhibition assays were performed against two chemically synthesized representative cyclic ^{CF3}K-peptides. Importantly, the inhibitory activity of the cyclic ^{CF3}K-peptides against SIRT2 correlated well with their binding activity toward SIRT2, and the peptides exhibited high inhibitory activity against SIRT2 (IC_{50} = 3-4 nM). On the other hand, the cyclic ^{CF3}K-peptides showed a similar IC_{50} value against SIRT2 to their linear versions (5-6 nM), and the authors showed that the R-I/V-^{CF3}K-RY short peptide segment contributed strongly to their SIRT2 inhibitory activity, whereas the constrained macrocyclic structure did not make much contribution. Similarly to the *in vitro*-selected Akt2 inhibitor peptides, *in vitro*-selected SIRT2 inhibitor peptides also showed isoform specificity against SIRT2 (IC_{50} = 3.2 nM) over SIRT1 (IC_{50} = 47 nM) and SIRT3 (IC_{50} = 480 nM).

By taking a different genetic code reprogramming method from Suga's group, Szostak's group demonstrated mRNA display selection from a highly modified non-standard peptide library containing 12 different proteinogenic amino acid analogs (Figure 5(a)) [128]. A peptide library containing 10 random residues between 2 Cys residues was expressed and displayed on encoding mRNAs in the reconstituted translation system with 8 proteinogenic amino acids and 12 nonproteinogenic amino acids. The peptide-mRNA fusion libraries were then cyclized by the addition of 1,3-di(bromomethyl)benzene after oligo-dT purification. After reverse transcription and Ni-NTA purification, the non-standard cyclic peptide library was subjected to selection against biotinylated thrombin protease, and active species were eluted selectively from streptavidin beads using the thrombin inhibitor hirudin. After 10 rounds of selection, several conserved sequences were observed, and importantly, the selected peptides had 50% nonproteinogenic amino acids in sequences that were distinct from the sequences obtained by selection using proteinogenic amino acids. The two most abundant peptides synthesized by the ribosomal translation system showed high affinity (K_d = 4.5 and 20 nM) and inhibitory activity (K_i^{app} = 23 and 35 nM) against thrombin, while the parent linear peptides did not. Moreover, the "proteinogenic analogs" of the two peptides did not bind to thrombin (K_d > 500 nM), suggesting that the incorporation of side-chain-modified amino acids may provide a different sequence space from that obtained with proteinogenic amino acids.

By using a similar approach, Hofmann et al. reported mRNA display selection from lanthionine-containing cyclic peptide libraries generated by the ribosomal incorporation of L-4-selenalysine (Figure 5(b)) [129]. 4-Selenalysine was reassigned to the AAA codon by adding 4-selenalysine instead of lysine to a reconstituted translation system, and a library containing nine random residues was generated. After oligo-dT purification, the 4-selenalysine residue was oxidized and converted to dehydroalanine, which was then reacted with the sulfhydryl group on the cysteine residue [130, 131]. After reverse transcription and Ni-NTA purification, the resulting cyclic lantipeptide libraries were used for selection against chemically biotinylated sortase A. After five rounds of selection, several conserved sequences were obtained, and the stereochemistry of the selected lantipeptides was determined through chemical peptide synthesis. While the peptides from one family bound sortase A with an affinity of 3–32 μM, none of the peptides from the other family bound it, suggesting that the sequences in the second family were selected by a different selection pressure from sortase A binding. Moreover, the best sortase A binding peptide did not inhibit its activity.

Taken together, genetic code reprogramming has been successfully combined with *in vitro* display for the selection of non-standard peptide libraries, and it has been shown that most non-standard peptides obtained by this method possess inhibitory activity *in vitro*. Future work will involve demonstrating the activity of the obtained non-standard peptides *in vivo* (not only in cultured cells, but also in an animal model).

7. Conclusions and Perspectives

In this paper, we have described the current strategies for the generation of genetically encoded libraries of non-standard peptides and the application of these strategies to the selection of novel functional non-standard peptides. The potential of non-standard peptides has markedly increased following the development of technologies that can generate highly diverse libraries of drug-like non-standard peptides. However, although the selection of several non-standard peptides has been demonstrated, some problems remain to be addressed. One is that there are some limitations to the types of nonproteinogenic amino acids that can be used in a ribosomal translation system. Although the translation system accepts a variety of side-chain-modified amino acids [22, 132], backbone-modified amino acids are less compatible with a ribosomal translation system. For example, Tan et al. showed that D-amino acids (D-alanine, D-phenylalanine) and β-amino acids (homo-L-alanine, homo-L-phenylalanine) were not compatible with their reconstituted translation system [133], and Hartman et al. also showed that several β-amino acids were not compatible with their reconstituted translation system [100]. We believe that more precise experiments will be required to determine the compatibility of D- and β-amino acids with ribosomal translation systems. Following this reevaluation of nonproteinogenic amino acids, further studies on the engineering of translation factors, such as the ribosome [134–136], elongation factor Tu [137, 138], and tRNA [139–141], will be required to overcome these limitations. The second problem is the low cell permeability of peptides. Many potential drug targets are located inside cells, requiring that non-standard peptides have good cell permeability. Since the *N*-alkyl group increases the permeability of a peptide, the incorporation of *N*-alkyl-amino acids into the genetic code might address

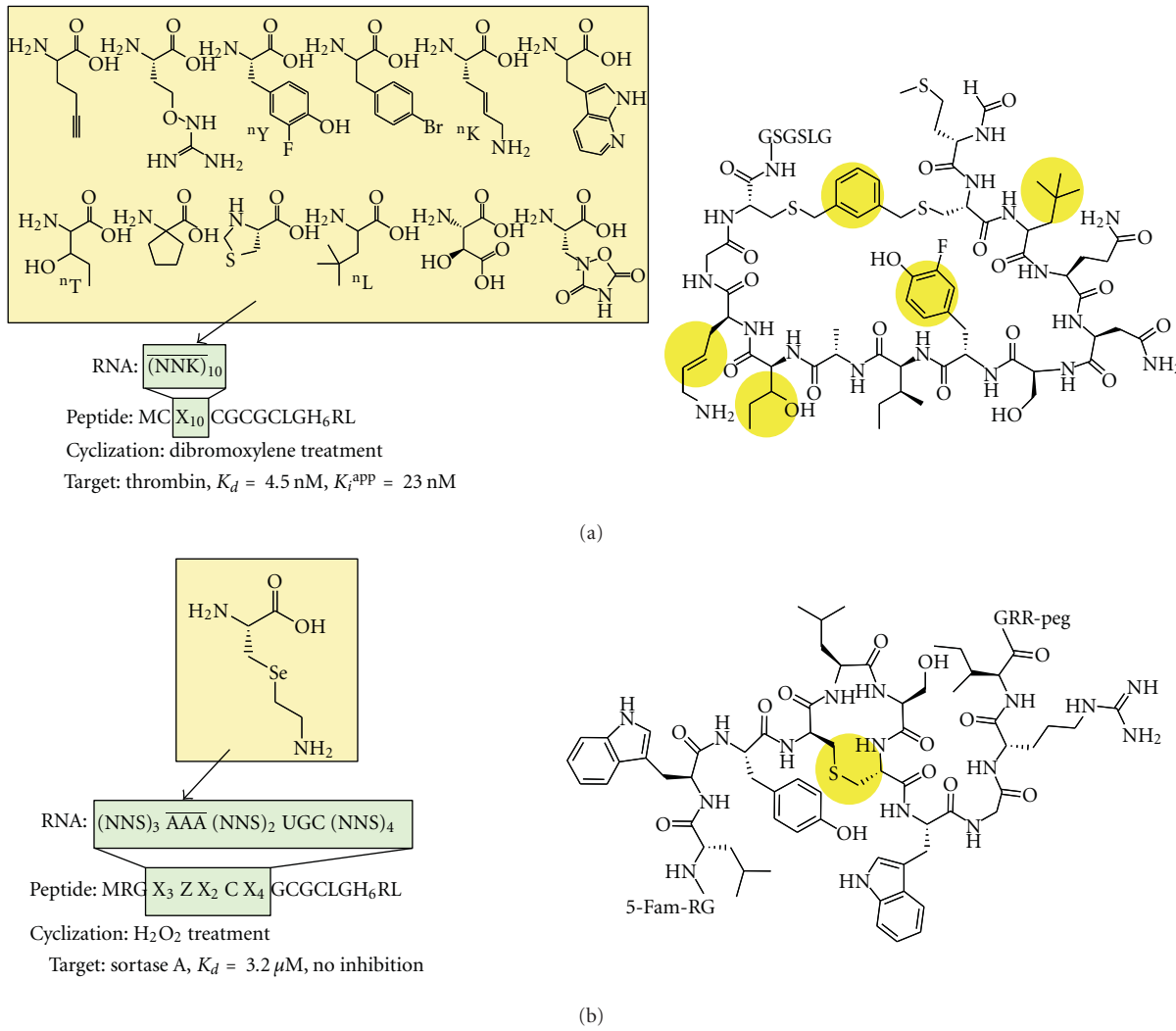


Figure 5: Selection of nonstandard peptides from libraries generated by aaRS-based genetic code reprogramming. Nonproteinogenic amino acids were charged onto tRNA by aaRS misacylation. The non-standard peptide libraries were expressed in a reconstituted cell-free translation system in the mRNA display format. (a) Selection of non-standard peptides from libraries containing 12 different nonproteinogenic amino acids (12 proteinogenic amino acid analogs). The nonproteinogenic amino acids were reassigned to NNK codons. The peptides were cyclized by the addition of 1,3-di(bromomethyl)benzene to react with the sulfhydryl groups on the two cysteine residues of the peptides. The cyclic peptide library, containing 10 random residues, was used for the selection of thrombin inhibitors. (b) Selection of sortase A binding lantipeptides. L-4-selenalysine was reassigned to the AAA codon. The L-4-selenalysine residue was oxidized and converted to dehydroalanine, which was used for cyclization of the peptide. The cyclic peptide library, containing nine random residues, was used for selection.

this problem [112, 113, 142, 143]. Another problem is the fact that the mRNA display method used for selection from highly diverse non-standard peptide libraries is still a laborious and time-consuming technique involving multiple steps. A method to prepare peptide/DNA or peptide/mRNA complex libraries continuously from the corresponding DNA libraries such as CIS display or ribosome display may allow for high-speed library preparation, even though these methods have not been applied to the generation of a highly diverse non-standard peptide library [7, 144]. Nevertheless, it is clear that more improvements of *in vitro* display techniques are required for the development of high-speed non-standard peptide selection.

Genetic code expansion and reprogramming strategies enable us to generate genetically encoded libraries of non-standard peptides. The value of non-standard peptide libraries will increase as these strategies are developed further. Selection from the genetically encoded non-standard peptide libraries will provide new ways to generate various novel functional peptides that are valuable for biological, therapeutic, and diagnostic applications.

Acknowledgments

H. Murakami received funding support from the Japan Society for the Promotion of Science and the Naito Foundation.

T. Kawakami received funding support from the Human Frontier Science Program long-term fellowship.

References

[1] S. M. Miller, R. J. Simon, S. Ng, R. N. Zuckermann, J. M. Kerr, and W. H. Moos, "Comparison of the proteolytic susceptibilities of homologous L-amino acid, D-amino acid, and N-substituted glycine peptide and peptoid oligomers," *Drug Development Research*, vol. 35, no. 1, pp. 20–32, 1995.

[2] A. Frankel, S. W. Millward, and R. W. Roberts, "Encodamers: unnatural peptide oligomers encoded in RNA," *Chemistry and Biology*, vol. 10, no. 11, pp. 1043–1050, 2003.

[3] Y. Sako, Y. Goto, H. Murakami, and H. Suga, "Ribosomal synthesis of peptidase-resistant peptides closed by a nonreducible inter-side-chain bond," *ACS Chemical Biology*, vol. 3, no. 4, pp. 241–249, 2008.

[4] O. Ovadia, S. Greenberg, J. Chatterjee et al., "The effect of multiple N-methylation on intestinal permeability of cyclic hexapeptides," *Molecular Pharmaceutics*, vol. 8, no. 2, pp. 479–487, 2011.

[5] Y. Yamagishi, I. Shoji, S. Miyagawa et al., "Natural product-like macrocyclic N-methyl-peptide inhibitors against a ubiquitin ligase uncovered from a ribosome-expressed de novo library," *Chemistry & Biology*, vol. 18, no. 12, pp. 1562–1570, 2011.

[6] G. P. Smith, "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface," *Science*, vol. 228, no. 4705, pp. 1315–1317, 1985.

[7] L. C. Mattheakis, R. R. Bhatt, and W. J. Dower, "An in vitro polysome display system for identifying ligands from very large peptide libraries," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 19, pp. 9022–9026, 1994.

[8] N. Nemoto, E. Miyamoto-Sato, Y. Husimi, and H. Yanagawa, "In vitro virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro," *FEBS Letters*, vol. 414, no. 2, pp. 405–408, 1997.

[9] R. W. Roberts and J. W. Szostak, "RNA-peptide fusions for the in vitro selection of peptides and proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 23, pp. 12297–12302, 1997.

[10] C. Heinis, T. Rutherford, S. Freund, and G. Winter, "Phage-encoded combinatorial chemical libraries based on bicyclic peptides," *Nature Chemical Biology*, vol. 5, no. 7, pp. 502–507, 2009.

[11] A. Angelini, L. Cendron, S. Chen et al., "Bicyclic peptide inhibitor reveals large contact interface with a protease target," *ACS Chemical Biology*, vol. 7, no. 5, pp. 817–821, 2012.

[12] V. Baeriswyl, H. Rapley, L. Pollaro et al., "Bicyclic peptides with optimized ring size inhibit human plasma kallikrein and its orthologues while sparing paralogous proteases," *ChemMedChem*, vol. 7, no. 7, pp. 1173–1176, 2012.

[13] T. Bosma, A. Kuipers, E. Bulten, L. de Vries, R. Rink, and G. N. Moll, "Bacterial display and screening of posttranslationally thioether-stabilized peptides," *Applied and Environmental Microbiology*, vol. 77, no. 19, pp. 6794–6801, 2011.

[14] S. Li and R. W. Roberts, "A novel strategy for in vitro selection of peptide-drug conjugates," *Chemistry and Biology*, vol. 10, no. 3, pp. 233–239, 2003.

[15] A. Angelini and C. Heinis, "Post-translational modification of genetically encoded polypeptide libraries," *Current Opinion in Chemical Biology*, vol. 15, no. 3, pp. 355–361, 2011.

[16] J. D. Bain, C. G. Glabe, T. A. Dix, A. R. Chamberlin, and E. S. Diala, "Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide," *Journal of the American Chemical Society*, vol. 111, no. 20, pp. 8013–8014, 1989.

[17] C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, and P. G. Schultz, "A general method for site-specific incorporation of unnatural amino acids into proteins," *Science*, vol. 244, no. 4901, pp. 182–188, 1989.

[18] T. G. Heckler, L. H. Chang, Y. Zama, T. Naka, M. S. Chorghade, and S. M. Hecht, "T4 RNA ligase mediated preparation of novel "chemically misacylated" tRNAPheS," *Biochemistry*, vol. 23, no. 7, pp. 1468–1473, 1984.

[19] J. D. Bain, E. S. Diala, C. G. Glabe et al., "Site-specific incorporation of nonnatural residues during in vitro protein biosynthesis with semisynthetic aminoacyl-tRNAs," *Biochemistry*, vol. 30, no. 22, pp. 5411–5421, 1991.

[20] T. Hohsaka, Y. Ashizuka, H. Murakami, and M. Sisido, "Incorporation of nonnatural amino acids into streptavidin through in vitro frame-shift suppression," *Journal of the American Chemical Society*, vol. 118, no. 40, pp. 9778–9779, 1996.

[21] H. Murakami, T. Hohsaka, Y. Ashizuka, and M. Sisido, "Site-directed incorporation of p-nitrophenylalanine into streptavidin and site-to-site photinduced electron transfer from a pyrenyl group to a nitrophenyl group on the protein framework," *Journal of the American Chemical Society*, vol. 120, no. 30, pp. 7520–7529, 1998.

[22] T. Hohsaka, D. Kajihara, Y. Ashizuka, H. Murakami, and M. Sisido, "Efficient incorporation of nonnatural amino acids with large aromatic groups into streptavidin in in vitro protein synthesizing systems," *Journal of the American Chemical Society*, vol. 121, no. 1, pp. 34–40, 1999.

[23] T. Hohsaka, Y. Ashizuka, H. Taira, H. Murakami, and M. Sisido, "Incorporation of nonnatural amino acids into proteins by using various four-base codons in an *Escherichia coli* in vitro translation system," *Biochemistry*, vol. 40, no. 37, pp. 11060–11064, 2001.

[24] T. Hohsaka, Y. Ashizuka, H. Murakami, and M. Sisido, "Five-base codons for incorporation of nonnatural amino acids into proteins," *Nucleic Acids Research*, vol. 29, no. 17, pp. 3646–3651, 2001.

[25] T. Hohsaka, Y. Ashizuka, H. Sasaki, H. Murakami, and M. Sisido, "Incorporation of two different nonnatural amino acids independently into a single protein through extension of the genetic code," *Journal of the American Chemical Society*, vol. 121, no. 51, pp. 12194–12195, 1999.

[26] M. Taki, T. Hohsaka, H. Murakami, K. Taira, and M. Sisido, "Position-specific incorporation of a fluorophore—quencher pair into a single streptavidin through orthogonal four-base codon/anticodon pairs," *Journal of the American Chemical Society*, vol. 124, no. 49, pp. 14586–14590, 2002.

[27] T. Ohtsuki, T. Manabe, and M. Sisido, "Multiple incorporation of non-natural amino acids into a single protein using tRNAs with non-standard structures," *FEBS Letters*, vol. 579, no. 30, pp. 6769–6774, 2005.

[28] D. Kajihara, R. Abe, I. Iijima, C. Komiyama, M. Sisido, and T. Hohsaka, "FRET analysis of protein conformational change through position-specific incorporation of fluorescent amino acids," *Nature Methods*, vol. 3, no. 11, pp. 923–929, 2006.

[29] M. W. Nowak, P. C. Kearney, J. R. Sampson et al., "Nicotinic receptor binding site probed with unnatural amino acid incorporation in intact cells," *Science*, vol. 268, no. 5209, pp. 439–442, 1995.

[30] D. L. Beene, D. A. Dougherty, and H. A. Lester, "Unnatural amino acid mutagenesis in mapping ion channel function," *Current Opinion in Neurobiology*, vol. 13, no. 3, pp. 264–270, 2003.

[31] L. Wang, A. Brock, B. Herberich, and P. G. Schultz, "Expanding the genetic code of *Escherichia coli*," *Science*, vol. 292, no. 5516, pp. 498–500, 2001.

[32] C. C. Liu and P. G. Schultz, "Adding new chemistries to the genetic code," *Annual Review of Biochemistry*, vol. 79, pp. 413–444, 2010.

[33] Q. Wang, A. R. Parrish, and L. Wang, "Expanding the genetic code for biological studies," *Chemistry and Biology*, vol. 16, no. 3, pp. 323–336, 2009.

[34] L. Davis and J. W. Chin, "Designer proteins: applications of genetic code expansion in cell biology," *Nature Reviews Molecular Cell Biology*, vol. 13, no. 3, pp. 168–182, 2012.

[35] C. C. Liu and P. G. Schultz, "Recombinant expression of selectively sulfated proteins in *Escherichia coli*," *Nature Biotechnology*, vol. 24, no. 11, pp. 1436–1440, 2006.

[36] H. Neumann, S. Y. Peak-Chew, and J. W. Chin, "Genetically encoding Nε-acetyllysine in recombinant proteins," *Nature Chemical Biology*, vol. 4, no. 4, pp. 232–234, 2008.

[37] D. P. Nguyen, M. M. Garcia Alai, P. B. Kapadnis, H. Neumann, and J. W. Chin, "Genetically encoding Nε-methyl-L-lysine in recombinant histones," *Journal of the American Chemical Society*, vol. 131, no. 40, pp. 14194–14195, 2009.

[38] D. P. Nguyen, M. M. G. Alai, S. Virdee, and J. W. Chin, "Genetically directing ε-N, N-dimethyl-l-lysine in recombinant histones," *Chemistry and Biology*, vol. 17, no. 10, pp. 1072–1076, 2010.

[39] D. Summerer, S. Chen, N. Wu, A. Deiters, J. W. Chin, and P. G. Schultz, "A genetically encoded fluorescent amino acid," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 26, pp. 9785–9789, 2006.

[40] J. Wang, J. Xie, and P. G. Schultz, "A genetically encoded fluorescent amino acid," *Journal of the American Chemical Society*, vol. 128, no. 27, pp. 8738–8739, 2006.

[41] W. Wang, J. K. Takimoto, G. V. Louie et al., "Genetically encoding unnatural amino acids for cellular and neuronal studies," *Nature Neuroscience*, vol. 10, no. 8, pp. 1063–1072, 2007.

[42] V. K. Lacey, A. R. Parrish, S. Han et al., "A fluorescent reporter of the phosphorylation status of the substrate protein STAT3," *Angewandte Chemie International Edition*, vol. 50, no. 37, pp. 8692–8696, 2011.

[43] H. S. Lee, J. Guo, E. A. Lemke, R. D. Dimla, and P. G. Schultz, "Genetic incorporation of a small, environmentally sensitive, fluorescent probe into proteins in Saccharomyces cerevisiae," *Journal of the American Chemical Society*, vol. 131, no. 36, pp. 12921–12923, 2009.

[44] N. Hino, Y. Okazaki, T. Kobayashi, A. Hayashi, K. Sakamoto, and S. Yokoyama, "Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid," *Nature Methods*, vol. 2, no. 3, pp. 201–206, 2005.

[45] N. Hino, M. Oyama, A. Sato et al., "Genetic incorporation of a photo-crosslinkable amino acid reveals novel protein complexes with GRB2 in mammalian cells," *Journal of Molecular Biology*, vol. 406, no. 2, pp. 343–353, 2011.

[46] J. W. Chin, A. B. Martin, D. S. King, L. Wang, and P. G. Schultz, "Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 17, pp. 11020–11024, 2002.

[47] J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, and P. G. Schultz, "Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*," *Journal of the American Chemical Society*, vol. 124, no. 31, pp. 9026–9027, 2002.

[48] I. S. Farell, R. Toroney, J. L. Hazen, R. A. Mehl, and J. W. Chin, "Photo-cross-linking interacting proteins with a genetically encoded benzophenone," *Nature Methods*, vol. 2, no. 5, pp. 377–384, 2005.

[49] I. Coin, M. H. Perrin, W. W. Vale, and L. Wang, "Photo-cross-linkers incorporated into G-protein-coupled receptors in mammalian cells: a ligand comparison," *Angewandte Chemie International Edition*, vol. 50, no. 35, pp. 8077–8081, 2011.

[50] A. Gautier, D. P. Nguyen, H. Lusic, W. An, A. Deiters, and J. W. Chin, "Genetically encoded photocontrol of protein localization in mammalian cells," *Journal of the American Chemical Society*, vol. 132, no. 12, pp. 4086–4088, 2010.

[51] A. Gautier, A. Deiters, and J. W. Chin, "Light-activated kinases enable temporal dissection of signaling networks in living cells," *Journal of the American Chemical Society*, vol. 133, no. 7, pp. 2124–2127, 2011.

[52] E. Arbely, J. Torres-Kolbus, A. Deiters, and J. W. Chin, "Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine," *Journal of the American Chemical Society*, vol. 134, no. 29, pp. 11912–11915, 2012.

[53] N. Wu, A. Deiters, T. A. Cropp, D. King, and P. G. Schultz, "A genetically encoded photocaged amino acid," *Journal of the American Chemical Society*, vol. 126, no. 44, pp. 14306–14307, 2004.

[54] A. Deiters, D. Groff, Y. Ryu, J. Xie, and P. G. Schultz, "A genetically encoded photocaged tyrosine," *Angewandte Chemie*, vol. 45, no. 17, pp. 2728–2731, 2006.

[55] E. A. Lemke, D. Summerer, B. H. Geierstanger, S. M. Brittain, and P. G. Schultz, "Control of protein phosphorylation with a genetically encoded photocaged amino acid," *Nature Chemical Biology*, vol. 3, no. 12, pp. 769–772, 2007.

[56] Y. S. Wang, B. Wu, Z. Wang et al., "A genetically encoded photocaged Nepsilon-methyl-L-lysine," *Molecular bioSystems*, vol. 6, no. 9, pp. 1557–1560, 2010.

[57] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, and S. Yokoyama, "Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode Nε-(o-Azidobenzyloxycarbonyl) lysine for site-specific protein modification," *Chemistry and Biology*, vol. 15, no. 11, pp. 1187–1197, 2008.

[58] D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, and J. W. Chin, "Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNACUA pair and click chemistry," *Journal of the American Chemical Society*, vol. 131, no. 25, pp. 8720–8721, 2009.

[59] D. P. Nguyen, T. Elliott, M. Holt, T. W. Muir, and J. W. Chin, "Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation," *Journal of the American Chemical Society*, vol. 133, no. 30, pp. 11418–11421, 2011.

[60] S. Virdee, P. B. Kapadnis, T. Elliott et al., "Traceless and site-specific ubiquitination of recombinant proteins," *Journal of the American Chemical Society*, vol. 133, no. 28, pp. 10708–10711, 2011.

[61] K. Lang, L. Davis, J. Torres-Kolbus, C. Chou, A. Deiters, and J. W. Chin, "Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction," *Nature Chemistry*, vol. 4, no. 4, pp. 298–304, 2012.

[62] K. Lang, L. Davis, S. Wallace et al., "Genetic encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic diels-alder reactions," *Journal of the American Chemical Society*, vol. 134, no. 25, pp. 10317–10320, 2012.

[63] J. L. Seitchik, J. C. Peeler, M. T. Taylor et al., "Genetically encoded tetrazine amino acid directs rapid site-specific in vivo bioorthogonal ligation with trans-cyclooctenes," *Journal of the American Chemical Society*, vol. 134, no. 6, pp. 2898–2901, 2012.

[64] C. Köhrer, E. L. Sullivan, and U. L. RajBhandary, "Complete set of orthogonal 21st aminoacyl-tRNA synthetase-amber, ochre and opal suppressor tRNA pairs: concomitant suppression of three different termination codons in an mRNA in mammalian cells," *Nucleic Acids Research*, vol. 32, no. 21, pp. 6200–6211, 2004.

[65] Z. Zhang, L. Alfonta, F. Tian et al., "Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 24, pp. 8882–8887, 2004.

[66] W. Wan, Y. Huang, Z. Wang et al., "A facile system for genetic incorporation of two different noncanonical amino acids into one protein in *Escherichia coli*," *Angewandte Chemie*, vol. 49, no. 18, pp. 3211–3214, 2010.

[67] J. C. Anderson, N. Wu, S. W. Santoro, V. Lakshman, D. S. King, and P. G. Schultz, "An expanded genetic code with a functional quadruplet codon," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 20, pp. 7566–7571, 2004.

[68] H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, and J. W. Chin, "Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome," *Nature*, vol. 464, no. 7287, pp. 441–444, 2010.

[69] T. S. Young, D. D. Young, I. Ahmad, J. M. Louis, S. J. Benkovic, and P. G. Schultz, "Evolution of cyclic peptide protease inhibitors," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 27, pp. 11052–11056, 2011.

[70] C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnon, and S. J. Benkovic, "Production of cyclic peptides and proteins in vivo," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 24, pp. 13638–13643, 1999.

[71] A. R. Horswill and S. J. Benkovic, "Cyclic peptides, a chemical genetics tool for biologists," *Cell Cycle*, vol. 4, no. 4, pp. 552–555, 2005.

[72] S. Li, S. Millward, and R. Roberts, "In vitro selection of mRNA display libraries containing an unnatural amino acid," *Journal of the American Chemical Society*, vol. 124, no. 34, pp. 9972–9973, 2002.

[73] A. Frankel, S. Li, S. R. Starck, and R. W. Roberts, "Unnatural RNA display libraries," *Current Opinion in Structural Biology*, vol. 13, no. 4, pp. 506–512, 2003.

[74] S. W. Millward, S. Fiacco, R. J. Austin, and R. W. Roberts, "Design of cyclic peptides that bind protein surfaces with antibody-like affinity," *ACS Chemical Biology*, vol. 2, no. 9, pp. 625–634, 2007.

[75] N. Muranaka, T. Hohsaka, and M. Sisido, "Four-base codon mediated mRNA display to construct peptide libraries that contain multiple nonnatural amino acids," *Nucleic Acids Research*, vol. 34, no. 1, p. e7, 2006.

[76] K. Wang, H. Neumann, S. Y. Peak-Chew, and J. W. Chin, "Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion," *Nature Biotechnology*, vol. 25, no. 7, pp. 770–777, 2007.

[77] Y. Huang, W. K. Russell, W. Wan, P. J. Pai, D. H. Russell, and W. Liu, "A convenient method for genetic incorporation of multiple noncanonical amino acids into one protein in *Escherichia coli*," *Molecular BioSystems*, vol. 6, no. 4, pp. 683–686, 2010.

[78] T. Mukai, A. Hayashi, F. Iraha et al., "Codon reassignment in the *Escherichia coli* genetic code," *Nucleic Acids Research*, vol. 38, no. 22, pp. 8188–8195, 2010.

[79] T. Mukai, T. Yanagisawa, K. Ohtake et al., "Genetic-code evolution for protein synthesis with non-natural amino acids," *Biochemical and Biophysical Research Communications*, vol. 411, no. 4, pp. 757–761, 2011.

[80] K. Ohtake, A. Sato, T. Mukai, N. Hino, S. Yokoyama, and K. Sakamoto, "Efficient decoding of the UAG triplet as a full-fledged sense codon enhances the growth of a prfA-deficient strain of *Escherichia coli*," *Journal of Bacteriology*, vol. 194, no. 10, pp. 2606–2613, 2012.

[81] D. B. Johnson, J. Xu, Z. Shen et al., "RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites," *Nature Chemical Biology*, vol. 7, no. 11, pp. 779–786, 2011.

[82] D. B. Johnson, C. Wang, J. Xu et al., "Release factor one is nonessential in *Escherichia coli*," *ACS Chemical Biology*, vol. 7, no. 8, pp. 1337–1344, 2012.

[83] F. J. Isaacs, P. A. Carr, H. H. Wang et al., "Precise manipulation of chromosomes in vivo enables genome-wide codon replacement," *Science*, vol. 333, no. 6040, pp. 348–353, 2011.

[84] G. F. Short III, S. Y. Golovine, and S. M. Hecht, "Effects of release factor 1 on in vitro protein translation and the elaboration of proteins containing unnatural amino acids," *Biochemistry*, vol. 38, no. 27, pp. 8808–8819, 1999.

[85] S. Sando, A. Ogawa, T. Nishi, M. Hayami, and Y. Aoyama, "In vitro selection of RNA aptamer against *Escherichia coli* release factor 1," *Bioorganic and Medicinal Chemistry Letters*, vol. 17, no. 5, pp. 1216–1220, 2007.

[86] M. Humenik, Y. Huang, I. Safronov, and M. Sprinzl, "Simultaneous and site-directed incorporation of an ester linkage and an azide group into a polypeptide by in vitro translation," *Organic and Biomolecular Chemistry*, vol. 7, no. 20, pp. 4218–4224, 2009.

[87] S. Sando, K. Kanatani, N. Sato, H. Matsumoto, T. Hohsaka, and Y. Aoyama, "A small-molecule-based approach to sense codon-templated natural-unnatural hybrid peptides. Selective silencing and reassignment of the sense codon by orthogonal reacylation stalling at the single-codon level," *Journal of the American Chemical Society*, vol. 127, no. 22, pp. 7998–7999, 2005.

[88] A. C. Forster, Z. Tan, M. N. L. Nalam et al., "Programming peptidomimetic syntheses by translating genetic codes designed de novo," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 11, pp. 6353–6357, 2003.

[89] B. Zhang, Z. Tan, L. G. Dickson, M. N. L. Nalam, V. W. Cornish, and A. C. Forster, "Specificity of translation for N-alkyl amino acids," *Journal of the American Chemical Society*, vol. 129, no. 37, pp. 11316–11317, 2007.

[90] A. C. Forster, "Low modularity of aminoacyl-tRNA substrates in polymerization by the ribosome," *Nucleic Acids Research*, vol. 37, no. 11, pp. 3747–3755, 2009.

[91] R. Gao and A. C. Forster, "Changeability of individual domains of an aminoacyl-tRNA in polymerization by the ribosome," *FEBS Letters*, vol. 584, no. 1, pp. 99–105, 2010.

[92] A. C. Forster, "Synthetic biology challenges long-held hypotheses in translation, codon bias and transcription," *Biotechnology Journal*, vol. 7, no. 7, pp. 835–845, 2012.

[93] K. Josephson, M. C. T. Hartman, and J. W. Szostak, "Ribosomal synthesis of unnatural peptides," *Journal of the American Chemical Society*, vol. 127, no. 33, pp. 11727–11735, 2005.

[94] H. Koide, S. Yokoyama, G. Kawai et al., "Biosynthesis of a protein containing a nonprotein amino acid by *Escherichia coli*: L-2-aminohexanoic acid at position 21 in human epidermal growth factor," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 85, no. 17, pp. 6237–6241, 1988.

[95] N. Budisa, "Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire," *Angewandte Chemie*, vol. 43, no. 47, pp. 6426–6463, 2004.

[96] T. L. Hendrickson, V. De Crécy-Lagard, and P. Schimmel, "Incorporation of nonnatural amino acids into proteins," *Annual Review of Biochemistry*, vol. 73, pp. 147–176, 2004.

[97] J. A. Johnson, Y. Y. Lu, J. A. Van Deventer, and D. A. Tirrell, "Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications," *Current Opinion in Chemical Biology*, vol. 14, no. 6, pp. 774–780, 2010.

[98] Y. Shimizu, A. Inoue, Y. Tomari et al., "Cell-free translation reconstituted with purified components," *Nature Biotechnology*, vol. 19, no. 8, pp. 751–755, 2001.

[99] M. C. T. Hartman, K. Josephson, and J. W. Szostak, "Enzymatic aminoacylation of tRNA with unnatural amino acids," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 12, pp. 4356–4361, 2006.

[100] M. C. Hartman, K. Josephson, C. W. Lin, and J. W. Szostak, "An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides," *PLoS ONE*, vol. 2, no. 10, p. e972, 2007.

[101] H. Murakami, A. Ohta, H. Ashigai, and H. Suga, "A highly flexible tRNA acylation method for non-natural polypeptide synthesis," *Nature Methods*, vol. 3, no. 5, pp. 357–359, 2006.

[102] H. Saito, D. Kourouklis, and H. Suga, "An in vitro evolved precursor tRNA with aminoacylation activity," *EMBO Journal*, vol. 20, no. 7, pp. 1797–1806, 2001.

[103] H. Murakami, H. Saito, and H. Suga, "A versatile tRNA aminoacylation catalyst based on RNA," *Chemistry and Biology*, vol. 10, no. 7, pp. 655–662, 2003.

[104] H. Xiao, H. Murakami, H. Suga, and A. R. Ferré-D'Amaré, "Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme," *Nature*, vol. 454, no. 7202, pp. 358–361, 2008.

[105] M. Ohuchi, H. Murakami, and H. Suga, "The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus," *Current Opinion in Chemical Biology*, vol. 11, no. 5, pp. 537–542, 2007.

[106] J. Morimoto, Y. Hayashi, K. Iwasaki, and H. Suga, "Flexizymes: their evolutionary history and the origin of catalytic function," *Accounts of Chemical Research*, vol. 44, no. 12, pp. 1359–1368, 2011.

[107] T. J. Kang, S. Yuzawa, and H. Suga, "Expression of histone H3 tails with combinatorial lysine modifications under the reprogrammed genetic code for the investigation on epigenetic markers," *Chemistry and Biology*, vol. 15, no. 11, pp. 1166–1174, 2008.

[108] Y. Sako, J. Morimoto, H. Murakami, and H. Suga, "Ribosomal synthesis of bicyclic peptides via two orthogonal inter-side-chain reactions," *Journal of the American Chemical Society*, vol. 130, no. 23, pp. 7232–7234, 2008.

[109] E. Nakajima, Y. Goto, Y. Sako, H. Murakami, and H. Suga, "Ribosomal synthesis of peptides with C-terminal lactams, thiolactones, and alkylamides," *ChemBioChem*, vol. 10, no. 7, pp. 1186–1192, 2009.

[110] Y. Yamagshi, H. Ahigai, Y. Goto, H. Murakami, and H. Suga, "Ribosomal synthesis of cyclic peptides with a fluorogenic oxidative coupling reaction," *ChemBioChem*, vol. 10, no. 9, pp. 1469–1472, 2009.

[111] Y. Goto, K. Iwasaki, K. Torikai, H. Murakami, and H. Suga, "Ribosomal synthesis of dehydrobutyrine- and methyllanthionine-containing peptides," *Chemical Communications*, no. 23, pp. 3419–3421, 2009.

[112] T. Kawakami, H. Murakami, and H. Suga, "Messenger RNA-programmed incorporation of multiple N-Methyl-Amino acids into linear and cyclic peptides," *Chemistry and Biology*, vol. 15, no. 1, pp. 32–42, 2008.

[113] T. Kawakami, H. Murakami, and H. Suga, "Ribosomal synthesis of polypeptoids and peptoid-peptide hybrids," *Journal of the American Chemical Society*, vol. 130, no. 50, pp. 16861–16863, 2008.

[114] T. Kawakami, A. Ohta, M. Ohuchi, H. Ashigai, H. Murakami, and H. Suga, "Diverse backbone-cyclized peptides via codon reprogramming," *Nature Chemical Biology*, vol. 5, no. 12, pp. 888–890, 2009.

[115] Y. Goto, A. Ohta, Y. Sako, Y. Yamagishi, H. Murakami, and H. Suga, "Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides," *ACS Chemical Biology*, vol. 3, no. 2, pp. 120–129, 2008.

[116] Y. Goto, H. Murakami, and H. Suga, "Initiating translation with D-amino acids," *RNA*, vol. 14, no. 7, pp. 1390–1398, 2008.

[117] Y. Goto and H. Suga, "Translation initiation with initiator tRNA charged with exotic peptides," *Journal of the American Chemical Society*, vol. 131, no. 14, pp. 5040–5041, 2009.

[118] Y. Ohshiro, E. Nakajima, Y. Goto et al., "Ribosomal synthesis of backbone-macrocyclic peptides containing γ-amino acids," *ChemBioChem*, vol. 12, no. 8, pp. 1183–1187, 2011.

[119] A. Ohta, H. Murakami, E. Higashimura, and H. Suga, "Synthesis of polyester by means of genetic code reprogramming," *Chemistry and Biology*, vol. 14, no. 12, pp. 1315–1322, 2007.

[120] A. Ohta, H. Murakami, and H. Suga, "Polymerization of alpha-hydroxy acids by ribosomes," *Chembiochem*, vol. 9, no. 17, pp. 2773–2778, 2008.

[121] P. R. Effraim, J. Wang, M. T. Englander et al., "Natural amino acids do not require their native tRNAs for efficient selection by the ribosome," *Nature Chemical Biology*, vol. 5, no. 12, pp. 947–953, 2009.

[122] A. C. Forster, V. W. Cornish, and S. C. Blacklow, "Pure translation display," *Analytical Biochemistry*, vol. 333, no. 2, pp. 358–364, 2004.

[123] C. J. Hipolito and H. Suga, "Ribosomal production and in vitro selection of natural product-like peptidomimetics: the FIT and RaPID systems," *Current Opinion in Chemical Biology*, vol. 16, no. 1-2, pp. 196–203, 2012.

[124] A. Ohta, Y. Yamagishi, and H. Suga, "Synthesis of biopolymers using genetic code reprogramming," *Current Opinion in Chemical Biology*, vol. 12, no. 2, pp. 159–167, 2008.

[125] K. Iwasaki, Y. Goto, T. Katoh, and H. Suga, "Selective thioether macrocyclization of peptides having the N-terminal 2-chloroacetyl group and competing two or three cysteine residues in translation," *Organic & Biomolecular Chemistry*, vol. 10, no. 30, pp. 5783–5786, 2012.

[126] Y. Hayashi, J. Morimoto, and H. Suga, "In vitro selection of Anti-Akt2 thioether-macrocyclic peptides leading to isoform-selective inhibitors," *ACS Chemical Biology*, vol. 7, no. 3, pp. 607–613, 2012.

[127] J. Morimoto, Y. Hayashi, and H. Suga, "Discovery of macrocyclic peptides armed with a mechanism-based warhead: isoform-selective inhibition of human deacetylase SIRT2," *Angewandte Chemie*, vol. 51, no. 14, pp. 3423–3427, 2012.

[128] Y. V. Guillen Schlippe, M. C. Hartman, K. Josephson, and J. W. Szostak, "In vitro selection of highly modified cyclic peptides that act as tight binding inhibitors," *Journal of the American Chemical Society*, vol. 134, no. 25, pp. 10469–10477, 2012.

[129] F. T. Hofmann, J. W. Szostak, and F. P. Seebeck, "In vitro selection of functional lantipeptides," *Journal of the American Chemical Society*, vol. 134, no. 19, pp. 8038–8041, 2012.

[130] F. P. Seebeck and J. W. Szostak, "Ribosomal synthesis of dehydroalanine-containing peptides," *Journal of the American Chemical Society*, vol. 128, no. 22, pp. 7150–7151, 2006.

[131] F. P. Seebeck, A. Ricardo, and J. W. Szostak, "Artificial lantipeptides from in vitro translations," *Chemical Communications*, vol. 47, no. 21, pp. 6141–6143, 2011.

[132] V. W. Cornish, D. Mendel, and P. G. Schultz, "Probing protein structure and function with an expanded genetic code," *Angewandte Chemie*, vol. 34, no. 6, pp. 621–633, 1995.

[133] Z. Tan, A. C. Forster, S. C. Blacklow, and V. W. Cornish, "Amino acid backbone specificity of the *Escherichia coli* translation machinery," *Journal of the American Chemical Society*, vol. 126, no. 40, pp. 12752–12753, 2004.

[134] L. M. Dedkova, N. E. Fahmi, S. Y. Golovine, and S. M. Hecht, "Enhanced D-amino acid incorporation into protein by modified ribosomes," *Journal of the American Chemical Society*, vol. 125, no. 22, pp. 6616–6617, 2003.

[135] L. M. Dedkova, N. E. Fahmi, S. Y. Golovine, and S. M. Hecht, "Construction of modified ribosomes for incorporation of D-amino acids into proteins," *Biochemistry*, vol. 45, no. 51, pp. 15541–15551, 2006.

[136] L. M. Dedkova, N. E. Fahmi, R. Paul et al., "beta-Puromycin selection of modified ribosomes for in vitro incorporation of beta-amino acids," *Biochemistry*, vol. 51, no. 1, pp. 401–415, 2012.

[137] Y. Doi, T. Ohtsuki, Y. Shimizu, T. Ueda, and M. Sisido, "Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system," *Journal of the American Chemical Society*, vol. 129, no. 46, pp. 14458–14462, 2007.

[138] H. S. Park, M. J. Hohn, T. Umehara et al., "Expanding the genetic code of *Escherichia coli* with phosphoserine," *Science*, vol. 333, no. 6046, pp. 1151–1154, 2011.

[139] T. J. Magliery, J. C. Anderson, and P. G. Schultz, "Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in *Escherichia coli*," *Journal of Molecular Biology*, vol. 307, no. 3, pp. 755–769, 2001.

[140] H. Taira, T. Hohsaka, and M. Sisido, "In vitro selection of tRNAs for efficient four-base decoding to incorporate non-natural amino acids into proteins in an *Escherichia coli* cell-free translation system," *Nucleic Acids Research*, vol. 34, no. 5, article e44, 2006.

[141] J. Guo, C. E. Melançon, H. S. Lee, D. Groff, and P. G. Schultz, "Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids," *Angewandte Chemie*, vol. 48, no. 48, pp. 9148–9151, 2009.

[142] A. O. Subtelny, M. C. T. Hartman, and J. W. Szostak, "Ribosomal synthesis of N-methyl peptides," *Journal of the American Chemical Society*, vol. 130, no. 19, pp. 6131–6136, 2008.

[143] A. O. Subtelny, M. C. T. Hartman, and J. W. Szostak, "Optimal codon choice can improve the efficiency and fidelity of N-methyl amino acid incorporation into peptides by in-vitro translation," *Angewandte Chemie*, vol. 50, no. 14, pp. 3164–3167, 2011.

[144] R. Odegrip, D. Coomber, B. Eldridge et al., "CIS display: in vitro selection of peptides from libraries of protein-DNA complexes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 9, pp. 2806–2810, 2004.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 740982, 15 pages
doi:10.1155/2012/740982

Review Article

Potential of Peptides as Inhibitors and Mimotopes: Selection of Carbohydrate-Mimetic Peptides from Phage Display Libraries

Teruhiko Matsubara

Department of Biosciences and Informatics, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kouhoku, Yokohama 223-8522, Japan

Correspondence should be addressed to Teruhiko Matsubara, matsubara@bio.keio.ac.jp

Received 1 June 2012; Revised 31 July 2012; Accepted 2 August 2012

Academic Editor: Hiroshi Murakami

Glycoconjugates play various roles in biological processes. In particular, oligosaccharides on the surface of animal cells are involved in virus infection and cell-cell communication. Inhibitors of carbohydrate-protein interactions are potential antiviral drugs. Several anti-influenza drugs such as oseltamivir and zanamivir are derivatives of sialic acid, which inhibits neuraminidase. However, it is very difficult to prepare a diverse range of sugar derivatives by chemical synthesis or by the isolation of natural products. In addition, the pathogenic capsular polysaccharides of bacteria are carbohydrate antigens, for which a safe and efficacious method of vaccination is required. Phage-display technology has been improved to enable the identification of peptides that bind to carbohydrate-binding proteins, such as lectins and antibodies, from a large repertoire of peptide sequences. These peptides are known as "carbohydrate-mimetic peptides (CMPs)" because they mimic carbohydrate structures. Compared to carbohydrate derivatives, it is easy to prepare mono- and multivalent peptides and then to modify them to create various derivatives. Such mimetic peptides are available as peptide inhibitors of carbohydrate-protein interactions and peptide mimotopes that are conjugated with adjuvant for vaccination.

1. Introduction

A variety of glycoconjugate carbohydrate structures on the cell surface are important for biological events [1]. Carbohydrate structures on the cell surface change according to cell status, for example, during development, differentiation, and malignant alteration. Several glycoconjugates, including stage-specific embryonic antigen (SSEA)-3, SSEA-4, and tumor-rejection antigen (TRA)-1-60, are used as molecular makers of pluripotency to control the quality of induced pluripotent stem (iPS) cells [2]. Carbohydrate-protein interactions are the first cell surface events in cell-cell communication, following which processes such as infection and signal transduction occur. However, the reasons for the changes in carbohydrate structures on the cell surface are not clear. In addition, most receptors for glycoconjugates have not been identified. To investigate the biological roles of carbohydrates, sets of carbohydrates and their corresponding carbohydrate-binding proteins are required.

Carbohydrate-binding proteins such as plant lectins, bacterial toxins, and anticarbohydrate antibodies are available for studying carbohydrate-protein interactions [3, 4]. However, the repertoire of carbohydrate structures recognized by these proteins is limited and insufficient to cover the majority of structures. In addition, because carbohydrates are ubiquitous components of cell membranes and bio(macro)molecules, the immune response stimulated by glycoconjugates is negligible [5, 6], that is, high affinity carbohydrate-specific IgG-isotype antibodies are not easily obtained. Even if anticarbohydrate antibodies are generated, IgG comprises no more than 28% of the antibodies (74 IgGs in a total of 268 antibodies, with the remainder being IgMs) [7]. Therefore, while anticarbohydrate antibodies of the IgG isotype are preferred for carbohydrate research, IgM-antibodies with low affinity have been often used. Moreover, obtaining pure and homogeneous carbohydrates (or glycoconjugates) is very difficult. This is because regioselective protection of the hydroxy groups of the monosaccharide is

required. Programmable one-pot oligosaccharide synthesis is widely performed using protected monosaccharides and/or oligosaccharides [8–10]. Enzyme-catalyzed oligosaccharide synthesis has been also developed [10–12]. Several oligosaccharides such as KH-1 antigen (nonasaccharide of Le^Y-Le^X), globo-H hexasaccharide, and the core pentamannosides have been prepared by automated solid-phase oligosaccharide synthesis [8]. However, due to the complicated procedures of carbohydrate preparation, a general methodology for their chemical synthesis is not yet established.

To compensate for the lack of synthetic carbohydrates and to overcome their inherent weak immunogenicity, short peptides that bind to carbohydrate-binding proteins have been identified from phage-display libraries (Figure 1). These peptides mimic carbohydrate structures [13] and are called "carbohydrate-mimetic peptides (CMPs)" or "peptide mimotopes." It is predicted that CMPs, as well as carbohydrates, are recognized by carbohydrate-binding proteins. Small molecules such as biotin and carbohydrate mimotope (Glycotope) mimicking peptides have been frequently identified, and a number of reviews focusing on different aspects of their properties and uses have been published [14–16]. In this paper, recent studies on the selection and application of CMPs are surveyed and summarized according to the classification of target carbohydrate-binding proteins.

2. Peptide Selection from Phage Display Libraries

Phage display is an efficient selection (and screening) system for the identification of target-specific peptides and proteins from a large number of candidates [20–22]. A filamentous virus (M13 and fd, etc.) that infects *E. coli* is frequently used in phage display technology. When DNA encoding foreign sequences is inserted into the coat protein (pIII or pVIII) region in the virus genome (M13 phage vector, etc.), the corresponding sequence is fused with the coat protein of the viral particle (Figure 2(a)) [20]. The foreign sequence is "displayed" on the viral particle and is able to interact with various types of target molecules.

In the case of peptide libraries, the length of the peptides is often 5–20 amino acids. There are two types of peptide library: linear peptide libraries and cyclic peptide libraries (Figure 2(b)). The randomized region of cyclic peptide libraries is surrounded by two cysteines (e.g., CX_7C) to restrict the peptide conformation via disulfide bonds. The diversity of a peptide library is often 10^8-10^9, which is sufficient to cover a combination of hexapeptide libraries (X_6; $20^6 = 6.4 \times 10^7$). Several kinds of peptide libraries (e.g., Ph.D. Phage Display Peptide Library Kits, New England Biolabs) and customizable phage vectors (Ph.D. Peptide Display Cloning System) are commercially available.

To isolate phage clones that have high affinity for a target molecule, a set of procedures called "affinity selection (biopanning)" is performed (Figure 2(c)). First, the target molecule is incubated with the phage library in order to bind to specific peptide sequences. After removal of excess phages by washing, the bound phages are eluted by incubation with a known ligand for the target or an acidic buffer. The phages are amplified by infection of hosts (*E. coli*), and the phage pool is subjected to another round of biopanning. By repeating these steps, target-binding phages are enriched, and, finally, phage clones are obtained. The peptides with high affinity for the target molecule are identified by DNA sequencing of individual phage clones. Huang and coworkers established a mimotope database MimoDB (http://immunet.cn/mimodb/) that contains the results of biopanning experiments including the phage libraries used and the peptide sequences identified [23, 24]. This database will help in the development of therapeutic molecules and the identification of superior peptide mimotopes for vaccination.

3. CMPs against Lectins

3.1. Monosaccharide-Mimetic Peptides. Most lectins recognize monosaccharides and disaccharides [4]. Concanavalin A (ConA) is a lectin from jack-bean (*Canavalia ensiformis*) that binds to α-mannose (α-Man) and α-glucose (α-Glc). ConA is a famous lectin that is commercially available for the biological investigation of glycoconjugates. The first CMPs were selected from a random peptide library against ConA simultaneously by Oldenburg et al. (octapeptide library) [25] and Scott et al. (hexapeptide library) [13] (Table 1). Peptides containing the consensus sequence, Tyr-Pro-Tyr (YPY), showed high affinity for ConA with a dissociation constant (K_d) of 46 μM, and the K_d for methyl α-Man was 89 μM. The peptides are considered to mimic the structure of carbohydrates because the ConA-peptide interaction was inhibited by α-Man.

To obtain Man/Glc-mimetic peptides, Yu et al. used three lectins, including ConA, *Lens culinaris* agglutinin (LCA) from lentil, and *Pisum sativum* agglutinin (PSA) from pea [31]. Two cyclic peptides, CNTPLTSRC and CSRILTAAC, were selected from a cyclic heptapeptide library, but these peptides did not contain the YPY motif. Docking simulation of the peptide-lectin interaction suggested that the cyclic peptides bound to an alternative binding site, not to the sugar-binding site that is recognized by YPY-containing peptides. In another screen using monosaccharide-binding lectins, Eggink and Hoober identified a GalNAc/Gal-mimetic dodecapeptide, VQATQSNQHTPR, that was selected against *Helix pomatia* (HPA) lectin [32]. A tetrameric dendrimer of the peptide, $[(VQATQSNQHTPR)_2\,K]_2\,K$, was synthesized chemically (Figure 3), which was shown to stimulate the secretion of interleukin (IL)-8 and IL-21 from human peripheral blood mononuclear cells (PBMCs).

3.2. Disaccharide-Mimetic Peptides. The Galα1-3Gal disaccharide is recognized by *Griffonia simplicifolia* I-B4 (GS-I-B4) and *Bandeiraea simplicifolia* isolectin B4 (BS-I-B4) (Figure 4). The Galα1-3Gal structure is a major carbohydrate antigen recognized by human anti-pig antibodies, and inhibitors of human natural antibodies may be useful in pig-to-human xenotransplantation. Kooyman et al. identified a peptide sequence, SSLRGF, that binds to GS-I-B4 lectin

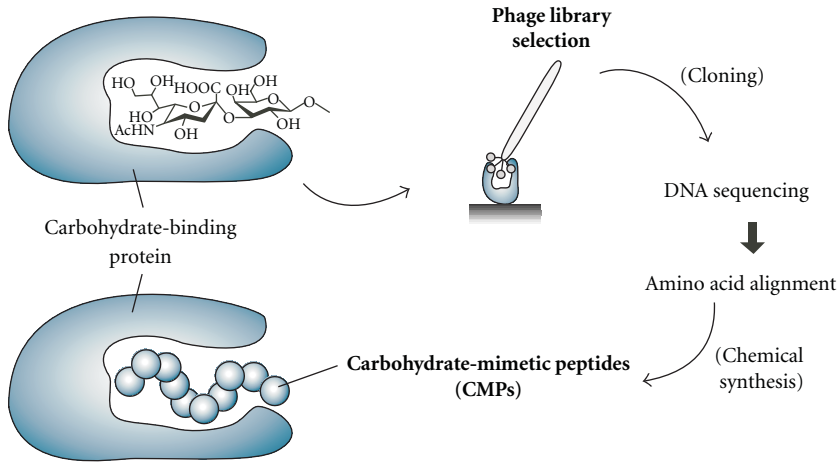


Figure 1: Identification of carbohydrate-mimetic peptides (CMPs) by affinity selection from a phage-display library. Selection is performed against carbohydrate-binding proteins. The peptides identified are chemically synthesized and recognized by the carbohydrate-binding protein.

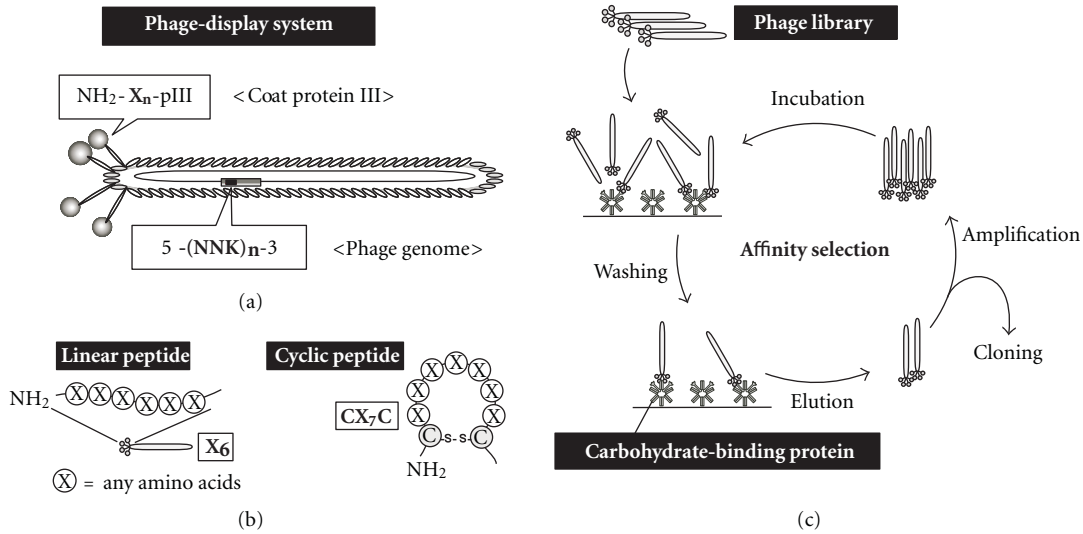


Figure 2: Phage-display system for affinity selection. (a) A typical filamentous phage carrying a peptide library. Foreign peptides (X_n) are displayed on the N-terminus of coat protein III (pIII) (type 3; M13 or fd phage). An oligonucleotide coding peptide library [-$(NNK)_n$-] is inserted into the phage genome. X = any amino acid; N = A, C, G, or T; K = G or T. (b) Linear (hexamer, *left*) and cyclic (heptamer, *right*) peptide libraries. (c) Schematic representation of the procedure for affinity selection (biopanning). The phage library is incubated with target receptors (carbohydrate-binding proteins), and unbound phages are removed by washing. Bound phages are eluted, amplified in *E. coli*, and subjected to the next cycle of biopanning. The cycle is repeated several times to enrich target-specific phages. Individual enriched phages are isolated and used for DNA sequencing.

from a hexapeptide library [27]. Zhan et al. identified a peptide, NCVSPYWCEPLAPSARA, by selection with BS-I-B4 lectin [28]. These peptides, SSLRGF and NCVSPYWCEPLAPSARA, inhibited the agglutination of pig red blood cells (RBCs) by human serum. Two peptides, FHENWPS and FHEFWPT, that inhibit the agglutination of RBCs were identified by selection against anti-Gal antibody by Lang et al. [42]. However, the peptides identified from three selections contained no obvious consensus sequence.

Influenza virus hemagglutinin (HA) recognizes sialyl-galactose structures (Neu5Ac-Gal) in glycoproteins and glycolipids on the cell surface in the initial stage of the infection process (Figure 4). Matsubara et al. identified CMPs from a pentadecapeptide library by selection with HAs of the H1 and H3 subtypes [17]. A HA-binding peptide, ARLSPTMVHPNGAQP, was identified from the first selection, and mutational sublibraries were prepared. A secondary selection was performed to improve the binding affinity for HAs, and the peptide was matured to peptide s2, ARLPRTMVHPKPAQP. The peptide was modified with a stearoyl group, and a molecular assembly of the alkylated peptides inhibited the infection of Madin-Darby

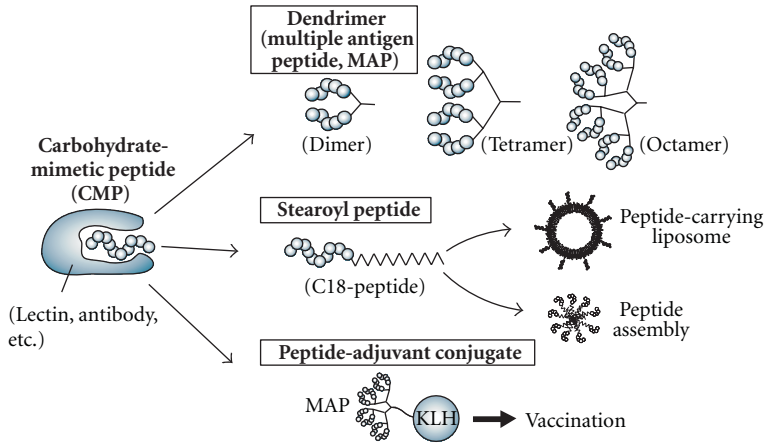


Figure 3: Representative chemical modifications of CMPs. To enhance the binding affinity, multiple CMPs are synthesized to give dimeric, tetrameric, and octameric dendrimers (multiple antigen peptide; MAP) (*upper*). The dendrimers are further conjugated with biotin, fluorescence groups, or adjuvants for vaccination. The peptide is modified with an alkyl group (stearic acid), enabling the peptide lipid to be incorporated into liposomes or to undergo self-assembly (*middle*). Monomeric CMP or CMP dendrimers are conjugated with adjuvants such as keyhole limpet hemocyanin (KLH), QS-21, and so forth (*lower*). The peptide-adjuvant conjugate is vaccinated into animals.

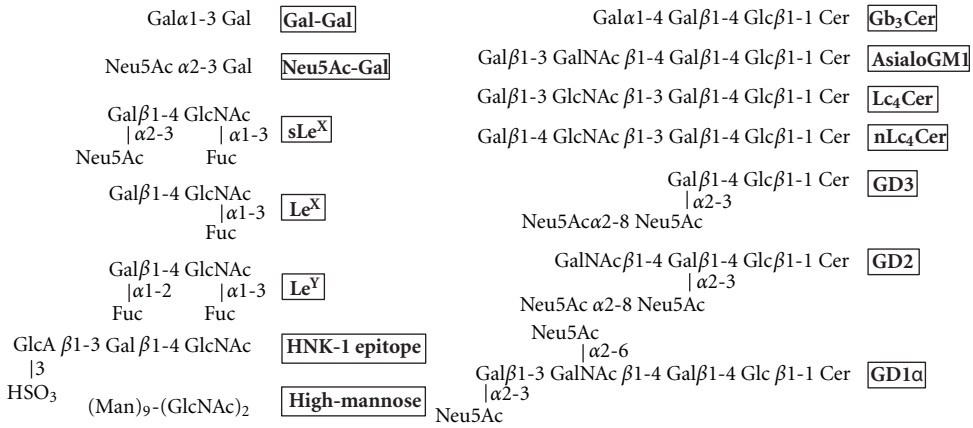


Figure 4: Oligosaccharide structures of carbohydrate antigens that are mimicked by peptides.

canine kidney cells by influenza virus (Figure 3). Finally, a pentapeptide fragment from the N-terminal of s2, ARLPR [s2(1–5)], was found to show the highest inhibitory activity. A docking study of the interaction between the peptide s2(1–5) and HA suggested that the peptide is recognized by the Neu5Ac-Gal receptor-binding pocket (Figure 5(a)). The figure indicates that three side chains of H3HA (Ser 136, Asn137, and Glu190) have the potential to interact with the peptide instead of Neu5Ac, and hydrophobic residues (Leu194, Leu226, and Trp222) are close to the peptide (Figure 5(b)).

4. CMPs against Oligosaccharide-Binding Antibodies

4.1. Oligosaccharide-Mimetic Peptides for Inhibition. Glycoproteins and glycosphingolipids have unique oligosaccharide structures at their nonreducing termini [1]. Cell-cell communication is performed by oligosaccharides that are recognized by families of cell adhesion proteins such as selectins and sialic acid-binding immunoglobulin- (Ig-) like lectins (siglecs). Pathogenic viruses, toxins, and bacteria also recognize oligosaccharide structures [3]. Because an abundant variety of oligosaccharide structures relates to many carbohydrate-protein interactions, oligosaccharide-mimetic peptides mediate many kinds of inhibitory activities.

The sialyl-Lewis^X (sLe^X) structure, Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAc, is recognized by E-selectin and is a famous carbohydrate antigen (Figure 4). sLe^X-mimetic peptides were identified by selection against E-selectin [29, 30] and anti-sLe^X antibody [36] (Tables 1 and 2). Martens et al. identified the HITWDQLWNVMN peptide and further optimized the sequence as DITWDQLWDLMK using a mutagenesis library [29]. The binding affinity of the synthetic peptide for E-selectin was improved 100-fold by this optimization (IC_{50} for sLe^X binding to E-selectin; from 420 nM to 4 nM). The DITWDQLWDLMK peptide inhibited

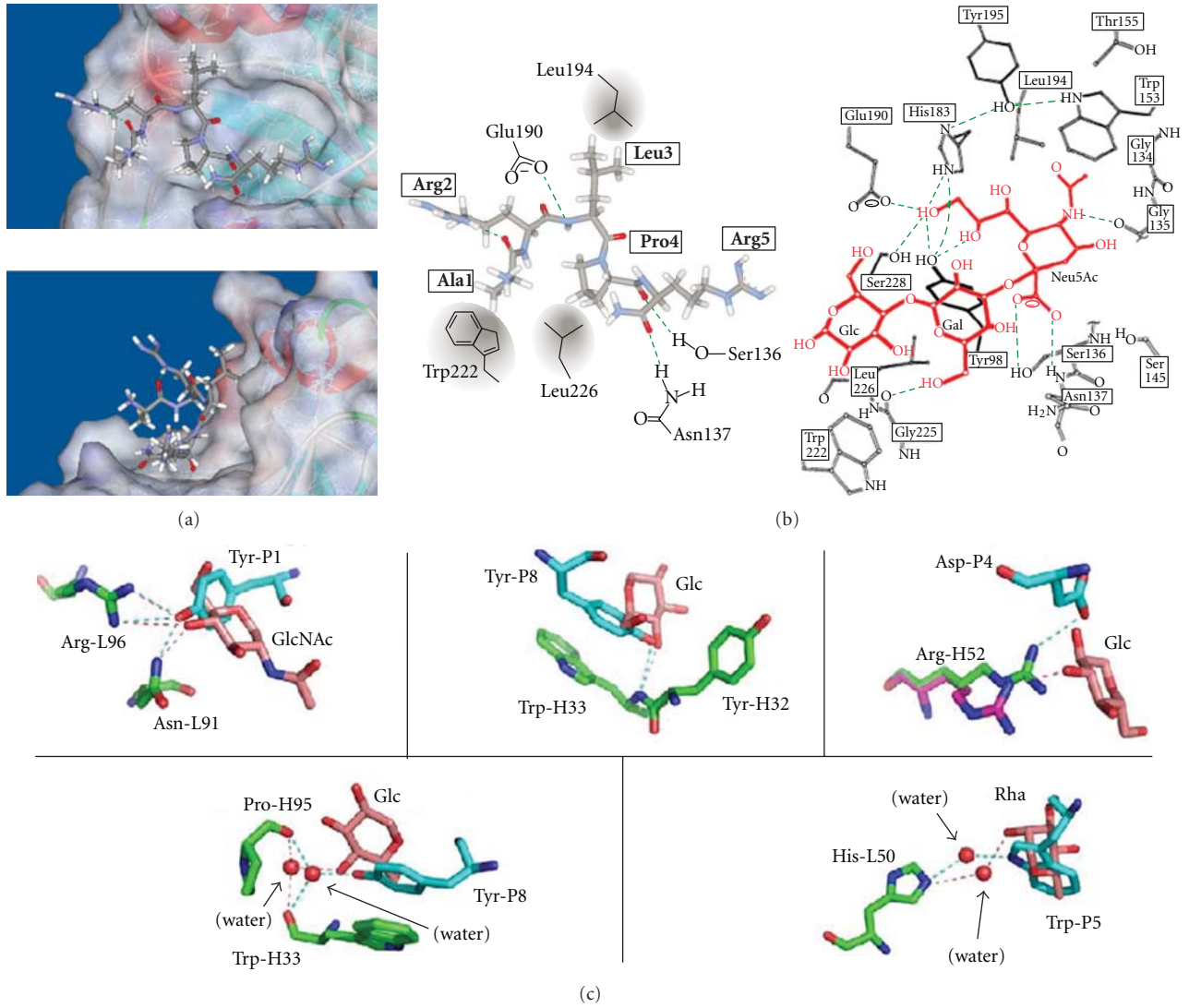


FIGURE 5: (a) Computer simulation of the interaction between peptide s2(1–5) and HA. A docking pose of the s2(1–5)-HA complex (*left*) and schematic diagram of the binding site of HA (*right*). The peptide is thought to be recognized by the Neu5Ac-Gal receptor-binding pocket. The peptide is shown as a stick model. Three potential hydrogen bonds (green dotted lines) between H3 and s2(1–5) are proposed (Glu190-Leu3, Ser136-Pro4, and Asn137-Arg5), which are similar to those in H3-Neu5Ac. Adapted from reference [17]. (b) Schematic diagram of the binding site of H3HA (Protein Data Bank entry, 1HGG). Neu5Acα2–3Gal-Glc (sialyllactose) is shown in red. Modified from [18]. (c) Comparison of the polar interactions shown in the oligosaccharide (*O*-antigen of *S. flexneri* serotype 2a) and peptide B1 (YLEDWIKYNNQK) complexes of monoclonal antibody F22-4. The peptide and oligosaccharide ligands are distinguished by carbon atoms shown in cyan and pink, respectively (P, peptide; Rha, rhamnose). The carbon atoms of the F22-4 residues are shown in green (H, heavy chain; L, light chain). Adapted from [19].

the adhesion of HL-60 cells and reduced neutrophil rolling on lipopolysaccharide- (LPS-) stimulated human umbilical vein endothelial cells. Qiu et al. designed WRY-containing peptides from the sLe^X-mimetic peptide sequences, but these peptides cross-reacted with anti-Lewis^Y antibody. Octameric multiple antigen peptides (MAPs) were conjugated with QS-21 adjuvant, which resulted in cytotoxic IgM and IgG antibodies (Figures 3 and 6). MAPs, in which peptides are attached to an octabranched amino acid backbone, are used to generate antibodies against a synthetic peptide, which is useful for the design of vaccines [94]. Katagihallimath et al. selected a cyclic CSRLNYLHC peptide against anti-Le^X antibody [37]. The trisaccharide Le^X structure is known as CD15 or SSEA-1, and this structure is expressed in the developing and adult murine central nervous system. The Le^X mimetic peptide inhibited CD24-induced neurite outgrowth.

Neutral glycosphingolipid Lc$_4$Cer-mimetic peptides showed unique activity [46] (Table 3). Lc$_4$Cer contains Galβ1-3GlcNAcβ1-3Galβ1-4Glc tetrasaccharide that is linked to ceramide (Figure 4), and Jack bean β-galactosidase digests a nonreducing terminus β-Gal to give Lc$_3$Cer. The Lc$_4$Cer-mimetic peptides inhibited digestion by

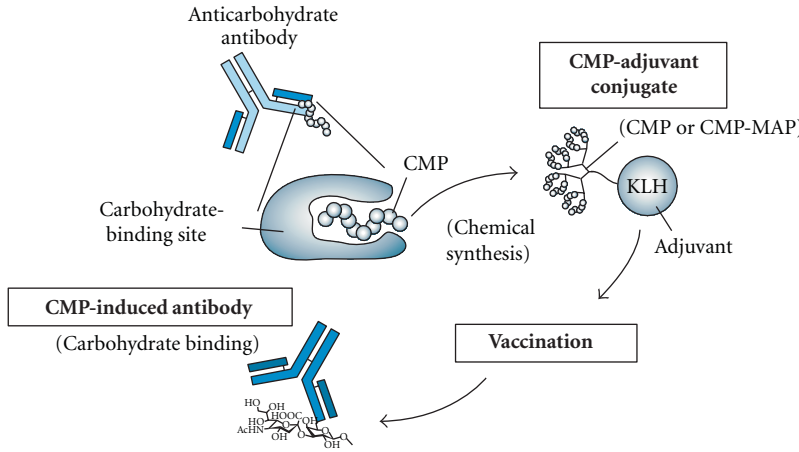


Figure 6: Procedure for obtaining CMP-induced antibodies by vaccination. A peptide mimotope (CMP) is conjugated with an adjuvant such as KLH and used for vaccination.

Table 1: Summary of the selection of CMPs with lectins.

Target lectins (abbreviations)	Peptide library	Peptide motif or representative sequences (peptide name)	Lectin-binding carbohydrate structures	References	Notes*
Concanavalin A (ConA)	X_8, X_6	YPY motif	Man; Glc	[13, 25, 26]	Inhibition of Man binding
Griffonia simplicifolia I-B4 isolectin (GS-I-B4)	X_6	SSLRGF	Galα1-3Gal	[27]	Inhibition of RBC agglutination
Bandeiraea simplicifolia I-B4 isolectin (BS-I-B4)	XCX_{15}	NCVSPYWCEPLAPSARA	Galα1-3Gal	[28]	Inhibition of RBC agglutination
E-selectin	X_{12}	DITWDQLWDLMK	Sialyl Lewis^X [Neu5Acα2-3Galβ1-4(Fucα1–3)GlcNAc]	[29]	Inhibition of cell adhesion, reduction of neutrophil rolling, and so forth
	X_7	IELLQAR		[30]	Octameric MAP, inhibition of HL-60, and B16 cell adhesion
Concanavalin A (ConA); *Lens culinaris* agglutinin (LCA); *Pisum sativum* agglutinin (PSA)	X_{12}, CX_7C	CNTPLTSRC; CSRILTAAC	Man; Glc	[31]	Inhibition of Man binding; docking study
Lectin from *Helix pomatia* (HPA)	X_{12}	VQATQSNQHTPRGGGS	*O*-linked α-GalNAc; Galβ1-3GalNAc; α-GlcNAc	[32]	Tetrameric dendrimer, stimulation of IL-8, and IL-21 secretion
Lipopolysaccharide (LPS) binding protein (LBP); CD14	X_{12}	FHRWPTWPLPSP (MP12)	Lipopolysaccharide	[33]	Inhibition of LPS-induced INF-α expression
Influenza virus hemagglutinin (HA)	X_{15}	ARLPRTMVHPKPAQP (s2); ARLPR [s2(1–5)]	Neu5Acα2-3Gal	[17]	*N*-stearoyl peptide; inhibition of flu infection

*RBC: red blood cell; IL: interleukin; INF: interferon.

β-galactosidase at a high concentration of enzyme, whereas the peptides enhanced the digestion of Lc_4Cer at lower concentration of enzyme. This unique activity of the peptides was also shown in the digestion of nLc_4Cer. This group also identified WHW-containing peptides such as WHWRHRIPLQLAAGR by selection with anti-GD1α antibody [47]. The ganglioside GD1α is cell adhesion molecule of murine metastatic large cell lymphoma (RAW117-H10 cells) that binds to endothelial cells. GD1α-mimetic peptides inhibited the adhesion between RAW117-H10 cells and hepatic sinusoidal endothelial (HSE) cells. Furthermore, the metastasis of RAW117-H10 cells to

Table 2: Summary of the selection of CMPs with oligosaccharide-binding antibodies.

Target antibodies (abbreviations)	Name of antibody	Peptide library	Peptide motif or representative sequences (peptide name)	Carbohydrate antigen	References	Notes*
Anti-Lewis^Y (Le^Y)	B3	X_8	APWLYGPA	Fucα1-2Galβ1-4(Fucα1-3)GlcNAc	[34, 35]	Induction of anti-Le^Y immune responses
	BR55-2; 15-6A	X_{15}	WRY-containing peptide		[36]	Cross-reaction with anti-Le^X
Anti-sialyl Lewis^X (sLe^X)	FH-6	X_{15}	WRY-containing peptide	Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAc	[36]	Cross-reaction with anti-Le^Y; octameric MAP-QS21
Anti-Lewis^X (Le^X)	L5	X_{12}, CX_7C	CSRLNYLHC	Galβ1-4(Fucα1-3)GlcNAc	[37]	Inhibition of CD24-induced neurite outgrowth
Antilipooligosaccharide (LOS)	—	X_7	SMYGSYN, APARQLP	LOS of group B *Neisseria meningitidis*	[38]	Peptide-DT
Antilipooligosaccharide (LOS)	—	X_{12}	NMMRFTSQPPNN and so forth	LOS of nontypeable *Haemophilus influenzae*	[39]	Peptide-KLH
Anti-β-1,2-oligomannoside	DJ2.8	X_7	FHENWPS	β-1,2-oligomannoside	[40]	Peptide-KLH
Anti-L2/HNK-1	L2-412	X_{15}	FLHTRLFVSDWYHTP, FLHTRLFV	SO_4-3GlcAβ1-3Galβ1-4GlcNAc	[41]	Promotion of neurite outgrowth
Anti-Gal	B	X_7, CX_7C	FHENWPS, FHEFWPT	Xenoreactive α-Gal antigenic epitope	[42]	Inhibition of RBC agglutination
Anti-GMDP	E6/1.2	X_{15}	RVPPRYHAKISPMVN	*N*-acetylglucosaminyl-β1,4-*N*-acetylmuramyl-alanyl-D-isoglutamine (GMDP)	[43]	Peptide-OVA
Antiglucitollysine	41; 226	X_{12}, CX_9C	CTSRXC motif	Glc-Lys	[44]	Inhibition of Glu-Lys binding
Antihigh-mannose oligosaccharides	2G12	$X_{15}CX$	ACPPSHVLDMRSGTCLAAEGK (2G12.1)	$Man_9GlcNAc_2$ (HIV-1 gp120)	[45]	X-ray analysis (no structural mimic)

*DT: diphtheria toxoid; KLH: keyhole limpet hemocyanin; OVA: ovalbumin.

Table 3: Summary of the selection of CMPs with glycolipid-binding antibodies.

Target antibodies	Name of antibody	Peptide library	Peptide motif or representative sequences (peptide name)	Glycolipid structures	References	Notes*
Anti-Lc_4Cer; anti-nLc_4Cer	AD117m; H11	X_{15}	VPPXFXXXY	Galβ1-3GlcNAcβ1-3Galβ1-4Glcβ1-1′Cer; Galβ1-4GlcNAcβ1-3Galβ1-4Glcβ1-1′Cer	[46]	Modulation of β-galactosidase activity
Anti-GD1α	KA17	X_{15}	WHWRHRIPLQLAAGR	Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcβ1-4Galβ1-4Glcβ1-1′Cer	[47, 48]	Inhibition of metastasis; peptide-liposome
Anti-asialo GM1	clone 10	X_7, CX_7C	KL/VWQXXX	Galβ1-3GalNAcβ1-4Galβ1-4Glcβ1-1′Cer	[49]	(Phage ELISA only)
Anti-GD3/GD2	ME36.1	X_{15}	WRY-containing peptide and so forth	Neu5Acα2-8Neu5Acα2-3Galβ1-4Glcβ1-1′Cer; GalNAcβ1-4(Neu5Acα2-8Neu5Acα2-3)Galβ1-4Glcβ1-1′Cer	[36, 50, 51]	Octameric MAP-QS21/KLH
Anti-GD2	14.18	$CX_{10}C$	CDGGWLSKGSWC; CGRLKMVPDLEC	GalNAcβ1-4(Neu5Acα2-8Neu5Acα2-3)Galβ1-4Glcβ1-1′Cer	[52–54]	Docking study; peptide-KLH
	14G2a	X_{15}	EDPSHSLGLDVALFM		[55, 56]	Molecular modeling; DNA vaccine; induction of $CD8^+$ T cell response
	14G2a	XCX_8CX	RCNPNMEPPRCF		[57, 58]	Inhibition of antibody binding to IMR-32 cells
Anti-GD3	4F6	X_{15}	LAPPRPRSELVFLSV (GD3P4)	Neu5Acα2-8Neu5Acα2-3Galβ1-4Glcβ1-1′Cer	[59]	Peptide-VSSP
Anti-Gb_3Cer	—	X_{12}	WHWTWLSEY	Galα1-4Galβ1-4Glcβ1-1′Cer	[60]	Neutralization of Shiga toxin
Anti-Neu5Gc-containing ganglioside (Neu5Gc-GM3)	1E10; chimeric P3; 1E10	X_9, X_{12}	KPPR, RRPR/K; LEICSYTPDEGC; KCGHHYCRQVDL	Neu5Gcα2-3Galβ1-4Glcβ1-1′Cer	[61, 62]	Inhibition of 1E10 binding to P3
Anti-phenolic glycolipid-I (PGL-I)	III603.8	X_7	W(T/R)LGPY(V/M)	*Mycobacterium leprae*-specific antigen	[63]	Does not bind to antibodies in serum

*MAP: multiple antigen peptide; VSSP: very small size proteoliposomes.

Table 4: Summary of the selection of CMPs with polysaccharide-binding antibodies.

Species	Carbohydrate antigen	Name of antibody	Peptide library	Peptide motif or representative sequences (peptide name)	References	Notes*
Cryptococcus neoformans	Polysaccharide (glucuronoxylomannan; GXM)	2H1	X_6, X_{10}, ADVA X_6 TPXW [M/L][M/L] X_6 AAG	(E)TPXWM/LM/L, W/YXWM/LY; GLQYTPSWMLVG (PA1); SYSWMYE (P601E); FGGETFTPDWMMEVAIDNE (P206.1)	[64–67]	Four motifs; X-ray analysis (PA1); peptide evolution (P206.1); peptide-KLH/TT
Streptococcus species	Capsular polysaccharide (type 3, group B)	S9	X_9	WENWMMGNA; FDTGAFDPDWPA	[68]	Group B streptococci (GBS); peptide-KLH/BSA/OVA
Streptococcus pneumoniae	Capsular polysaccharide (serotype 4)	mAb4	X_{15}	SGQARVLYSEFINAL (pep4)	[69]	DNA vaccine
	(serotype 8)	(human mAb IgA)	X_{12}	FHLPYNHNWFAL (PUB1)	[70]	Peptide-TT
	(serotype 6B, 9V)	206; F-5; Db3G9	X_{12}, X_{15}	MP7, 12, 55, 58	[71]	Peptide-KLH
Streptococcus pyogenes	Cell-wall polysaccharide (group A)	SA-3; Strep9; HGAC39; HGAC47; HGAC101	X_6, XCX_8CX, X_8CX_8, $X_{15}CX$, X_{15}	DRPVPY	[72]	Basis of peptide-carbohydrate cross-reactivity
Streptococcus agalactiae	(serotype A, B, C)	(IgG2, Ig polyclonal)	X_{12}	NPDHPRVPTFMA (2–8); LIPFHKHPHHRG (3-2)	[73]	DNA vaccine; MAP-CFA/IFA
Neisseria meningitidis	Capsular polysaccharide (serogroup A)	9C10	X_{15}	GEASGLCCRWSSLKGC	[74]	Peptide-proteasome
	(serogroup B)	HmenB3	X_{12}	NKVIWDRDWMYP	[75]	Peptide-BSA-CFA/IFA
	(serogroup B)	9-2-L3, 7, 9	CX_6C, CX_9C	CGAVIDDC	[76]	Peptide-KLH
	(serogroup B) [poly-α2–8 sialic acid (PSA)]	30H12	CX_9C	CSSVTAWTTGCG	[77, 78]	Enhanced migration of grafted neuroblasts in mouse brain
	(serogroup B)	Seam3	X_9, X_{12}	DYAWDQTHQDPAK (9M)	[79]	Peptide-KLH, DNA vaccine
	(serogroup B)	13D9	X_{15}	RGDKSRPPVWYVEGE	[80]	Phage vaccine
	(serogroups B and C)	(IgG2, Ig polyclonal)	X_{12}	EQEIFTNITDRV (G3)	[73]	DNA vaccine; MAP-CFA/IFA
	(serogroup C)	1E4	X_{15}	GFSYYRPPWIL (Pep2C)	[81]	Peptide-proteasome
	[*N*-propionyl derivative of CPS]	—	—	—	[82]	
Vibrio cholerae	Capsular polysaccharide (serogroup O139)	Vc1; Vc2; ICL12	six libraries (X_9, X_{12}, X_{28} etc.)	AEGEFSPGVWKAAFQGDKLPDPAK and so forth	[83]	Peptide-KLH/BSA
	(serogroup O1) Ogawa serotype	S-20-4; A-20-6	X_{12}, X_7, CX_7C	NHNYPPLSLLTF (4P-8)	[84]	Peptide-KLH/BSA; docking study
	(serogroup O1) Ogawa and Inab serotypes	72.1	X_{12}, X_7, CX_7C	ECLLLSKYCMPS (3ME-1); SMCMHGGAYCFP (3ME-2)	[85]	Peptide-KLH/BSA-CFA/IFA
Shigella flexneri	O-antigen of lipopolysaccharide (serotype 5a)	mIgAs C5; I3	X_9	YKPLGALTH; KVPPWARTA	[86]	Phage vaccine
	(serotype 2a)	F22-4	X_{12}	YLEDWIKYNNQK (B1)	[19]	X-ray analysis
—	Melanoma-associated chondroitin sulfate proteoglycan (MCSP)	763.74	X_6	VHINAH	[87]	Inhibition of MCSP binding

Table 4: Continued.

Species	Carbohydrate antigen	Name of antibody	Peptide library	Peptide motif or representative sequences (peptide name)	References	Notes*
Entamoeba histolytica	GPI-linked proteophosphoglycan antigens	EH5	six libraries (X_9, X_{12}, X_{28} etc.)	GTHPXL	[88]	Glycosylphosphatidylinositol (GPI); phage vaccine
Mycobacterium tuberculosis	Neutral polysaccharides	—	X_{12}	QEPLMGTVPIRAGGGS (P1)	[89]	Peptide oligomer vaccine
Mycobacterium tuberculosis	Mannosylated lipoarabinomannan	CS40	X_{12}	ISLTEWSMWYRH (B11)	[90]	Peptide-KLH-adjuvants
Burkholderia pseudomallei	Exopolysaccharide (EPS)	3VIE5; 4VA5	X_{12}, X_7, CX_7C	CYLPFQLSC; CHPLFDARC	[91]	Peptide-thyroglobulin
Brucella melitensis; *Brucella abortus*	Lipopolysaccharide	4F9; 11B2	X_9, X_{12}, CX_9CX, X_{15}	WTEIHDWEAAME	[92]	DNA vaccine
Staphylococcus aureus	Peptidoglycan	—	X_{12}	Sp-31	[93]	MAP vaccine

*BSA: bovine serum albumin; TT: tetanus toxoid; CFA: complete Freund's adjuvant; IFA: incomplete Freund's adjuvant.

lung and spleen was completely inhibited by the intravenous injection of the peptide. Subsequently, WHW was found to be a minimal sequence that mimics the GD1α structure [48]. To modify the liposome surface with the WHW peptide, the WHW tripeptide was conjugated to alkyl groups such as palmitoyl or stearoyl groups (Figure 3). Coating of liposomes with peptides is often performed in drug delivery systems. The WHW-modified liposomes inhibited the adhesion between RAW117-H10 cells and HSE cells.

Tryptophan/tyrosine-containing tripeptides (YPY for ConA, WRY for $sLe^{X(Y)}$, and WHW for GD1α) may comprise a key sequence that mimics oligosaccharide structure. Although Gb_3 (Galα1-4Galβ1-4Glc trisaccharide) is dissimilar to the disaccharide (Galβ1-3GlcNAcβ) structure of Lc_4 at the nonreducing terminus, Miura et al. identified a WHW-containing peptide (WHWTWLSEY) that mimics the Gb_3 structure [60]. Gb_3 is well known as a receptor for Shiga toxin (Stx). The Gb_3-mimetic peptide showed neutralization activity against Stxs (Stx-1 and Stx-2) in a HeLa cell cytotoxicity assay. The binding affinity of the Gb_3-mimetic peptide for Stx-1 was also investigated by surface plasmon resonance analysis (K_d = 1.4 nM).

4.2. Oligosaccharide-Mimetic Peptides for Vaccination. The immunogenicity of oligosaccharides is weak because oligosaccharides are ubiquitous components of cell membranes in tissues throughout the human body. When antioligosaccharide antibodies are generated, they attack these tissues and cause the risk of autoimmune disease. For example, lipopolysaccharides of *Campylobacter jejuni* isolated from GBS patients contain ganglioside-like epitopes such as GM1, GM1b, GD1a, and GalNAc-GDla, and these epitopes induce Guillain-Barre syndrome [95]. However, this low immunogenicity interferes with the preparation of antioligosaccharide antibodies that are useful for the investigation of glycoconjugate function.

To improve the binding affinity, specificity and cytotoxicity of antibodies, oligosaccharide-mimetic peptides are applied as peptide mimotopes of carbohydrate antigens for vaccination (Figure 6). Oligosaccharide-mimetic peptides were identified by selection against $Le^{X(Y)}$ [34, 35, 37], $sLe^{X(Y)}$ [36, 50], GD2 [36, 50–56], GD3 [36, 50, 59], lipooligosaccharide (LOS) [38, 39], β-1,2-oligomannoside [40], *N*-acetylglucosaminyl-β1,4-*N*-acetylmuramyl-alanyl-D-isoglutamine (GMDP) [43], and high-mannose oligosaccharide ($Man_9GlcNAc_2$ for HIV-1 gp120) [45]. The oligosaccharide-mimetic peptides were chemically synthesized and conjugated with adjuvant. To enhance the immunogenicity of the peptides, MAPs were prepared and resulted in dimeric, tetrameric, and octameric dendrimers (Figure 3). The peptide-adjuvant conjugates were vaccinated, with the adjuvants used being keyhole limpet hemocyanin (KLH) [39, 40, 53, 54], QS-21 [36, 50, 54], diphtheria toxoid (DT) [38], ovalbumin (OVA) [43], or very small size proteoliposomes (VSSP) [59] (Figure 6, Tables 2 and 3). In some cases, DNA vaccination was also performed [55, 56]. The CMP-induced antibodies are able to bind to peptide mimotopes and carbohydrate antigens.

5. CMPs against Polysaccharide-Binding Antibodies

Most polysaccharide-mimetic peptides to be applied for vaccination are identified as peptide mimotopes of carbohydrate antigens (Figure 6). Capsular polysaccharides of microorganisms are carbohydrate antigens, and it is known that these polysaccharides cause meningoencephalitis in immunocompromised patients, particularly those with AIDS (polysaccharide from *Cryptococcus neoformans*), pneumonia and bacteremia (*Streptococcus pneumoniae*), bacterial meningitis (*Neisseria meningitidis*), cholera (*Vibrio cholerae*), tuberculosis (*Mycobacterium tuberculosis*), and so forth (Table 4). These peptide mimotopes are potential antigens for safe vaccination and are expected to produce highly cytotoxic antibodies.

The typical methodology for vaccination uses a CMP-conjugated adjuvant. Valadon et al. identified CMPs that bind to anticryptococcal polysaccharide (glucuronoxylomannan, GXM) monoclonal antibody 2H1 [64]. The CMPs shared four motifs, for example, (E)TPXWM/LM/L and W/YXWM/LYE, and the dodecapeptide, GLQYTPSWMLVG (PA1) was found to bind 2H1 with a K_d of 295 nM [64]. The three-dimensional structure of 2H1 has been solved in a complex with PA1 [65]. The peptide PA1 was improved by selection from a PA1-based sublibrary, which identified the peptide P206-1 (FGGETFTPDWMMEVAIDNE) [66]. The affinity of peptide 206-1 for 2H1 was 80-fold higher than that of PA1 (K_d of 3.7 nM). Immunization of mice with P206-1-tetanus toxoid (TT), but not PA1 or P601E (DGASYSWMYEA), induced an anti-GXM response [66, 67].

Although antibodies against the capsular polysaccharide of the same species (e.g., *Neisseria meningitidis* serogroup B) were used, the CMPs identified were different and shared no consensus motif [73, 75–80] (Table 4). This may be due to the different antibodies used (HmenB3, 9-2-L3, 30H12, Seam3, or 13D9), different primary peptide libraries (CX_6C, X_9, CX_9C, X_{12}, or X_{15}), or different selection conditions. Harris et al. also concluded that the CMPs identified by each antibody possessed distinct consensus motifs [72]. A variety of peptide-conjugating adjuvants such as KLH, TT, BSA, OVA, proteasome, and thyroglobulin have been used. In some cases, phage particles were directly used for vaccination [80, 86, 88], and a high level of the IgG_{2a} subtype in the response against CMPs was shown [80].

Theillet et al. clarified the structural mimicry of *O*-antigen oligosaccharide by CMPs [19]. Figure 5(c) shows a structural representation of the antibody-peptide complex in which the sugar chains were replaced by amino acids. Glc and GlcNAc were replaced by Tyr or Asp, and one or more hydrogen bonds are indicated. On the other hand, high-mannose oligosaccharide-mimic peptide (2G12-1 peptide) binds to a neighboring pocket of the oligosaccharide (Table 2) [45]. The binding site for the DVFYPYPYASGS peptide, which was selected against ConA, was different from the mannose/trimannose-binding site [26]. However, the peptide inhibits α-mannopyranoside binding to ConA [25], indicating that this

peptide shows functional mimicry rather than structural mimicry.

6. Conclusion

Anticarbohydrate antibodies are necessary for clarifying the biological functions of carbohydrates, the detection of carbohydrates during etiological diagnosis, and therapy for carbohydrate-related diseases [7, 96]. Due to the difficulty in obtaining homogeneous glycoconjugates and carbohydrate-binding proteins, phage display libraries have been applied for the identification of peptide mimotopes. In this paper, the selection of CMPs was classified according to the types of target carbohydrates. The first selection was performed against lectins, and then the selections were performed against anticarbohydrate antibodies. To apply the peptide mimotopes for vaccination, this methodology is becoming more widespread.

Acknowledgment

The preparation of this paper was supported in part by the Keio Gijuku Fukuzawa Memorial Fund for the Advancement of Education and Research.

References

[1] A. Varki, "Biological roles of oligosaccharides: all of the theories are correct," *Glycobiology*, vol. 3, no. 2, pp. 97–130, 1993.

[2] K. Takahashi, K. Tanabe, M. Ohnuki et al., "Induction of pluripotent stem cells from adult human fibroblasts by defined factors," *Cell*, vol. 131, no. 5, pp. 861–872, 2007.

[3] H. Lis and N. Sharon, "Lectins: carbohydrate-specific proteins that mediate cellular recognition," *Chemical Reviews*, vol. 98, no. 2, pp. 637–674, 1998.

[4] W. I. Weis and K. Drickamer, "Structural basis of lectin-carbohydrate recognition," *Annual Review of Biochemistry*, vol. 65, pp. 441–473, 1996.

[5] G. A. Nores, R. D. Lardone, R. Comín, M. E. Alaniz, A. L. Moyano, and F. J. Irazoqui, "Anti-GM1 antibodies as a model of the immune response to self-glycans," *Biochimica et Biophysica Acta*, vol. 1780, no. 3, pp. 538–545, 2008.

[6] S. F. Slovin, S. J. Keding, and G. Ragupathi, "Carbohydrate vaccines as immunotherapy for cancer," *Immunology and Cell Biology*, vol. 83, no. 4, pp. 418–428, 2005.

[7] R. Kannagi and S. Hakomori, "A guide to monoclonal antibodies directed to glycotopes," *Advances in Experimental Medicine and Biology*, vol. 491, pp. 587–630, 2001.

[8] C.-H. Hsu, S.-C. Hung, C.-Y. Wu, and C.-H. Wong, "Toward automated oligosaccharide synthesis," *Angewandte Chemie*, vol. 50, no. 50, pp. 11872–11923, 2011.

[9] T. J. Boltje, T. Buskas, and G. J. Boons, "Opportunities and challenges in synthetic oligosaccharide and glycoconjugate research," *Nature Chemistry*, vol. 1, no. 8, pp. 611–622, 2009.

[10] S. I. Nishimura, "Combinatorial syntheses of sugar derivatives," *Current Opinion in Chemical Biology*, vol. 5, no. 3, pp. 325–335, 2001.

[11] S. M. Hancock, M. D. Vaughan, and S. G. Withers, "Engineering of glycosidases and glycosyltransferases," *Current Opinion in Chemical Biology*, vol. 10, no. 5, pp. 509–519, 2006.

[12] A. M. Daines, B. A. Maltman, and S. L. Flitsch, "Synthesis and modifications of carbohydrates, using biotransformations," *Current Opinion in Chemical Biology*, vol. 8, no. 2, pp. 106–113, 2004.

[13] J. K. Scott, D. Loganathan, R. B. Easley, X. Gong, and I. J. Goldstein, "A family of concanavalin A-binding peptides from a hexapeptide epitope library," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 12, pp. 5398–5402, 1992.

[14] M. N. Fukuda, "Peptide-displaying phage technology in glycobiology," *Glycobiology*, vol. 22, no. 3, pp. 318–325, 2012.

[15] F. C. Dudak, I. H. Boyaci, and B. P. Orner, "The discovery of small-molecule mimicking peptides through phage display," *Molecules*, vol. 16, no. 1, pp. 774–789, 2011.

[16] R. Cortese, F. Felici, G. Galfre, A. Luzzago, P. Monaci, and A. Nicosia, "Epitope discovery using peptide libraries displayed on phage," *Trends in Biotechnology*, vol. 12, no. 7, pp. 262–267, 1994.

[17] T. Matsubara, A. Onishi, T. Saito et al., "Sialic acid-mimic peptides as hemagglutinin inhibitors for anti-influenza therapy," *Journal of Medicinal Chemistry*, vol. 53, no. 11, pp. 4441–4449, 2010.

[18] N. K. Sauter, J. E. Hanson, G. D. Glick et al., "Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography," *Biochemistry*, vol. 31, no. 40, pp. 9609–9621, 1992.

[19] F. X. Theillet, F. A. Saul, B. Vulliez-Le Normand et al., "Structural mimicry of O-antigen by a peptide revealed in a complex with an antibody raised against *Shigella flexneri* serotype 2a," *Journal of Molecular Biology*, vol. 388, no. 4, pp. 839–850, 2009.

[20] G. P. Smith and V. A. Petrenko, "Phage display," *Chemical Reviews*, vol. 97, no. 2, pp. 391–410, 1997.

[21] G. P. Smith, "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface," *Science*, vol. 228, no. 4705, pp. 1315–1317, 1985.

[22] J. K. Scott and G. P. Smith, "Searching for peptide ligands with an epitope library," *Science*, vol. 249, no. 4967, pp. 386–390, 1990.

[23] B. Ru, J. Huang, P. Dai et al., "MimoDB: a new repository for mimotope data derived from phage display technology," *Molecules*, vol. 15, no. 11, pp. 8279–8288, 2010.

[24] J. Huang, B. Ru, P. Zhu et al., "MimoDB 2.0: a mimotope database and beyond," *Nucleic Acids Research*, vol. 40, no. D1, pp. D271–D277, 2012.

[25] K. R. Oldenburg, D. Loganathan, I. J. Goldstein, P. G. Schultz, and M. A. Gallop, "Peptide ligands for a sugar-binding protein isolated from a random peptide library," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 12, pp. 5393–5397, 1992.

[26] D. Jain, K. Kaur, B. Sundaravadivel, and D. M. Salunke, "Structural and functional consequences of peptide-carbohydrate mimicry: crystal structure of a carbohydrate-mimicking peptide bound to concanavalin A," *Journal of Biological Chemistry*, vol. 275, no. 21, pp. 16098–16102, 2000.

[27] D. L. Kooyman, S. B. Mcclellan, W. Parker et al., "Identification and characterization of a galactosyl peptide mimetic. Implications for use in removing xenoreactive anti-a gal antibodies," *Transplantation*, vol. 61, no. 6, pp. 851–855, 1996.

[28] J. Zhan, Z. Xia, L. Xu, Z. Yan, and K. Wang, "A peptide mimetic of Gal-α1,3-Gal is able to block human natural antibodies," *Biochemical and Biophysical Research Communications*, vol. 308, no. 1, pp. 19–22, 2003.

[29] C. L. Martens, S. E. Cwirla, R. Y. W. Lee et al., "Peptides which bind to E-selectin and block neutrophil adhesion," *Journal of Biological Chemistry*, vol. 270, no. 36, pp. 21129–21136, 1995.

[30] M. N. Fukuda, C. Ohyama, K. Lowitz et al., "A peptide mimic of E-selectin ligand inhibits sialyl Lewis X-dependent lung colonization of tumor cells," *Cancer Research*, vol. 60, no. 2, pp. 450–456, 2000.

[31] L. Yu, P. S. Yu, E. Yee Yen Mui et al., "Phage display screening against a set of targets to establish peptide-based sugar mimetics and molecular docking to predict binding site," *Bioorganic and Medicinal Chemistry*, vol. 17, no. 13, pp. 4825–4832, 2009.

[32] L. L. Eggink and J. K. Hoober, "A biologically active peptide mimetic of N-acetylgalactosamine/galactose," *BMC Research Notes*, vol. 2, article 23, 2009.

[33] Z. Xu, G. S. Qian, Q. Li, Q. J. Feng, G. M. Wu, and K. L. Li, "Screening of mimetic peptides for CD14 binding site with LBP and antiendotoxin activity of mimetic peptide in vivo and in vitro," *Inflammation Research*, vol. 58, no. 1, pp. 45–53, 2009.

[34] R. Hoess, U. Brinkmann, T. Handel, and I. Pastan, "Identification of a peptide which binds to the carbohydrate-specific monoclonal antibody B3," *Gene*, vol. 128, no. 1, pp. 43–49, 1993.

[35] Q. Lou and I. Pastan, "A Lewis(y) epitope mimicking peptide induces anti-Lewis(y) immune responses in rabbits and mice," *Journal of Peptide Research*, vol. 53, no. 3, pp. 252–260, 1999.

[36] J. Qiu, P. Luo, K. Wasmund, Z. Steplewski, and T. Kieber-Emmons, "Towards the development of peptide mimotopes of carbohydrate antigens as cancer vaccines," *Hybridoma*, vol. 18, no. 1, pp. 103–112, 1999.

[37] N. Katagihallimath, A. Mehanna, D. Guseva, R. Kleene, and M. Schachner, "Identification and validation of a Lewisx glycomimetic peptide," *European Journal of Cell Biology*, vol. 89, no. 1, pp. 77–86, 2010.

[38] B. M. Charalambous and I. M. Feavers, "Peptide mimics elicit antibody responses against the outer-membrane lipooligosaccharide of group B *Neisseria meningitidis*," *FEMS Microbiology Letters*, vol. 191, no. 1, pp. 45–50, 2000.

[39] Y. Hou and X. X. Gu, "Development of peptide mimotopes of lipooligosaccharide from nontypeable Haemophilus influenzae as vaccine candidates," *Journal of Immunology*, vol. 170, no. 8, pp. 4373–4379, 2003.

[40] T. Jouault, C. Fradin, F. Dzierszinski et al., "Peptides that mimic Candida albicans-derived β-1,2-linked mannosides," *Glycobiology*, vol. 11, no. 8, pp. 693–701, 2001.

[41] M. Simon-Haldi, N. Mantei, J. Franke, H. Voshol, and M. Schachner, "Identification of a peptide mimic of the L2/HNK-1 carbohydrate epitope," *Journal of Neurochemistry*, vol. 83, no. 6, pp. 1380–1388, 2002.

[42] J. Lang, J. Zhan, L. Xu, and Z. Yan, "Identification of peptide mimetics of xenoreactive α-Gal antigenic epitope by phage display," *Biochemical and Biophysical Research Communications*, vol. 344, no. 1, pp. 214–220, 2006.

[43] A. G. Laman, A. O. Shepelyakovskaya, I. A. Berezin et al., "Identification of pentadecapeptide mimicking muramyl peptide," *Vaccine*, vol. 25, no. 15, pp. 2900–2906, 2007.

[44] G. Rojas, A. Pupo, M. Del Rosario Aleman, and N. Santiago Vispo, "Preferential selection of Cys-constrained peptides from a random phage-displayed library by anti-glucitollysine antibodies," *Journal of Peptide Science*, vol. 14, no. 11, pp. 1216–1221, 2008.

[45] A. Menendez, D. A. Calarese, R. L. Stanfield et al., "A peptide inhibitor of HIV-1 neutralizing antibody 2G12 is not a structural mimic of the natural carbohydrate epitope on gp120," *FASEB Journal*, vol. 22, no. 5, pp. 1380–1392, 2008.

[46] T. Taki, D. Ishikawa, H. Hamasaki, and S. Handa, "Preparation of peptides which mimic glycosphingolipids by using phage peptide library and their modulation on β-galactosidase activity," *FEBS Letters*, vol. 418, no. 1-2, pp. 219–223, 1997.

[47] D. Ishikawa, H. Kikkawa, K. Ogino, Y. Hirabayashi, N. Oku, and T. Taki, "GD1α-replica peptides functionally mimic GD1α, an adhesion molecule of metastatic tumor cells, and suppress the tumor metastasis," *FEBS Letters*, vol. 441, no. 1, pp. 20–24, 1998.

[48] M. Takikawa, H. Kikkawa, T. Asai et al., "Suppression of GD1α ganglioside-mediated tumor metastasis by liposomalized WHW-peptide," *FEBS Letters*, vol. 466, no. 2-3, pp. 381–384, 2000.

[49] J. X. Qiu and D. M. Marcus, "Use of peptide ligands to analyze the fine specificity of antibodies against asialo GM1," *Journal of Neuroimmunology*, vol. 100, no. 1-2, pp. 58–63, 1999.

[50] T. Kieber-Emmons, P. Luo, J. Qiu, T. Y. Chang, M. Blaszczyk-Thurin, and Z. Steplewski, "Vaccination with carbohydrate peptide mimotopes promotes anti-tumor responses," *Nature Biotechnology*, vol. 17, no. 7, pp. 660–665, 1999.

[51] A. Wondimu, T. Zhang, T. Kieber-Emmons et al., "Peptides mimicking GD2 ganglioside elicit cellular, humoral and tumor-protective immune responses in mice," *Cancer Immunology, Immunotherapy*, vol. 57, no. 7, pp. 1079–1089, 2008.

[52] E. Förster-Waldl, A. B. Riemer, A. K. Dehof et al., "Isolation and structural analysis of peptide mimotopes for the disialoganglioside GD2, a neuroblastoma tumor antigen," *Molecular Immunology*, vol. 42, no. 3, pp. 319–325, 2005.

[53] A. B. Riemer, E. Förster-Waldl, K. H. Brämswig et al., "Induction of IgG antibodies against the GD2 carbohydrate tumor antigen by vaccination with peptide mimotopes," *European Journal of Immunology*, vol. 36, no. 5, pp. 1267–1274, 2006.

[54] M. Bleeke, S. Fest, N. Huebener et al., "Systematic amino acid substitutions improved efficiency of GD2-peptide mimotope vaccination against neuroblastoma," *European Journal of Cancer*, vol. 45, no. 16, pp. 2915–2921, 2009.

[55] E. Bolesta, A. Kowalczyk, A. Wierzbicki et al., "DNA vaccine expressing the mimotope of GD2 ganglioside induces protective GD2 cross-reactive antibody responses," *Cancer Research*, vol. 65, no. 8, pp. 3410–3418, 2005.

[56] A. Wierzbicki, M. Gil, M. Ciesielski et al., "Immunization with a mimotope of GD2 ganglioside induces CD8+ T cells that recognize cell adhesion molecules on tumor cells," *Journal of Immunology*, vol. 181, no. 9, pp. 6644–6653, 2008.

[57] I. Horwacik, D. Czaplicki, K. Talarek et al., "Selection of novel peptide mimics of the GD2 ganglioside from a constrained phage-displayed peptide library," *International Journal of Molecular Medicine*, vol. 19, no. 5, pp. 829–839, 2007.

[58] I. Horwacik, M. Kurciński, M. Bzowska et al., "Analysis and optimization of interactions between peptides mimicking the GD2 ganglioside and the monoclonal antibody 14G2a," *International Journal of Molecular Medicine*, vol. 28, no. 1, pp. 47–57, 2011.

[59] I. Popa, D. Ishikawa, M. Tanaka, K. Ogino, J. Portoukalian, and T. Taki, "GD3-replica peptides selected from a phage peptide library induce a GD3 ganglioside antibody response," *FEBS Letters*, vol. 580, no. 5, pp. 1398–1404, 2006.

[60] Y. Miura, A. Sakaki, M. Kamihira, S. Iijima, and K. Kobayashi, "A globotriaosylceramide (Gb3Cer) mimic peptide isolated from phage display library expressed strong neutralization to

Shiga toxins," *Biochimica et Biophysica Acta*, vol. 1760, no. 6, pp. 883–889, 2006.

[61] A. Perez, E. S. Mier, N. S. Vispo, A. M. Vazquez, and R. P. Rodríguez, "A monoclonal antibody against NeuGc- containing gangliosides contains a regulatory idiotope involved in the interaction with B and T cells," *Molecular Immunology*, vol. 39, no. 1-2, pp. 103–112, 2002.

[62] A. López-Requena, C. M. De Acosta, E. Moreno et al., "Gangliosides, Ab1 and Ab2 antibodies. I. Towards a molecular dissection of an idiotype-anti-idiotype system," *Molecular Immunology*, vol. 44, no. 4, pp. 423–433, 2007.

[63] J. H. Youn, H. J. Myung, A. Liav et al., "Production and characterization of peptide mimotopes of phenolic glycolipid-I of *Mycobacterium leprae*," *FEMS Immunology and Medical Microbiology*, vol. 41, no. 1, pp. 51–57, 2004.

[64] P. Valadon, G. Nussbaum, L. F. Boyd, D. H. Margulies, and M. D. Scharff, "Peptide libraries define the fine specificity of anti-polysaccharide antibodies to *Cryptococcus neoformans*," *Journal of Molecular Biology*, vol. 261, no. 1, pp. 11–22, 1996.

[65] A. C. M. Young, P. Valadon, A. Casadevall, M. D. Scharff, and J. C. Sacchettini, "The three-dimensional structures of a polysaccharide binding antibody to *Cryptococcus neoformans* and its complex with a peptide from a phage display library: implications for the identification of peptide mimotopes," *Journal of Molecular Biology*, vol. 274, no. 4, pp. 622–634, 1997.

[66] D. O. Beenhouwer, R. J. May, P. Valadon, and M. D. Scharff, "High affinity mimotope of the polysaccharide capsule of *Cryptococcus neoformans* identified from an evolutionary phage peptide library," *Journal of Immunology*, vol. 169, no. 12, pp. 6992–6999, 2002.

[67] P. Valadon, G. Nussbaum, J. Oh, and M. D. Scharff, "Aspects of antigen mimicry revealed by immunization with a peptide mimetic of *Cryptococcus neoformans* polysaccharide," *Journal of Immunology*, vol. 161, no. 4, pp. 1829–1836, 1998.

[68] S. H. Pincus, M. J. Smith, H. J. Jennings, J. B. Burritt, and P. M. Glee, "Peptides that mimic the group B streptococcal type III capsular polysaccharide antigen," *Journal of Immunology*, vol. 160, no. 1, pp. 293–298, 1998.

[69] G. B. Lesinski, S. L. Smithson, N. Srivastava, D. Chen, G. Widera, and M. A. J. Westerink, "A DNA vaccine encoding a peptide mimic of *Streptococcus pneumoniae* serotype 4 capsular polysaccharide induces specific anti-carbohydrate antibodies in Balb/c mice," *Vaccine*, vol. 19, no. 13-14, pp. 1717–1726, 2001.

[70] U. K. Buchwald, A. Lees, M. Steinitz, and L. A. Pirofski, "A peptide mimotope of type 8 pneumococcal capsular polysaccharide induces a protective immune response in mice," *Infection and Immunity*, vol. 73, no. 1, pp. 325–333, 2005.

[71] C. M. Smith, C. L. Passo, A. Scuderi et al., "Peptide mimics of two pneumococcal capsular polysaccharide serotypes (6B and 9V) protect mice from a lethal challenge with *Streptococcus pneumoniae*," *European Journal of Immunology*, vol. 39, no. 6, pp. 1527–1535, 2009.

[72] S. L. Harris, L. Craig, J. S. Mehroke et al., "Exploring the basis of peptide-carbohydrate crossreactivity: evidence for discrimination by peptides between closely related anti-carbohydrate antibodies," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 6, pp. 2454–2459, 1997.

[73] Y. Wu, Q. Zhang, D. Sales, A. E. Bianco, and A. Craig, "Vaccination with peptide mimotopes produces antibodies recognizing bacterial capsular polysaccharides," *Vaccine*, vol. 28, no. 39, pp. 6425–6435, 2010.

[74] M. C. Grothaus, N. Srivastava, S. L. Smithson et al., "Selection of an immunogenic peptide mimic of the capsular polysaccharide of *Neisseria meningitidis* serogroup A using a peptide display library," *Vaccine*, vol. 18, no. 13, pp. 1253–1263, 2000.

[75] I. Park, I. H. Choi, S. J. Kim, and J. S. Shin, "Peptide mimotopes of *Neisseria meningitidis* group B capsular polysaccharide," *Yonsei Medical Journal*, vol. 45, no. 4, pp. 755–758, 2004.

[76] V. Lauvrak, G. Berntzen, U. Heggelund et al., "Selection and characterization of cyclic peptides that bind to a monoclonal antibody against meningococcal L3,7,9 lipopolysaccharides," *Scandinavian Journal of Immunology*, vol. 59, no. 4, pp. 373–384, 2004.

[77] P. Torregrossa, L. Buhl, M. Bancila et al., "Selection of poly-α 2,8-sialic acid mimotopes from a random phage peptide library and analysis of their bioactivity," *Journal of Biological Chemistry*, vol. 279, no. 29, pp. 30707–30714, 2004.

[78] P. Marino, J. C. Norreel, M. Schachner, G. Rougon, and M. C. Amoureux, "A polysialic acid mimetic peptide promotes functional recovery in a mouse model of spinal cord injury," *Experimental Neurology*, vol. 219, no. 1, pp. 163–174, 2009.

[79] C. Lo Passo, A. Romeo, I. Pernice et al., "Peptide mimics of the group B meningococcal capsule induce bactericidal and protective antibodies after immunization," *Journal of Immunology*, vol. 178, no. 7, pp. 4417–4423, 2007.

[80] T. Menéndez, N. F. Santiago-Vispo, Y. Cruz-Leal et al., "Identification and characterization of phage-displayed peptide mimetics of *Neisseria meningitidis* serogroup B capsular polysaccharide," *International Journal of Medical Microbiology*, vol. 301, no. 1, pp. 16–25, 2011.

[81] D. M. Prinz, S. L. Smithson, and M. A. J. Westerink, "Two different methods result in the selection of peptides that induce a protective antibody response to *Neisseria meningitidis* serogroup C," *Journal of Immunological Methods*, vol. 285, no. 1, pp. 1–14, 2004.

[82] G. R. Moe and D. M. Granoff, "Molecular mimetics of *Neisseria meningitidis* serogroup B polysaccharide," *International Reviews of Immunology*, vol. 20, no. 2, pp. 201–220, 2001.

[83] S. Falklind-Jerkérus, F. Felici, C. Cavalieri et al., "Peptides mimicking *Vibrio cholerae* O139 capsular polysaccharide elicit protective antibody response," *Microbes and Infection*, vol. 7, no. 15, pp. 1453–1460, 2005.

[84] M. N. Dharmasena, D. A. Jewell, and R. K. Taylor, "Development of peptide mimics of a protective epitope of *Vibrio cholerae* Ogawa O-antigen and investigation of the structural basis of peptide mimicry," *Journal of Biological Chemistry*, vol. 282, no. 46, pp. 33805–33816, 2007.

[85] M. N. Dharmasena, S. J. Krebs, and R. K. Taylor, "Characterization of a novel protective monoclonal antibody that recognizes an epitope common to *Vibrio cholerae* Ogawa and Inaba serotypes," *Microbiology*, vol. 155, no. 7, pp. 2353–2364, 2009.

[86] A. Phalipon, A. Folgori, J. Arondel et al., "Induction of anti-carbohydrate antibodies by phage library-selected peptide mimics," *European Journal of Immunology*, vol. 27, no. 10, pp. 2620–2625, 1997.

[87] M. Geiser, D. Schultz, A. Le Cardinal, H. Voshol, and C. García-Echeverría, "Identification of the human melanoma-associated chondroitin sulfate proteoglycan antigen epitope recognized by the antitumor monoclonal antibody 763.74 from a peptide phage library," *Cancer Research*, vol. 59, no. 4, pp. 905–910, 1999.

[88] H. Melzer, P. Fortugno, E. Mansouri et al., "Antigenicity and immunogenicity of phage library-selected peptide mimics of

the major surface proteophosphoglycan antigens of *Entamoeba histolytica*," *Parasite Immunology*, vol. 24, no. 6, pp. 321–328, 2002.

[89] G. Gevorkian, E. Segura, G. Acero et al., "Peptide mimotopes of *Mycobacterium tubercolosis* carbohydrate immunodeterminants," *Biochemical Journal*, vol. 387, no. 2, pp. 411–417, 2005.

[90] A. Barenholz, A. H. Hovav, Y. Fishman, G. Rahav, J. M. Gershoni, and H. Bercovier, "A peptide mimetic of the mycobacterial mannosylated lipoarabinomannan: characterization and potential applications," *Journal of Medical Microbiology*, vol. 56, no. 5, pp. 579–586, 2007.

[91] J. B. Legutki, M. Nelson, R. Titball, D. R. Galloway, A. Mateczun, and L. W. Baillie, "Analysis of peptide mimotopes of *Burkholderia pseudomallei* exopolysaccharide," *Vaccine*, vol. 25, no. 45, pp. 7796–7805, 2007.

[92] C. Beninati, M. Garibaldi, C. L. Passo et al., "Immunogenic mimics of *Brucella* lipopolysaccharide epitopes," *Peptides*, vol. 30, no. 10, pp. 1936–1939, 2009.

[93] Y. Chen, B. Liu, D. Yang et al., "Peptide mimics of peptidoglycan are vaccine candidates and protect mice from infection with *Staphylococcus aureus*," *Journal of Medical Microbiology*, vol. 60, no. 7, pp. 995–1002, 2011.

[94] J. P. Tam, "Synthetic peptide vaccine design: synthesis and properties of a high-density multiple antigenic peptide system," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 85, no. 15, pp. 5409–5413, 1988.

[95] N. Yuki, "Infectious origins of, and molecular mimicry in, Guillain-Barré and Fisher syndromes," *Lancet Infectious Diseases*, vol. 1, no. 1, pp. 29–37, 2001.

[96] T. Feizi, "Demonstration by monoclonal antibodies that carbohydrate structures of glycoproteins and glycolipids are onco-developmental antigens," *Nature*, vol. 314, no. 6006, pp. 53–57, 1985.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 156482, 13 pages
doi:10.1155/2012/156482

Review Article

Artificial Specific Binders Directly Recovered from Chemically Modified Nucleic Acid Libraries

Yuuya Kasahara and Masayasu Kuwahara

Graduate School of Engineering, Gunma University, 1-5-1 Tenjin-cho, Kiryu 376-8515, Japan

Correspondence should be addressed to Masayasu Kuwahara, mkuwa@gunma-u.ac.jp

Received 8 July 2012; Accepted 19 August 2012

Academic Editor: Hiroshi Murakami

Specific binders comprised of nucleic acids, that is, RNA/DNA aptamers, are attractive functional biopolymers owing to their potential broad application in medicine, food hygiene, environmental analysis, and biological research. Despite the large number of reports on selection of natural DNA/RNA aptamers, there are not many examples of direct screening of chemically modified nucleic acid aptamers. This is because of (i) the inferior efficiency and accuracy of polymerase reactions involving transcription/reverse-transcription of modified nucleotides compared with those of natural nucleotides, (ii) technical difficulties and additional time and effort required when using modified nucleic acid libraries, and (iii) ambiguous efficacies of chemical modifications in binding properties until recently; in contrast, the effects of chemical modifications on biostability are well studied using various nucleotide analogs. Although reports on the direct screening of a modified nucleic acid library remain in the minority, chemical modifications would be essential when further functional expansion of nucleic acid aptamers, in particular for medical and biological uses, is considered. This paper focuses on enzymatic production of chemically modified nucleic acids and their application to random screenings. In addition, recent advances and possible future research are also described.

1. Introduction

RNA/DNA aptamers, which are specific for a broad spectrum of targets, can be artificially created by systematic evolution of ligands by exponential enrichment (SELEX) methods [1, 2]. Large-scale chemical synthesis of RNA/DNA aptamers is possible, and synthesizing them is less expensive than producing antibodies; therefore, they have been considered as alternatives to therapeutic antibodies. Although RNA/DNA aptamers do not cause antibody-dependent cell-mediated cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC), their specific binding abilities are expected to neutralize actions on the target and relieve symptoms. Indeed, the first example of an aptamer drug, "Macugen (pegaptanib sodium injection)" is being used for age-related macular degeneration (AMD) therapy [3]. Pegaptanib is a RNA-based aptamer that involves 2′-fluoropyrimidine nucleotides (U, C) and 2′-methoxy purine nucleotides (A, G) to remain intact under physiological conditions. In addition, a branched polyethylene glycol strand (40 kDa) and 3′-thymidylic acid are introduced at its 5′ and 3′ ends, respectively. The 5′-end modification is known to prolong circulation time *in vivo* as well as to enhance nuclease resistance. Pegaptanib tightly binds to the vascular endothelial growth factor (VEGF) in a Ca^{2+}-dependent fashion with a dissociation constant (K_d) of 200 pM, while the corresponding aptamer, which lacks the 5′- and 3′-end capping, has much higher affinity ($K_d = 49 \pm 6$ pM at 37°C in phosphate-buffered saline containing 2 mM Ca^{2+}). Incidentally, the K_d value of the anti-VEGF antibody, "Avastin (bevacizumab)," which is used for cancer therapies, is 1.1 nM at 25°C. The natural type of anti-VEGF RNA aptamers also shows high-binding affinity at a picomolar range ($K_d = 140 \pm 4$ pM at 37°C in phosphate-buffered saline containing no Ca^{2+}) [4], indicating that the effects of chemical modifications on binding affinity are not significant, considering the different Ca^{2+} concentrations used. In contrast, the effects on biostability are remarkable; pegaptanib was found to be stable after incubation at ambient temperature for 18 h in human plasma containing ethylenediaminetetraacetic acid,

whereas unmodified oligoribonucleic acids are known to degrade within a few minutes *in vivo* [5].

Owing to the limited tolerance for modified substrates of the RNA polymerase (T7 RNA polymerase) used for SELEX, the 2′-methoxy (–OMe) groups need to be replaced with 2′-hydroxy (–OH) groups of natural purine nucleotides after obtaining the precursor from a modified RNA library involving 2′-fluoro (–F) analogs of uridine and cytidine and natural adenosine and guanosine (Figure 1). The post-SELEX modifications have been successful in rendering nuclease resistance but required considerable time and effort because binding affinities could be markedly decreased or eliminated, depending on the position of the replacement. To overcome this problem, T7 RNA polymerase double-mutant Y639F/H784A was used for enzymatic preparation of the modified RNA library in the SELEX processes, and 2′-OMe RNA aptamers specific to VEGF have been successfully screened directly [6]. One of the 2′-OMe RNA aptamers that could be minimized to 23-mer (which is an unusual short length) was found to be quite stable, and no degradation was observed after incubation at 37°C for 96 h in plasma. Despite being successful for direct screening, structural minimizing, and biostability enhancing, these aptamers were found to have binding affinities in a low nanomolar range that were inferior to those of pegaptanib and its precursors.

This may be because the potential binding ability of the chemical library used was inherently low, and/or the unusual polymerase reaction would cause unfavorable critical biases in the sequences of the chemical library constructed. Conversely, it may also be possible that differences in the selection outcomes would not be sufficient to clarify their causes because only a part of all possible sequences were screened. This is a characteristic difficulty in SELEX when chemical modification is involved. Regardless of this difficulty, a polymerase reaction involving modified nucleotides is a key step that should be improved and optimized to construct desirable direct screening systems for modified RNA/DNA aptamers when the SELEX methods are applied.

2. Enzymatic Modified RNA/DNA Polymerization

2.1. Kinetics of Modified Substrate Triphosphates Incorporation for SELEX. RNA/DNA polymerases incorporate substrate triphosphates (NTPs/dNTPs) corresponding to the type of bases on the template strand and successively add them to the 3′ end of the extending strand to form 3′,5′-phosphodiester linkage. Some polymerases are known to accept chemically modified NTPs/dNTPs as substrates and can produce nucleic acid polymers containing foreign functionalities. Such polymerase reactions are applied to DNA sequencing [7–10], fluorophore, and redox labeling [11, 12], expanding the genetic alphabet [13–15], and preparing library for SELEX [16–48]. Unlike enzymatic functional labeling of DNA, for which modified dNTP is often used in the presence of the corresponding natural dNTP to increase product yields, modified NTP/dNTP is generally used in the absence of the corresponding natural NTP/dNTP when modified nucleic acid libraries are prepared. This is because the natural nucleotide needs to be completely replaced with the corresponding modified nucleotide at all sites incorporated into the extending strand. In general, total replacement could decrease the product yield because the catalytic efficiencies of the polymerase may be affected by the modifications not only for substrate triphosphate but also on the extending strand and template. Previously reported kinetic studies using a base modified nucleotide showed that the reaction efficiencies of single modified nucleotide incorporation are drastically decreased when the modifications exist on the 3′ terminus of the extending strand, although the single incorporation of the modified substrate proceeds smoothly at almost the same rate as the corresponding natural substrate [37]. The results indicate that the successive incorporation of modified nucleotides is the most difficult aspect of strand extension. Therefore, the inefficiency of modified RNA/DNA polymerization could naturally bias the outcomes of the selection; it could unintentionally lead to the exclusion of the sequences with the highest binding affinity. To reduce this influence, reactions are often conducted under very high enzyme and/or substrate concentrations to achieve large reaction velocities. However, it should be noted that such conditions are prone to result in a high frequency of misincorporations. As the solution strategy, polymerase variants, triphosphate analogs, and their combinations that improve the reaction efficiency have been developed and are still being studied. Furthermore, when modified DNA is used in SELEX, the modified DNA is normally amplified indirectly by a polymerase chain reaction (PCR) to prepare the next library. After affinity selection, the selected modified DNA is reverse transcribed and PCR amplified to natural DNA, and then transcribed to modified DNA, even when PCR amplification was available for the modification.

2.2. Library Preparation. T7 RNA polymerase and its variants have primarily been used for SELEX using modified RNA. These polymerases could accept NTP analogs of 2′-thio (–SH) [17], 2′-amino ($–NH_2$) [19], 2′-azido ($–N_3$) [29], 2′-hydroxymethyl ($–CH_2OH$) [33], and 4′-thio (–S–) [39] in addition to 2′-F and 2′-OMe as substrates. Furthermore, NTP analogs with base modification (e.g., C5-modified uridines and cytidines) and with phosphate-modified nucleosides (e.g., 5′-(α-thio)triphosphates and 5′-(α-borano)triphosphates) [21, 30], which are available for modified RNA polymerization, have also been reported. It is known that polymerization by T7 RNA polymerases starts with the generation of purine-rich oligonucleotide with a length of approximately 10 residues in the initiation step, and the composition of the leader oligonucleotide, preferably with guanosine residues, greatly influences the transcript yields [49]. In the initiation step, the polymerases are sensitive to modifications of the 2′-hydroxyl group, while recognition of the 2′-position is tolerated during the elongation step [50]. Therefore, reactions are often performed using modified NTPs plus a low ratio of natural GTP to increase transcript yields. For example, a highly modified 2′-OMe RNA library was provided in sufficient

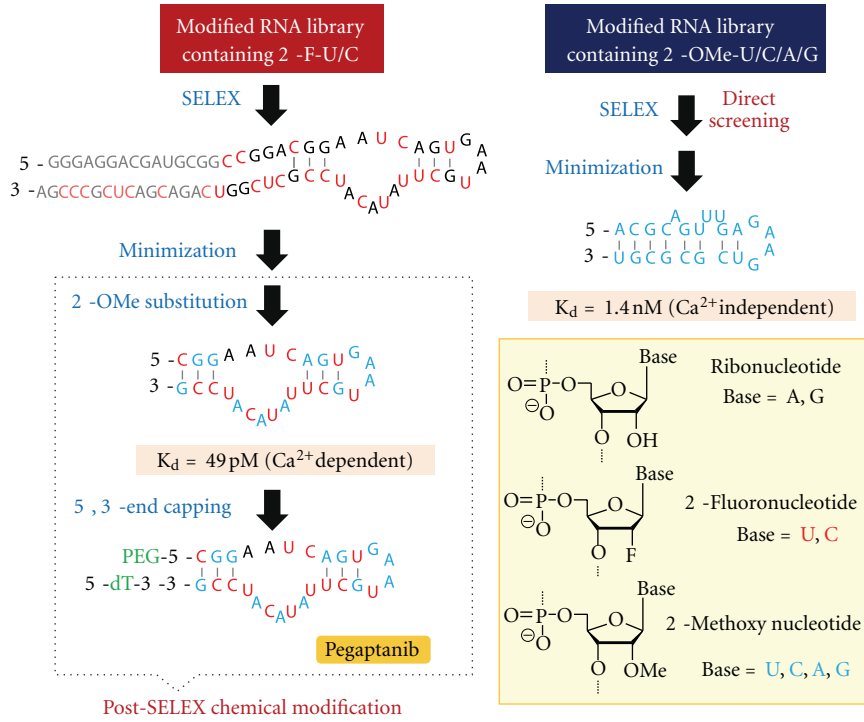


Figure 1: Preparation scheme for chemically modified nucleic acid aptamers that bind to VEGF. High nuclease-resistant 2′-methoxy nucleotides were introduced through Post-SELEX modification process (left), and fully modified 2′-OMe RNA aptamers were directly selected from a library of 2′-OMe transcripts (right).

yield by a reaction containing four 2′-OMe NTPs and natural GTP, which was catalyzed by the T7 RNA polymerase variant (Y639F/H784A) and initiated by the formation of the leader oligonucleotide with a sequence of GGGAGAGGAGAGAA [6]. Thus, preparations of modified RNA libraries typically require a low ratio of natural GTP or both GTP and ATP, and polymerase mutation could reduce the ratio.

For SELEX using modified DNA, certain thermophilic DNA polymerases, for example, *Pwo*, *Pfu*, *Vent(exo-)*, *Deep Vent(exo-)*, and *KOD Dash*, that belong to the evolutional family B, were found to be preferable to other types of DNA polymerases [31]. Especially in successive incorporations of modified nucleotides, those polymerases were found to exhibit much superior performance than family A DNA polymerases such as *Taq*, *Tth*, and *thermo sequenase*. In addition, a family D DNA polymerase derived from *Pyrococcus horikoshii* did not show any tolerance for chemical modification in the experiments using C5-substituted pyrimidine nucleoside triphosphates [41]. The efficiency of enzymatic production using modified dNTP varies depending on the site where the substituent is introduced. As for base modification, dNTP analogs with pyrimidine substituted at the 5th position and purine substituted at the 7th position of the base moiety tend to be acceptable for DNA polymerases and act as good substrates [27, 32]. Modified purine nucleotide analogs at the 8th position can also be incorporated but with lower efficiency [10, 42]. In addition, sugar modifications such as 2′-fluoro, 2′-fluoro-D-arabino, and 2′-*O*,4′-*C*-methylene (BNA/LNA) were also found to be acceptable [44–46]. Furthermore, phosphate-modified dNTPs with 5′-(*α*-thio)triphosphates and 5′-(*α*-borano)triphosphates were found to work as alternative substrates. Certain types of DNA polymerases were also found to accept some modifications of the leaving group of the phosphate moiety [9], although those analogs have been applied for advanced DNA sequencing and not for SELEX.

Among the modifications, DNA polymerases could endure substitution at the base moieties, and various functional groups could be introduced into those positions with relatively high efficiency. When artificially created protein-like functional nucleic acids were considered, researchers would first envisage introducing proteinous amino acids into nucleic acids. Indeed, base-modified dNTP analogs bearing various proteinous amino acids or their side chains have been reported to date, and their substrate properties in polymerase reactions such as PCR and primer extension have been investigated (Figure 2) [38]. For example, PCR assays using *KOD Dash* DNA polymerase showed that triphosphates containing amino acyl group with basic (Arg, His, Lys), aromatic (Phe, Trp), aliphatic (Leu, Pro), and neutral hydrophilic (Gln, Ser, Thr) side chains act as good substrates, while those with acidic (Asp, Glu) and thiol (Cys) side chains act as poor substrates. Production of DNA-containing cysteinyl residue necessitated the addition of dithiothreitol as a reduction reagent. To introduce plural functionalities with high density, it was found that four natural nucleotides (A, G, C, T) are totally replaced with four base-modified nucleotides by the addition of manganese chloride and betaine [32, 35]. Those

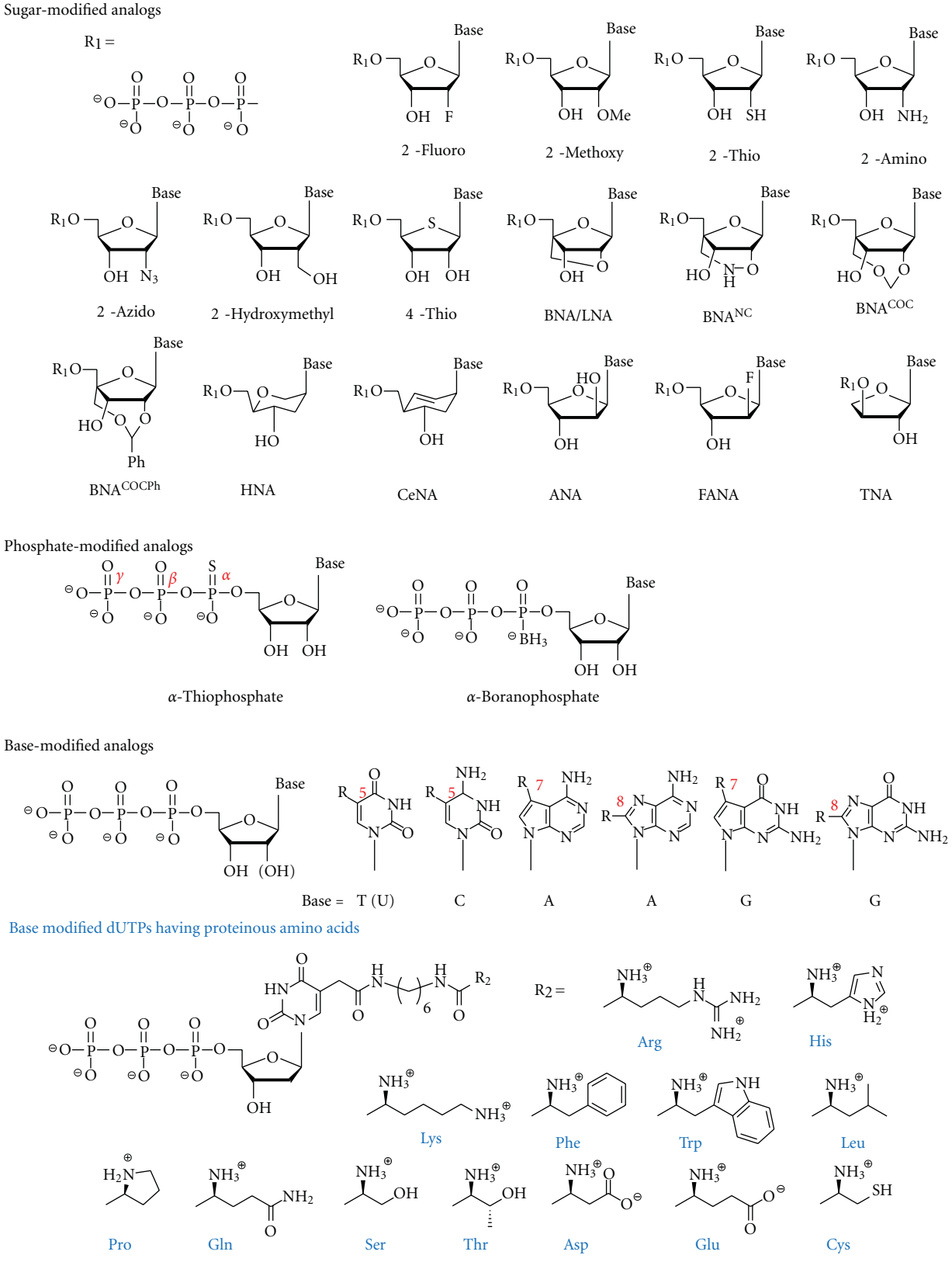


Figure 2: Examples of modified nucleoside triphosphates that act as substrates for polymerase reactions.

additives could improve efficiency and yield in the enzymatic production of modified DNA, although they could raise the frequency of misincorporation at the same time [51, 52].

3. Efficacies of Introduced Foreign Functional Groups

3.1. Biostability and Binding Property. In modified RNA/DNA aptamers created by direct screening of modified libraries, the effectiveness of modified groups introduced on biostability has been sufficiently supported by many examples. In particular, RNA-based aptamers with modified sugar, in which the 2′-hyrdoxyl group is substituted, dramatically enhanced nuclease resistances. In addition to pegaptanib, for example, a modified RNA aptamer with 2′-NH_2-U/C specific to human neutrophil elastase (HNE) retained its intact form in serum with a degradation half-life of approximately 20 h, while the corresponding natural RNA mostly degraded just within 5 min [53]. Moreover, thioaptamers, RNA/DNA-based aptamers with modified phosphate, that is, α-phosphorothioate, which typically displays increased stability in the biological milieu, have been developed [54]. On the other hand, reports that imply positive effects of modification in binding properties are limited. That is, chemical modifications do not always result in raising binding affinity and specificity as researchers expected (Sections 3.2 and 3.3). Regarding the beneficial aspects of modification, we previously reported that a base-modified DNA aptamer, which was recovered from a modified DNA library containing 5-(2-(6-aminohexylamino)-2-oxoethyl-2′-deoxyuridine, can bind with the R-isomer of a thalidomide derivative with high enantioselectivity [55]. Moreover, another base-modified DNA aptamer, selected from a DNA library containing arginine residues, could clearly distinguish the dicarboxylate moiety of D-glutamic acid from that of L-isomer [56]. Also, Li et al. reported that a base-modified DNA aptamer carrying the boronic acid moiety, which was obtained by implementing thorough counter selections, can sensitively recognize glycosylation sites of fibrinogen as a glycoprotein [57]. Although the impact on binding affinity seems to be limited, these examples merit attention because of their implications for the possibility of expanding binding modes in molecular recognition with introduced foreign functional groups. Thus, the use of base-modified DNA for selection library would certainly be one of the most promising strategies to improve binding properties, because of the aforementioned functional expandability; repertories of modification in bases, which could be available for enzymatic incorporation, are much broader than those in the other parts. Furthermore, the number of candidates for polymerase variants capable of catalyzing modified nucleotide polymerization with high efficiency is expected to be much greater in DNA polymerases than in RNA polymerases.

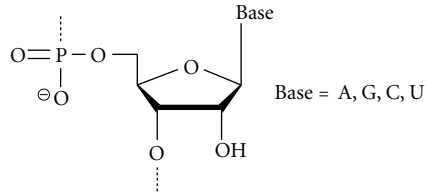
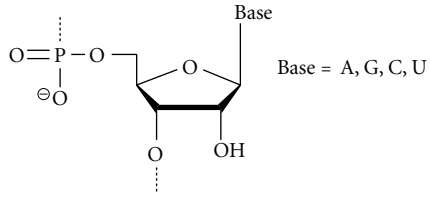
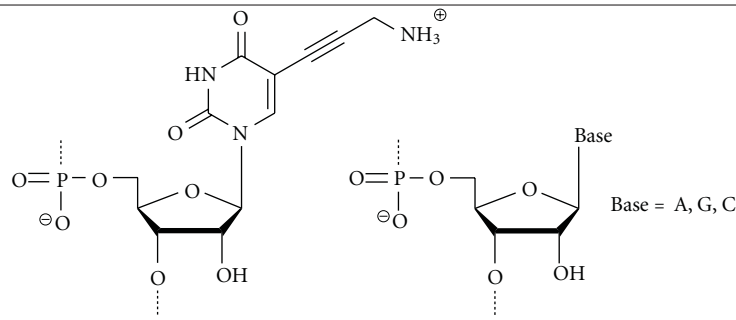
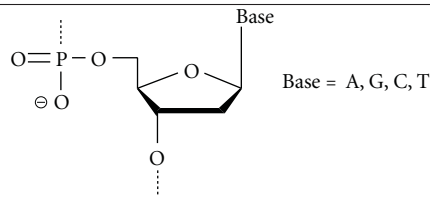
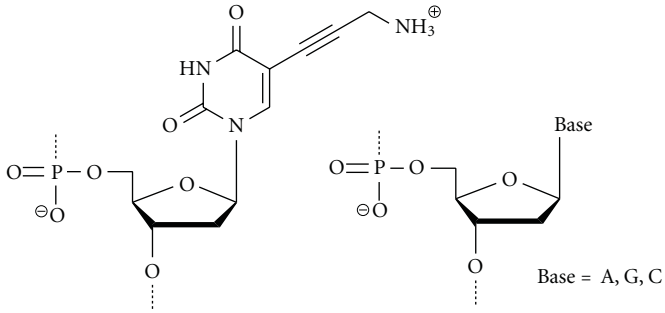
3.2. Efficacy of Modification on Binding to Small Molecular Targets. To probe superiorities in terms of introducing foreign functionality, random screenings from nucleic acid libraries of compounds having different chemical structures should be performed using the same target, and those selection outcomes should be collated carefully. One small molecular target that has generated significant attention is adenosine-5′-triphosphate (ATP). To date, ATP-binding aptamers have been obtained from libraries of RNA, DNA, modified RNA, and modified DNA [58–62]. However, neither the benefit of modification nor the contribution expected from the chemical structure of foreign functionality to binding affinity and specificity is likely to be substantiated by those SELEX experiments (Table 1).

RNA aptamers were reported by Sazani et al. and Sassanfar and Szostak, respectively, and a modified RNA aptamer was reported by Vaish et al.; those aptamers were found to bind to ATP in a 1 : 1 stoichiometry, depending on the Mg^{2+} concentration. The K_d value of Sassanfar's RNA aptamer was found to be 6–8 μM at 5 mM Mg^{2+} and 0.7 μM at 20 mM Mg^{2+}, while that of Sazani's RNA aptamer was 11 μM at 10 mM Mg^{2+} and 4.8 μM at 30 mM Mg^{2+}. The modified RNA aptamer contains 5-(3-aminopropyl) uridine, and its binding affinity was significantly improved as the Mg^{2+} concentration increased up to 3 mM, with a K_d value estimated to be between 0.45 and 1.1 μM at 6 mM Mg^{2+}. Although there were some differences in the selection protocol, no notable differences in binding affinities between natural and modified RNA aptamers were observed, with K_d values for the modified RNA aptamers ranging from approximately 10^{-7} to 10^{-6}.

Electrostatic interactions between the introduced amino group and the triphosphate moiety of ATP were expected; however, the modified RNA aptamer does not clearly distinguish ATP from ADP, AMP, or adenosine (K_d = 1.02, 1.01, 2.18, and 3.91 μM, resp.) under the same conditions, although the foreign functional group is essential for target binding. Similarly, significant binding specificities to the triphosphate moiety in Sassanfar's RNA were not observed, with K_d values for ATP and those related targets being similar within the range of 2-3 μM. In contrast, surprisingly, Sazani's RNA aptamer exhibits superior binding specificity for the moiety; its binding affinity is 64-fold lower because of a lack of γ-phosphate in ATP and is 1100-fold lower because of lack of both β- and γ-phosphates. These results indicate that the introduction of a cationic functional group could be alternated with that of native functionalities in RNA and divalent metal ions in solution. Conversely, the modified RNA aptamer exhibits superior specificity for the base moiety in comparison to Sazani's RNA aptamer; the former shows 150-fold weaker binding to inosine-5′-triphosphate (ITP) and does not bind to GTP, CTP, or UTP, while the latter shows 66-, 19-, 680-, and 610-fold weaker binding to ITP, GTP, CTP, and UTP, respectively.

A DNA aptamer reported by Huizenga and Szostak binds to ATP with an apparent K_d value of 6 ± 3 μM, which was calculated assuming a 1 : 1 binding ratio and shows a similar affinity to AMP and adenosine. However, a sequence containing a representative motif of Huizenga's DNA aptamer was found to cooperatively bind two ATP molecules with a binding ratio of 1 : 2 and a K_d of

Table 1: Comparison of binding affinities in ATP-binding natural/modified nucleic acid aptamers.

Aptamer type	Chemical structure	Binding affinity K_d	References
Natural RNA		6–8 μM (5 mM Mg^{2+}) 0.7 μM (20 mM Mg^{2+})	[58]
Natural RNA		11 μM (10 mM Mg^{2+}) 4.8 μM (30 mM Mg^{2+})	[59]
Modified RNA		0.45–1.1 μM (6 mM Mg^{2+})	[60]
Natural DNA		9 ± 2 μM^2	[61]
Modified DNA		6 ± 1 μM^2	[62]

$9 \pm 2\,\mu M^2$. Interestingly, a modified DNA aptamer containing 5-(3-aminopropyl)-2′-deoxyuridine also forms a 1 : 2 complex with ATP, with a K_d value of $6 \pm 1\,\mu M^2$. Unlike the ATP-binding-modified RNA aptamer, the DNA aptamer with thymidine replacing the modified dU was found to retain the binding affinity, although the affinity for ATP was lowered by approximately 2-fold ($K_d = 13 \pm 4\,\mu M^2$). This indicates that the introduction of functionality did not dramatically influence the structures or activities of the aptamers selected.

3.3. Efficacy of Modifications on Binding to Protein Targets. Thrombin (FIIa) is a multifunctional serine protease that plays key roles in hemostasis, thrombosis, and inflammation.

Therefore, developing nucleic acid aptamers for antithrombin therapy has been of great interest since the earliest SELEX studies. To date, a number of SELEX experiments have been performed to obtain thrombin-binding aptamers (TBAs) using natural or modified nucleic acid libraries (Table 2) [63–69].

A minimized 20 mer RNA aptamer reported by Kubik et al. binds to human thrombin with a K_d value of 9.3 ± 1.0 nM. White et al. reported a minimized 25 mer RNA aptamer with superior binding affinity ($K_d = 0.54 \pm 0.1$ nM), which was recovered by a method named toggle SELEX. A modified RNA aptamer comprising all four 4′-thionucleotides (A, G, C, U) was obtained with a K_d value of 7.2 nM. Another type of modified TBA consists of natural 2′-deoxy purine nucleotides (A, G) and 2′-OMe pyrimidine nucleotides (C, U), having libraries constructed with Y639F mutant T7 RNA polymerase, exhibiting a binding affinity in the nanomolar range ($K_d = 26$ nM). Minimized 15 mer and 29 mer DNA aptamers were reported by Bock et al. and Tasset et al., respectively; their respective K_d values were 113 ± 20 nM and 0.5 nM. A modified DNA aptamer containing 5-(1-pentynyl)-2′-dU exhibited weaker binding affinity with a K_d value of 400 nM. However, the experimental data indicated that these chemical modifications did not result in superior binding affinity.

Two electropositive domains (i.e., exosite-1 and exosite-2) that are displayed on the surface of thrombin are known to be TBA-binding sites [71]. Binding to these domains, which are distal from the catalytic site, could affect enzyme activity. The 15 mer DNA aptamer formed a two-tiered G-quadruplex and recognized exosite-1 via van der Waals forces and hydrogen bonding [72]. Although it was considered that the 29 mer DNA aptamer forms a two-tiered G-quadruplex containing structure based on its primary sequence, its binding site was found to be exosite-2 [73].

Natural RNA aptamers form hairpin structures and also bind to exosite-2. The 4′-thio TBA and the pentynyl TBA were predicted to have hairpin structures. With regard to the 2′-OMe TBA, the formation of potential stacked G-quartets was suggested by its multiple contiguous guanines. The binding sites for these modified TBAs have not been thoroughly investigated. Inferences based on NMR and X-ray structural studies of natural RNA/DNA aptamer-thrombin complexes [74, 75] may lead to the determination of appropriate modifications that are effective for improving affinity.

4. Recent Advances

There are many successful examples of postmodifications, although, in many cases, unsatisfactory outcomes have resulted from the direct screening of modified RNA/DNA libraries. This indicates that chemical modifications should improve the binding properties if the proper functionalities are chosen. Recently, Vaught et al. demonstrated that screening using a modified DNA library containing 5-tryptaminocarbonyl-dU (TrpdU) could provide modified DNA aptamers that were specific for so-called "difficult protein targets," which were previously intractable with SELEX using a natural DNA library (Table 3) [70]. A modified DNA aptamer involving TrpdU was found to bind tightly to necrosis factor receptor superfamily member 9 (TNFRSF9), a difficult protein target, with a K_d value of approximately 5 nM, whereas no DNA aptamer was recovered that could bind to TNFRSF9 from a natural DNA library using the same selection protocol.

Subsequently, significantly larger and more systematically conceived screenings were performed for 13 difficult protein targets using three different modified DNA libraries that contained TrpdU, 5-isobutylaminocarbonyl-dU (IbdU), and 5-benzylaminocarbonyl-dU (BndU), as well as a natural DNA library [76]. All screenings for these 13 protein targets were successful when a library containing TrpdU was used, whereas five and seven screenings failed when libraries containing IbdU and BndU, respectively, were used. No enrichment was observed in any of the screenings for these targets using a natural DNA library.

The apparent K_d values of the enriched TrpdU-containing libraries were between several ten's of pM and several nM, which meant that sufficient enrichments of active species were achieved; some isolated aptamers exhibited extremely high-binding affinities at the level of several pM although the lowest K_d values differ with protein targets. These modified DNA aptamers were named Slow Off-rate Modified Aptamers (SOMAmers) because they were selected so as to have slow dissociation rates as a characteristic feature of their binding kinetics, which could improve binding specificity. SOMAmer technologies have enabled to expand over protein targets for which SELEX using natural RNA/DNA libraries did not yield high-affinity aptamers, which apparently illustrates the superiority of these chemical modifications.

During the preparation of modified DNA libraries for SOMAmer selection, *KOD XL* was used (essentially the same as *KOD Dash*). *KOD XL* comprises a mixture of approximately 2 : 98 of wild-type *KOD* DNA polymerase and *KOD(exo-)* DNA polymerase, which has no or less 3′,5′ exonuclease activity. We have focused on the high-catalytic activity and high fidelity of *KOD* DNA polymerase and its variants and first demonstrated its utility and application to PCR and SELEX involving modified nucleotides [25, 37, 56]. Recently, in our collaboration with Toyobo Co., Ltd., the catalytic properties of eight new *KOD* DNA polymerase variants in modified nucleotide polymerizations were assessed using base-/sugar-modified nucleoside triphosphates [77]. Among these *KOD* variants, one achieved efficient successive incorporation of bridged nucleotides with a 2′-$ONHCH_2$-4′ linkage; this is much more bulky but exhibits far superior biostability than the prototype BNA/LNA.

Thus, to expand the functional repertories of chemical modifications, both exploring favorable combinations of polymerases and substrate triphosphate analogs and genetically modifying polymerases have been performed. Developing highly efficient enzymatic polymerizations of artificially designed nucleic acid analogs with odd chemical structures, such as a bridged nucleic acid (BNA)

Table 2: Comparison of binding affinities in thrombin-binding natural/modified nucleic acid aptamers.

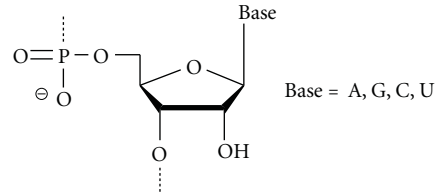
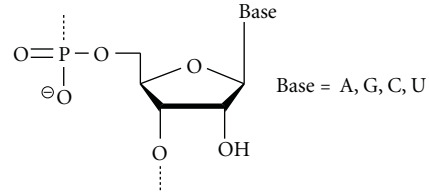
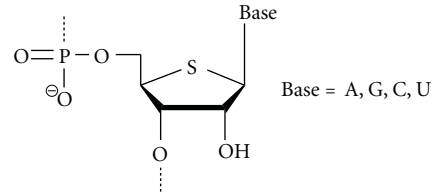
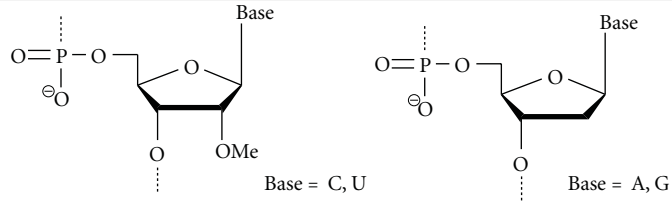
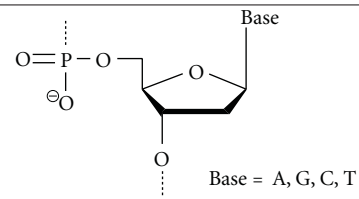
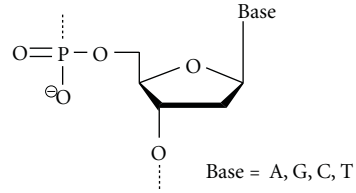
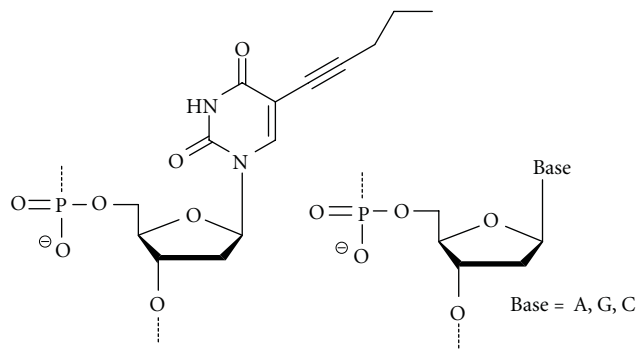
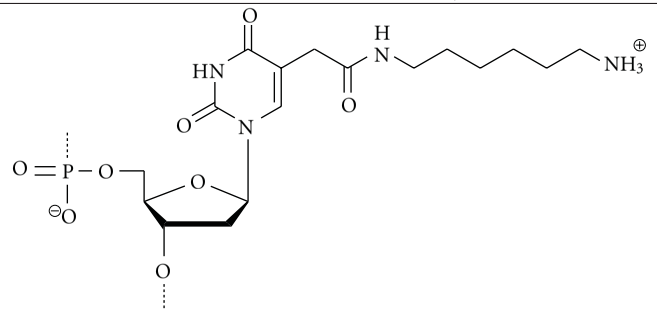
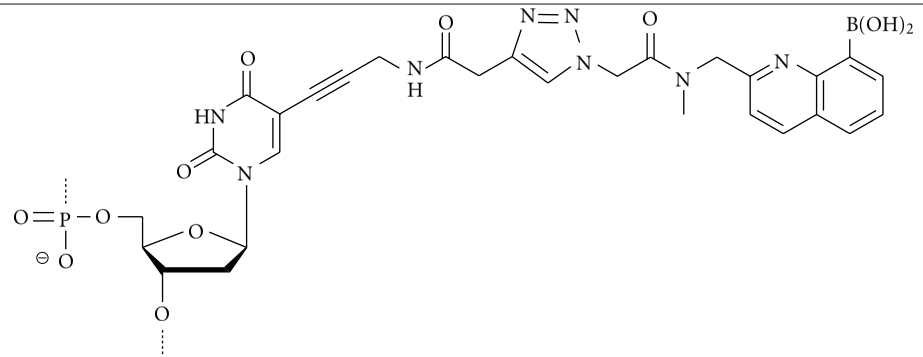
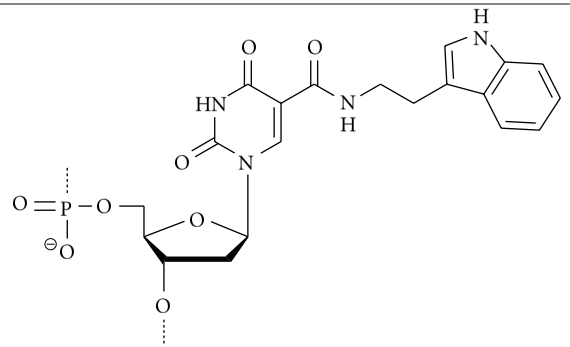
Aptamer type	Chemical structure	Binding affinity K_d	References
Natural RNA		9.3 ± 1.0 nM	[63]
Natural RNA		0.54 ± 1.0 nM	[64]
Modified RNA		7.2 nM	[65]
Modified RNA/DNA hybrid		26 nM	[66]
Natural DNA		113 nM	[67]
Natural DNA		0.5 nM	[68]
Modified DNA		400 nM	[69]

Table 3: Modified DNA aptamers containing C5-substituted thymidine having foreign functionalities, which were directly recovered from chemical libraries.

Target	Chemical structure of modified thymidine	Binding affinity K_d	Polymerase used	References
Thalidomide derivative		1.0 μM for *R*-isomer	*KOD Dash*	[55]
Fibrinogen		6.2 ± 1.4 nM	*Taq*	[57]
TNFRSF9		5 nM	*KOD XL*	[70]

[45], a glycol nucleic acid (GNA) [40], a peptide nucleic acid (PNA) [78], and a phosphorodiamidate morpholino oligomer (PMO) [79], would be expected to provide not only greatly enhanced affinities, specificities, and biostabilities but could also unexpected new functions for aptamers. Very recently, Pinherio et al. have developed evolved polymerases by a selection strategy named compartmentalized self-tagging (CST) [80], which is an improved methodology of compartmentalized self-replication (CSR) [81]. Using the evolved polymerases, six artificial biopolymers, 1,5-anhydrohexitol nucleic acid (HNA) [22], cyclohexenyl nucleic acids (CeNA) [36], arabinonucleic acid (ANA) [23], 2′-fluoroarabinonucleic acid (FANA) [24], α-L-threofuranosyl nucleic acid (TNA) [48], and BNA/LNA, were confirmed to be transcribed and reverse transcribed. Furthermore, HNA aptamers specific to HIV transactivating response (TAR) RNA and hen egg lysozyme (HEL) were successfully recovered by the traditional SELEX method using the acquired HNA polymerase although the binding affinities of the best aptamers were found to lie between the middle and high nanomolar range (K_d = 28 and 107 nM, resp.).

5. Conclusion

In over two decades of studies on nucleic acid aptamers, the effectiveness of employing chemically modified libraries for SELEX on binding properties was not evident until the development of SOMAmer, whereas that on biostability had been clearly evident in the early research era. Meanwhile, we could not conclusively prove that the chemical modifications may induce enhanced binding affinity, but we were successful in exemplifying functional expression, that is, high enantioselectivity, owing to the introduced foreign functionalities. Furthermore, first, we have focused on *KOD* DNA polymerase and its variants from various polymerases and demonstrated their feasibility for the direct screening of modified nucleic acid aptamers. Generally, the choice of polymerase has practical significance for our research purpose; not only catalytic efficiency but also the fidelity of polymerase greatly affects suitable enrichments during selection rounds because polymerase invariably causes misincorporation at a certain rate even with natural substrates. Despite the difficulties, it has been demonstrated that the successful engineering of polymerase can dramatically broaden the diversity of chemical structures for a selection library.

One of the greatest concerns in the further development of this research is to gain systematic knowledge about the inter- and intramolecular interactions involving foreign functionalities for eventually deriving guidelines for the rational design of artificial specific binders. Nowadays, methodologies for the random screening of nucleic acid aptamers have become rapid and convenient, for example, capillary electrophoresis-SELEX [82] and microfluidic SELEX [83]; random screening can be implemented in systematic selections by using a varied modified library in order to address the challenging issue.

Acknowledgments

This work was supported in part by a Grant for Industrial Technology Research from the New Energy and Industrial Technology Development Organization (NEDO) of Japan. The authors are grateful for contributions from researches of Emeritus Professor Hiroaki Sawai, Professor Hiroaki Ozaki, Dr. Astushi Shoji, and Dr. Junichi Nagashima of Gunma University and Emeritus Professor Takeshi Imanishi, Professor Satoshi Obika of Osaka University, and Dr. Akio Sugiyama of Toyobo Co., Ltd. Y. Kasahara is grateful for a Research Fellowship from the Japan Society for the Promotion of Science (JSPS) for Young Scientists.

References

[1] A. D. Ellington and J. W. Szostak, "In vitro selection of RNA molecules that bind specific ligands," *Nature*, vol. 346, no. 6287, pp. 818–822, 1990.

[2] C. Tuerk and L. Gold, "Systemic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase," *Science*, vol. 249, no. 4968, pp. 505–510, 1990.

[3] J. Ruckman, L. S. Green, J. Beeson et al., "2'-fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165): inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain," *Journal of Biological Chemistry*, vol. 273, no. 32, pp. 20556–20567, 1998.

[4] N. Janjić, "Inhibition of receptor binding by high-affinity RNA ligands to vascular endothelial growth factor," *Biochemistry*, vol. 33, no. 34, pp. 10450–10456, 1994.

[5] C. E. Tucker, L. S. Chen, M. B. Judkins, J. A. Farmer, S. C. Gill, and D. W. Drolet, "Detection and plasma pharmacokinetics of an anti-vascular endothelial growth factor oligonucleotide-aptamer (NX1838) in rhesus monkeys," *Journal of Chromatography B*, vol. 732, no. 1, pp. 203–212, 1999.

[6] P. E. Burmeister, S. D. Lewis, R. F. Silva et al., "Direct in vitro selection of a 2′-O-methyl aptamer to VEGF," *Chemistry and Biology*, vol. 12, no. 1, pp. 25–33, 2005.

[7] J. Ju, C. Ruan, C. W. Fuller, A. N. Glazer, and R. A. Mathies, "Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 10, pp. 4347–4351, 1995.

[8] B. Gharizadeh, T. Nordström, A. Ahmadian, M. Ronaghi, and P. Nyrén, "Long-read pyrosequencing using pure 2′-deoxyadenosine-5′-O′-(1-thiotriphosphate) Sp-isomer," *Analytical Biochemistry*, vol. 301, no. 1, pp. 82–90, 2002.

[9] J. Eid, A. Fehr, J. Gray et al., "Real-time DNA sequencing from single polymerase molecules," *Science*, vol. 323, no. 5910, pp. 133–138, 2009.

[10] T. Kajiyama, M. Kuwahara, M. Goto, and H. Kambara, "Optimization of pyrosequencing reads by superior successive incorporation efficiency of improved 2′-deoxyadenosine-5′-triphosphate analogs," *Analytical Biochemistry*, vol. 416, no. 1, pp. 8–17, 2011.

[11] Z. Zhu, J. Chao, H. Yu, and A. S. Waggoner, "Directly labeled DNA probes using fluorescent nucleotides with different length linkers," *Nucleic Acids Research*, vol. 22, no. 16, pp. 3418–3422, 1994.

[12] A. Anne, B. Blanc, and J. Moiroux, "Synthesis of the first ferrocene-labeled dideoxynucleotide and its use for 3′-redox end-labeling of 5′-modified single-stranded oligonucleotides," *Bioconjugate Chemistry*, vol. 12, no. 3, pp. 396–405, 2001.

[13] J. A. Piccirilli, S. E. Moroney, and S. A. Benner, "A C-nucleotide base pair: methylpseudouridine-directed incorporation of formycin triphosphate into RNA catalyzed by T7 RNA polymerase," *Biochemistry*, vol. 30, no. 42, pp. 10350–10356, 1991.

[14] S. Matsuda, A. A. Henry, and F. E. Romesberg, "Optimization of unnatural base pair packing for polymerase recognition," *Journal of the American Chemical Society*, vol. 128, no. 19, pp. 6369–6375, 2006.

[15] M. Kimoto, T. Mitsui, R. Yamashige, A. Sato, S. Yokoyama, and I. Hirao, "A new unnatural base pair system between fluorophore and quencher base analogues for nucleic acid-based imaging technology," *Journal of the American Chemical Society*, vol. 132, no. 43, pp. 15418–15426, 2010.

[16] T. Ono, M. Scalf, and L. M. Smith, "2'-Fluoro modified nucleic acids: polymerase-directed synthesis, properties and stability to analysis by matrix-assisted laser desorption/ionization mass spectrometry," *Nucleic Acids Research*, vol. 25, no. 22, pp. 4581–4588, 1997.

[17] K. Raines and P. A. Gottlieb, "Enzymatic incorporation of 2'-thio-CTP into the HDV ribozyme," *RNA*, vol. 4, no. 3, pp. 340–345, 1998.

[18] K. Sakthivel and C. F. Barbas III, "Expanding the potential of DNA for binding and catalysis: highly functionalized dUTP derivatives that are substrates for thermostable DNA polymerases," *Angewandte Chemie*, vol. 37, no. 20, pp. 2872–2875, 1998.

[19] R. Padilla and R. Sousa, "Efficient synthesis of nucleic acids heavily modified with non-canonical ribose 2'-groups using a mutant T7 RNA polymerase (RNAP)," *Nucleic Acids Research*, vol. 27, no. 6, pp. 1561–1563, 1999.

[20] N. K. Vaish, A. W. Fraley, J. W. Szostak, and L. W. McLaughlin, "Expanding the structural and functional diversity of RNA: analog uridine triphosphates as candidates for in vitro selection of nucleic acids," *Nucleic Acids Research*, vol. 28, no. 17, pp. 3316–3322, 2000.

[21] M. L. Andreola, C. Calmels, J. Michel, J. J. Toulmé, and S. Litvak, "Towards the selection of phosphorothioate aptamers: optimizing in vitro selection steps with phosphorothioate nucleotides," *European Journal of Biochemistry*, vol. 267, no. 16, pp. 5032–5040, 2000.

[22] K. Vastmans, S. Pochet, A. Peys et al., "Enzymatic incorporation in DNA of 1,5-anhydrohexitol nucleotides," *Biochemistry*, vol. 39, no. 42, pp. 12757–12765, 2000.

[23] A. M. Noronha, C. J. Wilds, C. N. Lok et al., "Synthesis and biophysical properties of arabinonucleic acids (ANA): circular dichroic spectra, melting temperatures, and ribonuclease H susceptibility of ANA·RNA hybrid duplexes," *Biochemistry*, vol. 39, no. 24, pp. 7050–7062, 2000.

[24] K. U. Schoning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, and A. Eschenmoser, "Chemical etiology of nucleic acid structure: the α-threofuranosyl-(3'→2') oligonucleotide system," *Science*, vol. 290, no. 5495, pp. 1347–1351, 2000.

[25] H. Sawai, A. N. Ozaki, F. Satoh, T. Ohbayashi, M. M. Masud, and H. Ozaki, "Expansion of structural and functional diversities of DNA using new 5-substituted deoxyuridine derivatives by PCR with superthermophilic *KOD Dash* DNA polymerase," *Chemical Communications*, no. 24, pp. 2604–2605, 2001.

[26] O. Thum, S. Jäger, and M. Famulok, "Functionalized DNA: a new replicable biopolymer," *Angewandte Chemie*, vol. 40, no. 21, pp. 3990–3993, 2001.

[27] S. E. Lee, A. Sidorov, T. Gourlain et al., "Enhancing the catalytic repertoire of nucleic acids: a systematic study of linker length and rigidity," *Nucleic Acids Research*, vol. 29, no. 7, pp. 1565–1573, 2001.

[28] H. A. Held and S. A. Benner, "Challenging artificial genetic systems: thymidine analogs with 5-position sulfur functionality," *Nucleic Acids Research*, vol. 30, no. 17, pp. 3857–3869, 2002.

[29] R. Padilla and R. Sousa, "A Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs," *Nucleic Acids Research*, vol. 30, no. 24, p. e138, 2002.

[30] S. M. Lato, N. D. S. Ozerova, K. He, Z. Sergueeva, B. R. Shaw, and D. H. Burke, "Boron-containing aptamers to ATP," *Nucleic Acids Research*, vol. 30, no. 6, pp. 1401–1407, 2002.

[31] M. Kuwahara, Y. Takahata, A. Shoji, A. N. Ozaki, H. Ozaki, and H. Sawai, "Substrate properties of C5-substituted pyrimidine 2′-deoxynucleoside 5′-triphosphates for thermostable DNA polymerases during PCR," *Bioorganic and Medicinal Chemistry Letters*, vol. 13, no. 21, pp. 3735–3738, 2003.

[32] T. Tasara, B. Angerer, M. Damond et al., "Incorporation of reporter molecule-labeled nucleotides by DNA polymerases. II. High-density labeling of natural DNA," *Nucleic Acids Research*, vol. 31, no. 10, pp. 2636–2646, 2003.

[33] J. B. J. Pavey, A. J. Lawrence, I. A. O'Neil, S. Vortler, and R. Cosstick, "Synthesis and transcription studies on 5′-triphosphates derived from 2′-C-branched-uridines: 2′-homouridine-5′-triphosphate is a substrate for T7 RNA polymerase," *Organic and Biomolecular Chemistry*, vol. 2, no. 6, pp. 869–875, 2004.

[34] T. Ohbayashi, M. Kuwahara, M. Hasegawa, T. Kasamatsu, T. Tamura, and H. Sawai, "Expansion of repertoire of modified DNAs prepared by PCR using *KOD Dash* DNA polymerase," *Organic and Biomolecular Chemistry*, vol. 3, no. 13, pp. 2463–2468, 2005.

[35] S. Jäger, G. Rasched, H. Kornreich-Leshem, M. Engeser, O. Thum, and M. Famulok, "A versatile toolbox for variable DNA functionalization at high density," *Journal of the American Chemical Society*, vol. 127, no. 43, pp. 15071–15082, 2005.

[36] V. Kempeneers, M. Renders, M. Froeyen, and P. Herdewijn, "Investigation of the DNA-dependent cyclohexenyl nucleic acid polymerization and the cyclohexenyl nucleic acid-dependent DNA polymerization," *Nucleic Acids Research*, vol. 33, no. 12, pp. 3828–3836, 2005.

[37] M. Kuwahara, J. I. Nagashima, M. Hasegawa et al., "Systematic characterization of 2′-deoxynucleoside- 5′-triphosphate analogs as substrates for DNA polymerases by polymerase chain reaction and kinetic studies on enzymatic production of modified DNA," *Nucleic Acids Research*, vol. 34, no. 19, pp. 5383–5394, 2006.

[38] M. Kuwahara, K. Hanawa, K. Ohsawa, R. Kitagata, H. Ozaki, and H. Sawai, "Direct PCR amplification of various modified DNAs having amino acids: convenient preparation of DNA libraries with high-potential activities for in vitro selection," *Bioorganic and Medicinal Chemistry*, vol. 14, no. 8, pp. 2518–2526, 2006.

[39] N. Inoue, A. Shionoya, N. Minakawa, A. Kawakami, N. Ogawa, and A. Matsuda, "Amplification of 4′-thioDNA in the presence of 4′-thio-dTTP and 4′-thio-dCTP, and 4′-thioDNA-directed transcription in vitro and in mammalian cells," *Journal of the American Chemical Society*, vol. 129, no. 50, pp. 15424–15425, 2007.

[40] C. H. Tsai, J. Chen, and J. W. Szostak, "Enzymatic synthesis of DNA on glycerol nucleic acid templates without stable duplex formation between product and template," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 37, pp. 14598–14603, 2007.

[41] H. Sawai, J. Nagashima, M. Kuwahara, R. Kitagata, T. Tamura, and I. Matsui, "Differences in substrate specificity of C(5)-substituted or C(5)-unsubstituted pyrimidine nucleotides by DNA polymerases from thermophilic bacteria, archaea, and phages," *Chemistry and Biodiversity*, vol. 4, no. 9, pp. 1979–1995, 2007.

[42] P. Čapek, H. Cahová, R. Pohl, M. Hocek, C. Gloeckner, and A. Marx, "An efficient method for the construction of functionalized dna bearing amino acid groups through cross-coupling reactions of nucleoside triphosphates followed by primer extension or PCR," *Chemistry*, vol. 13, no. 21, pp. 6196–6203, 2007.

[43] C. G. Peng and M. J. Damha, "Polymerase-directed synthesis of 2′-deoxy-2′-fluoro-β-D- arabinonucleic acids," *Journal of the American Chemical Society*, vol. 129, no. 17, pp. 5310–5311, 2007.

[44] R. N. Veedu, B. Vester, and J. Wengel, "Polymerase chain reaction and transcription using locked nucleic acid nucleotide triphosphates," *Journal of the American Chemical Society*, vol. 130, no. 26, pp. 8124–8125, 2008.

[45] M. Kuwahara, S. Obika, J. I. Nagashima et al., "Systematic analysis of enzymatic DNA polymerization using oligo-DNA templates and triphosphate analogs involving 2′,4′-bridged nucleosides," *Nucleic Acids Research*, vol. 36, no. 13, pp. 4257–4265, 2008.

[46] R. N. Veedu, B. Vester, and J. Wengel, "Efficient enzymatic synthesis of LNA-modified DNA duplexes using *KOD* DNA polymerase," *Organic and Biomolecular Chemistry*, vol. 7, no. 7, pp. 1404–1409, 2009.

[47] V. Borsenberger, M. Kukwikila, and S. Howorka, "Synthesis and enzymatic incorporation of modified deoxyuridine triphosphates," *Organic and Biomolecular Chemistry*, vol. 7, no. 18, pp. 3826–3835, 2009.

[48] H. Yu, S. Zhang, and J. C. Chaput, "Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor," *Nature Chemistry*, vol. 4, no. 3, pp. 183–187, 2012.

[49] J. F. Milligan, D. R. Groebe, G. W. Witherell, and O. C. Uhlenbeck, "Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates," *Nucleic Acids Research*, vol. 15, no. 21, pp. 8783–8798, 1987.

[50] S. O. Gudima, D. A. Kostyuk, O. I. Grishchenko, V. L. Tunitskaya, L. V. Memelova, and S. N. Kochetkov, "Synthesis of mixed ribo/deoxyribopolynucleotides by mutant T7 RNA polymerase," *FEBS Letters*, vol. 439, no. 3, pp. 302–306, 1998.

[51] M. F. Goodman, S. Keener, S. Guidotti, and E. W. Branscomb, "On the enzymatic basis for mutagenesis by manganese," *Journal of Biological Chemistry*, vol. 258, no. 6, pp. 3469–3475, 1983.

[52] R. A. Beckman, A. S. Mildvan, and L. A. Loeb, "On the fidelity of DNA replication: manganese mutagenesis in vitro," *Biochemistry*, vol. 24, no. 21, pp. 5810–5817, 1985.

[53] Y. Lin, Q. Qiu, S. C. Gill, and S. D. Jayasena, "Modified RNA sequence pools for in vitro selection," *Nucleic Acids Research*, vol. 22, no. 24, pp. 5229–5234, 1994.

[54] X. Yang, S. Fennewald, B. A. Luxon, J. Aronson, N. K. Herzog, and D. G. Gorenstein, "Aptamers containing thymidine 3'-O-phosphorodithioates: synthesis and binding to nuclear factor-κB," *Bioorganic and Medicinal Chemistry Letters*, vol. 9, no. 23, pp. 3357–3362, 1999.

[55] A. Shoji, M. Kuwahara, H. Ozaki, and H. Sawai, "Modified DNA aptamer that binds the (R)-isomer of a thalidomide derivative with high enantioselectivity," *Journal of the American Chemical Society*, vol. 129, no. 5, pp. 1456–1464, 2007.

[56] K. Ohsawa, T. Kasamatsu, J. I. Nagashima et al., "Arginine-modified DNA aptamers that show enantioselective recognition of the dicarboxylic acid moiety of glutamic acid," *Analytical Sciences*, vol. 24, no. 1, pp. 167–172, 2008.

[57] M. Li, N. Lin, Z. Huang et al., "Selecting aptamers for a glycoprotein through the incorporation of the boronic acid moiety," *Journal of the American Chemical Society*, vol. 130, no. 38, pp. 12636–12638, 2008.

[58] M. Sassanfar and J. W. Szostak, "An RNA motif that binds ATP," *Nature*, vol. 364, no. 6437, pp. 550–553, 1993.

[59] P. L. Sazani, R. Larralde, and J. W. Szostak, "A small aptamer with strong and specific recognition of the triphosphate of ATP," *Journal of the American Chemical Society*, vol. 126, no. 27, pp. 8370–8371, 2004.

[60] N. K. Vaish, R. Larralde, A. W. Fraley, J. W. Szostak, and L. W. McLaughlin, "A novel, modification-dependent ATP-binding aptamer selected from an RNA library incorporating a cationic functionality," *Biochemistry*, vol. 42, no. 29, pp. 8842–8851, 2003.

[61] D. E. Huizenga and J. W. Szostak, "A DNA aptamer that binds adenosine and ATP," *Biochemistry*, vol. 34, no. 2, pp. 656–665, 1995.

[62] T. R. Battersby, D. N. Ang, P. Burgstaller et al., "Quantitative analysis of receptors for adenosine nucleotides obtained via in vitro selection from a library incorporating a cationic nucleotide analog," *Journal of the American Chemical Society*, vol. 121, no. 42, pp. 9781–9789, 1999.

[63] M. F. Kubik, A. W. Stephens, D. Schneider, R. A. Marlar, and D. Tasset, "High-affinity RNA ligands to human α-thrombin," *Nucleic Acids Research*, vol. 22, no. 13, pp. 2619–2626, 1994.

[64] R. White, C. Rusconi, E. Scardino et al., "Generation of species cross-reactive aptamers using "toggle" SELEX," *Molecular Therapy*, vol. 4, no. 6, pp. 567–574, 2001.

[65] N. Minakawa, M. Sanji, Y. Kato, and A. Matsuda, "Investigations toward the selection of fully-modified 4′-thioRNA aptamers: optimization of in vitro transcription steps in the presence of 4′-thioNTPs," *Bioorganic and Medicinal Chemistry*, vol. 16, no. 21, pp. 9450–9456, 2008.

[66] P. E. Burmeister, C. Wang, J. R. Killough et al., "2′-deoxy purine, 2′-O-methyl pyrimidine (dRmY) aptamers as candidate therapeutics," *Oligonucleotides*, vol. 16, no. 4, pp. 337–351, 2006.

[67] L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas, and J. J. Toole, "Selection of single-stranded DNA molecules that bind and inhibit human thrombin," *Nature*, vol. 355, no. 6360, pp. 564–566, 1992.

[68] D. M. Tasset, M. F. Kubik, and W. Steiner, "Oligonucleotide inhibitors of human thrombin that bind distinct epitopes," *Journal of Molecular Biology*, vol. 272, no. 5, pp. 688–698, 1997.

[69] J. A. Latham, R. Johnson, and J. J. Toole, "The application of a modified nucleotide in aptamer selection: novel thrombin aptamers containing 5-(1-pentynyl)-2'-deoxyuridine," *Nucleic Acids Research*, vol. 22, no. 14, pp. 2817–2822, 1994.

[70] J. D. Vaught, C. Bock, J. Carter et al., "Expanding the chemistry of DNA for in vitro selection," *Journal of the American Chemical Society*, vol. 132, no. 12, pp. 4141–4151, 2010.

[71] M. T. Stubbs and W. Bode, "The clot thickens: clues provided by thrombin structure," *Trends in Biochemical Sciences*, vol. 20, no. 1, pp. 23–28, 1995.

[72] I. R. Krauss, A. Merlino, C. Giancola, A. Randazzo, L. Mazzarella, and F. Sica, "Thrombin-aptamer recognition: a revealed ambiguity," *Nucleic Acids Research*, vol. 39, no. 17, pp. 7858–7867, 2011.

[73] S. M. Nimjee, S. Oney, Z. Volovyk et al., "Synergistic effect of aptamers that inhibit exosites 1 and 2 on thrombin," *RNA*, vol. 15, no. 12, pp. 2105–2111, 2009.

[74] S. B. Long, M. B. Long, R. R. White, and B. A. Sullenger, "Crystal structure of an RNA aptamer bound to thrombin," *RNA*, vol. 14, no. 12, pp. 2504–2512, 2008.

[75] K. Padmanabhan and A. Tulinsky, "An ambiguous structure of a DNA 15-mer thrombin complex," *Acta Crystallographica Section D*, vol. 52, no. 2, pp. 272–282, 1996.

[76] L. Gold, D. Ayers, J. Bertino et al., "Aptamer-based multiplexed proteomic technology for biomarker discovery," *PLoS ONE*, vol. 5, no. 12, Article ID 15004, 2010.

[77] M. Kuwahara, Y. Takano, Y. Kasahara et al., "Study on suitability of *KOD* dna polymerase for enzymatic production of artificial nucleic acids using base/sugar modified nucleoside triphosphates," *Molecules*, vol. 15, no. 11, pp. 8229–8240, 2010.

[78] P. Wittung, P. E. Nielsen, O. Buchardt, M. Egholm, and B. Norden, "DNA-like double helix formed by peptide nucleic acid," *Nature*, vol. 368, no. 6471, pp. 561–563, 1994.

[79] J. Summerton, D. Stein, S. B. Huang, P. Matthews, D. Weller, and M. Partridge, "Morpholino and phosphorothioate antisense oligomers compared in cell-free and in-cell systems," *Antisense and Nucleic Acid Drug Development*, vol. 7, no. 2, pp. 63–70, 1997.

[80] V. B. Pinheiro, A. I. Taylor, C. Cozens et al., "Synthetic genetic polymers capable of heredity and evolution," *Science*, vol. 336, no. 6079, pp. 341–344, 2012.

[81] F. J. Ghadessy, J. L. Ong, and P. Holliger, "Directed evolution of polymerase function by compartmentalized self-replication," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 8, pp. 4552–4557, 2001.

[82] S. D. Mendonsa and M. T. Bowser, "In vitro evolution of functional DNA using capillary electrophoresis," *Journal of the American Chemical Society*, vol. 126, no. 1, pp. 20–21, 2004.

[83] X. Lou, J. Qian, Y. Xiao et al., "Micromagnetic selection of aptamers in microfluidic channels," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 9, pp. 2989–2994, 2009.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 962652, 11 pages
doi:10.1155/2012/962652

Research Article

Imaging mRNA Expression in Live Cells via PNA·DNA Strand Displacement-Activated Probes

Zhenghui Wang,[1] Ke Zhang,[1] Karen L. Wooley,[1,2] and John-Stephen Taylor[1]

[1] *Department of Chemistry, Washington University, St. Louis, MO 63130, USA*
[2] *Department of Chemistry, Texas A&M University, P.O. Box 30012, College Station, TX 77842-3012, USA*

Correspondence should be addressed to John-Stephen Taylor, taylor@wustl.edu

Received 16 June 2012; Accepted 30 July 2012

Academic Editor: Eriks Rozners

Probes for monitoring mRNA expression *in vivo* are of great interest for the study of biological and biomedical problems, but progress has been hampered by poor signal to noise and effective means for delivering the probes into live cells. Herein we report a PNA·DNA strand displacement-activated fluorescent probe that can image the expression of iNOS (inducible nitric oxide synthase) mRNA, a marker of inflammation. The probe consists of a fluorescein labeled antisense PNA annealed to a shorter DABCYL^{plus}-labeled DNA which quenches the fluorescence, but when the quencher strand is displaced by the target mRNA the fluorescence is restored. DNA was used for the quencher strand to facilitate electrostatic binding of the otherwise netural PNA strand to a cationic shell crosslinked knedel-like (cSCK) nanoparticle which can deliver the PNA·DNA duplex probe into cells with less toxicity and greater efficiency than other transfection agents. RAW 264.7 mouse macrophage cells transfected with the iNOS PNA·DNA probe via the cSCK showed a 16 to 54-fold increase in average fluorescence per cell upon iNOS stimulation. The increase was 4 to 7-fold higher than that for a non-complementary probe, thereby validating the ability of a PNA·DNA strand displacement-activated probe to image mRNA expression *in vivo*.

1. Introduction

There has been great interest in developing real-time fluorescent imaging agents for mRNA expression *in vivo* that are based on antisense oligodeoxynucleotides and analogs [1–3]. There are two main problems in getting such systems to work well. The first is to deliver the agents efficiently into the cytoplasm, and the second is to minimize background signal from unbound probe. The main problem with getting nucleic acids and analogs into the cytoplasm is that they are membrane impermeable, thereby requiring the use of a physical, chemical, or biochemical device or agent [4]. Many mRNA-imaging studies have used microinjection, electroporation, or pore forming agents such as streptolysin O (SLO), but such agents would be unsuitable for *in vivo* work. Others have made use of cell-penetrating peptides, or transfection agents, but these often result in endocytosis and trapping of the probe in endosomes which reduces the amount of probes in the cytoplasm and can lead to nonspecific background signal. To reduce the background signal from unbound probe, probes have been designed to emit fluorescence only in the presence of target mRNA by a variety of strategies. Among these are molecular beacons, binary and dual FRET probes, strand-displacement probes, quenched autoligating probes, FIT-probes, and nucleic-acid-triggered probe activation [5, 6].

One general approach to activatable probes makes use of a fluorophore-quencher pair, typified by the molecular beacon strategy [5, 7]. Molecular beacons consist of a fluorescent molecule and a quencher that are conjugated to both ends of an antisense nucleic acid sequence which may or may not have a short complementary stem. When free in solution, the fluorophore component is quenched by either FRET, in which case the energy of the excited fluorophore is transferred to a quencher by a through space mechanism [8]; or by "contact quenching," in which a fluorophore and a quencher are close enough that they can form a nonfluorescent complex [9]. Upon binding to the target RNA,

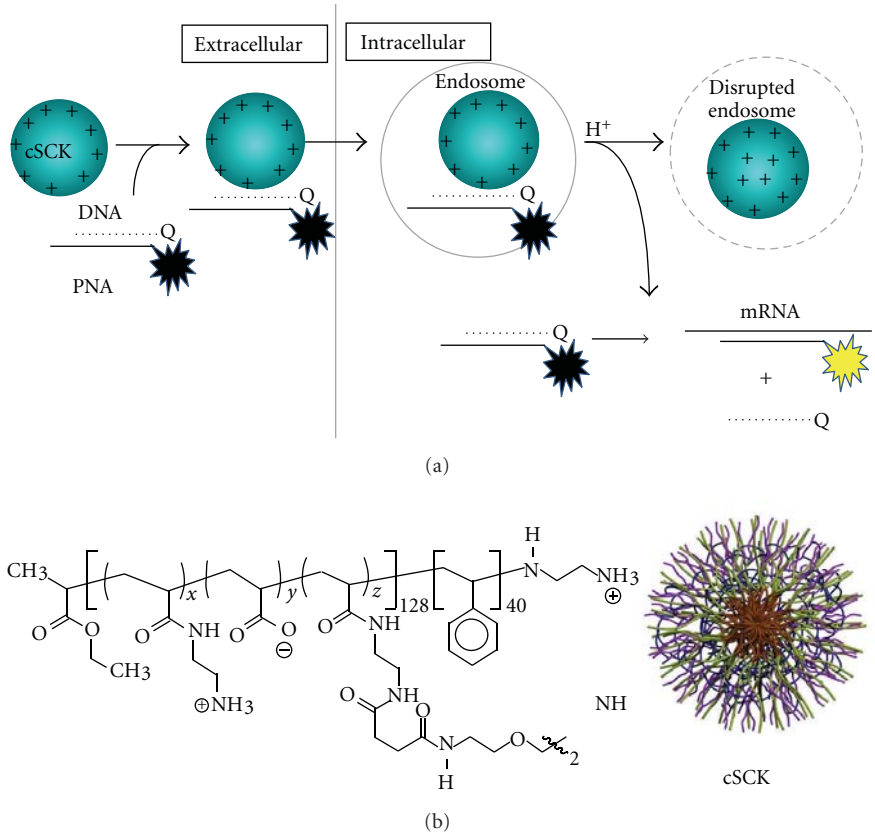


Figure 1: Schematic representation of cSCK-mediated delivery of strand-displacement-activated PNA·DNA probes for imaging mRNA in living cells. (a) The probes consist of a fluorescently labeled nondegradable antisense PNA (peptide nucleic acid) hybridized to a shorter negatively charged complementary DNA strand bearing a quencher, leaving a short single-stranded section of PNA (the toehold). The nonfluorescent PNA•DNA duplex probe is then electrostatically bound to the cationic-shell-crosslinked knedel-like nanoparticle (cSCK). The positive nature of the cSCK facilitates its endocytosis, and the presence of unprotonated amines facilitates disruption of the endosome by the proton-sponge effect which enables the strand-displacement probe to escape into the cytoplasm. Binding of the toe-hold portion of the PNA to the target mRNA sequence then facilitates strand displacement of the quenching DNA strand by branch migration and results in restoration of fluorescence to the PNA strand. (b) Structure of the cSCK formed by crosslinking the block copolymers following micellization $x \approx 122$, $y \approx 0$, $z \approx 6$.

the fluorophore and quencher are physically held apart by duplex formation, and fluorescence is restored. While this is an elegant system, it suffers from background fluorescence due to nonspecific binding events that lead to separation of the fluorophore and the quencher.

A bimolecular version of a molecular beacon, often referred to as a strand-displacement probe, makes use of an antisense oligonucleotide conjugated to a fluorescent probe that is annealed to a shorter complementary oligonucleotide conjugated to a quencher [10–12] (Figure 1(a)). In this system, the duplex region is much longer and much more stable than the generally short duplex stem used in molecular beacons. Despite the high stability, rapid strand exchange can take place because the short section of single strand on the probe strand can hybridize to the target RNA and facilitate the thermodynamically favorable displacement of the quencher strand through branch migration. The rate of strand-displacement depends on the single-strand length ("toehold") while the extent of reaction will depend on the difference in length between the fluorescent and quenching probes [13]. The larger the difference, the longer the unpaired section and the faster the rate for displacing the shorter strand by the target mRNA and the more complete the displacement.

While there are numerous studies using molecular beacons for imaging of gene expression *in vivo*, there have only been a few reports of the use of strand-displacement probes. Hnatowich and coworkers constructed a probe from a 25-mer phosphorodiamidate morpholino (MORF) oligomer conjugated to a Cy5.5 and a complementary 18-mer cDNA conjugated to a BHQ3 quencher. They showed that this probe could image a complementary biotinylated 25-mer MORF oligomer immobilized on streptavidin polystyrene microspheres that were intramuscularly implanted into a mouse [14]. The same group also utilized a probe consisting of a 25-mer phosphorothioate DNA bearing Cy5.5 and a 10-mer complementary ODN with the BHQ3 quencher to image the KB-G2 tumor in mice which overexpresses the multi-drug-resistant mdr1 mRNA [15]. In another approach, Mirkin and coworkers developed "nanoflares" in

which antisense ODNs to a target mRNA are conjugated to a gold nanoparticle and then hybridized to a shorter strand of complementary DNA bearing Cy5 which is quenched by the gold nanoparticle. When taken up by cells containing the target mRNA, the Cy5-bearing ODN becomes displaced resulting in fluorescence activation [16]. In their design, however, the fluorescent reporter becomes displaced by the mRNA making it unable to report on the location of the mRNA within the cell. DNA-based probes also suffer from premature intracellular degradation, which generates a high background signal.

All previous studies of the strand displacement-activated probes have made use of either DNA, phosphorothioate, or phosphorodiamidate morpholino, and none have made use of PNA. PNAs have a number of properties that make them ideal for strand-displacement probe technology. They are very resistant to chemical and enzymatic degradation, bind with higher affinity to RNA than DNA, and able to invade regions of RNA with secondary structure [17, 18]. They also do not activate RNAse H degradation of the target RNA and protect a complementary ODN from degradation. We have also shown that PNA·ODN duplexes can be efficiently delivered into cells by cationic-shell-crosslinked nanoparticles (cSCKs) (Figure 1(b)) through favorable electrostatic interactions, and remain highly bioactive [19, 20]. The cSCKs are also much less cytotoxic and efficient than the commonly used Lipofectamine.

To determine whether or not PNA·ODN hybrids delivered by a cSCK can be used as strand-displacement-activated fluorescent probes to monitor gene expression within living cells, we used iNOS as a model target system. iNOS is an important biomarker for inflammation and is greatly upregulated in response to environmental stimuli such as gamma interferon (γ-IFN) or lipopolysaccharide (LPS) [21, 22]. We have also previously determined a number of antisense accessible sites on iNOS mRNA that could be used as target sites by a modified reverse transcriptase random oligonucleotide library PCR method [23]. Herein we show that PNA·ODN-strand-displacement-activated fluorescence probes can be used to monitor iNOS mRNA expression in living cells by confocal microscopy following delivery by cationic shell crosslinked knedel-like nanoparticles.

2. Materials and Methods

2.1. General. The cSCK nanoparticles were prepared as previously described [24]. Anhydrous N,N-dimethylformamide (DMF), diisopropylethylamine (DIPEA), trifluoroacetic acid (TFA), meta-cresol, dichloromethane (DCM)., N-methylpyrrolidone (NMP), dimethyl sulfoxide (DMSO), and (5,6)- fluorescein-N-succinimidyl ester (FAM-NHS ester) were purchased from Sigma-Aldrich (St Louis, MO). DABCYL^{plus}-N-succinimidyl ester (DABCYL^{plus}-NHS) was purchased from Anaspec Inc (Fremont, CA). PNA monomers were purchased from PolyOrg Inc (Leominster, MA). Fmoc-protected amino acids were purchased from EMD chemicals (Gibbstown, NJ). 2-(1H-7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyl uronium hexafluorophosphate (HATU) was purchased from GenScript (Piscataway, NJ). Fmoc-PAL-PEG-PS resin for the solid-phase-PNA synthesis was purchased from Applied Biosystems (Carlsbad, CA). The PNAs were synthesized by solid-phase Fmoc chemistry on an Expedite 8909 DNA/PNA synthesizer on a 2 μmol scale. All the oligodeoxynucleotides (ODN) and amino-modified ODNs were purchased from Integrated DNA technologies (Coralville, IA). The crude FAM-PNAs and DNA-DABCYLs were purified by a reversed-phase high-performance liquid chromatography (HPLC) on a Beckman Gold System with a UV array detector and a Varian Microsorb-MV column (C-18, 5 μm, 300 Å pore size, 4.6 $\times$ 250 mm internal diameter and length). For the FAM-PNAs, a step gradient of 0–10% (2 min), 10–60% (20 min), 60–100% (20 min), and 100–0% (5 min) of solvent B (0.1% TFA in acetonitrile) in solvent A (0.1% TFA in water) was used. For the DNA-DABCYLs, a step gradient of 0% (1 min), 0–40% (5 min), 40–80% (24 min), 80–100% (3 min), and 100–0% (3 min), of solvent B (50 mM triethylammonium acetate (TEAA) in 1 : 1 water : acetonitrile) in solvent A (50 mM TEAA in water) was used. The purified PNAs and DNAs were verified by MALDI-TOF on an AppliedBiosystems 4700 mass spectrometer. The concentration of the DNAs was determined from the absorbance at 260 nm taken on a Bausch and Lomb Spectronic 1001 spectrophotometer. The concentration of the PNAs was determined from the absorbance at 260 nm at 70°C to eliminate hypochromicity due to secondary structure. For the DNAs, the molar extinction coefficient provided by the manufacturer was used. For the PNAs, the molar extinction coefficient was estimated using 13.7, 11.7, 6.6, and 8.6 mL/μmol·cm for A, G, C, and T, respectively.

2.2. PNA-Fluorescein Synthesis and Purification. A 23-mer PNA probe antisense to the bases starting at position 480 of iNOS mRNA (FAM-iNOS-PNA) and a control probe with the same length but targeting HeLa pLuc 705 splice correction site (FAM-pLuc-PNA) were synthesized on an Expedite 8909 DNA/PNA synthesizer. After removal of the Fmoc-protecting group at the amino end of the PNA, the resin was dried with nitrogen gas and was shaken overnight with 200 μL of 0.02 M FAM-NHS ester (2 eq) in DMSO, together with 2 eq DIPEA at room temperature. The resin was then washed sequentially with DMF and DCM and dried under nitrogen. The PNA was then cleaved from the support with 250 μL TFA/m-cresol (4 : 1) mixture for 2–4 h. The cleavage mixture was separated from the support and the PNA precipitated by adding 1 mL cold diethyl ether and centrifuging for 10 min. The product was dried on a hot block at 55°C and dissolved in water containing 0.1% TFA. The FAM-PNAs were purified by HPLC and characterized by MALDI mass spectrometry (See Supplementary Material available online at doi:10.1155/2012/962652), UV and fluorescence spectroscopy. The overall yield for FAM-PNAs was about 5%.

2.3. DNA-DABCYL Synthesis and Purification. Regular and 3′-end-modified ODNs were purchased from IDT Inc. and purified by HPLC. The 17-mer DNAs modified

with an amino linker at the 3′-end (50 nmol) were shaken overnight with 10 eq of DABCYL^{plus}-NHS ester in 10 mM Na_2CO_3/$NaHCO_3$ buffer (adjusted to pH 8.5 with hydrochloride acid). The products were purified by gel electrophoresis on a 20% polyacrylamide gel. Bands containing the desired product were eluted with 0.5 M ammonium acetate, 10 mM $MgCl_2$, 1 mM EDTA, and 0.1% SDS, precipitated with 3 volumes of ethanol, cooled to −20°C for 30 min, and collected after centrifugation for 30 min. The DNA-DABCYLs were characterized by MALDI mass spectrometry and UV spectroscopy.

2.4. In Vitro mRNA Transcription. The PCMV-SPORT6 vector containing the iNOS mRNA gene was purchased from American Type Culture Collection (ATCC, Manassas, VA). LB media was inoculated with *E. coli* containing the vector at 37°C for 18 h after which the plasmid was isolated from the *E. coli* by using HiPure Plasmid Maxprep kit (Invitrogen). The plasmid was then digested by Xhol (Promega) to form linear DNA, which was purified by phenol extraction, ethanol precipitation and was characterized by electrophoresis on a 1% agarose gel stained with ethidium bromide. The linear DNA was then transcribed into iNOS mRNA using the RiboMAX SP6 large scale RNA transcription kit (Promega) following the manufacturer's protocol. The integrity of iNOS mRNA was verified on the 1% w/v agarose gel. All aqueous solutions used in this process were prepared with diethylpyrocarbonate- (DEPC-) treated water and the mRNA was stored at −80°C in water with 2 μL (80 U) RNaseOUT recombinant RNase inhibitor (Invitrogen).

2.5. Displacement by Complementary DNA in Solution. FAM-iNOS-PNA (0.2 μM) and iNOS-DNA-DABCYL (0.4 μM) were heated at 95°C for 3 min in a buffer containing 100 mM Tris, 5 mM $MgCl_2$, and annealed at room temperature. The iNOS-DNA was then added to final concentrations of 0.1, 0.2, 0.4, and 2 μM. After the fluorescence intensity reached its maximum value, each sample was then incubated for another 15 min and the fluorescent emission spectrum was collected with excitation at 488 nm. A similar procedure was followed for the FAM-pLuc-PNA·pLuc-DNA-DABCYL probe. The strand-displacement rate at 37°C was monitored by the increase in fluorescence at 525 nm as a function of time with the excitation at 488 nm.

2.6. Displacement by In Vitro Transcribed mRNA. FAM-iNOS-PNA (0.2 μM) and complementary iNOS-DNA-DABCYL (0.4 μM) were first annealed in 0.1 M KCl, 5 mM $MgCl_2$, 10 mM Na-Hepes buffer (pH 7.11) and then *in vitro* transcribed iNOS mRNA was added to give final concentrations of 0.1, 0.2, and 0.4 μM, respectively. The mixtures were heated at 65°C for 1 min and incubated at 37°C for 15 min. After the fluorescence intensity reached its maximum value, each sample was then incubated for another 15 min, and the fluorescent emission spectrum was collected with-excitation at 488 nm. All solutions were prepared with DEPC-treated dd water. A similar procedure was followed for the FAM-pLuc-PNA·pLuc-DNA-DABCYL probe. The strand-displacement kinetics were carried out in 10 mM Tris-HCl buffer (pH 7.15) by first heating the iNOS mRNA at 65°C for 1.5 min and letting cool to 37°C, after which 1 μL of RNaseOUT (40 U, Invitrogen) was added. The prehybridized 1 : 2 FAM-iNOS-PNA·iNOS-DNA-DABCYL probe (1 μM in PNA) was added to the mRNA solution to a final concentration of 0.05 μM in the PNA and 0.025, 0.05, 0.1, and 0.25 μM in iNOS mRNA. The fluorescence of the samples was monitored at 525 nm by Varian Eclipse Fluorimeter at 37°C with excitation at 488 nm as a function of time. All aqueous solutions were prepared using DEPC-treated dd water.

2.7. Quantitative RT PCR to Quantify iNOS mRNA Copy Numbers in RAW 264.7 Cells. RAW 264.7 cells were seeded on 10 mm Petri dish plates (Corning Inc, Lowell, MA) and grown until 70% confluence. Selected plates were then treated with 1 μg/mL LPS and 300 ng/mL γ-IFN for 18, 6, and 0 h (without LPS and γ-IFN), respectively. Cells in each plate were counted with a hemocytometer and spun down in a centrifuge. Total RNA from each sample was extracted with the TRizol reagent (Invitrogen, CA) following the manufacturer's protocol and quantified by measuring UV absorbance at 260 nm. After treatment with Turbo DNase (RNase free), 0.5 μg of each total RNA sample was reverse-transcribed into cDNA using SuperScript II reverse transcriptase (Invitrogen), following the manufacture's procedure. Briefly, 0.5 μg of each total RNA sample was mixed with 300 ng random primers and 1 μL dNTP (10 mM each) to make a solution of 12 μL. The mixture was incubated at 65°C for 5 min and quickly chilled on ice. Then 4 μL 5 × first-strand buffer, 2 μL 0.1 M DTT, and 1 μL RNaseOUT were added, and the mixture was incubated at 25°C for 2 min. Then 1 μL of the SuperScript II RT was added to the mixture, incubated at 25°C for 10 min and then at 42°C for another 50 min. The reaction was inactivated at 70°C for 15 min, and the cDNA product was diluted 2500-fold for RT-PCR reaction. To generate the cDNA standard, 0.5 μg mRNA prepared previously was reverse transcribed into cDNA using the same kit with exactly the same procedure. The resulting cDNA product was serially diluted by a factor of ten. The cDNAs and standards were then mixed with Power SYBR Green RT-PCR master mix (Invitrogen) and the RT-PCR was performed on a Steponeplus real-time PCR system with the following profile: 1 cycle of 50°C for 2 min, 95°C for 15 min, then 40 cycles of 95°C for 15 s, 60°C for 30 s, and 72°C for 45 s. The primers used to amplify iNOS cDNA were d(TGGTGGTGACAAGCACATTT) and d(AAGGCCAAACACAGCATACC), and for the GAPDH cDNA, the primers were d(TGGAGAAACCTGCCAAG-TATG) and d(GTTGAAGTCGCAGGAGACAAC). Each well contained 25 μL of reaction mixture including 2.5 μL forward primer, 2.5 μL reverse primer, 2.5 μL double distilled water, 5 μL cDNA template and 12.5 μL Power SYBR Green RT-PCR master mix. The threshold cycle C_T was automatically set by the machine. The standard-curve method was used to determine the absolute copy number of the iNOS mRNA in cells. The comparative C_T ($\triangle\triangle C_T$) method was used to calculate the relative

Strand-displacement probes

FAM-iNOS-PNA	FAM-CCAAGTGAAATCCGATGTGGCCT
iNOS-DNA-DABCYL	3 -DABCYL-GGTTCACTTTAGGCTAC-5
iNOS-DNA	3 -TTCACTTTAGGCTACACCGGA-5
FAM-pLuc-PNA	FAM-CCACCTCTTACCTCAGTTACAAT
pLuc-DNA-DABCYL	3 -DABCYL-GGTGGAGAATGGAGTCA-5
pLuc-DNA	3 -GGTGGAGAATGGAGTCA-5

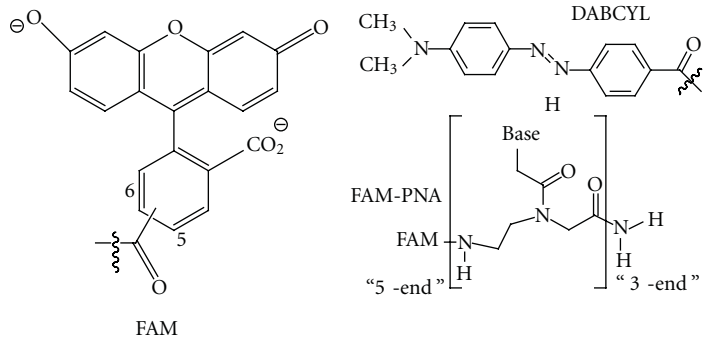


Figure 2: Sequences used in this study. FAM is directly linked to the amino-terminus of the PNA, while a derivative of DABCYL is linked to the 3′-terminus of the DNA through an amino linker.

increase of the iNOS mRNA level compared to the GAPDH mRNA.

2.8. Imaging iNOS mRNA Expression in Living Cells. RAW 264.7 cells were seeded on 10 mm glass-bottom dish (MatTek) at 5×10^4 per well and incubated overnight until they reached 70% confluence. The cells were then washed with PBS and incubated in 1 mL media containing 1 μg/mL LPS, 0.3 μg/mL γ-IFN for 18 h at 37°C in a humidified atmosphere with 5% CO_2. As a control, cells were incubated under the same conditions without LPS and γ-IFN. FAM-PNA·DNA-DABCYL (1 : 1.25) probes were annealed in 25 μL OPTI-MEM for each sample and mixed with cSCK nanoparticles. The mixtures were incubated at room temperature for 20 min to let the cSCK associate with the probes and then were added to 75 μL DMEM medium containing 10% FBS and without antibiotics. Cells were then washed with PBS and incubated with the 100 μL medium containing the cSCK complexes. To maintain iNOS mRNA induction, LPS and γ-IFN were added again at 1 μg/mL and 0.3 μg/mL concentration, respectively. The final concentration of the FAM-PNA·DNA-DABCYL probes was 0.4 μM in PNA, and the cSCK was 9.7 μg/mL for an N/P ratio 8 : 1. After 24 h of incubation, fluorescent images of the cells were collected on a Nikon A-1 confocal microscope. The fluorescent images were processed by image J software. The mean fluorescence per cell was calculated by integrating the signal intensity of the regions of interest, then dividing by the number of cells.

3. Results and Discussion

3.1. Design and Synthesis of the Strand-Displacement Probes. The strand-displacement probes were designed to have a longer antisense PNA conjugated to the fluorophore and a shorter sense DNA conjugated to the quencher to insure that the fluorophore-bearing PNA would both kinetically and thermodynamically favor hybridization to the target mRNA (Figure 2). We chose to image iNOS mRNA because it is a biomarker for inflammation that is dramatically elevated upon treatment of cells or tissue with γ-interferon and LPS (lipopolysaccharide). The PNA sequence used for the construction of the fluorescent probe was selected from a number of PNAs that we had previously demonstrated to bind to *in vitro* transcribed and endogenous iNOS mRNA, and to suppress iNOS expression *in vivo* [23]. The antisense accessible sites on the iNOS mRNA were identified by an RT-ROL (reverse transcriptase-random oligonucleotide library) method that we had improved upon [25]. Transfection of selected PNA·ODN duplexes with Lipofectamine confirmed the ability of these PNAs to inhibit gene expression. *In vitro* binding assays with in vitro transcribed mRNA confirmed that a number of these sites bound both antisense ODNs and PNAs with high affinity [26]. From these, we chose the 23-mer PNA480 sequence that targets nucleotides 473–494 on iNOS mRNA. The specificity of the antisense iNOS 23-mer sequence was assessed by BLAST (basic local alignment search tool) which revealed that the next best mRNA targets were complementary to only 14 bases of the 23-mer nucleoredoxin-like, protein 1-like, and myosin VA (Myo5a) mRNAs (See Supplementary Material). The length of the quenching strand was therefore chosen to be 17 nucleotides so that the PNA·ODN duplex would be less stable than the targeted PNA·mRNA duplex, and more stable than the non-target PNA-RNA duplexes. This length would also leave a 6-nucleotide toehold for binding to the mRNA target and initiating strand displacement by branch migration.

We chose fluorescein as the fluorophore and DABCYL^{plus} as the quencher on the complementary strand as this is

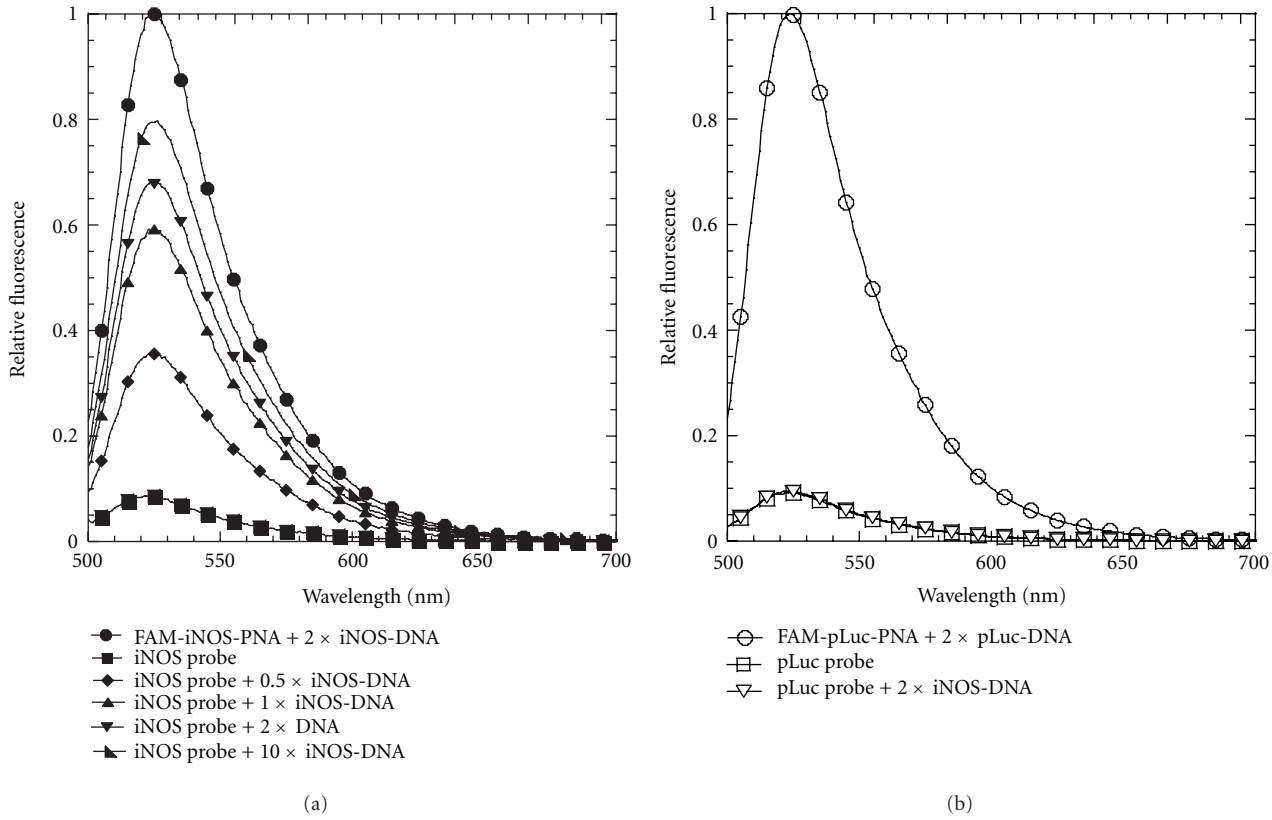


Figure 3: Fluorescence activation of the strand-displacement probes by DNA. (a) Fluorescence spectra of the iNOS probe (0.2 μM FAM-iNOS-PNA annealed to 0.4 μM iNOS-DNA-DABCYL) following addition of 0.1, 0.2, 0.4, and 2 μM iNOS DNA. Positive control: 0.2 μM FAM-iNOS-PNA annealed to 0.4 μM iNOS-DNA. (b) Fluorescence spectrum of the pLuc probe (0.2 μM FAM-pLuc-PNA annealed to 0.4 μM pLuc-DNA-DABCYL) following addition of 0.4 μM iNOS-DNA. Positive control: 0.2 μM pLuc-FAM-PNA annealed to 0.4 μM pLuc-DNA. Excitation wavelength: 488 nm, temperature: 37°C.

a common fluorophore/quencher combination [27, 28]. DABCYL^{plus} is a more soluble version of DABCYL and though its structure is proprietary, appears to involve the addition of an ethylene sulfonate chain as deduced from its molecular weight. Since it is known that a G opposite to fluorescein can also quench up to 90% of its fluorescence [29], we designed the PNA probe to have a C at the amino end (equivalent to the 5′ end of DNA), to be complementary to a G at the 3′-end of the quencher DNA strand to enhance the quenching efficiency. Because there is an A in the target iNOS mRNA at this position, we did not expect any quenching from the target mRNA. As a control, we synthesized a 23-mer PNA that is antisense to an mRNA splice correcting site in a pLuc 705 HeLa cell line which we have previously used to demonstrate the ability of cSCKs to deliver PNA•DNA hybrids into this cell line. BLAST analysis indicated that there are no mRNAs sequences greater than 13 nt in mice that could activate this probe. The probes were prepared by automated solid phase Fmoc synthesis, purified by HPLC, and characterized by MALDI (See Supplementary Material).The T_m of the antisense and mismatched FAM-PNA·DNA-DABCYL duplexes was determined by temperature dependent fluorescence measurements to be about 68°C under physiological conditions and almost completely duplex at 37°C (See Supplementary Material).

3.2. Fluorescence Activation by Complementary DNA. The PNA•DNA strand-displacement probes were first tested with a 21-mer ODN identical to the mRNA target sequence (iNOS-DNA) (Figure 3(a)). This sequence was truncated at the 3′-end to avoid introducing complementary Gs that might have quenched some of the fluorescence emission. As a positive control for the maximal amount of fluorescence achievable, FAM-iNOS-PNA was hybridized with iNOS-DNA in the absence of the iNOS-DNA-DABCYL strand. When FAM-iNOS-PNA was hybridized with a 2-fold amount of the iNOS-DNA-DABCYL in the absence of iNOS-DNA, 90% of the maximal fluorescence was quenched. Upon adding 1 equivalent of iNOS-DNA to this FAM-iNOS-PNA·iNOS-DNA-DABCYL probe (iNOS probe), almost 60% of the maximal fluorescence could be recovered. Increasing the amount of iNOS-DNA 10-fold increased the fluorescent recovery to about 80%. On the other hand, adding 2 equivalents of the iNOS-DNA to the noncomplementary FAM-pLuc-PNA·pLuc-DNA-DABCYL probe (pLuc probe) did not lead to any recovery of fluorescence

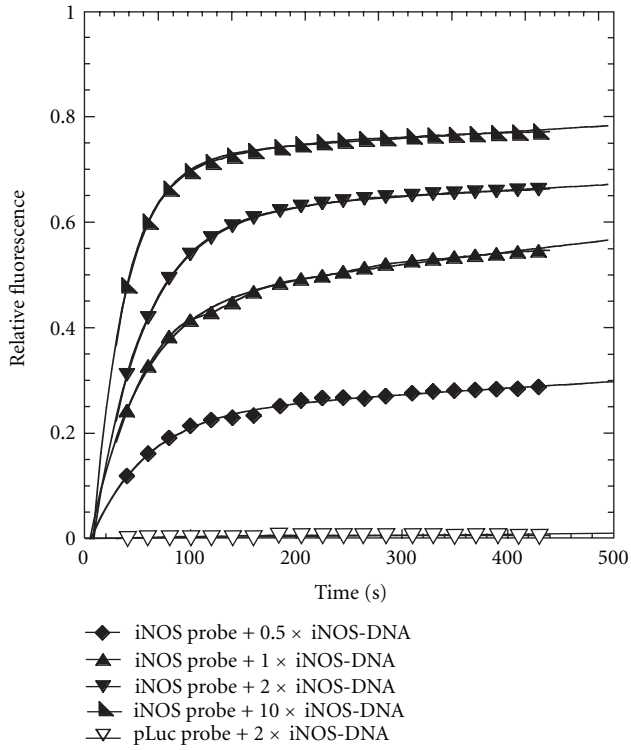


Figure 4: Kinetics of fluorescence activation of the strand-displacement probes by iNOS-DNA. Fluorescence emission at 525 nm of the iNOS probe (0.2 μM FAM-iNOS-PNA annealed to 0.4 μM iNOS-DNA-DABCYL) or the pLuc probe (0.2 μM FAM-pLuc-PNA annealed to 0.4 μM pLuc-DNA-DABCYL) in the presence of the indicated amount of iNOS-DNA. Excitation wavelength 488 nm. Temperature: 37°C.

(Figure 3(b)). When the strand-displacement reaction with the iNOS probe and iNOS-DNA was followed as a function of time about 80% of the maximal fluorescence was achieved in less than 10 min (Figure 4). The fluorescence recovery could be best fit to a biexponential where the major component (about 75%) occurred with a rate constant of about 0.02 s^{-1} while the slower component had a rate constant of about about 0.001 s^{-1}. The origin of the slower phase is not understood at the moment. The results clearly show that the strand-displacement probe is able to effectively detect a complementary nucleic acid target in solution.

3.3. Fluorescence Activation by In Vitro Transcribed mRNA. Unlike the 21-mer iNOS-DNA target, *in vitro* transcribed iNOS mRNA is about 4000 nucleotides, and adopts a complicated folded structure. Studies in our lab have previously shown that the iNOS mRNA is accessible to the iNOS-PNA used for the iNOS probe [23], and that the 18-mer carboxy terminal 18-mer section, TGAAATCCGATGTGGCCT, has a high binding affinity (86 ± 26 pM) for annealed *in vitro* transcribed iNOS mRNA [26]. Also, siRNA knockdown and PNA antisense inhibition of iNOS expression suggested that the 480 site was also accessible *in vivo* [23]. The mRNA was transcribed from a cDNA clone *in vitro* and characterized by agarose gel electrophoresis (See Supplementary Material). To demonstrate that the *in vitro* transcribed iNOS mRNA has the correct sequence and could displace the quencher strand without interference from its folded structure, the iNOS probe was heated together with varying concentrations of the mRNA to 65°C for 1 min to unfold the mRNA and then cooled to 37°C for 15 min. With 0.5 to 1 equivalents of iNOS mRNA, there was about 50% recovery of fluorescence, and at 2 equivalents, about 70% demonstrating that the target mRNA sequence was indeed present and accessible after heating (Figure 5(a)). When the same procedure was carried out with the pLuc strand displacement probe no increase in fluorescence was observed, again showing the specificity of the strand displacement reaction (Figure 5(b)).

We then investigated the ability of the probe to be activated by the full length iNOS mRNA transcript at 37°C. Initial studies with directly transcribed mRNA at 37°C were not very reproducible, so the samples were annealed first to insure that the results would be reproducible and could be correlated with independent PNA-binding measurements that were also carried out on annealed mRNA. Thus, the mRNA was first heated to 65°C for 1.5 min and then annealed at 37°C for 15 min in 10 mM Tris buffer. The iNOS probe was similarly annealed at a high concentration (1 μM) and then 20-fold diluted into the mRNA solution. The fluorescence of the mixtures was monitored as a function of time and iNOS mRNA concentration at 37°C (Figure 6). When the concentration of iNOS mRNA increased from 25 nM to 250 nM, corresponding to 0.5 to 5 times the concentration of the probe, an unexpected rapid jump in fluorescence was observed, followed by an increase in the fluorescence intensity of the mixture. The pLuc probe with two equivalents of mRNA, also showed a rapid jump in fluorescence, but there was no further increase in fluorescence with time suggesting that the jump in fluorescence was due to some experimental artifact. We have not been able to establish the origin of the initial jump in fluorescence with the addition of the mRNA and it was not observed in the DNA experiment. The portion of the curve following the initial rapid rise could be fit to the same type of biexponential curve as with the DNA experiment with two approximately equal phases with rate constants of about 0.006 s^{-1} and 0.0005 s^{-1}. The maximum increase in fluorescence following the rapid jump with 10-fold excess iNOS mRNA was only about 33% of that observed for a sample in which the strand displacement probe was heated and cooled with the mRNA. The lower amount of fluorescence may be due to the tertiary structure of the mRNA at 37°C which could reduce the binding affinity, and/or to the presence of multiple folded mRNAs, some of which are more kinetically accessible than others. Such folded structures, as well as protein binding, could affect the accessibility of an antisense probe *in vivo*.

3.4. Copy Number of iNOS mRNA in Cells. mRNAs are usually expressed at very low levels inside cells, ranging from tens to thousands of copies per cell [30]. The low copy

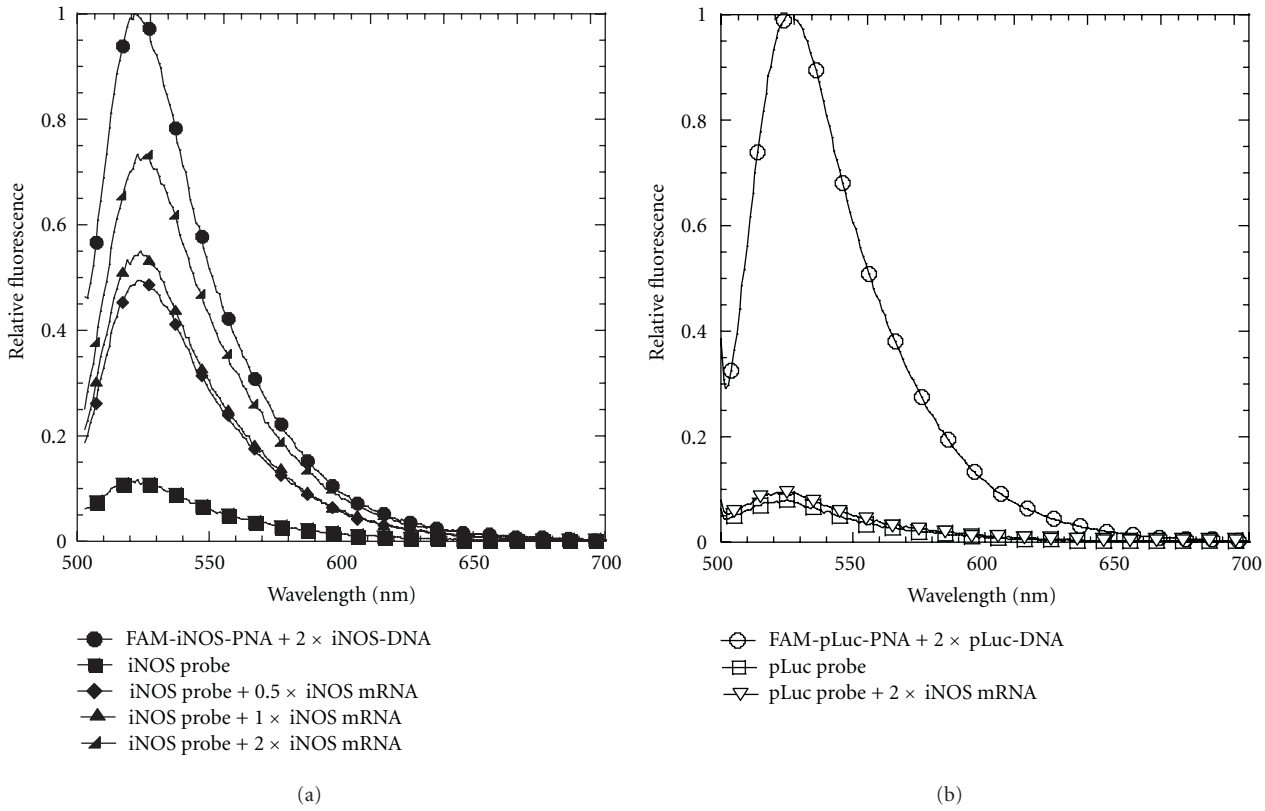


Figure 5: Fluorescence activation of the strand displacement probes by iNOS mRNA.(a) Fluorescence spectra of the iNOS probe (0.2 μM iNOS-FAM-PNA annealed to 0.4 μM iNOS-DNA-DABCYL) in the presence of 0.1, 0.2, and 0.4 μM iNOS mRNA. Positive control: 0.2 μM iNOS-FAM-PNA annealed to 0.4 μM iNOS-DNA. (b) Fluorescence spectra of the pLuc probe (0.2 μM FAM-pLuc-PNA annealed to 0.4 μM pLuc-DNA-DABCYL) in the presence of 0.4 μM iNOS mRNA. Excitation wavelength 488 nm. Temperature: 37°C.

number of mRNAs can be a problem for *in vivo* mRNA imaging because the signal generated will be very low and hard to be distinguished from background noise. So far, antisense imaging by fluorescently labeled probes are still limited to relatively abundant transcripts [2]. Normally, the expression level of iNOS is very low, but becomes greatly stimulated by LPS and γ-IFN, making it a good system for testing and validating antisense imaging probes. To our best knowledge, the actual copy number of iNOS mRNA inside cells before or after stimulation has not been reported. To determine the copy numbers for iNOS mRNA, we performed quantitative RT-PCR on nonstimulated RAW 264.7 cells and cells stimulated with LPS/γ-IFN for 6 and 18 h. We chose RAW264.7 cells for these studies because this is a mouse macrophage cell line which is well known to elevate iNOS expression in response to LPS/γ-IFN [31]. Furthermore, the cells primarily responsible for iNOS induction in acute lung injury (ALI) are alveolar macrophages, and we plan to ultimately extend our studies to mouse models of ALI [32]. The *in vitro* transcribed iNOS mRNA was used to generate a standard curve, and the housekeeping gene glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as an internal control to determine the relative increase of iNOS mRNA (See Supplementary Material). Using the standard curve, the copy number for unstimulated cells was estimated to be 760 per cell, but rose 70-fold to about 53,000 after 6 h of stimulation, and 100-fold to 76,000 after eighteen hours. The $\triangle\triangle C_T$ method using GAPDH as an internal reference also showed a 96-fold increase for the iNOS mRNA after 18 h of stimulation, confirming the results obtained from the standard-curve method. The large change in copy number, and high mRNA level after stimulation makes iNOS mRNA an ideal target for development and validation of antisense imaging agents.

3.5. Imaging of iNOS mRNA Expression in Living Cells. Intracellular delivery of nucleic acids has always been a major obstacle for *in vivo* antisense imaging due to their membrane impermeability. We have found that PNAs can be efficiently delivered into cells by hybridizing the PNA with negatively charged DNA and then forming an electrostatic complex with cSCK (cationic-shell-crosslinked knedel-like nanoparticle) [19, 20, 24]. In addition to being able to form the electrostatic complex with the PNA•DNA duplex, the positively charged shell of the cSCK nanoparticle also facilitates entry into cells via endocytosis, and escape of the PNA•DNA duplex from the endosome by the proton sponge effect. Figure 7 shows the results of confocal imaging of live RAW cells following with optimized concentrations of both the probes and cSCK nanoparticles. For cells treated

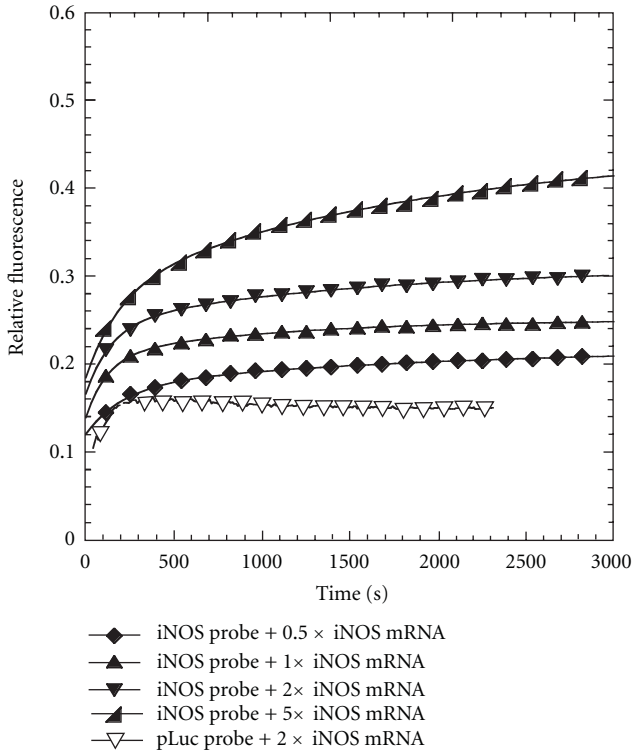


Figure 6: Kinetics of fluorescence activation by strand-displacement probes by iNOS mRNA. Fluorescence emission at 525 nm of the iNOS probe (0.05 μM FAM-iNOS-PNA annealed to 0.1 μM iNOS-DNA-DABCYL) or pLuc probe (0.05 μM FAM-pLuc-PNA annealed to 0.1 μM pLuc-DNA-DABCYL) with the indicated amount of iNOS mRNA. Excitation wavelength: 488 nm, temperature: 37°C.

with LPS/γ-IFN, and the iNOS probe, there was bright fluorescence inside the cytoplasm, indicating hybridization of the probes to the mRNA. For cells not treated with LPS/γ-IFN and cells treated with LPS/γ-IFN but with pLuc probe, there was much less observable fluorescence. Quantification of the fluorescence shows that there was a 16.6 ± 7.9-fold increase in the average fluorescence of the iNOS probes per cell that were stimulated with LPS/γ-IFN relative to the cells that were not stimulated, which is consistent with the expected difference in iNOS mRNA expression level.

The average fluorescence/cell for the stimulated cells treated with pLuc probe, however, showed a 4.1 ± 2.3-fold increase in fluorescence compared to that for the iNOS probe in unstimulated cells. One possible explanation is that LPS/γ-IFN treatment might have caused an increased internalization of the probes which would lead to an increase in background fluorescence compared to unstimulated cells. Figure 7 shows that stimulated cells are about two times larger in diameter than unstimulated cells which could explain the increase in background signal. LPS/γ-IFN stimulation may also lead to an increase in degradation rate of the probes within the cells that could increase the background signal. The same experiment was repeated one month later with similar, if not better results (See Supplementary Material). In the second experiment, a 56 ± 24-fold increase in average fluorescence per cell was observed for the iNOS probe upon stimulation, while an 8 ± 4.2-fold increase was observed for the pLuc probe. The difference in the fluorescence per cell between the iNOS and pLuc probes in the stimulated cells in the second experiment (7-fold) was also greater than that observed in the first experiment (4-fold). This second set of results, together with results from an initial experiment preceding the first experiment indicate that the results are reproducible but that there may be experiment-to-experiment variability.

There are many other factors that could contribute to the lower-than-expected difference in fluorescence from the probes between the stimulated and unstimulated cells, such as a difference in accessibility to the targeted mRNA in stimulated and nonstimulated cells due to different protein interactions and ribosomal activity. There is also a possibility that the change in expression level of iNOS mRNA determined by RT-PCR does not properly reflect the change in expression level in the presence of nanoparticle in the cytoplasm, where the probes appear to be. We saw no fluorescence in the nucleus, either suggesting that the probes are not entering the nucleus or the mRNA is inaccessible in the nucleus. The former explanation is more likely, as unpublished experiments carried out with similar but unquenched probes do not appear to enter the nucleus. Since it has been recently reported that there can be differences in the level of a particular gene transcript in the cytoplasm and the nucleus [33], it is possible that the increase in cytoplasmic iNOS expression measured by the displacement probes is less than what is being measured by RT-PCR for the whole cell.

4. Conclusion

We have showed that the strand-displacement-activated PNA probes function *in vitro* and can be efficiently delivered by cSCK nanoparticles to image iNOS mRNA in living cells. The iNOS probes showed a 17-to-56-fold increase in average fluorescent signal per cell upon stimulation of cells, but the signal was only 4-to-7-fold greater than the signal seen for the noncomplementary pLuc probe. The observed increase in iNOS probe fluorescence intensity compared to unstimulated cells is much less than the expected value of about 100 determined by RT-PCR, which may be due to off target activation of the nontargeted probe, and/or activation of the nontargeted probe resulting from degradation of the quencher strand. The difference could also be due to differences in mRNA expression detected by the strand displacement probes in the cytoplasm, compared to that detected by RT-PCR in the whole cell. Nonetheless, this class of PNA-based strand-displacement probes combined with cSCK nanoparticle delivery looks promising for live-cell mRNA imaging, and merits further study and optimization. In the future, the quencher strand could be made more stable through the use of nondegradable nucleic acid analogs, and the probes shifted farther to the red for *in vivo* studies.

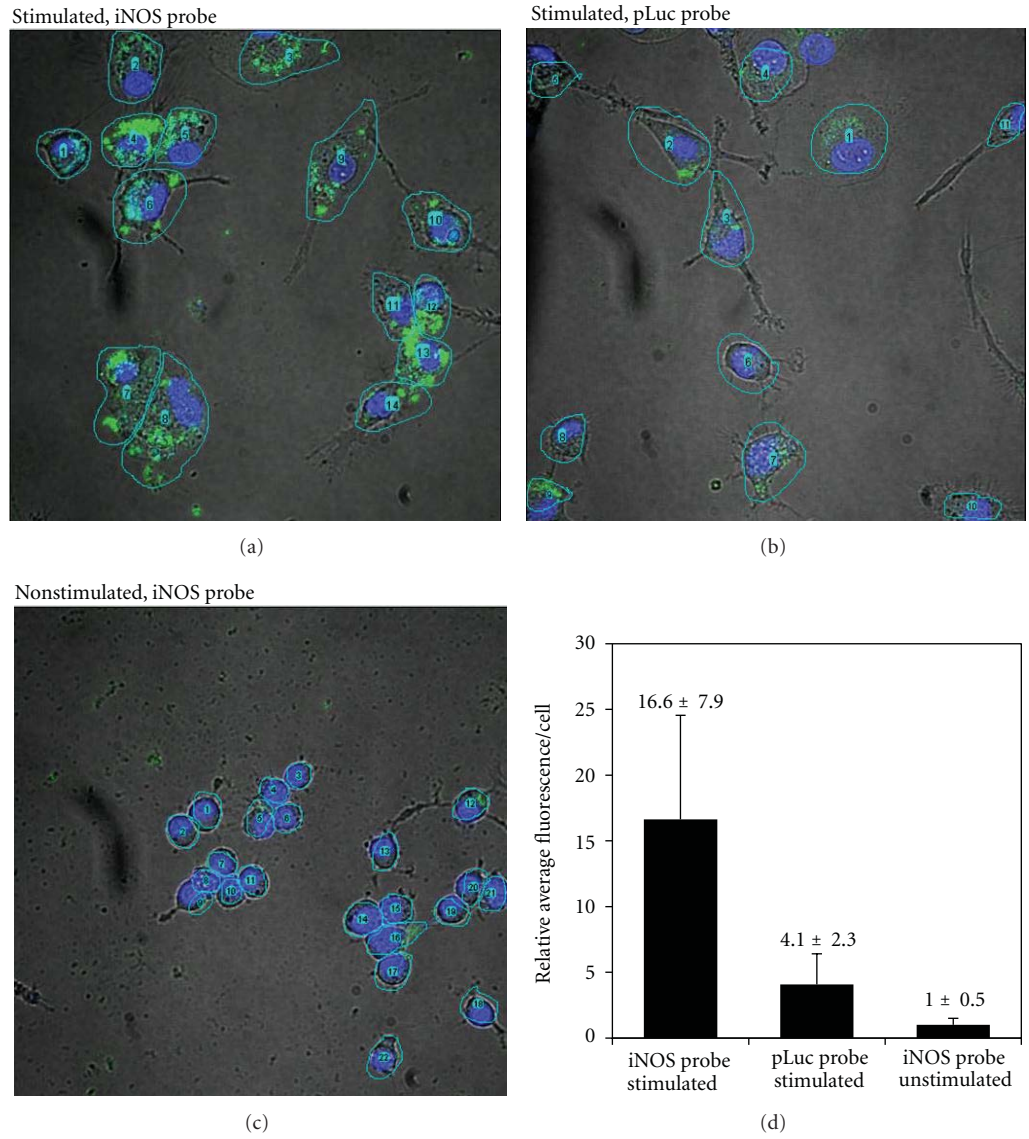


Figure 7: Live cell imaging of iNOS mRNA expression with the strand-displacement probes. Z-stack projection of confocal fluorescent images of RAW 264.7 cells and the quantitative analysis of fluorescence in selected regions of interests (ROIs) 24 h after transfection. The iNOS probe (0.4 μM FAM-iNOS-PNA annealed to 0.5 μM iNOS-DNA-DABCYL) or pLuc probe (0.4 μM FAM-pLuc-PNA annealed to 0.5 μM pLuc-DNA-DABCYL) was delivered with 9.7 μg/mL cSCK nanoparticle at an N/P ratio of 8 : 1. Green: FAM signal. Blue: Hoechst nuclear stain.

Acknowledgments

This material is based upon work supported by the National Heart Lung and Blood Institute of the National Institutes of Health as a Program of Excellence in Nanotechnology (HHSN268201000046C) and by the Washington University NIH Mass Spectrometry Resource (Grant no. P41 RR000954).

References

[1] S. Tyagi, "Imaging intracellular RNA distribution and dynamics in living cells," *Nature Methods*, vol. 6, no. 5, pp. 331–338, 2009.

[2] G. Bao, J. R. Won, and A. Tsourkas, "Fluorescent probes for live-cell RNA detection," *Annual Review of Biomedical Engineering*, vol. 11, pp. 25–47, 2009.

[3] B. A. Armitage, "Imaging of RNA in live cells," *Current Opinion in Chemical Biology*, vol. 15, no. 6, pp. 806–812, 2011.

[4] A. K. Chen, M. A. Behlke, and A. Tsourkas, "Efficient cytosolic delivery of molecular beacon conjugates and flow cytometric analysis of target RNA," *Nucleic Acids Research*, vol. 36, no. 12, article e69, 2008.

[5] P. Santangelo, N. Nitin, and G. Bao, "Nanostructured probes for RNA detection in living cells," *Annals of Biomedical Engineering*, vol. 34, no. 1, pp. 39–50, 2006.

[6] D. M. Kolpashchikov, "Binary probes for nucleic acid analysis," *Chemical Reviews*, vol. 110, no. 8, pp. 4709–4723, 2010.

[7] Y. Li, X. Zhou, and D. Ye, "Molecular beacons: an optimal multifunctional biological probe," *Biochemical and Biophysical Research Communications*, vol. 373, no. 4, pp. 457–461, 2008.

[8] K. J. Livak, S. J. A. Flood, J. Marmaro, W. Giusti, and K. Deetz, "Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization," *PCR Methods and Applications*, vol. 4, no. 6, pp. 357–362, 1995.

[9] M. K. Johansson, H. Fidder, D. Dick, and R. M. Cook, "Intramolecular dimers: a new strategy to flourescence quenching in dual-labeled oligonucleotide probes," *Journal of the American Chemical Society*, vol. 124, no. 24, pp. 6950–6956, 2002.

[10] M. S. Ellwood, M. Collins, E. F. Fritsch et al., "Strand displacement applied to assays with nucleic acid probes," *Clinical Chemistry*, vol. 32, no. 9, pp. 1631–1636, 1986.

[11] L. E. Morrison, T. C. Halder, and L. M. Stols, "Solution-phase detection of polynucleotides using interacting fluorescent labels and competitive hybridization," *Analytical Biochemistry*, vol. 183, no. 2, pp. 231–244, 1989.

[12] Q. Li, G. Luan, Q. Guo, and J. Liang, "A new class of homogeneous nucleic acid probes based on specific displacement hybridization," *Nucleic Acids Research*, vol. 30, no. 2, p. E5, 2002.

[13] D. Y. Zhang and E. Winfree, "Control of DNA strand displacement kinetics using toehold exchange," *Journal of the American Chemical Society*, vol. 131, no. 47, pp. 17303–17314, 2009.

[14] J. He, M. Rusckowski, Y. Wang et al., "Optical pretargeting of tumor with fluorescent MORF oligomers," *Molecular Imaging and Biology*, vol. 9, no. 1, pp. 17–23, 2007.

[15] M. Liang, X. Liu, D. Cheng et al., "Optical antisense tumor targeting in vivo with an improved fluorescent DNA duplex probe," *Bioconjugate Chemistry*, vol. 20, no. 6, pp. 1223–1227, 2009.

[16] D. S. Seferos, D. A. Giljohann, H. D. Hill, A. E. Prigodich, and C. A. Mirkin, "Nano-flares: probes for transfection and mRNA detection in living cells," *Journal of the American Chemical Society*, vol. 129, no. 50, pp. 15477–15479, 2007.

[17] S. Shakeel, S. Karim, and A. Ali, "Peptide nucleic acid (PNA)—a review," *Journal of Chemical Technology and Biotechnology*, vol. 81, no. 6, pp. 892–899, 2006.

[18] P. E. Nielsen, "Peptide Nucleic Acids (PNA) in chemical biology and drug discovery," *Chemistry and Biodiversity*, vol. 7, no. 4, pp. 786–804, 2010.

[19] H. Fang, K. Zhang, G. Shen, K. L. Wooley, and J. S. A. Taylor, "Cationic shell-cross-linked knedel-Like (cSCK) nanoparticles for highly efficient PNA delivery," *Molecular Pharmaceutics*, vol. 6, no. 2, pp. 615–626, 2009.

[20] K. Zhang, H. Fang, G. Shen, J. S. A. Taylor, and K. L. Wooley, "Well-defined cationic shell crosslinked nanoparticles for efficient delivery of DNA or peptide nucleic acids," *Proceedings of the American Thoracic Society*, vol. 6, no. 5, pp. 450–457, 2009.

[21] F. Aktan, "iNOS-mediated nitric oxide production and its regulation," *Life Sciences*, vol. 75, no. 6, pp. 639–653, 2004.

[22] A. Pautz, J. Art, S. Hahn, S. Nowag, C. Voss, and H. Kleinert, "Regulation of the expression of inducible nitric oxide synthase," *Nitric Oxide*, vol. 23, no. 2, pp. 75–93, 2010.

[23] H. Fang, Y. Shen, and J. S. Taylor, "Native mRNA antisense-accessible sites library for the selection of antisense oligonucleotides, PNAs, and siRNAs," *RNA*, vol. 16, no. 7, pp. 1429–1435, 2010.

[24] K. Zhang, H. Fang, Z. Wang, J. S. A. Taylor, and K. L. Wooley, "Cationic shell-crosslinked knedel-like nanoparticles for highly efficient gene and oligonucleotide transfection of mammalian cells," *Biomaterials*, vol. 30, no. 5, pp. 968–977, 2009.

[25] H. Fang, X. Yue, X. Li, and J. S. Taylor, "Identification and characterization of high affinity antisense PNAs for the human unr (upstream of N-ras) mRNA which is uniquely overexpressed in MCF-7 breast cancer cells," *Nucleic Acids Research*, vol. 33, no. 21, pp. 6700–6711, 2005.

[26] R. Shrestha, Y. Shen, K. A. Pollack, J. S. Taylor, and K. L. Wooley, "Dual peptide nucleic acid- and peptide-functionalized shell cross-linked nanoparticles designed to target mRNA toward the diagnosis and treatment of acute lung injury," *Bioconjugate Chemistry*, vol. 23, no. 3, pp. 574–585, 2012.

[27] H. Kuhn, V. V. Demidov, B. D. Gildea, M. J. Fiandaca, J. C. Coull, and M. D. Frank-Kamenetskii, "PNA beacons for duplex DNA," *Antisense and Nucleic Acid Drug Development*, vol. 11, no. 4, pp. 265–270, 2001.

[28] I. A. Nazarenko, S. K. Bhatnagar, and R. J. Hohman, "A closed tube format for amplification and detection of DNA based on energy transfer," *Nucleic Acids Research*, vol. 25, no. 12, pp. 2516–2521, 1997.

[29] I. Nazarenko, R. Pires, B. Lowe, M. Obaidy, and A. Rashtchian, "Effect of primary and secondary structure of oligodeoxyribonucleotides on the fluorescent properties of conjugated dyes," *Nucleic Acids Research*, vol. 30, no. 9, pp. 2089–2095, 2002.

[30] M. R. Lewis and F. Jia, "Antisense imaging: and miles to go before we sleep?" *Journal of Cellular Biochemistry*, vol. 90, no. 3, pp. 464–472, 2003.

[31] T. Noda and F. Amano, "Differences in nitric oxide synthase activity in a macrophage-like cell line, RAW264.7 cells, treated with lipopolysaccharide (LPS) in the presence or absence of interferon-γ (IFN-γ): possible heterogeneity of iNOS activity," *Journal of Biochemistry*, vol. 121, no. 1, pp. 38–46, 1997.

[32] C. Altmann, A. Andres-Hernando, R. H. McMahan et al., "Macrophages mediate lung inflammation in a mouse model of ischemic acute kidney injury," *American Journal of Renal Physiology*, vol. 302, no. 4, pp. F421–F432, 2012.

[33] H. W. Trask, R. Cowper-Sal-Lari, M. A. Sartor et al., "Microarray analysis of cytoplasmic versus whole cell RNA reveals a considerable number of missed and false positive mRNAs," *RNA*, vol. 15, no. 10, pp. 1917–1928, 2009.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 707323, 7 pages
doi:10.1155/2012/707323

Research Article

Superior Silencing by 2′,4′-BNA^{NC}-Based Short Antisense Oligonucleotides Compared to 2′,4′-BNA/LNA-Based Apolipoprotein B Antisense Inhibitors

Tsuyoshi Yamamoto,[1,2] Hidenori Yasuhara,[1,2] Fumito Wada,[1,2] Mariko Harada-Shiba,[2] Takeshi Imanishi,[3] and Satoshi Obika[1]

[1] *Applied Biopharmaceutical Sciences, Graduate School of Pharmaceutical Sciences, Osaka University, 1-6 Yamadaoka, Suita, Osaka 565-0871, Japan*
[2] *Department of Molecular Innovation in Lipidology, National Cerebral and Cardiovascular Center Research Institute, 5-7-1 Fujishirodai, Suita, Osaka 565-8565, Japan*
[3] *BNA Inc, 7-7-20 Saito-Asagi, Ibaraki, Osaka 567-0085, Japan*

Correspondence should be addressed to Satoshi Obika, obika@phs.osaka-u.ac.jp

Received 13 June 2012; Revised 17 August 2012; Accepted 18 August 2012

Academic Editor: Masayasu Kuwahara

The duplex stability with target mRNA and the gene silencing potential of a novel bridged nucleic acid analogue are described. The analogue, 2′,4′-BNA^{NC} antisense oligonucleotides (AONs) ranging from 10- to 20-nt-long, targeted apolipoprotein B. 2′,4′-BNA^{NC} was directly compared to its conventional bridged (or locked) nucleic acid (2′,4′-BNA/LNA)-based counterparts. Melting temperatures of duplexes formed between 2′,4′-BNA^{NC}-based antisense oligonucleotides and the target mRNA surpassed those of 2′,4′-BNA/LNA-based counterparts at all lengths. An *in vitro* transfection study revealed that when compared to the identical length 2′,4′-BNA/LNA-based counterpart, the corresponding 2′,4′-BNA^{NC}-based antisense oligonucleotide showed significantly stronger inhibitory activity. This inhibitory activity was more pronounced in shorter (13-, 14-, and 16-mer) oligonucleotides. On the other hand, the 2′,4′-BNA^{NC}-based 20-mer AON exhibited the highest affinity but the worst IC_{50} value, indicating that very high affinity may undermine antisense potency. These results suggest that the potency of AONs requires a balance between reward term and penalty term. Balance of these two parameters would depend on affinity, length, and the specific chemistry of the AON, and fine-tuning of this balance could lead to improved potency. We demonstrate that 2′,4′-BNA^{NC} may be a better alternative to conventional 2′,4′-BNA/LNA, even for "short" antisense oligonucleotides, which are attractive in terms of drug-likeness and cost-effective bulk production.

1. Introduction

Recently designed and synthesized high-performance modified-nucleic-acids (HiPerNAs) such as 2′-*O*-methyl RNA (2′-OMe), 2′-*O*-methoxyethyl RNA (MOE), and 2′,4′-bridged nucleic acid/locked nucleic acid (2′,4′-BNA/LNA) have improved performance compared to phosphorothioate antisense oligonucleotides (AONs). HiPerNAs overcome the systemic antisense effects of these earlier antisense oligonucleotides and show promise as antisense therapeutics for the treatment of a variety of diseases [1–5]. However, more potent and less toxic AONs are required, since several clinical trials of AON drugs carrying HiPerNAs have been recently terminated due to the lack of efficacy or because of safety concerns. In addition, toxicity and delivery problems remain [6–8].

We previously described a unique modified nucleic acid, 2′,4′-bridged nucleic acid (2′,4′-BNA; also known as LNA) [9, 10]. Its high therapeutic efficacy is based on the extraordinarily high target binding of the original 2′,4′-BNA/LNA-based AON. 2′,4′-BNA/LNA-based AON is widely accepted as one of the most promising antisense drugs, so fine-tuning the structure of BNA is the key for further improving the therapeutic potency and toxicological properties of

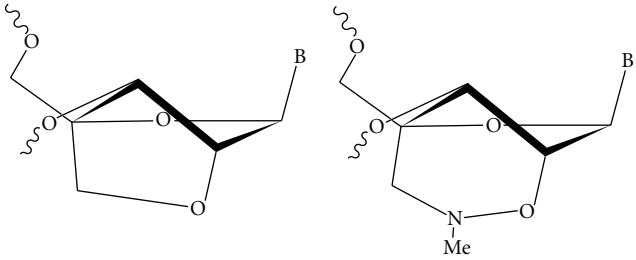


Figure 1: Structures of the BNAs. The chemical structure on the left is the original BNA, 2′,4′-BNA/LNA, and the structure on the right is 2′,4′-BNA^{NC}.

AON. In this context, Seth and coworkers developed BNA analogues and elucidated their potency and safety *in vivo*. They reported BNA or LNA analogues with 2′,4′-BNA/LNA-like binding affinities and biological activities with increased nuclease resistance and reduced toxicity [11–14]. We have also reported the synthesis and physicochemical properties of several novel BNAs, including 2′,4′-BNA^{COC} and 2,′4′-BNA^{NC} [15–17], and recently demonstrated the biological activity of a 2′,4′-BNA^{NC} -based AON. Our results indicate that 2′,4′-BNA^{NC} may be a better candidate as an antisense therapeutic [7]. 2′,4′-BNA^{NC} is a six-membered bridged structure containing a hydrophilic aminooxy moiety. A wide range of functional groups can be easily introduced, and the introduction of nitrogen atoms would improve the stability of the duplex by reducing repulsions between phosphates in the backbone [18, 19] (Figure 1). 2′,4′-BNA^{NC}-modified oligonucleotides have very high RNA affinity, similar to or even higher than their 2′,4′-BNA/LNA counterparts. Moreover, 2′,4′-BNA^{NC}-modified oligonucleotides are more resistant to endonucleolytic cleavage by nucleases than their 2′,4′-BNA/LNA counterparts [17]. Although there is limited information regarding the biological activity or therapeutic potency of 2′,4′-BNA^{NC}-based AONs, we have demonstrated the high systemic effects and safety of 2′,4′-BNA/LNA- and 2′,4′-BNA^{NC}-based 20-nucleotide-long (20-nt-long) AONs that target PCSK9 mRNA [7]. Additionally, Prakashet al. independently showed the high potency and the nontoxicity of 2′,4′-BNA^{NC}-based AONs [11].

Straarup et al. recently shortened the length of 2′,4′-BNA/LNA-based phosphorothioate AONs to eliminate the latent potency of 2′,4′-BNA/LNA drugs [20]. These short phosphorothioate AONs contain central 6- to 10-nt-long DNA regions flanked by terminal 2- to 4-nt-long 2′,4′-BNA/LNA segments. These short (12- to 14-nt) AONs would be beneficial in terms of target specificity. The introduction of only small numbers of modifications into short 2′,4′-BNA/LNA-based AONs can greatly increase target affinity. Thus, short 2′,4′-BNA/LNA phosphorothioate AONs can minimize length-dependent disadvantages such as phosphorothioate-related protein binding and RNase H inactivation [21–23] while maintaining satisfactory affinity and specificity. Additionally, short AONs are easier to produce on a bulk scale and could exhibit more drug-like characteristics. Based on the assumption that the strand-shortening strategy is also applicable to 2′,4′-BNA^{NC}-based AONs, we shortened 2′,4′-BNA^{NC}-based phosphorothioate AONs and directly compared their silencing activities against the corresponding 2′,4′-BNA/LNA-based apolipoprotein B (apoB)-targeting AONs.

2. Materials and Methods

2.1. Oligonucleotides. A series of 2′,4′-BNA/LNA-based antisense 10- to 20-nucleotide-long phosphorothioate gapmers reported previously by Straarup et al. [20] were prepared and used in this study. These AONs were designed with complementary target sites for both cynomolgus monkey and human apoB mRNA sequences. The 10- to 16-nt-long AONs can also target murine apoB mRNA (GenBank accession number NM_000384 and NM_009693 for human and mouse apoB mRNA, resp.). Additionally, we prepared 2′,4′-BNA^{NC}-based counterparts in which all the 2′,4′-BNA/LNAs were substituted by 2′,4′-BNA^{NC}. The synthesis of 2′,4′-BNA^{NC} with pyrimidine bases was previously reported [16, 17]; the synthesis of 2′,4′-BNA^{NC} with purine bases is currently being optimized and will be reported elsewhere. All the modified oligonucleotides were synthesized by Gene Design, Inc. (Ibaraki, Osaka, Japan) using standard phosphoramidite procedures and purified using HPLC.

2.2. Thermal Melting Study of Duplexes. UV melting experiments were carried out using a SHIMAZU UV-1650 spectrometer equipped with a T_m analysis accessory. Equimolar amounts of two single-stranded oligonucleotides were dissolved in 10 mM sodium phosphate buffer (pH 7.2) containing 100 mM NaCl to give a final strand concentration of 4.0 μM. The mixture was annealed by heating at 90°C followed by slow cooling to room temperature. The melting profile was recorded at 260 nm in the forward and reverse direction from 5 to 90°C at a scan rate of 0.5°C/min.

2.3. In Vitro Transfection Procedures. For AON transfection experiments, Huh-7 cells were seeded at 15×10^4 cells per well in 12-well plates. AONs were transfected by using Lipofectamine 2000 (Invitrogen, Carlsbad, CA) according to the manufacturer's procedures. After a 4-hour transfection, cells were washed with PBS, fresh medium was added, and the cells were incubated for an additional 20 hours at 37°C. After incubation, cells were collected and subjected to analyses.

2.4. mRNA Quantification Procedures. Total RNA was isolated from cultured cells using an RNeasy Mini Kit (Qiagen) according to the manufacturer's procedure. Gene expression was evaluated by a two-step quantitative reverse transcription-PCR method. Reverse transcription of RNA samples was performed by using a High Capacity cDNA Reverse-Transcription Kit (Applied Biosystems, Foster City, CA), and quantitative PCR was performed using a Fast TaqMan Gene Expression Assay (Applied Biosystems). The mRNA levels of target genes were normalized to the GAPDH

Table 1: Oligonucleotides used in this study.

ID	Sequence	T_m (°C)	IC_{50} (nM)
ApoB-LNA-20	5′-TTCAGcattggtattCAGTG-3′	75 ± 0.8	7.9 ± 1.7
ApoB-LNA-16	5′-CAGcattggtatTCAG-3′	64 ± 0.8	2.9 ± 0.4[a]
ApoB-LNA-14	5′-AGCattggtatTCA-3′	61 ± 0.2	1.4 ± 0.5[b]
ApoB-LNA-13	5′-GCattggtatTCA-3′	58 ± 0.2	2.8 ± 0.8[a]
ApoB-LNA-10	5′-CattggtatT-3′	25 ± 1.4	N.D.
ApoB-NC-20	5′-*TTCAG*cattggtatt*CAGTG*-3′	79 ± 1.0	11.4 ± 3.0
ApoB-NC-16	5′-*CAG*cattggtat*TCAG*-3′	69 ± 0.8	1.3 ± 0.3[b]
ApoB-NC-14	5′-*AGC*attggtat*TCA*-3′	64 ± 0.3	0.6 ± 0.1
ApoB-NC-13	5′-*GC*attggtat*TCA*-3′	61 ± 0.1	0.9 ± 0.1
ApoB-NC-10	5′-*C*attggtat*T*-3′	32 ± 0.4	N.D.

2′,4′-BNA/LNA was shown in uppercase and 2′,4′-BNA^{NC} was in italic. Natural DNA was shown in lowercase. All the linkages are phosphorothioated. We measured T_m and IC_{50} values of all entries. T_m values were determined in three independent experiments (±SD). Nondetectable IC_{50} values, due to low potency, were marked ND. [a,b]Pairs of two IC_{50} values with superscript letters are NOT statistically significant.

mRNA level. The following primer sets were used for quantitative PCR. For human apoB and GAPDH, assay IDs of Hs01071209_m1 and Hs02758991_g1 were used, respectively.

2.5. Western Blotting. Two days after transfection, the cultures were subjected to centrifugation at 4°C, 10,000 rpm for 15 min. Each supernatant was collected into an Amicon Ultra-4 Centrifugal Filter Ultracel PL-10k (Millipore) and centrifuged at 4°C at 3,000 rpm for 1 h, and then each supernatant was added to individual Vivaspin 500 units (Sartorius Stedim Biotech) and centrifuged at 4°C at 3,000 rpm for 0.5 h. Each sample (9 μL) was added to 9 μL of Novex Tris-Glycine SDS Sample Buffer (2x) (Invitrogen) and applied to a 3–8% NuPAGE Tris-Acetate Gel (Invitrogen). Electrophoresis was performed at 180 V for 130 min. The separated proteins were transferred to a PVDF membrane (Millipore) at 220 mA for 120 min. Membranes were then incubated with 10 mL of blocking buffer (Blocking One; Nacalai Tesque) for 12 h at 4°C. Membranes were successively incubated with primary antibody of anti-human ApoB antibody (R&D Systems) for 80 min at room temperature. Then, each membrane was washed with PBS containing 0.1% tween (PBST) 4 times. Membranes were incubated with goat anti-mouse IgG-HRP antibody (Santa Cruz Biotechnology) for 80 min at room temperature. Chemiluminescent detection was performed using an ECL Advance Western Blot Detection Kit (Amersham Biosciences) according to the manufacture's procedure. Bands were visualized using an LAS-4000 mini (Fujifilm).

2.6. Statistics. Application of linear regression techniques to plots of the expression of apoB mRNA levels versus the logarithm of transfection concentration allows estimation of the coefficient (slope) and intercept values using the following equation: apoB mRNA = Coefficient∗log (Concentration) + Intercept. These estimates can be useful guides for the efficacies of the AONs. Linear regression analysis was applied to 500 bootstrap sample sets obtained from three independent cellular assays. The coefficient and intercept of each regression line were compared using Tukey's test or one-sample t-test.

3. Results and Discussion

To better understand the effect of strand shortening on 2′,4′-BNA^{NC}-based AONs, a series of 2′,4′-BNA^{NC}-based antisense oligonucleotides of 10- to 20-nt-long phosphorothioate gapmers and 2′,4′-BNA/LNA-counterparts were synthesized (Table 1). Formation of a stable duplex with the target mRNA is a minimum essential step of the onset of an antisense effect. We first evaluated the thermal stability of the duplex formed between each modified AON and the single-stranded 20-nt-long oligoribonucleotide complementary to ApoB-LNA-20 and ApoB-NC-20. As expected, the longer the strand or the greater the number of modifications, the higher the T_m value. Moreover, the T_m values of 2′,4′-BNA^{NC}-based AONs surpassed those of their 2′,4′-BNA/LNA-based counterparts for any given length (Table 1), in good agreement with previous reports [11, 17]. Note that the exact T_m values of LNAs in Table 1 are different from those given by Straarup et al., due to differences in composition of the measurement buffer solutions and in the length of the complementary RNAs between the two studies.

We next used *in vitro* mRNA silencing assays to estimate the potency of 2′,4′-BNA^{NC}-based AONs and to compare their potency directly to the corresponding 2′,4′-BNA/LNA-based AONs. We used the Huh-7 human hepatoma cell line, which expresses high levels of apoB mRNA in cells and secrets its protein into the medium. Each AON was introduced using standard lipofection procedures. All the AONs, except the 10-mers, ApoB-LNA-10, and ApoB-NC-10, reduced apoB mRNA and protein expression (and hence secreted protein) levels in a dose-dependent manner in the cells and culture medium, respectively, (Figures 2(a) and 2(b)). ApoB-LNA-10 did not reduce apoB mRNA levels even at concentrations above 64 nM. This may be because ApoB-LNA-10 did not bind target mRNA at 37°C due to

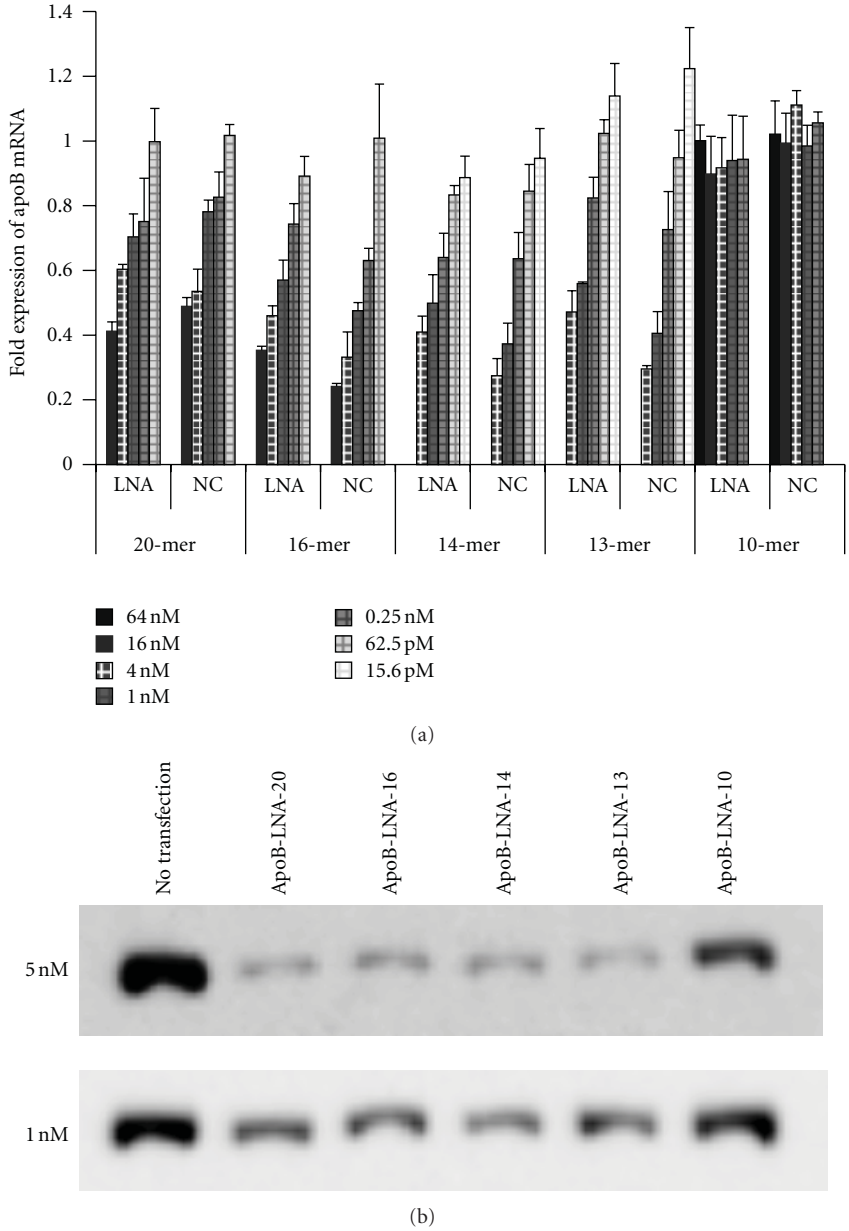


Figure 2: *In vitro* silencing properties of BNA-based AONs. (a) Various concentrations (15.6 pM–64 nM) of AONs were introduced into Huh-7 cells using Lipofectamine 2000. After 24-h incubation, the cells were collected, and the expression levels of apoB mRNA were determined. Data represent means ± SD. (b) Reduction of Apo B protein levels in the culture medium following transfection was confirmed by western blotting.

lack of affinity. ApoB-NC-10 also did not reduce apoB mRNA expression, despite its higher T_m value compared to that of ApoB-LNA-10. Application of linear regression techniques to plots of the expression of apoB mRNA levels versus the logarithm of transfection concentration allowed estimation of the coefficient (slope) and intercept values. Statistical comparison of these parameters between arms with the identical length would provide useful guides for the efficacies of the AONs. One-sample *t*-tests revealed that the coefficients (slopes) of the 10-mer AONs had statistically insignificant downward slopes, suggesting that these 10-mers have little or no silencing effect (Figure 3).

A length-dependent decrease in potency in 14- to 20-nt-long 2′,4′-BNA/LNA and 2′,4′-BNA^{NC}-based AONs was observed (Table 1 and Figure 1). In each series of AONs, the ApoB-LNA-14 and ApoB-NC-14 were the most potent, whereas the ApoB-LNA-13 was the most potent AON reported by Straarup et al. [20]. IC_{50} values of the 15-, 14-, 13-, and 12-mer AONs were so close to each other (≈0.5 nM) that Straarup et al. confirmed the order by

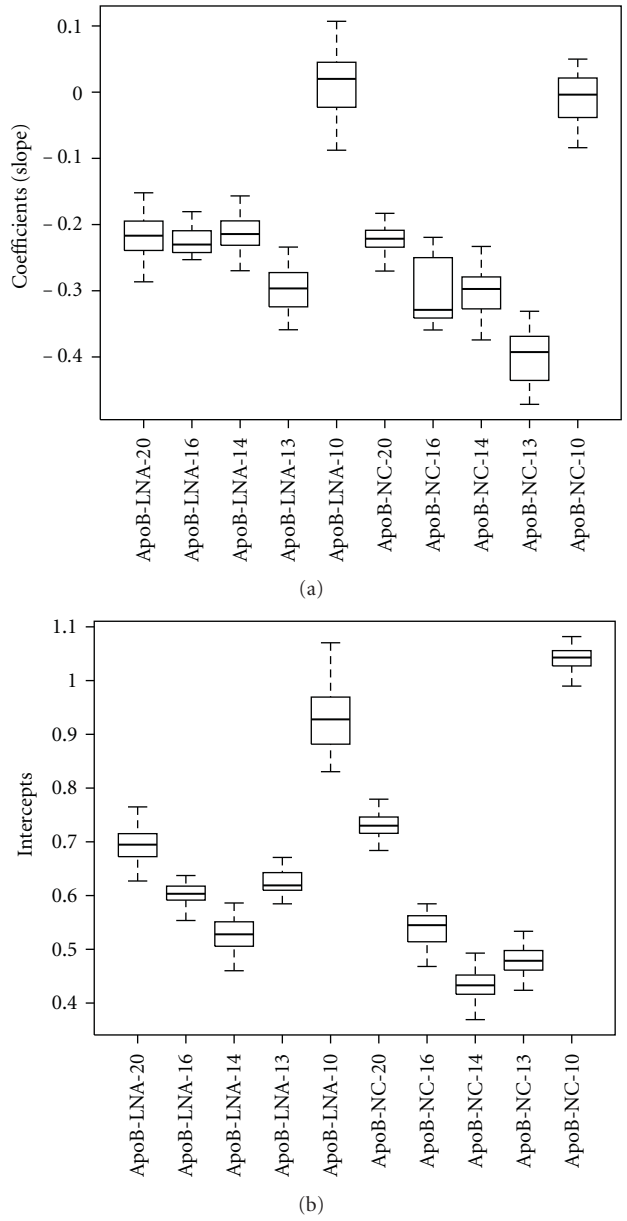


Figure 3: Boxplots of coefficients and intercepts of regression lines obtained from cellular assay data. Bold lines in the box indicate medians. Top and bottom lines of the box indicate upper and lower quartile points, respectively. Top and bottom lines in contact with a dotted line indicate the highest and lowest data point, respectively. Open circles indicate outliers.

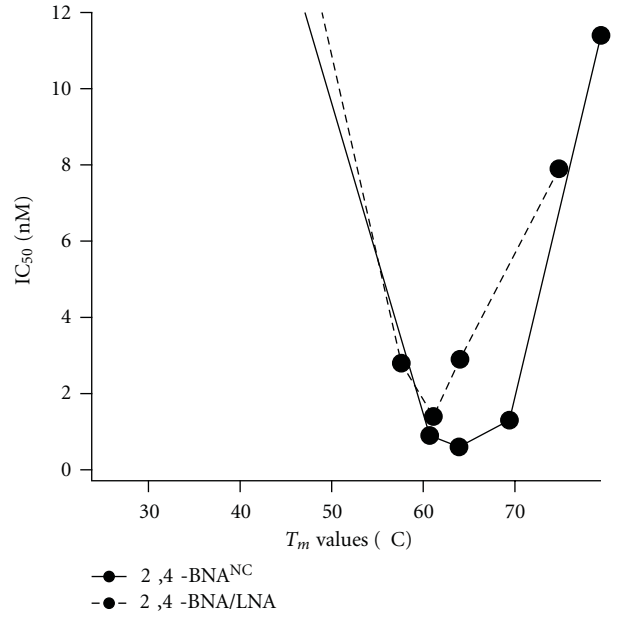


Figure 4: Relationship between target affinity and antisense potency. IC_{50} as a function of T_m values of 2′,4′-BNA^{NC}-based AONs (solid line) and 2′,4′-BNA/LNA-based AONs (dotted line).

conducting an *in vivo* silencing study and showed that 12- and 13-mer AONs are the most potent. In contrast, we confirmed the order of *in vitro* silencing effects of all the entries in Table 1 by using larger 12-well culture dishes. These differences in experimental conditions (*in vivo* versus *in vitro*) may explain the differences in the IC_{50} values of 2′,4′-BNA/LNA-based AONs and their order of potency. An *in vivo* study might be necessary to estimate the true order. Nevertheless, we observed in both chemistries that shorter AONs (16-, 14-, and 13-mer) are statistically significantly more potent than 20-mer AONs. Because the 10-mers did not show any activity, activity is positively correlated with binding affinity and indicates the presence of a "threshold affinity." However, longer AONs with higher affinities did not exhibit higher activities. This suggests the presence of a "length penalty" [20, 22] and the presence of an "optimal affinity," which might be a more appropriate description than threshold affinity. The variables that independently govern "length penalty" and "optimal affinity" remain largely unknown. However, plasma and intracellular proteins preferentially bind to the phosphorothioate internucleotide linkages, and these linkages are known to control RNase H activity [23–26]. Thus, the number of phosphorothioate linkages is partly related to the "length penalty."

Surprisingly, when compared to the identical length 2′,4′-BNA/LNA-based counterpart, the corresponding 2′,4′-BNA^{NC}-based AON showed stronger inhibitory activity. The differential in inhibitory activity is more pronounced in shorter (13-, 14-, and 16-mers) AONs than in the 20-mer (Table 1 and Figure 4). Indeed, as shown in Figure 3, statistical comparison using the Tukey test of the coefficients and intercepts of the regression lines, which indicate how efficiently and strongly AONs reduce apoB mRNA, revealed that 2′,4′-BNA^{NC}-based 13-, 14-, and 16-mers exhibit significantly stronger inhibitory activity than their 2′,4′-BNA/LNA counterparts. These potency differences could not be explained simply by the higher T_m of the 2′,4′-BNA^{NC}-based AONs compared to their 2′,4′-BNA/LNA-based counterpart. 2′,4′-BNA^{NC}-based AONs might exhibit less "length penalty" than their 2′,4′-BNA/LNA-based counterpart. A weaker affinity for protein binding or lower RNase H inhibitory activity of 2′,4′-BNA^{NC}-based AONs could cause this potency difference. On the other hand,

Figure 4 shows a plot of IC_{50} versus T_m, in which ApoB-NC-20, which has the highest affinity, exhibited worse IC_{50}value than ApoB-LNA-20. These findings suggest that very high affinity possibly undermines antisense potency, although binding affinity to the target generally correlates positively with potency in the case of traditional small molecule drugs. The precise onset mechanism giving rise to this phenomenon remains unclear due to the lack of experimental data. However, the U- or V-shaped curves in Figure 4 clearly indicate that the potency of AONs is the result of a delicate balance of reward term and penalty term. Thus, fine adjustment to the optimal affinity and elimination of the penalties such as excess protein binding and RNase H inhibitory activity could result in superior efficacy, represented as the lowest point on the U- or V-shaped curve.

In conclusion, we have shown that 2′,4′-BNA^{NC}-based AONs targeting apoB mRNA have higher binding affinities to the target RNA than do 2′,4′-BNA/LNA-based AONs. Additionally, *in vitro* transfection studies revealed the superior silencing effect of short 2′,4′-BNA^{NC}-based AONs (<20-nt-long), indicating that 2′,4′-BNA^{NC} may have advantageous properties as short antisense drugs. We are currently investigating the potential and safety of 2′,4′-BNA^{NC}-based AONs as therapeutic drugs.

Acknowledgments

This work was supported by the advanced research for medical products mining program of the National Institute of Biomedical Innovation (NIBIO) and a research grant from the Ministry of Health, Labor, and Welfare (H23-seisakutansaku-ippan-004) of Japan. T. Yamamoto acknowledges the Research Fellowship for Young Scientists from the Japan Society for the Promotion of Science (JSPS).

References

[1] E. R. Rayburn and R. Zhang, "Antisense, RNAi, and gene silencing strategies for therapy: mission possible or impossible?" *Drug Discovery Today*, vol. 13, no. 11-12, pp. 513–521, 2008.

[2] T. Yamamoto, M. Nakatani, K. Narukawa, and S. Obika, "Antisense drug discovery and development," *Future Medicinal Chemistry*, vol. 3, no. 3, pp. 339–365, 2011.

[3] F. Akdim, M. E. Visser, D. L. Tribble et al., "Effect of mipomersen, an apolipoprotein B synthesis inhibitor, on low-density lipoprotein cholesterol in patients with familial hypercholesterolemia," *American Journal of Cardiology*, vol. 105, no. 10, pp. 1413–1419, 2010.

[4] M. E. Visser, J. J. P. Kastelein, and E. S. G. Stroes, "Apolipoprotein B synthesis inhibition: results from clinical trials," *Current Opinion in Lipidology*, vol. 21, no. 4, pp. 319–323, 2010.

[5] T. S. Crooke, *Antisense Drug Technologies: Principles, Strategies, and Applications*, CRC Press, 2007.

[6] E. E. Swayze, A. M. Siwkowski, E. V. Wancewicz et al., "Antisense oligonucleotides containing locked nucleic acid improve potency but cause significant hepatotoxicity in animals," *Nucleic Acids Research*, vol. 35, no. 2, pp. 687–700, 2007.

[7] M. H. -S. Tsuyoshi Yamamoto, M. Nakatani, S. Wada et al., "Cholesterol-lowering action of BNA-based antisense oligonucleotides targeting PCSK9 in atherogenic diet-induced hypercholesterolemic mice," *Molecular Therapy-Nucleic Acids*, vol. 1, article e22, 2012.

[8] P. J. White, F. Anastasopoulos, C. W. Pouton, and B. J. Boyd, "Overcoming biological barriers to in vivo efficacy of antisense oligonucleotides," *Expert reviews in molecular medicine*, vol. 11, article e10, 2009.

[9] S. Obika, S. M. A. Rahman, A. Fujisaka, Y. Kawada, T. Baba, and T. Imanishi, "Bridged nucleic acids: development, synthesis and properties," *Heterocycles*, vol. 81, no. 6, pp. 1347–1392, 2010.

[10] T. Imanishi and S. Obika, "BNAs: novel nucleic acid analogs with a bridged sugar moiety," *Chemical Communications*, no. 16, pp. 1653–1659, 2002.

[11] T. P. Prakash, A. Siwkowski, C. R. Allerson et al., "Antisense oligonucleotides containing conformational constrained 2′, 4′-(N-Methoxy)aminomethylene and 2′, 4′-aminooxymethylene and 2′-O,4′-C-aminomethylene bridged nucleoside analogues show improved potency in animal models," *Journal of Medicinal Chemistry*, vol. 53, no. 4, pp. 1636–1650, 2010.

[12] P. P. Seth, C. R. Allerson, A. Berdeja et al., "An exocyclic methylene group acts as a bioisostere of the 2′-oxygen atom in lna," *Journal of the American Chemical Society*, vol. 132, no. 42, pp. 14942–14950, 2010.

[13] P. P. Seth, A. Siwkowski, C. R. Allerson et al., "Short antisense oligonucleotides with novel 2′, 4′ conformationaly restricted nucleoside analogues show improved potency without increased toxicity in animals," *Journal of Medicinal Chemistry*, vol. 52, no. 1, pp. 10–13, 2009.

[14] P. P. Seth, G. Vasquez, C. A. Allerson et al., "Synthesis and biophysical evaluation of 2′, 4′-constrained 2′O-methoxyethyl and 2′, 4′-constrained 2′O-ethyl nucleic acid analogues," *Journal of Organic Chemistry*, vol. 75, no. 5, pp. 1569–1581, 2010.

[15] Y. Mitsuoka, T. Kodama, R. Ohnishi, Y. Hari, T. Imanishi, and S. Obika, "A bridged nucleic acid, 2′, 4′-BNACOC: synthesis of fully modified oligonucleotides bearing thymine, 5-methylcytosine, adenine and guanine 2′, 4′-BNACOC monomers and RNA-selective nucleic-acid recognition," *Nucleic Acids Research*, vol. 37, no. 4, pp. 1225–1238, 2009.

[16] K. Miyashita, S. M. A. Rahman, S. Seki, S. Obika, and T. Imanishi, "N-Methyl substituted 2′, 4′-BNANC: a highly nuclease-resistant nucleic acid analogue with high-affinity RNA selective hybridization," *Chemical Communications*, no. 36, pp. 3765–3767, 2007.

[17] S. M. A. Rahman, S. Seki, S. Obika, H. Yoshikawa, K. Miyashita, and T. Imanishi, "Design, synthesis, and properties of 2′, 4′-BNANC: a bridged nucleic acid analogue," *Journal of the American Chemical Society*, vol. 130, no. 14, pp. 4886–4896, 2008.

[18] S. K. Singh, R. Kumar, and J. Wengel, "Synthesis of 2′-amino-LNA: a novel conformationally restricted high-affinity oligonucleotide analogue with a handle," *Journal of Organic Chemistry*, vol. 63, no. 26, pp. 10035–10039, 1998.

[19] D. Honcharenko, O. P. Varghese, O. Plashkevych, J. Barman, and J. Chattopadhyaya, "Synthesis and structure of novel conformationally constrained 1′, 2′-azetidine-fused bicyclic pyrimidine nucleosides: their incorporation into oligo-DNAs

and thermal stability of the heteroduplexes," *Journal of Organic Chemistry*, vol. 71, no. 1, pp. 299–314, 2006.

[20] E. M. Straarup, N. Fisker, M. Hedtjärn et al., "Short locked nucleic acid antisense oligonucleotides potently reduce apolipoprotein B mRNA and serum cholesterol in mice and nonhuman primates," *Nucleic Acids Research*, vol. 38, no. 20, pp. 7100–7111, 2010.

[21] A. A. Levin, "A review of issues in the pharmacokinetics and toxicology of phosphorothioate antisense oligonucleotides," *Biochimica et Biophysica Acta*, vol. 1489, no. 1, pp. 69–84, 1999.

[22] T. A. Watanabe, R. S. Geary, and A. A. Levin, "Plasma protein binding of an antisense oligonucleotide targeting human ICAM-1 (ISIS 2302)," *Oligonucleotides*, vol. 16, no. 2, pp. 169–180, 2006.

[23] W. Y. Gao, F. S. Han, C. Storm, W. Egan, and Y. C. Cheng, "Phosphorothioate oligonucleotides are inhibitors of human DNA polymerases and RNase H: implications for antisense technology," *Molecular Pharmacology*, vol. 41, no. 2, pp. 223–229, 1992.

[24] L. Benimetskaya, J. L. Tonkinson, M. Koziolkiewicz et al., "Binding of phosphorothioate oligodeoxynucleotides to basic fibroblast growth factor, recombinant soluble CD4, laminin and fibronectin is P-chirality independent," *Nucleic Acids Research*, vol. 23, no. 21, pp. 4239–4245, 1995.

[25] D. A. Brown, S. H. Kang, S. M. Gryaznov et al., "Effect of phosphorothioate modification of oligodeoxynucleotides on specific protein binding," *Journal of Biological Chemistry*, vol. 269, no. 43, pp. 26801–26805, 1994.

[26] M. A. Guvakova, L. A. Yakubov, I. Vlodavsky, J. L. Tonkinson, and C. A. Stein, "Phosphorothioate oligodeoxynucleotides bind to basic fibroblast growth factor, inhibit its binding to cell surface receptors, and remove it from low affinity binding sites on extracellular matrix," *Journal of Biological Chemistry*, vol. 270, no. 6, pp. 2620–2627, 1995.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 371379, 9 pages
doi:10.1155/2012/371379

Research Article

In Vitro Selection of Fab Fragments by mRNA Display and Gene-Linking Emulsion PCR

Takeshi Sumida, Hiroshi Yanagawa, and Nobuhide Doi

Department of Biosciences and Informatics, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan

Correspondence should be addressed to Nobuhide Doi, doi@bio.keio.ac.jp

Received 29 May 2012; Revised 1 August 2012; Accepted 1 August 2012

Academic Editor: Hiroshi Murakami

In vitro selection by display methods has been an effective tool for engineering recombinant antibodies. mRNA display based on a cell-free translation system has the advantages of larger library sizes and quicker selection procedures compared with cell-based display methods such as phage display. However, mRNA display has been limited to select single-chain polypeptides such as scFvs due to its characteristic of linking a nascent polypeptide with its encoding mRNA on the ribosome. Here we demonstrated a new way of selecting heterodimeric Fab fragments by using mRNA display combined with emulsion PCR. We designed a pair of complementary 5′ UTR sequences that can link the Fab heavy and light chain genes together by overlap-extension PCR in water-in-oil emulsions. We confirmed that two mRNA-displayed polypeptides for heavy and light chain of a model Fab fragment were associated into the active form and that a specific Fab fragment gene was enriched over 100-fold per round of a model affinity selection followed by the gene-linking emulsion PCR. We further performed directed evolution of Fab fragments with higher binding activity from a randomized Fab fragment library.

1. Introduction

In vitro selection by display methods has been an effective tool in the field of protein engineering and especially has been used to engineer recombinant antibodies for various biological applications [1]. Phage display has been widely used in the industry due to its feasibility to select Fab fragments [2]. The Fab fragment of an immunoglobulin is a heterodimer of the N-terminal half of a heavy (H) chain and a complete light (L) chain. Because the Fab is more native-like than the single-chain Fv (scFv), which is the other commonly used recombinant antibody format for *in vitro* selection, the Fab fragment format makes it able to select more practical antibodies [3]. Other than phage display, cell-free translation-based methods such as ribosome display [4] and mRNA display [5] are being used for *in vitro* selection of antibodies due to its advantage of permitting speedier selection from larger size libraries than cell-based methods. However, these cell-free translation-based methods are limited to select scFvs due to its characteristic of linking a nascent polypeptide with its encoding mRNA on the ribosome.

To overcome this limit, we have recently developed a bicistronic DNA display to select Fab fragments in a cell-free translation system [6]. Bicistronic DNA display relies on *in vitro* compartmentalization in water-in-oil emulsions [7], and the man-made cell-like compartments make it possible to display oligomeric proteins in a cell-free translation system. Although bicistronic DNA display has made it possible to select Fab fragments in a cell-free translation system, it has some disadvantages compared with mRNA display. First, the initial library size of bicistronic DNA display is three orders of magnitude less than that of mRNA display. Second, the linkage between the DNA and protein is a streptavidin-biotin complex, making it less stable compared with the covalent bond in mRNA display.

In this study we combined emulsion PCR [8–11] with mRNA display in order to be able to select Fab fragments by mRNA display. Since mRNA display is capable of selecting candidates from a more diverse library and designing a more

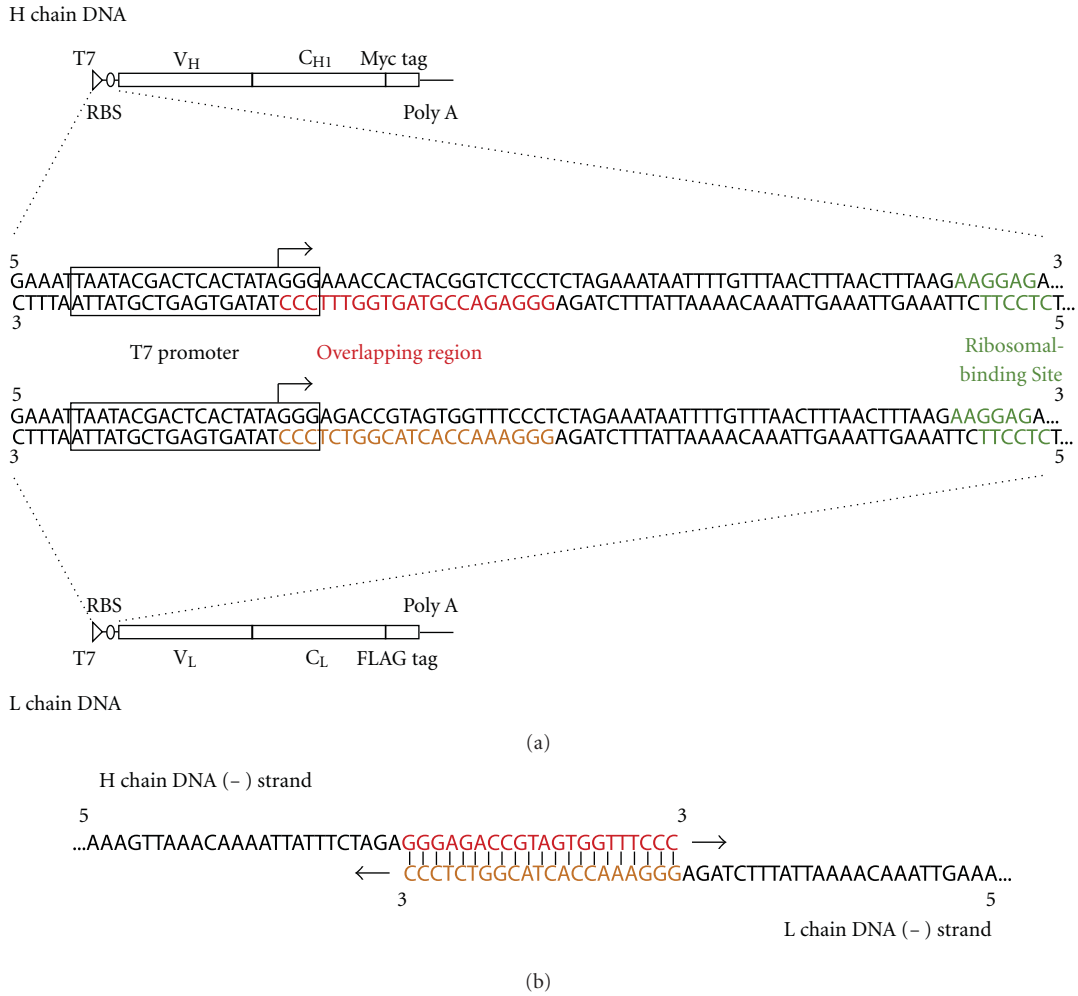


Figure 1: The DNA construct of the Fab fragments for mRNA display. (a) From the 5′ end it consists of a T7 promoter (T7), ribosomal-binding site (RBS), variable region and constant region of the H chain or L chain, epitope tag, and a poly A tail. The H chain epitope tag is a Myc tag and the L chain is a FLAG tag. The poly A tail is for purification of mRNA-displayed molecules by biotin-oligo dT in combination with streptavidin beads. (Middle) Details of the linkable 5′ UTR. Between the T7 promoter (boxed) and RBS (green), 21 bases from the beginning of T7 promoter transcription start point (arrow), in other words +1 to +21 of the H chain DNA (red) and L chain DNA (orange) are designed so that they overlap with each other during overlap-extension PCR. (b) The reverse-transcribed DNA strands overlap at the overlapping region (red) to form an H chain gene and L-chain-gene-linked DNA.

flexible selection strategy compared with bicistronic DNA display, this new method would provide a new option for selecting Fab fragments in a cell-free translation system.

2. Results and Discussion

2.1. Strategy. A Fab fragment consists of an H chain and an L chain, and by applying mRNA display, an mRNA-displayed H chain and an mRNA-displayed L chain can each be made. If these two mRNA-displayed molecules dimerize, they will form an mRNA-displayed Fab fragment. However, in this case, the correspondence of the selected H and L chains cannot be determined because the two genes are different RNA molecules and will be amplified separately after affinity selection. Applying overlap-extension PCR in water-in-oil emulsion from a single Fab molecule and linking these two genes together to amplify them at once will overcome this problem. Thus, we have designed a pair of complementary 5′ UTR sequences that can be linked together by overlap-extension PCR (Figure 1). The whole DNA construct for this strategy consists of a linkable 5′ UTR with a T7 promoter and ribosomal binding site; an ORF with the variable region, constant region, and an affinity tag, and at the 3′ end there are 25 adenines for mRNA-based purification by oligo-dT resin.

The scheme for *in vitro* selection of Fab fragments using mRNA display and emulsion PCR is shown in Figure 2. Firstly, mRNA-displayed H and L chains are separately prepared by *in vitro* translation of puromycin-ligated mRNA templates. Both the H chain and L chain are subsequently purified by oligo-dT resin in order to remove all free proteins that could not form an mRNA-displayed molecule, avoiding a Fab fragment existing with only one chain being

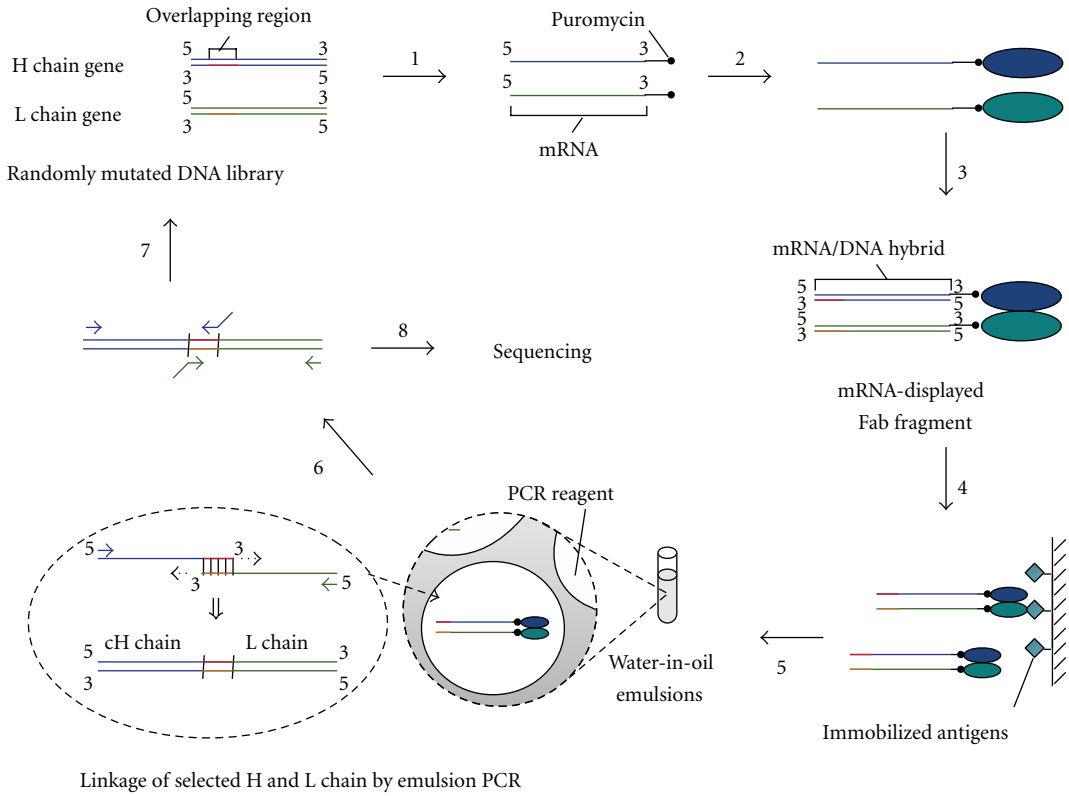


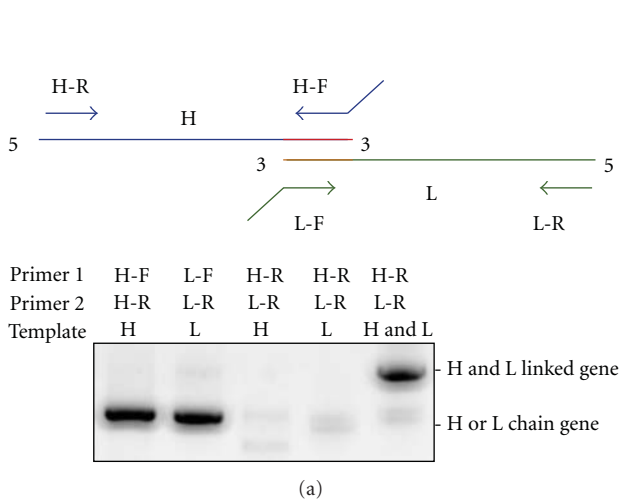
Figure 2: Scheme of *in vitro* selection of Fab fragments using mRNA display and emulsion PCR. Step 1: A randomly mutated DNA library of an H chain gene and an L chain gene is separately prepared. Each DNA library is transcribed and puromycin is ligated to the 3′ end to make an mRNA template. Step 2: The mRNA template is translated to form an mRNA-displayed molecule. Step 3: The molecules are purified and subsequently reverse is transcribed to make the mRNA portion a DNA hybrid. The H and L chain molecules are combined to form an mRNA-displayed Fab fragment. Step 4: The mRNA-displayed Fab fragments are subjected to *in vitro* antigen selection. Step 5: The selected mRNA-displayed Fab fragments are recovered and mixed with PCR reagents. The mixture is emulsified and the corresponding H and L chain genes are linked together by overlap extension PCR. Step 6: The linked DNA is amplified by PCR again with different primers to regenerate the H and L chain genes. Step 7: The regenerated genes are reamplified by DNA shuffling to make the templates for the next round of selection. Step 8: After a suitable number of rounds of selection, the selected DNA is cloned and sequenced to identify the selected Fab fragments.

mRNA-displayed. After a reverse transcription step in order to make the mRNA portion of the mRNA-displayed molecule an mRNA/DNA hybrid, the H and L chains are mixed together to form mRNA-displayed Fab fragments. These mRNA-displayed Fab fragments are selected by the target antigen and then eluted under conditions that the oligomeric structure is maintained. The eluted mRNA-displayed Fab fragments are subjected to emulsion PCR and a single mRNA-displayed Fab fragment is trapped inside a single micelle where the H and L chain genes are linked together by overlap-extension PCR. The linked DNA is either amplified by PCR to regenerate the H and L chain genes for further selection or sequenced to identify the selected Fab fragments.

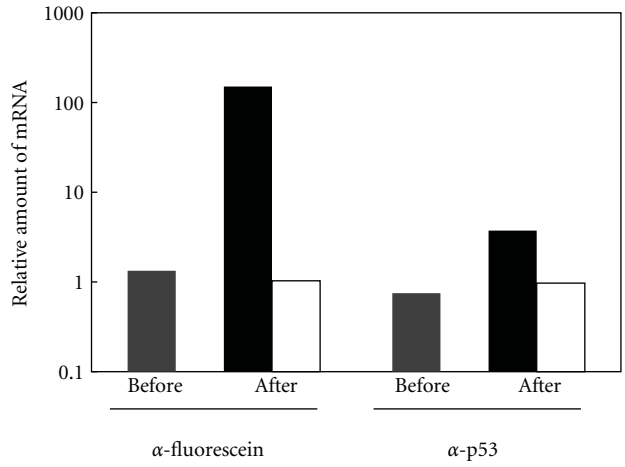
2.2. Proof-of-Principle Experiments. In order for the scheme depicted in Figure 2 to work, we confirmed the following three points: (i) the H and L chain genes overlap properly by emulsion PCR; (ii) an mRNA-displayed H chain and an mRNA-displayed L chain form a Fab fragment with binding activity; (iii) genes do not crossover during emulsion PCR and the corresponding H chain and L chain genes are properly linked together.

Figure 3(a) shows that when there are only reverse primers for the H and L chains, the DNA amplifies only when both the H and L chain genes exist and does not when there is only one of the genes. This proves that the linkable 5′ UTR is properly designed to overlap and only Fab fragments with both mRNA-displayed H and L chains can be specifically amplified. Therefore, unwanted gene amplification of unspecific H chain or L chain binders that does not form Fab fragments can be removed, exhibiting the advantage of this overlapping method compared to regular PCR amplification.

Next, to show that an mRNA-displayed H chain and mRNA-displayed L chain form a Fab fragment with binding activity, mRNA-displayed anti-fluorescein Fab and anti-p53 Fab fragments were used in a model affinity selection experiment. Confirmed by pull-down assays and western blotting, these antibodies do not bind unless they form a Fab fragment (data not shown). Equal molar amounts of mRNA-displayed anti-fluorescein Fab and anti-p53 Fab fragments were subjected to *in vitro* selection against antigen



(a)



(b)

		L chain	
		Fluorescein	p53
H chain	Fluorescein	11	0
	p53	0	2

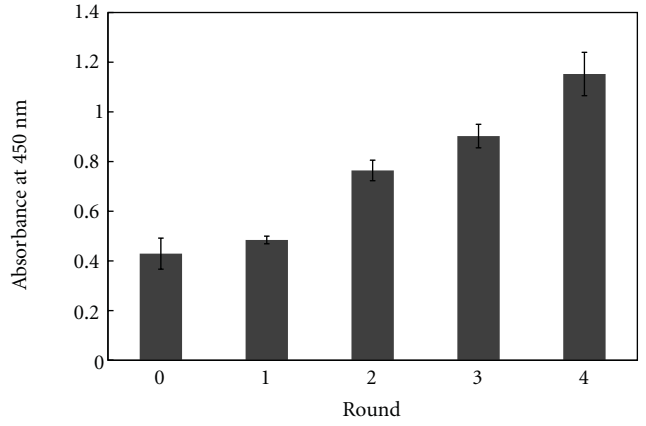
(c)

Figure 3: Model experiments of *in vitro* selection of Fab fragments using mRNA display and emulsion PCR. (a) Confirmation of the H and L chain gene linking by PCR. The H chain or/and L chain gene of anti-fluorescein Fab fragments were amplified by PCR with forward and reverse primers of the corresponding gene or only reverse primers of both genes and analyzed on a 1.5% agarose gel. Primers H-F, H-R, L-F and L-R are UnivOL1-F, Myc-R, UnivOL2-F, and FLAG-R, respectively (Table 1). (b) Antigen-specific enrichment of mRNA-displayed Fab fragments. A single round of affinity selection was carried out for antigen-immobilized and nonimmobilized beads from a mixture of equally amounted anti-p53 and anti-fluorescein Fab fragment genes. The amount of each gene before and after selection was quantified by quantitative PCR. The amount of the positive control gene divided by that of the negative control gene is plotted for both before selection (gray) and after selection by antigen-immobilized bead (black) and nonimmobilized beads (white). (c) Confirmation of H and L chain gene linking by cloning and sequencing after affinity selection. mRNA-displayed anti-fluorescein Fab fragments and mRNA-displayed anti-p53 Fab fragments were mixed in a ratio of 1 : 50 and then used for affinity selection against fluorescein-immobilized beads. After selection, a total of 13 clones were sequenced, and the correspondence of the H and L chains was confirmed.

p53 or fluorescein, and the amounts of the molecules before and after selection were quantified by quantitative PCR. As expected, each of the Fab fragment genes was enriched when they were selected against their antigens (Figure 3(b)). The anti-fluorescein Fab fragment gene showed an approximate 110-fold enrichment and the anti-p53 Fab fragment gene showed an approximate 5-fold enrichment. The difference in the enrichment efficiency is probably due to the difference in the dissociation constant; the anti-p53 Fab fragment has a weaker affinity towards its antigen. On the other hand, both antibodies did not show any enrichment when their antigens were not present during *in vitro* selection, showing that both mRNA-displayed Fab fragments bind to their antigens with specificity.

Finally, to confirm that genes do not crossover during emulsion PCR and the corresponding H chain and L chain genes are properly linked together, the DNA after affinity selection was cloned and confirmed by capillary DNA sequencing. The mRNA-displayed anti-fluorescein Fab fragment and the mRNA-displayed anti-p53 Fab fragment were mixed in a ratio of 1 : 50 and then used for affinity selection against fluorescein-immobilized beads. Out of the 13 sequenced DNA clones after selection, 11 clones were H and L chain-linked anti-fluorescein and 2 clones were H and L chain-linked anti-p53 ($P < 0.001$), resulting in a conclusion that the corresponding H chain and L chain genes are properly linked together and the genes do not crossover during emulsion PCR (Figure 3(c)).

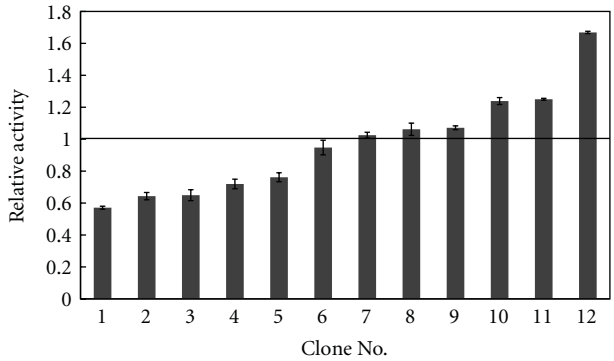
2.3. Affinity Selection from a Randomized Fab Fragment Library. Finally, we applied the mRNA display and emulsion PCR procedure for selection from a randomized Fab fragment library. We constructed an anti-p53 Fab fragment library with random point mutations by error-prone PCR and DNA shuffling from the wild type. When a fraction of the library was analyzed by DNA sequencing, the H chain and L chain had an average of 2.9 base/gene and 3.1 base/gene mutation, respectively. From this library, 4 rounds of affinity selection were performed under the condition of gradually decreasing amounts of immobilized antigen (round 1, 400 nM; round 2, 40 nM; round 3, 4 nM; round 4, 0.4 nM). After 4 rounds of selection, the total binding activity of *in-vitro*-translated products of the library at each round was analyzed by ELISA (Figure 4(a)). The binding activity gradually increased in successive rounds of selection,



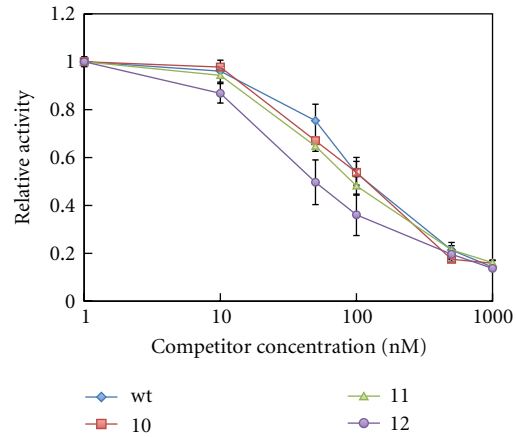
(a)

Clone no.	H chain			L chain	
1	**A98V**	H164N	S187R	L11S	F121S
2	K206R			Q17K	R46H
3	**A98V**	S187R		–	
4	**A98V**	S187R	K206Q	–	
5	A88T	S187R		–	
6	S161N	S187R		–	
7	S120P	S187R	K206R	–	
8	S187R			–	
9	K206R			–	
10	K206R			V85A	
11	S187R			**K53R**	
12	–			**S28N**	P143L

(b)



(c)



(d)

Figure 4: Affinity selection of randomly mutated anti-p53 Fab fragments. (a) Binding activity of the library after each round of selection. Random mutations were introduced to anti-p53 Fab fragments and 4 rounds of affinity selection were carried out under decreasing antigen concentration (round 1, 400 nM; round 2, 40 nM; round 3, 4 nM; round 4, 0.4 nM). After each round of selection, the fraction of the mutant Fab fragment library that binds to p53 was monitored by ELISA. Round 0 represents the randomly mutated initial library before selection. (b) Amino acid mutations of the selected variants. Bold type represents the mutations in the CDR and the minus (−) represents no amino acid mutations. (c) Binding activity of the selected variants after the 4th round of selection. The antigen binding activities of the selected variants were measured by ELISA. The absorbance at 450 nm of the wild type was used for normalization. (d) Competitive ELISA for estimating their affinities of variants with higher binding activity than the wild type. Competitive ELISA was preformed with 0–1000 nM free p53 for clone number 10–12 and the wild type (wt). The signals were normalized to that of 0 nM competitor concentration.

indicating that specific binders have been enriched in the library.

The library after the 4th round of selection was cloned and sequenced (Figure 4(b)). The binding activities of arbitrarily chosen 12 clones were analyzed and approximately 60% of those clones had similar or higher binding activities than the wild type (Figure 4(c)). The clones that had similar binding activities as the wild type (numbers 6–9) had mutations only in the constant region and since the constant region does not influence the affinity of most antibodies [3, 6] it is likely that these are neutral mutations. Among these mutations, S187R and K206R were found in several other clones and, these may be fixed neutral mutations in an early round of selection by random genetic drift. The clones that had higher binding activity than the wild type (numbers 10–12) all had mutations in the variable region of the L chain, either in the CDR or close to the CDR. Since the CDR directly contacts with the antigen, mutations in and near the CDR affect the affinity of the antibody [5, 12–14]. Analysis by competitive ELISA confirmed that all these clones specifically competed against free p53 (Figure 4(d)), and from the Scatchard plot, the K_d of clone 10, 11, 12 and wild type were estimated to be around 100 nM, 90 nM, 60 nM, and 140 nM, respectively.

These results demonstrated that the combination of mRNA display and emulsion PCR was able to eliminate inactive Fab fragments from the randomized Fab fragment library and select Fab fragment candidates according to the designed procedure. However, the affinities of the selected mutants were not so high. Only mutations in the L chain had a positive influence on the affinity, and mutations in the H chain were either neutral (S187R, K206R) or slightly harmful

(A98V). Mutants with A98V still have intermediate binding activity and the rather mild selection pressure of decreasing antigen concentration in this study may have let such a mutation survive through selection. Gradually lowering the antigen concentration has been demonstrated as a strategy for affinity maturation of some antibodies [15, 16] but introducing more stringent selection pressure such as off-rate selection [5, 12–14] may have produced a better result. Further, the difference in the cell-free translation system may have affected the result as well. In this study, a reconstituted PURE *E. coli* translation system [17] was used for mRNA display of antibody fragments for the first time, but the efficiency of the mRNA being linked to the protein (~10% of the total mRNA library; data not shown) is lower than our previous study based on a wheat germ translation system [5]. Optimizing the cell-free translation system for our mRNA display system should produce better results as well.

2.4. Future Perspectives. A big advantage of our method in this study may be the large library size of mRNA display, which is the largest among all display methods, and the use of next generation sequencers would be able to pull out the full potential of this method. The newest Roche 454 Sequencer can sequence approximately 10^6 reads of approximately 900 bp, long enough to cover both of the linked variable regions. Also, a recent study by phage display and deep sequencing has revealed that one round of selection is enough for identifying positive clones [18]. Although, 10^6 reads are not enough to cover the whole selected Fab fragment library, combining it with a microfluidic chip for high enrichment efficiency per round of selection [19] may make it possible to obtain unique high affinity binders. Further possibilities may be considered when the specifications of the next generation sequencers improve even more. The speed of improvement for this technology is remarkable, and when it becomes possible to sequence the whole selected Fab fragment library (around 10^8–10^9 molecules), it should allow selection of low affinity antibodies that would usually be lost in a typical selection of repetitious rounds, expanding the variety of potentially effective antibody candidates.

Other possible applications by our method described in this study would be proteome analyses, such as massively parallel detection of protein/protein interactions. Recently, Nirantar and Ghadessy demonstrated a way to identify various protein/protein interactions by library versus library two-hybrid screening using emulsion PCR [20]. A similar strategy can be carried out by our method by pulling-down *in-vitro*-translated mRNA-displayed protein complexes with an affinity tag and incorporating them into emulsions. Furthermore, the immense flexibility of our cell-free translation-based method would allow selection of not only protein/protein complexes, but also RNA/protein complexes. This concept can be completed by simply changing one side of the complex from an mRNA-displayed molecule to an ordinary RNA molecule, in other words merging SELEX [21], a common way for *in vitro* selection of RNA aptamers, with this mRNA display and emulsion PCR method. An example of RNA/protein complex selection would be selection of a ribonucleic peptide aptamer. In previous studies, it took two steps to make a high-affinity ribonucleic peptide aptamer against ATP, first step by SELEX [22] and second step by phage display [23]. Since our method can do selection against RNA and peptide at once, it may be possible to obtain a high-affinity binder in only one step. We have confirmed that emulsion RT-PCR can be carried out and incorporated into this method (data not shown), and selection of RNA/protein complexes by *in vitro* display methods shall be carried out in the near future.

In conclusion, other than *in vitro* selection of Fab fragments, our method in this study has the possibilities of being able to carry out proteomics applications and RNA/protein complex selection. This variety of possible applications shows the potential convenience of *in vitro* selection by mRNA display and gene-linking emulsion PCR.

3. Materials and Methods

3.1. DNA Construction. The oligonucleotide sequences used in this study are listed on Table 1. The H chain and L chain genes of both anti-fluorescein Fab and anti-p53 Fab fragments were constructed by PCR with KOD-plus Neo DNA polymerase (Toyobo) from plasmids including these Fab fragment genes [6] using primers Universal-OL1 and Myc-R or Universal-OL2 and FLAG-R, respectively. The PCR products were cloned into pCR2.1-TOPO vector (Invitrogen) and confirmed by an ABI PRISM 3100 genetic analyzer (Applied Biosystems). DNA templates of the H chain and L chain genes for mRNA display were prepared by PCR from these plasmids using primers Universal-OL1 and Myc25A-R, or Universal-OL2 and FLAG25A-R, respectively.

3.2. Library Construction. A randomized anti-p53 Fab fragment library was constructed by introducing point mutation to the wild type with the combination of error-prone PCR and StEP [24]. Mutazyme II DNA polymerase (Stratagene) was used for error-prone PCR and 15 ng of either the H chain gene or the L chain gene constructed above was amplified for 30 cycles according to the protocol. After error-prone PCR, StEP was performed using *Ex Taq* DNA polymerase (Takara) as follows: denaturation at 95°C for 2 min, 80 cycles of 95°C for 30 sec and 55°C for 5 sec, followed by 95°C for 30 sec, 60°C for 30 sec, and 72°C for 15 min. The PCR products were then resolved by agarose gel electrophoresis, extracted and purified with the QIAquick gel extraction kit (Qiagen). Error-prone PCR and StEP was performed once more each under the same condition. At every step, primers Universal-OL1 and Myc25A-R or Universal-OL2 and FLAG25A-R were used for the H chain and L chain genes, respectively.

3.3. In Vitro Transcription and Translation. The DNA templates were transcribed by T7 Ribomax Express Large Scale RNA production system (Promega) and purified by using the RNeasy mini kit (Qiagen). The transcribed mRNAs were then ligated with a puromycin linker by T4 RNA ligase (Takara) as described in previous studies [5]. Translation was done by PURE system S-S (PostGenome Institute) and

Table 1: Oligonucleotide sequences.

Name	Sequence (5′ to 3′)
Universal-OL1	GAAATTAATACGACTCACTATA**GGGAAACCACTACGGTCTCCC**TCTAGAAATAATTTTGTTTAACTTTAAGAAG GAGATATACCA
Myc-R	AAGGTCTTCTTCACTAATCAGTTTCTGCTC
Universal-OL2	GAAATTAATACGACTCACTATA**GGGAGACCGTAGTGGTTTCCC**TCTAGAAATAATTTTGTTTAACTTTAAGAAG GAGATATACCA
FLAG-R	CTTGTCGTCATCGTCCTTGTAGTC
Myc25A-R	TTTTTTTTTTTTTTTTTTTTTTTTTAAGGTCTTCTTCACTAATCAGTTTCTGCTC
FLAG25A-R	TTTTTTTTTTTTTTTTTTTTTTTTTCTTGTCGTCATCGTCCTTGTAGTC
PolyT oligo	TTTTTTTTTTTTTTTTTTTTTTTTT
UnivOL1-F	GAAATTAATACGACTCACTATAGGGAAACCACTACGGTC
UnivOL2-F	GAAATTAATACGACTCACTATAGGGAGACCGTAGTGGTT
Fluo-F	CAGGATCGAGTGGGTCAAAC
Fluo-R	GGTGCTGGAGCTTTTGTCTG
p53-F	GGCAGAGCTTGTAAGGTCAG
p53-R	CAATCCATCCAATCCACTCC

The bold type resembles the sequences of the overlapping region for overlap-extension PCR.
Poly T oligo is labeled with a photo-cleavable biotin at the 5′ end.

5 pmol each of H and L chain templates were translated separately in a 25 μL scale for 2 hours at 37°C to form mRNA-displayed molecules. The molecules were then diluted into 175 μL of Hybridization Buffer (1 M NaCl, 100 mM Tris-HCl, pH 7.4, 10 mM EDTA, and 0.25% Triton X-100) and mixed with 100 pmol of poly-dT oligonucleotide with a biotinylated photo-cleavable linker and a high-capacity NeutrAvidin agarose resin (Pierce). The resin mixture was mixed gently at 4°C for 1 hour and subsequently washed with PBST (PBS with 0.1% Tween 20) three times. The resin was then mixed with ReverTra Ace reverse transcriptase (Toyobo) and incubated at 42°C for 30 min. After washing with PBST three times, 40 μL of elution buffer (PBS with 10% Solution A of PURE system S-S) was added to the resin and exposed to UV radiation at >300 nm to elute the mRNA-displayed molecules as previously described [25]. The purified mRNA-displayed H and L chains were mixed and then incubated overnight at 4°C to form mRNA-displayed Fab fragments.

3.4. Gene-linking Emulsion PCR. Emulsions were prepared by stirring 50 μL of PCR reagents containing the selected mRNA-displayed Fab fragments, KOD-plus Neo DNA polymerase, and primers Myc-R and FLAG-R into 950 μL of mineral oil-surfactant mixture [mineral-oil (Nacalai Tesque) containing 0.45% Span 85 (Nacalai Tesque), 0.04% Tween 20 (Sigma), and 0.01% Triton X-100 (Nacalai Tesque)] at 2300 rpm for 30 sec at 4°C. The emulsions were dispensed into 0.2 ml PCR tubes 80-μL each and PCR was performed with a T1 thermocycler (Biometra). The PCR program was as follows: denaturation at 94°C for 2 min; 40–50 cycles of 98°C for 10 sec, 60°C for 30 sec, and 68°C for 1 min; final extension at 68°C for 4 min. A 70 μL aliquot of the top layer was collected from each PCR tube, moved to a 1.7 mL tube, mixed with 100 μL of ddH_2O, and centrifuged at 15,000 rpm for 10 min at 40°C to break the emulsions. Approximately 90% of the aqueous layer from the bottom was recovered and purified by briefly mixing it with 1 mL of mineral oil and centrifuging the mixture at 15,000 rpm for 2 min. From the purified aqueous layer, a 120 μL aliquot was recovered and purified by using the QIAquick PCR purification kit (Qiagen). Subsequently, the PCR products were resolved by agarose gel electrophoresis, extracted, and purified with the QIAquick gel extraction kit.

3.5. Affinity Selection. Biotinylated antigen (either fluorescein (Sigma) or p53 C-terminal peptide (SKKGQSYSRH)) or biotin was added to 100 μL of Magnotex-SA beads (Takara) dispersed in PBST, gently mixed at 4°C for 1 hour. After washing with PBST three times, 100 μL of blocking buffer (DIG wash and block buffer set; Roche) was added and gently mixed at 4°C for another hour and finally washed with PBST to prepare antigen-immobilized beads or mock beads, respectively. To these antigen-immobilized beads or mock beads 5 μL of sonicated salmon sperm DNA (Stratagene), 5 μL of yeast tRNA (Invitrogen) and 40 μL mRNA-displayed Fab fragments were added. After having been gently mixed for 1 hr at 4°C for binding, the beads were washed with PBST three times and heated at 70°C for 15 min to elute the bound mRNA-displayed Fab fragments. Subsequently, emulsion PCR was performed as described above to link the corresponding H and L chain genes together.

To regenerate the genes, PCR was performed with KOD-plus Neo DNA polymerase using primers UnivOL1-F and Myc-R or UnivOL2-F and FLAG-R for the H chain and L chain genes, respectively. To prepare the DNA templates for the next round of selection, the regenerated H chain and L chain genes were amplified by StEP using *Ex Taq* DNA polymerase and the same primers and program described above with exception that the cycle numbers ranged from 60–80 cycles. The PCR products were then resolved by agarose

gel electrophoresis, extracted and purified with the QIAquick gel extraction kit.

After several rounds of selection, the selected DNA was cloned using a TOPO XL PCR cloning kit (Invitrogen) and sequenced with the ABI PRISM 3100 genetic analyzer.

3.6. Quantitative PCR. The DNA amount of each gene was quantified by real-time PCR using SYBR premix *Ex Taq* DNA polymerase (Takara) and Lightcycler (Roche). Primers Fluo-F and Fluo-R were used for anti-fluorescein Fab fragment genes, and primers p53-F and p53-R were used for anti-p53 Fab fragment genes.

3.7. ELISA. Fab fragments were expressed by transcribing and translating the H chain and L chain gene DNA with the PURE system S-S. Meanwhile, a p53 C-terminal peptide-immobilized plate was prepared by adding 100 pmol of biotinylated p53 C-terminal peptide and 100 μL of TBST (TBS with 0.1% Tween 20) to a Streptavidin C8 transparent plate (Nunc) and shaking it for 1 hour at room temperature. The plate was then washed with TBST 10 times and blocked with 200 μL of blocking buffer by shaking it for 1 hour at room temperature. Separately, the expressed Fab fragments were diluted into 100 μL of TBST containing 0–1000 nM p53 C-terminal peptide as a competitor and preincubated at 4°C for 1 hour. Then, the samples were added to the fluorescein-immobilized plate and shaken for 5 min. After a washing step, 100 μl of TBST with 0.1% anti-FLAG M2 peroxidase conjugate (Sigma) was added and shaking was continued for 1 hour. The plate was washed for the last time and 100 μL of TMB (Nacalai Tesque) was added. The plate was shaken for 20 min and then 100 μL of 1 N H_2SO_4 was added to stop the reaction. The absorbance at 450 nm was measured (reference wavelength: 655 nm). The K_d values of the selected clones were estimated from the Scatchard plot.

Acknowledgments

This work was supported by a Grant-in-Aid for JSPS Fellows (23-2677), a Grant-in-Aid for Scientific Research (19360377, 22360351), and a Strategic Research Foundation Grant-Aided Project for Private Universities (S0801008) from the MEXT of Japan. T. Sumida is supported by a scholarship from Keio-Kougakukai.

References

[1] H. R. Hoogenboom, "Selecting and screening recombinant antibody libraries," *Nature Biotechnology*, vol. 23, no. 9, pp. 1105–1116, 2005.

[2] H. R. Hoogenboom, A. D. Griffiths, K. S. Johnson, D. J. Chiswell, P. Hudson, and G. Winter, "Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains," *Nucleic Acids Research*, vol. 19, no. 15, pp. 4133–4137, 1991.

[3] V. Quintero-Hernández, V. R. Juárez-González, M. Ortíz-León, R. Sánchez, L. D. Possani, and B. Becerril, "The change of the scFv into the Fab format improves the stability and in vivo toxin neutralization capacity of recombinant antibodies," *Molecular Immunology*, vol. 44, no. 6, pp. 1307–1315, 2007.

[4] J. Hanes and A. Plückthun, "In vitro selection and evolution of functional proteins by using ribosome display," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 10, pp. 4937–4942, 1997.

[5] I. Fukuda, K. Kojoh, N. Tabata et al., "In vitro evolution of single-chain antibodies using mRNA display," *Nucleic Acids Research*, vol. 34, no. 19, article e127, 2006.

[6] T. Sumida, N. Doi, and H. Yanagawa, "Bicistronic DNA display for in vitro selection of Fab fragments," *Nucleic Acids Research*, vol. 37, no. 22, article e147, 2009.

[7] D. S. Tawfik and A. D. Griffiths, "Man-made cell-like compartments for molecular evolution," *Nature Biotechnology*, vol. 16, no. 7, pp. 652–656, 1998.

[8] M. Nakano, J. Komatsu, S. I. Matsuura, K. Takashima, S. Katsura, and A. Mizuno, "Single-molecule PCR using water-in-oil emulsion," *Journal of Biotechnology*, vol. 102, no. 2, pp. 117–124, 2003.

[9] T. Kojima, Y. Takei, M. Ohtsuka, Y. Kawarasaki, T. Yamane, and H. Nakano, "PCR amplification from single DNA molecules on magnetic beads in emulsion: application for high-throughput screening of transcription factor targets," *Nucleic Acids Research*, vol. 33, no. 17, article e150, 2005.

[10] R. Williams, S. G. Peisajovich, O. J. Miller, S. Magdassi, D. S. Tawfik, and A. D. Griffiths, "Amplification of complex gene libraries by emulsion PCR," *Nature Methods*, vol. 3, no. 7, pp. 545–550, 2006.

[11] F. Diehl, M. Li, Y. He, K. W. Kinzler, B. Vogelstein, and D. Dressman, "BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions," *Nature Methods*, vol. 3, no. 7, pp. 551–559, 2006.

[12] L. Jermutus, A. Honegger, F. Schwesinger, J. Hanes, and A. Plückthun, "Tailoring in vitro evolution for protein affinity or stability," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 1, pp. 75–80, 2001.

[13] C. Zahnd, S. Spinelli, B. Luginbühl, P. Amstutz, C. Cambillau, and A. Plückthun, "Directed *in Vitro* evolution and crystallographic analysis of a peptide-binding single chain antibody fragment (scFv) with low picomolar affinity," *Journal of Biological Chemistry*, vol. 279, no. 18, pp. 18870–18877, 2004.

[14] B. Luginbühl, Z. Kanyo, R. M. Jones et al., "Directed evolution of an anti-prion protein scFv fragment to an affinity of 1 pM and its structural interpretation," *Journal of Molecular Biology*, vol. 363, no. 1, pp. 75–97, 2006.

[15] G. Thom, A. C. Cockroft, A. G. Buchanan et al., "Probing a protein-protein interaction by in vitro evolution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 20, pp. 7619–7624, 2006.

[16] H. Persson, H. Wallmark, A. Ljungars, J. Hallborn, and M. Ohlin, "In vitro evolution of an antibody fragment population to find high-affinity hapten binders," *Protein Engineering, Design and Selection*, vol. 21, no. 8, pp. 485–493, 2008.

[17] Y. Shimizu, T. Kanamori, and T. Ueda, "Protein synthesis by pure translation systems," *Methods*, vol. 36, no. 3, pp. 299–304, 2005.

[18] P. A. C. 'T Hoen, S. M. G. Jirka, B. R. Ten Broeke et al., "Phage display screening without repetitious selection rounds," *Analytical Biochemistry*, vol. 421, no. 2, pp. 622–631, 2012.

[19] N. Tabata, Y. Sakuma, Y. Honda et al., "Rapid antibody selection by mRNA display on a microfluidic chip," *Nucleic Acids Research*, vol. 37, no. 8, article e64, 2009.

[20] S. R. Nirantar and F. J. Ghadessy, "Compartmentalized linkage of genes encoding interacting protein pairs," *Proteomics*, vol. 11, no. 7, pp. 1335–1339, 2011.

[21] C. Tuerk and L. Gold, "Systemic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase," *Science*, vol. 249, no. 4968, pp. 505–510, 1990.

[22] T. Morii, M. Hagihara, S. I. Sato, and K. Makino, "In vitro selection of ATP-binding receptors using a ribonucleopeptide complex," *Journal of the American Chemical Society*, vol. 124, no. 17, pp. 4617–4622, 2002.

[23] S. I. Sato, M. Fukuda, M. Hagihara, Y. Tanabe, K. Ohkubo, and T. Morii, "Stepwise molding of a highly selective ribonucleopeptide receptor," *Journal of the American Chemical Society*, vol. 127, no. 1, pp. 30–31, 2005.

[24] H. Zhao and W. Zha, "In vitro 'sexual' evolution through the PCR-based staggered extension process (StEP)," *Nature Protocols*, vol. 1, no. 4, pp. 1865–1871, 2006.

[25] J. Olejnik, S. Sonar, E. Krzymañska-Olejnik, and K. J. Rothschild, "Photocleavable biotin derivatives: a versatile approach for the isolation of biomolecules," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 16, pp. 7590–7594, 1995.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 591025, 9 pages
doi:10.1155/2012/591025

Research Article

Inhibition of HIV Replication by Cyclic and Hairpin PNAs Targeting the HIV-1 TAR RNA Loop

Gregory Upert,[1] Audrey Di Giorgio,[1] Alok Upadhyay,[2] Dinesh Manvar,[2] Nootan Pandey,[2] Virendra N. Pandey,[2] and Nadia Patino[1]

[1] *Institut de Chimie de Nice, UMR CNRS 7272, Université de Nice-Sophia Antipolis, 28 Avenue de Valrose, F06100 Nice, France*
[2] *Department of Biochemistry and Molecular Biology, New Jersey Medical School, University of Medicine and Dentistry of New Jersey, 185 South Orange Avenue, Newark, NJ 07103, USA*

Correspondence should be addressed to Virendra N. Pandey, pandey@umdnj.edu and Nadia Patino, nadia.patino@unice.fr

Received 30 May 2012; Accepted 12 July 2012

Academic Editor: Eriks Rozners

Human immunodeficiency virus-1 (HIV-1) replication and gene expression entails specific interaction of the viral protein Tat with its transactivation responsive element (TAR), to form a highly stable stem-bulge-loop structure. Previously, we described triphenylphosphonium (TPP) cation-based vectors that efficiently deliver nucleotide analogs (PNAs) into the cytoplasm of cells. In particular, we showed that the TPP conjugate of a linear 16-mer PNA targeting the apical stem-loop region of TAR impedes Tat-mediated transactivation of the HIV-1 LTR *in vitro* and also in cell culture systems. In this communication, we conjugated TPP to cyclic and hairpin PNAs targeting the loop region of HIV-1 TAR and evaluated their antiviral efficacy in a cell culture system. We found that TPP-cyclic PNAs containing only 8 residues, showed higher antiviral potency compared to hairpin PNAs of 12 or 16 residues. We further noted that the TPP-conjugates of the 8-mer cyclic PNA as well as the 16-mer linear PNA displayed similar antiviral efficacy. However, cyclic PNAs were shown to be highly specific to their target sequences. This communication emphasizes on the importance of small constrained cyclic PNAs over both linear and hairpin structures for targeting biologically relevant RNA hairpins.

1. Introduction

The transcriptional transactivation of the HIV-1 genome requires a specific interaction between the highly conserved TAR RNA hairpin fragment with the viral Tat protein and cellular factors (PTEFb-cyclin T1-CDK9 kinase complex). Both the six-nucleotide loop and the three-nucleotide bulge of TAR RNA (Figure 1(a)) are involved in the formation of this complex [1–3]. Therefore, molecules that can bind to the bulge or the loop of TAR are of great therapeutic interest, since disruption of the ternary complex formation leads to abortive mRNA synthesis and, consequently, to inhibition of viral replication.

During the last decade, a wide number of TAR ligands have been described [4, 5]. Among them, one can cite R06 aptamers (such as $R06_{24}$ or $R06_{18}$, Figure 1(b)), which were identified initially by *in vitro* selection [6]. These aptamers are folded RNA stem-loop structures which recognize the mini-TAR fragment (Figure 1(a)) not only on the basis of sequence complementarity, as classical antisense oligomers, but also on the basis of the tertiary structure of their target. This leads to highly stable and specific loop-loop complexes, also called "kissing complexes." The key features for the establishment of such complexes are the hairpin structure of R06 aptamers as well as the octameric loop constituted by the 5′-UCCCAG-3′ sequence complementary to the TAR hexaloop, flanked by a G and a A residues. Although these two G/A residues are not directly involved in the loop-loop interaction, they were shown to be crucial for the formation of a stable kissing complex [7–10].

In a cellular compartment, RNA aptamers are rapidly degraded by nucleases, limiting their potential as therapeutic agents. Thus, several chemically-modified R06 derivatives were prepared with the view of improving both the pharmacological properties and TAR affinity. N3- > P5 phosphoramidate [11, 12], 2-O-methyl RNA [13, 14], and some

hexitol nucleic acids (HNA)/RNA mixmers [15] were shown to display an improved nuclease resistance while maintaining a similar TAR-binding constant. TAR-binding properties of R06 analogs containing LNA residues were also studied [10, 16–19]. While the fully modified LNA version of R06 proved to be a poor TAR ligand, some chimeric LNA/DNA, and LNA/2′-OMe RNA aptamers displayed binding properties of interest. However, the identification of such chimeric aptamers is laborious, because it requires a systematic screening of all possible combinations, as no rule dictates the number and positions at which LNA nucleotides have to be incorporated to allow a strong loop-loop interaction.

Concerning the biological activity of these aptamer analogs, although some of them were shown to inhibit specifically Tat-mediated transcription in cell-free assays [12, 13, 15, 20, 21] or in cell assays when transfected with cationic lipids [17], none of them was evaluated as anti-HIV agents. However, it was shown that, when expressed endogenously in HeLa cells, the RNA aptamer R06 was able to inhibit HIV replication [22], highlighting the antiviral potential of nuclease resistant molecules that recognize the TAR loop through both their primary sequence and their tertiary structure.

Based on these results, we previously devised small synthetic constrained structures derived from the R06 aptamer derivatives, and reported that they were able to interact with the TAR loop through "kissing-like" complexes of high affinity [23]. These structures are constituted by an octameric PNA (Figure 2) 5′-GTCCCAGA-3′ sequence identical to the one found in R06 aptamers, head-to-tail cyclized *via* polyamide linkers of different length (**1a–c**, Figure 2(b)). We chose to introduce PNAs as RNA mimics since they are highly stable in biological media and they hybridize strongly with complementary RNA sequences [24]. A limitation to their *in vivo* applications is their poor ability to cross cell membranes. Thus, a lysine residue was incorporated in cyclic PNAs to allow their subsequent conjugation with cell penetrating moieties [25–27].

Here, we report the synthesis of cyclic PNAs (**1–3**, Figures 2(b) and 3) and hairpin PNAs (**4–6**, Figure 3) conjugated to a triphenylphosphonium (TPP)-based cell penetrating vector (Figure 4(a)). This vector is constituted by a TPP lipophilic cation capable of transporting PNAs across the lipid bilayer [28, 29], bound *via* a disulfide bridge, to a mercaptoethoxycarbonyl moiety connected to the PNA. Intracellularly, the reduction of the disulfide bond leads to a spontaneous decomposition that releases the PNAs (Figure 4(b)) [30]. The cyclic PNAs exhibited potent anti HIV-1 activity in comparison to other derivatives, confirming the therapeutic potency of these conjugates.

2. Material and Methods

2.1. General Methods. Reagents and solvents were obtained from commercial sources and used without further purification unless indicated. Analytical HPLC chromatograms were obtained using an HP1100 UV detector set at 260 nm and a Beckman Ultrasphere RP-C18 (5 μm) column at 55°C. Purifications using semipreparative HPLC were done on the same instrument using a Phenomenex Jupiter column RP-C18 (5 μm). Elution solvent A was water (0.1% TFA); elution solvent B was acetonitrile (0.1% TFA). ESI mass spectra were recorded with a Bruker Esquire 3000 plus. Concentrations of cyclic PNAs, hairpin PNAs, and TPP conjugates were determined by UV spectroscopy, using the usual extinction coefficients [31]. The mini-TAR RNA fragment used for thermal denaturation studies was purchased at Dharmacon Inc. (Lafayette, USA). Thermal denaturations of mini-TAR/PNAs complexes were carried out on a Varian Cary 300 Scan spectrophotometer.

2.2. Chemistry

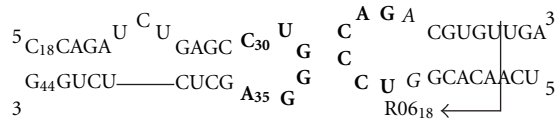
2.2.1. Synthesis of Cyclic PNA 3 and Hairpin PNAs 4–6. These PNAs were synthesized in Merrifield vessels on MBHA resin (100–200 mesh, 0.63 mmol/g, Merck Schuchardt OHG, Hohenbrunn, Germany), on a 100-μmol scale. Elongation was carried out starting from Boc/Z protected PNA monomers, using HBTU as the coupling reagent, and NMP as solvent. Compound 3 was synthesized as previously described for cyclic PNAs **1a–c** and **2** [23]. The lysine residue at the 5′-end of hairpin PNAs **4–6** was introduced after elongation, by means of Boc-Lys(2-Cl-Z)-OH and HBTU as the activator. Acetylation of the lysine residue was performed after Boc deprotection (TFA/TIS 10%, 4 mL for 15 min), using an Ac_2O/pyridine/NMP 1/1/8 v/v/v solution (2 × 4 mL for 15 min). Compounds **4–6** were deprotected and cleaved from the resin using a TFMSA/TFA/TIS solution (1 : 8 : 1) for 4 h, then precipitated in cold anhydrous diethyl ether. Crude products were isolated by centrifugation (3,000 min^{-1}, −4°C), washed twice with diethyl ether (10 mL), and purified by semipreparative HPLC using the following method: 55°C, A/B 100/0 for 7 min, then from 100/0 to 50/50 for 45 min, with a flow rate of 2 mL/min.

2.2.2. c(Lys-TCCCAG-Gln)$_{n=4}$ (PNA 3). HPLC (A/B 100/0 for 7 min, then from 100/0 at 50/50 for 45 min): *tR* = 17.5 min; MS (ESI, positive mode): (calculated for $C_{79}H_{108}N_{38}O_{22}$: 1941.9) *m*/*z* = 1943.3 $(M+H)^+$, 971.9 $(M+2H)^{2+}/2$, 648.5 $(M+3H)^{3+}/3$.

2.2.3. Ac-Lys-CGGTCCCAGACG-NH$_2$ (PNA 4). HPLC (A/B 100/0 for 7 min, then from 100/0 at 50/50 for 45 min): *tR* = 21.3 min; MS (ESI, positive mode): (calculated for $C_{136}H_{175}N_{73}O_{37}$: 3424.41) *m*/*z* = 1713.4 $(M+2H)^{2+}/2$, 1142.9 $(M+3H)^{3+}/3$, 857.5 $(M+4H)^{4+}/4$, 686.5 $(M+5H)^{5+}/5$.

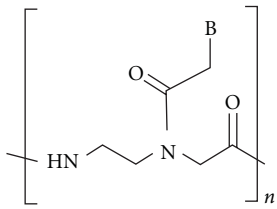
2.2.4. Ac-Lys-CGCGGTCCCAGACGCG-NH$_2$ (PNA 5). HPLC (A/B 100/0 for 7 min, then from 100/0 at 50/50 for 45 min): *tR* = 23.5 min; MS (ESI, positive mode): (calculated for $C_{178}H_{227}N_{97}O_{49}$: 4508.83), *m*/*z* = 2255.5 $(M+2H)^{2+}/2$ 1504.4 $(M+3H)^{3+}/3$, 1128.7 $(M+4H)^{4+}/4$, 903.3 $(M+5H)^{5+}/5$, 753.2 $(M+6H)^{6+}/6$.

2.2.5. Ac-Lys-CGTCCCAGCG-NH$_2$ (PNA 6). HPLC (A/B 100/0 for 7 min, then from 100/0 at 50/50 for 45 min):



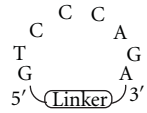
(a) (b)

Figure 1: Sequence and secondary structure of (a) HIV-1 mini-TAR RNA, (b) $R06_{24}$ and $R06_{18}$ aptamers reported in this study. Bold bases indicate complementarity between aptamer and TAR loops. The crucial G and A residues flanking the R06 aptamers loop are in italics.



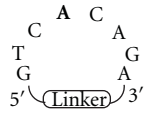
B = A, T, C, G

(a)



Linker: 3′-Gln-NH$(CH_2)_n$CO**Lys**-5′

n = 3: PNA **1a** ; n = 4: PNA **1b** ; n = 5: PNA **1c**



Linker: 3′-Gln-NH$(CH_2)_4$CO**Lys**-5′

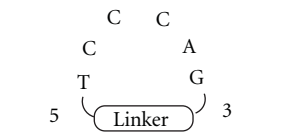
PNA **2**

(b)

PNA	1a	1b	1c	2	$R06_{18}$
Mini-TAR/PNA complex Tm (C)	42.4 ± 0.3	43.4 ± 0.4	41 ± 0.3	<10	36.5 ± 0.3

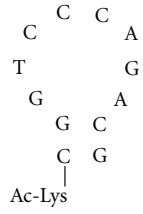
(c)

Figure 2: General structures of: (a) PNAs (b) cyclic PNAs **1a–c**, **2**; in bold, mismatched residue; (c) Melting temperatures of cyclic PNAs **1a–c**, **2**/TAR and $R06_{18}$ aptamer/TAR complexes, obtained from [23].

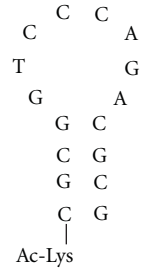


Linker: 3 -Gln-NH$(CH_2)_4$COLys-5

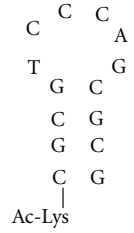
PNA **3**



PNA **4**

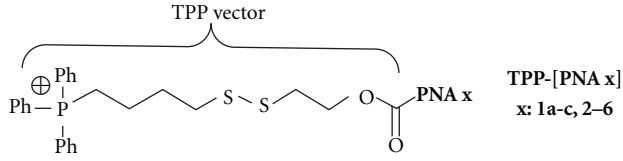


PNA **5**

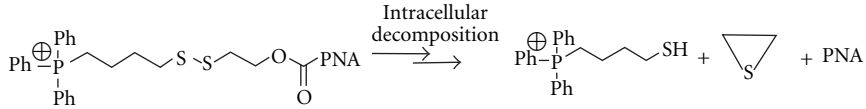


PNA **6**

Figure 3: General structures of cyclic and hairpin PNAs synthesized in the present study.



(a)



(b)

TPP-[PNA]$_{TAR}$: TPP-[5 -TCCCAGGCTCAGATCT-3]

(c)

Figure 4: (a) Structures of the TPP conjugates of PNAs **1a–c**, **2–6**. (b) Schematic representation of the intracellular degradation mode of TPP-PNA conjugates (from [28]). (c) Structure of the reference compound used in this study (16-mer antisense TPP-$[PNA]_{TAR}$).

tR = 19.9 min; MS (ESI, positive mode): (calculated for $C_{113}H_{148}N_{60}O_{32}$: 2858.8), *m/z* = 1430.7 $(M+2H)^{2+}/2$, 953.5 $(M+3H)^{3+}/3$, 715.5 $(M+4H)^{4+}/4$.

2.2.6. Synthesis of TPP-[PNA] Conjugates. An NMP solution containing TEA (10 eq.), NaN_3 (4 eq.) and compound 7 (4 eq.) was added at 0°C to pure PNAs **1a–c**, **2–6** (1 eq.) dissolved in a NMP/DMSO 1/1 v/v solution to obtain a 5 mM final concentration. The mixture was stirred for 30 min at 0°C, then for 12 h at room temperature. Cold diethyl ether (10-fold volume) was added to the mixture. Crude products were isolated by centrifugation (3,000 min^{-1}, −4°C) and washed twice with diethyl ether (10-fold volume). These products were further purified by a semipreparative HPLC method: A/B 100/0 for 7 min, then from 100/0 to 20/80 for 45 min, at 55°C, with a flow rate of 2 mL/min, to give, after lyophilization, TPP-[PNAs **1a–c**, **2–6**].

2.2.7. TPP-[PNA 1a]. HPLC (1 mL/min, A/B 100/0 for 7 min, then from 100/0 to 20/80 in 45 min): *tR* = 21.6 min; MS (ESI, positive mode): (calculated for $C_{126}H_{159}N_{51}O_{29}PS_2^+$: 2946.17) *m/z* = 1473.4 $(M+H)^{2+}/2$ 982.4 $(M+2H)^{3+}/3$, 736.9 $(M+3H)^{4+}/4$, 589.7 $(M+4H)^{5+}/5$.

2.2.8. TPP-[PNA 1b]. HPLC (1 mL/min, A/B 100/0 for 7 min, then from 100/0 to 20/80 in 45 min): *tR* = 23.8 min; MS (ESI, positive mode): (calculated for $C_{127}H_{161}N_{51}O_{29}PS_2^+$: 2960.19) *m/z* = 988.0 $(M+2H)^{3+}/3$, 741.3 $(M+3H)^{4+}/4$.

2.2.9. TPP-[PNA 1c]. HPLC (1 mL/min, A/B 100/0 for 7 min, then from 100/0 to 20/80 in 45 min): *tR* = 22.4 min; MS (ESI, positive mode): (calculated for $C_{128}H_{163}N_{51}O_{29}PS_2^+$: 2974.21) *m/z* = 992.6 $(M+2H)^{3+}/3$, 744.7 $(M+3H)^{4+}/4$, 596.1 $(M+4H)^{5+}/5$.

2.2.10. TPP-[PNA 2]. HPLC (1 mL/min, A/B from 100/0 to 20/80 in 30 min): *tR* = 13.2 min; MS (ESI, positive mode): (calculated for $C_{127}H_{160}N_{54}O_{28}PS_2^+$: 2985.20) *m/z* = 1505.4 $(M+Na)^{2+}/2$.

2.2.11. TPP-[PNA 3]. HPLC (1 mL/min, A/B from 100/0 to 20/80 in 30 min): *tR* = 12.4 min; MS (ESI, positive mode): (calculated for $C_{104}H_{134}N_{38}O_{24}PS_2^+$: 2395.52) *m/z* = 1209.23 $(M+Na)^{2+}/2$.

2.2.12. TPP-[PNA 4]. HPLC (1 mL/min, A/B from 100/0 to 20/80 in 30 min): *tR* = 12.7 min; MS (ESI, positive mode): (calculated for $C_{161}H_{202}N_{73}O_{39}PS_2^+$: 3878.52), *m/z* = 1939.5 $(M+H)^{2+}/2$, 1293.6 $(M+2H)^{3+}/3$, 970.9 $(M+3H)^{4+}/4$.

2.2.13. TPP-[PNA 5]. HPLC (1 mL/min, from A/B 100/0 to 20/80 in 30 min): *tR* = 12.8 min; MS (ESI, positive mode): (calculated for $C_{202}H_{253}N_{98}O_{51}PS_2^+$: 4962.95) *m/z* = 1654.3 $(M+2H)^{3+}/3$, 1241.5 $(M+3H)^{4+}/4$, 994.0 $(M+4H)^{5+}/5$.

2.2.14. TPP-[PNA 6]. HPLC (1 mL/min, from A/B 100/0 to 20/80 in 30 min): *tR* = 12.8 min; MS (ESI, positive mode): (calculated for $C_{138}H_{174}N_{60}O_{34}PS_2^+$: 3312.32) *m/z* = 1656.1 $(M+H)^{2+}/2$, 1105.2 $(M+2H)^{3+}/3$.

2.3. Thermal Denaturation Studies. One nmol of mini-TAR was solubilized in 250 μL (4 μM concentration) of *R* buffer solution at pH 7.3, that buffer containing cacodylate (20 mM), NaCl (20 mM), KCl (140 mM), and $MgCl_2$ (0.3 mM). The solution was heated at 90°C for 2 min, immediately cooled at 4°C, and maintained at this temperature for 10 min, then kept at 20°C. For preparing hairpin PNAs **4–6**, a solution of each compound in *R* buffer (4 μM) was heated for 3 min at 95°C, then cooled to 20°C with a rate of 0.5°C/min [31]. Individual compounds **3**, **4–6** and mini-TAR in *R* buffer (2 μM final concentration of each) were mixed, then incubated at 5°C for 1 h. Thermal denaturation was generated by increasing the temperature from 5°C, to 90°C at 0.4°C/min, then followed by UV absorption (260 nm). Melting temperatures (T_m) were determined as the maximum of the first derivative of the melting curves.

2.4. Transfection and Production of HIV-1 Virions. For production of highly infectious pseudotyped HIV-1 virions, 293T cells grown in complete Dulbecco's modified Eagle's medium (DMEM) were cotransfected with pHIV-1JR-CSF-lucenv(−) and pVSV-G, using a calcium phosphate transfection system (Invitrogen Carlsbad, CA, USA) [32, 33]. The culture supernatant was saved at 24, 48, and 72 h after transfection, then pooled and analyzed for p24 antigen using the ELISA p24 antigen kit (ZeptomMetrix, Buffalo, NY, USA). The pseudotyped HIV-1 virions were then isolated from the culture supernatant by filtration through a 0.45 μm pore size PVDF membrane (Millipore Bedford, MA, USA) and then by ultracentrifugation at 70,000 g for 45 min. The viral pellet was resuspended in complete Dulbecco's medium and stored at −80°C.

2.5. Anti-HIV-1 Activity in CEM Cells. CEM CD4+ lymphocytes 12D7 were grown in RPMI-1640 medium supplemented with 10% fetal calf serum and 4 mM L-glutamine at 37°C in 5% CO_2 containing humidified air [34]. Early- to mid-log-phase cells were harvested and washed with an equal volume of PBS without Ca^{2+} and Mg^{2+}. Approximately 10^6 cells suspended in 250 μL of RPMI-1640 medium were incubated with pseudovirions (equivalent to 25 ng of p24) by gentle rocking for 2 h in the presence of polybrene (10 μg/mL). The infected cells were washed with PBS and resuspended in 1 mL of complete RPMI medium in a 24-well plate containing increasing amounts of individual TPP-[PNAs **1a–c**, **2–6**] (0.5 μM–5 μM). After 48 h, the cells were harvested, washed with PBS, and lysed in 1 × passive lysis buffer (Promega) with gentle shaking on a rocker for 15 min at room temperature. The lysed cells were centrifuged at 15,000 rpm for 15 min and an aliquot of the clear lysate was added to 100 μL luciferase assay reporter (Promega) in 96-well plate Fluotrac 200 (Greiner Labortechnik, Germany). The luciferase activity was measured on a Packard Top Count Luminometer. The total light unit was normalized by total protein content in the cell lysate. Total protein

was quantified using the DC protein assay kit (Bio-Rad Laboratories, Hercules, CA, USA). Median dose effects (IC_{50}) for individual TPP-[PNAs **1a–c**, **2–6**] were determined using CalcuSyn software (Biosoft, Cambridge, UK) [35, 36].

3. Results and Discussion

Previously, we have shown that cyclic PNAs **1a–c** tightly bind to TAR (Figure 2(c)), with a higher affinity than that of a R06 aptamer ($R06_{18}$, Figure 1) and that they were highly specific for TAR despite the limited number of bases constituting them, since the introduction of a single mismatch in the PNA sequence was strongly deleterious for TAR binding. Indeed, compound **2** (Figure 2(b)), in which the C4 residue was replaced by an A4 residue, showed no affinity for TAR [23]. The first goal of the present study was to assess whether these PNA structures, which are cyclized in a covalent way, are more advantageous for targeting the TAR loop than hairpin structures in which the loop is not covalently closed. Although PNAs are among the best nucleic acid mimics, no PNA analogue of R06 aptamers has been reported so far. Thus, we have prepared hairpin PNAs (compounds **4** and **5**, Figure 3) containing the same octameric PNA sequence than in cyclic PNAs **1a–c**, closed by two and four G-C pairs, respectively, and measured their affinity for TAR. The second goal of this study was to determine whether, as for R06 aptamer derivatives, the G and A PNA residues flanking the loop sequence are necessary for the establishment of stable loop-loop complexes. For this purpose, we synthesized the G- and A-deleted cyclic PNA **3** and hairpin PNA **6** (Figure 3), and studied their interaction with TAR. Finally, in order to evaluate the ability of both cyclic (**1a–c** and **2-3**) and hairpin (**4–6**) PNAs to inhibit HIV replication in infected cells, we conjugated them to a cell-penetrating vector, *via* their lysine residue (Figure 4(a)). The vector chosen in this study is an intracytoplasmic biodegradable triphenylphosphonium (TPP)-based moiety, which was shown to allow the uptake and release of a "naked" PNA into cytoplasm (i.e., without any residual TPP moiety attached to PNAs, Figure 4(b)) [30]. For antiviral activity studies, a previously described TPP conjugate of a 16-mer PNA_{TAR} targeting the apical stem-loop of TAR was taken as a reference compound (Figure 4(c)) [30].

3.1. Chemistry. The synthesis of compounds **1a–c** and **2** was previously reported [23]. Cyclic PNA **3** was prepared following a solid-phase strategy *via* on-resin cyclization, using a glutamic acid-anchored MBHA resin, as for cyclic PNAs **1a–c** and **2**. Hairpin PNAs **4–6** were synthesized on a MBHA resin, using standard procedures. Briefly, the elongation was performed using Boc/Z-protected PNA monomers and HBTU as coupling reagent. The lysine residue at the 5′-end was introduced after elongation by means of Boc-Lys(2-Cl-Z)-OH and HBTU. After Boc cleavage under acidic conditions (TFA/TIS 10%), the α-amino group of the lysine residue was acetylated using an Ac_2O/pyridine/NMP solution. Hairpin PNAs **4–6** were obtained after deprotection and cleavage from the resin using a TFMSA/TFA/TIS solution and purification by semipreparative HPLC. Their structures were confirmed by ESI-MS experiments (see experimental protocols in Supplementry Material available online at doi:10.1155/2012/591025).

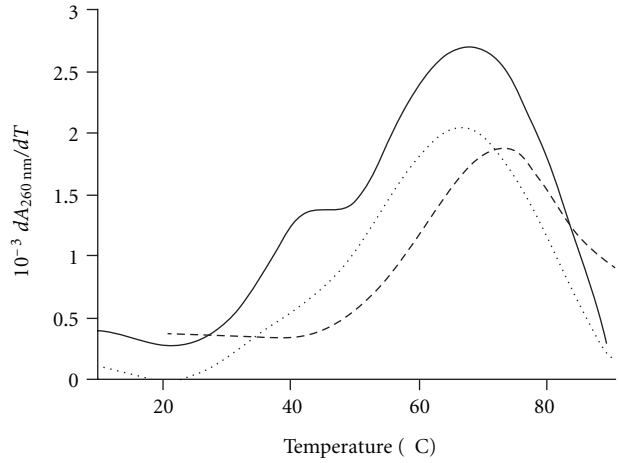


Figure 5: First derivative plot of melting curves followed by UV spectroscopy for hairpin PNA **4** alone (dashed line), TAR RNA alone (dotted line) and of hairpin PNA **4**/TAR complex (plain line).

The TPP-conjugates of cyclic and hairpin PNAs were obtained in almost quantitative yields from their corresponding cyclic **1a–c**, **2-3** and hairpin **4–6** precursors, in one step, using an excess of the key para-nitrophenyl carbonate reagent **7** in the presence of sodium azide, as previously described [28]. The TPP-conjugates were purified by semipreparative HPLC, with an RP-C18 column and water (0.1% TFA) and acetonitrile (0.1% TFA) as the elution solvents. Their structures were confirmed by ESI-MS experiments in which the corresponding spectra displayed characteristic $(M+nH)^{(n+1)+}/(n+1)$ peaks (n = 1 to 5) (see experimental protocols in Supplementry Material).

3.2. Thermal Denaturation Study. The affinity of compounds **3–6** for the mini-TAR RNA fragment was evaluated by thermal denaturation monitored by UV absorption spectroscopy (λ_{max} = 260 nm) in *R* buffer, as previously reported for cyclic PNAs **1a–c** and **2** and $R06_{18}$ aptamer [23]. Melting temperatures (T_m) of the TAR/cyclic PNA **3** and TAR/hairpin PNAs(**4–6**) complexes are summarized in Table 1, together with the melting temperatures of hairpin PNAs **4–6** alone (i.e., without TAR).

Thermal denaturation studies of hairpin PNAs **4** and **5** alone exhibited a single transition at, respectively, 73.0°C and 84.0°C (Table 1), independently of PNA concentration, indicating that they fold to form highly stable hairpins [31]. The difference between their T_m values (ΔT_m = 11°C) reflects the higher stability of the double strand in PNA **5** than in PNA **4**, due to the presence of two additional canonical CG pairs in PNA **5** relative to PNA **4**. The melting profiles obtained with mixtures of mini-TAR and hairpin PNAs **4** or **5** displayed two transitions (e.g., see Figure 5). The highest one forms a broad peak, resulting from the

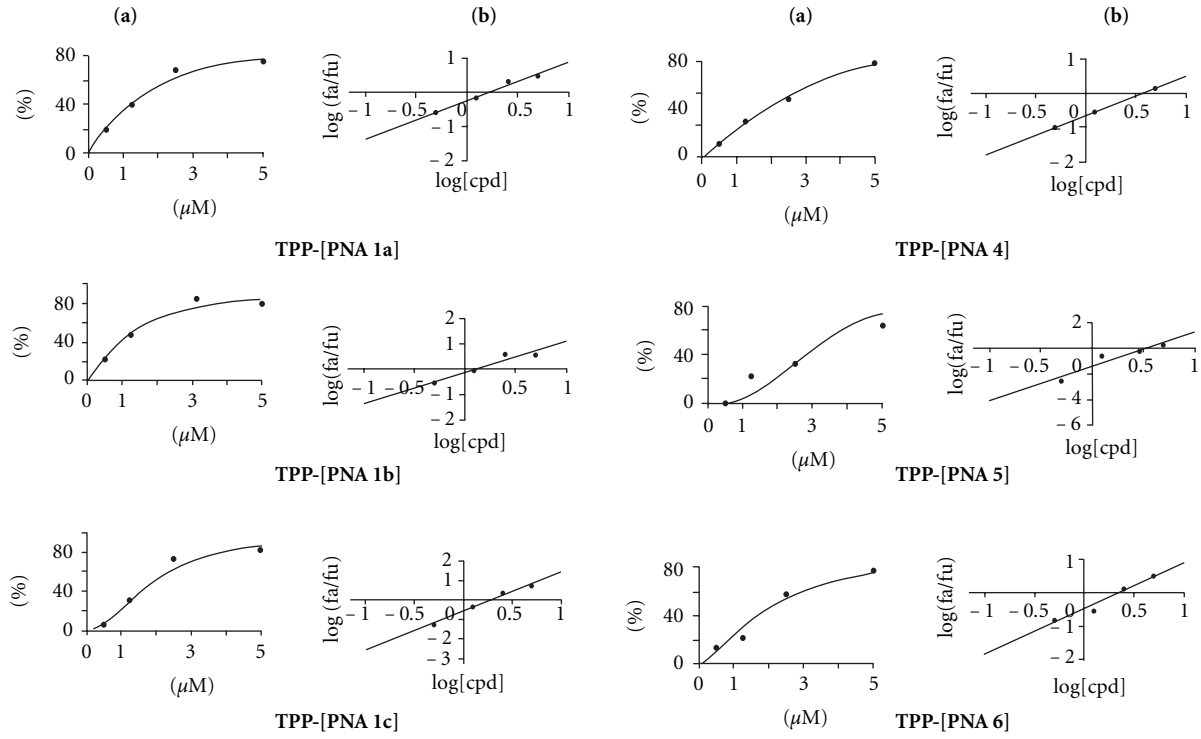


Figure 6: Dose-effect curves and median-effect plots for antiviral activity of TPP-conjugates. CEM cells were infected with VSV-G pseudotyped S1 strain of HIV-1 and grown in the presence of increasing concentrations of test conjugates. After 48 h, cells were harvested and assayed for luciferase expression using Luciferase assay kit (Promega). The effect of conjugate concentrations on the luciferase expression was analyzed using CalcuSyn software (Biosoft). The IC_{50} values, determined from the ratio of log of luciferase expression in treated cells (fa) and untreated cells (fu) as a function of the log concentration of conjugates was $1.0 > M$; the linear correlation coefficient, r, ranged from 0.943 to 0.985. (a) Dose-effect curves for antiviral activity of TPP conjugates. (b) Median-effect plot.

Table 1: Melting temperatures (T_m°C) of Mini-TAR/PNA complexes and of hairpin PNAs alone.

PNA	Mini-TAR/PNA complex T_m (°C)	Hairpin PNA alone T_m(°C)
3	<10	—
4	45.1 ± 0.6	73.0 ± 0.3
5	52.4 ± 0.2	84.0 ± 0.5
6	38.4 ± 0.3	29.8 ± 0.6

Experiments were performed in R buffer (pH 7.3) containing cacodylate (20 mM), NaCl (20 mM), KCl (140 mM), and MgCl2 (0.3 mM). Individual PNAs and mini-TAR (2 μM final concentration of each) were gently mixed and incubated at 5°C for 1 h. Thermal denaturation was achieved by increasing the temperature from 5°C to 90°C at the linear gradient of 0.4°C/min. The changes in UV absorption at 260 nm were monitored. Melting temperatures (T_m) were determined as the maximum of the first derivative of the melting curves.

overlapping of the two melting transitions concerning mini-TAR (T_m = 69.5°C) and hairpin PNAs **4** or **5**. The transition at the lowest temperature corresponds to the melting of the TAR/PNA **4** or **5** complex (45.1°C and 52.4°C, resp.). These values are significantly higher than those obtained with cyclic PNAs **1a–c**($\Delta T_m \geq +2$ and ≥ 10°C, resp.), emphasizing the higher stability of the complexes formed with hairpin PNA **4** and **5** over cyclic PNAs. Nevertheless, both TAR/hairpin PNAs and TAR/cyclic PNAs complexes are more stable than TAR/RNA $R06_{18}$ complex (T_m = 36.5°C) and than most of TAR/$R06_{24}$ analogues ones, analyzed in the same conditions (for NP-DNA [11], 2′-OMe RNA [13] and HNA/RNA [15] $R06_{24}$ derivatives: ($T_m \approx 30$°C; for the best LNA/DNA $R06_{24}$ derivative: ($T_m > 40$°C [16]).

Concerning the 5′-G and 3′-A deleted cyclic PNA **3**, the thermal denaturation study clearly demonstrates that it does not bind to TAR, highlighting the importance of the G and A flanking residues for the formation of a stable complex between TAR and cyclic PNAs. Conversely, the 5′-G and 3′-A deleted hairpin PNA **6** is able to smoothly interact with TAR. However, the melting temperature of the corresponding complex (38.4°C) is higher than the melting temperature of the hairpin PNA **6** itself (29.8°C). Thus, it is likely that the formation of the complex with TAR occurs, at least in part, on the unfolded form of **6**.

It has been earlier shown for TAR/TAR RNA aptamer complexes that the presence of G and A loop closing residues is a key structural determinant conferring a high stability both to the RNA aptamer alone and the TAR/RNA aptamer complex. Substituting the GA pair by the AU one (or GC, CA…) decreases the T_m of both the RNA aptamer and the TAR/RNA aptamer complex by 17°C and 14°, respectively [9]. Comparing PNA 5 and 6 shows that the presence of the GA pair also leads to a drastic increase in the stability

of both the PNA hairpin alone and the TAR/PNA complex (ΔT_m of 54° and 14°C, resp.). NMR [8] and molecular dynamics studies [9] have shown that the presence of these two residues increases the stability of both the aptamer and the kissing complex by increasing the stacking at the stem-loop junctions, *via* the stabilization of two hydrogen-bond base pairs located at these stem-loop junctions of the kissing complex: the intramolecular G-A pair itself, *via* hydrogen bonding of N1-N1 carbonyl-amino type, and the intermolecular A-U pair, *via* the classical Watson-Cricks network. By contrast, when the loop of the aptamer is closed by the classical AU pair, the very high tension in the loop causes the opening of both this AU intramolecular pair and of the intermolecular AU one, leading to a less stable kissing complex. It is possible that such events also occur in the case of PNA **4–6** hairpins and corresponding TAR complexes but for the moment, no proof supports this hypothesis.

It is also possible that larger loop size of PNA **4/5** with a nonhydrogen bonding GA pair may offer greater stability to stem region as compared to small loop size of PNA 6 with a hydrogen-bonding GT pair closing the loop. The smaller loop size may cause strain on the stability of the stem region.

3.3. HIV-1 Inhibition in Cell Culture by Anti-TAR PNA. In order to evaluate the *in vitro* antiviral efficacy, we incubated CEM CD4+ lymphocytes, infected with highly infectious VSV-G pseudotyped HIV-1 virions expressing the firefly luciferase reporter gene [30], with varying concentrations of individual TPP-[PNA **1a–c**, **2–6**]. Similar experiments were carried out with unconjugated PNAs **1a–c** and **4–6**. The TPP-conjugate of a 16-mer antisense $[PNA]_{TAR}$ targeting the apical stem-loop of TAR was taken as a reference compound [30] (Figure 4(c)). We have previously showed that this compound inhibited HIV replication in infected cells at a micromolar concentration. To measure the effect of TPP-PNA conjugates on HIV-1 production in CEM cells, we monitored the expression of the firefly luciferase reporter gene, cloned instead of the nef gene in the HIV-1 virions. After 48 h incubation followed by cells lysis, the extracts were normalized for total protein content and analyzed for quantitative levels of luciferase expression. An arbitrary value of 100 was assigned to the luciferase activity obtained in infected cells in the absence of compounds; values relative to this control value were given to the other samples. Median dose effects (IC_{50}) for individual compoundwere then determined using CalcuSyn software (Figure 6). As expected, no antiviral activity was detected for the unconjugated PNAs **1a–c** and **4–6**, probably as a consequence of their poor cellular permeation (data not shown). Median dose effects (IC_{50}) obtained for individual TPP-[PNA **1a–c**, **2–6**] are summarized in Table 2.

In all cases, except TPP-conjugates of PNAs **2** and **3**, a substantial decrease in HIV-1 replication was observed as the concentration of TPP-conjugate was increased, the IC_{50} values ranging from 1.24 to 3.70 μM (Table 2). The micromolar inhibitory effects measured for TPP conjugates of cyclic PNAs **1a–c** are very encouraging, because they are similar to those obtained with the heavier TPP-conjugate of the antisense 16-mer $[PNA]_{TAR}$ (Figure 4(c)). As previously noticed for the complexes stability (T_m from to 41°C to 43.4°C, Table 1), the length of the linker closing cyclic PNAs **1a–c** (n = 3, 4, 5, Figure 2(b)) has little influence on the antiviral activity of their TPP conjugates (IC_{50} from 1.24 to 1.94 μM). In addition, it appears that the antiviral activity of TPP-[PNA **1a–c**] conjugates is specific. Indeed, TPP-[PNA **2**] and TPP-[PNA **3**], which, respectively, derive from the mismatched cyclic PNA **2** and from the GA-deleted cyclic PNA **3**, have no effect on viral replication. These results, together with the fact that no interaction between mini-TAR and these two cyclic PNAs was detected, tend to further demonstrate that the inhibition of HIV replication by TPP-[PNAs **1a–c**] is related to the formation of cyclic PNAs **1a–c**/TAR complexes and thus, to the inhibition of the Tat/TAR/cellular factors complex formation.

Table 2: IC_{50} values for TPP-[PNA] conjugates.

TPP-[PNA] conjugates	IC_{50} (μM)	r[a]	m[b]
TPP-[PNA **1a**]	1.63	0.983	1.15
TPP-[PNA **1b**]	1.24	0.953	1.23
TPP-[PNA **1c**]	1.94	0.985	1.99
TPP-[PNA **2**]	—[c]	—	—
TPP-[PNA **3**]	—[c]	—	—
TPP-[PNA **4**]	3.30	0.943	2.65
TPP-[PNA **5**]	2.10	0.974	1.39
TPP-[PNA **6**]	3.70	0.990	1.18
$TPP\text{-}[PNA]_{TAR}$[d]	1.00	0.970	—

[a]Represents linear correlation coefficient of the median effect plot.
[b]Represents measurements of the sigmoidicity of the dose-effect curve.
[c]No inhibition up to 50 μM.
[d]Obtained from [28].

Concerning TPP-hairpin PNAs (**4–6**) conjugates, it can be noted that the antiviral activity increases with the stability of the TAR/hairpin PNAs complexes [(TPP-[PNA **6**] T_m = 38.4°C, IC_{50} = 3.70 μM < TPP-[PNA **4**] (T_m = 45.1°C, IC_{50} = 3.30 μM < TPP-[PNA **5**] T_m = 52.4°C, IC_{50} = 2.10 μM).] However, the complexes stability is not the only factor that explains biological activity, since cyclic PNAs **1a–c** conjugates (T_m values from 41°C to 43.4°C) are more effective for the inhibition of HIV replication in cells (IC_{50} 1.24–1.94 μM) than the corresponding hairpin PNAs **4-5** conjugates.

Altogether, these results emphasize the advantage of cyclic PNA structures over hairpin ones for inhibiting HIV-1, through the targeting of the TAR RNA loop.

4. Conclusion

We demonstrated that the small cyclic PNAs targeting the HIV-1 TAR RNA loop inhibit viral replication when conjugated to a cell penetrating vector, as efficiently as do higher molecular weight compounds, such as hairpin PNAs or an anti-TAR 16-mer PNA antisense targeting both the stem and loop of TAR. Furthermore, despite their short PNA sequence, they are highly specific for their RNA target, since the introduction of a single mismatch in the PNA sequence is detrimental both for TAR binding and HIV

inhibition. In addition, these results, combined with the high PNA stability in biological media, indicate that such cyclic compounds hold potential as new anti-HIV agents. On the other hand, these results emphasize the advantage of using small constrained cyclic structures over both linear antisense oligonucleotides and hairpin ones for targeting biologically relevant RNA hairpins.

Acknowledgments

This research was supported by grants from Agence Nationale de Recherches sur le SIDA (ANRS) (to N. Patino) and the Conseil Général des Alpes-Maritimes (to N. Patino) and from NIAID/NIH Grant AI042520-09 (to V. N. Pandey).

References

[1] Q. Zhou, D. Chen, E. Pierstorff, and K. Luo, "Transcription elongation factor P-TEFb mediates Tat activation of HIV-1 transcription at multiple stages," *EMBO Journal*, vol. 17, no. 13, pp. 3681–3691, 1998.

[2] S. Richter, H. Cao, and T. M. Rana, "Specific HIV-1 TAR RNA loop sequence and functional groups are required for human cyclin T1-Tat-TAR ternary complex formation," *Biochemistry*, vol. 41, no. 20, pp. 6391–6397, 2002.

[3] S. Richter, Y. H. Ping, and T. M. Rana, "TAR RNA loop: a scaffold for the assembly of a regulatory switch in HIV replication," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7928–7933, 2002.

[4] M. Yang, "Discoveries of Tat-TAR interaction inhibitors for HIV-1," *Current Drug Targets*, vol. 5, no. 4, pp. 433–444, 2005.

[5] S. N. Richter and G. Palù, "Inhibitors of HIV-1 Tat-mediated transactivation," *Current Medicinal Chemistry*, vol. 13, no. 11, pp. 1305–1315, 2006.

[6] F. Ducongé and J. J. Toulmé, "In vitro selection identifies key determinants for loop-loop interactions: RNA aptamers selective for the TAR RNA element of HIV-1," *RNA*, vol. 5, no. 12, pp. 1605–1614, 1999.

[7] F. Ducongé, C. Di Primo, and J. J. Toulmé, "Is a closing "GA pair" a rule for stable loop-loop RNA complexes?" *Journal of Biological Chemistry*, vol. 275, no. 28, pp. 21287–21294, 2000.

[8] H. Van Melckebeke, M. Devany, C. Di Primo et al., "Liquid-crystal NMR structure of HIV TAR RNA bound to its SELEX RNA aptamer reveals the origins of the high stability of the complex," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 27, pp. 9210–9215, 2008.

[9] F. Beaurain, C. Di Primo, J. J. Toulmé, and M. Laguerre, "Molecular dynamics reveals the stabilizing role of loop closing residues in kissing interactions: comparison between TAR-TAR* and TAR-aptamer," *Nucleic Acids Research*, vol. 31, no. 14, pp. 4275–4284, 2003.

[10] F. Darfeuille, S. Reigadas, J. B. Hansen, H. Orum, C. Di Primo, and J. J. Toulmé, "Aptamers targeted to an RNA hairpin show improved specificity compared to that of complementary oligonucleotides," *Biochemistry*, vol. 45, no. 39, pp. 12076–12082, 2006.

[11] F. Darfeuille, C. Cazenave, S. Gryaznov, F. Ducongé, C. Di Primo, and J. J. Toulmé, "RNA and N3′ → P5′ kissing aptamers targeted to the trans-activation responsive (TAR) RNA of the human immunodeficiency virus-1," *Nucleosides, Nucleotides and Nucleic Acids*, vol. 20, no. 4–7, pp. 441–449, 2001.

[12] F. Darfeuille, A. Arzumanov, S. Gryaznov, M. J. Gait, C. Di Primo, and J. J. Toulmé, "Loop-loop interaction of HIV-1 TAR RNA with N3′ → P5′ deoxyphosphoramidate aptamers inhibits in vitro Tat-mediated transcription," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 15, pp. 9709–9714, 2002.

[13] F. Darfeuille, A. Arzumanov, M. J. Gait, C. Di Primo, and J. J. Toulmé, "2′-O-methyl-RNA hairpins generate loop-loop complexes and selectively inhibit HIV-1 tat-mediated transcription," *Biochemistry*, vol. 41, no. 40, pp. 12186–12192, 2002.

[14] L. Zapata, K. Bathany, J. M. Schmitter, and S. Moreau, "Metal-assisted hybridization of oligonucleotides, evaluation of circular 2′-O-Me RNA as ligands for the TAR RNA target," *European Journal of Organic Chemistry*, no. 6, pp. 1022–1028, 2003.

[15] G. Kolb, S. Reigadas, C. Boiziau et al., "Hexitol nucleic acid-containing aptamers are efficient ligands of HIV-1 TAR RNA," *Biochemistry*, vol. 44, no. 8, pp. 2926–2933, 2005.

[16] F. Darfeuille, J. B. Hansen, H. Orum, C. Di Primo, and J. J. Toulmé, "LNA/DNA chimeric oligomers mimic RNA aptamers targeted to the TAR RNA element of HIV-1," *Nucleic Acids Research*, vol. 32, no. 10, pp. 3101–3107, 2004.

[17] C. Di Primo, I. Rudloff, S. Reigadas, A. A. Arzumanov, M. J. Gait, and J. J. Toulmé, "Systematic screening of LNA/2′-O-methyl chimeric derivatives of a TAR RNA aptamer," *FEBS Letters*, vol. 581, no. 4, pp. 771–774, 2007.

[18] I. Lebars, T. Richard, C. Di Primo, and J. J. Toulmé, "LNA derivatives of a kissing aptamer targeted to the trans-activating responsive RNA element of HIV-1," *Blood Cells, Molecules, and Diseases*, vol. 38, no. 3, pp. 204–209, 2007.

[19] I. Lebars, T. Richard, C. Di primo, and J. J. Toulmé, "NMR structure of a kissing complex formed between the TAR RNA element of HIV-1 and a LNA-modified aptamer," *Nucleic Acids Research*, vol. 35, no. 18, pp. 6103–6114, 2007.

[20] J. J. Toulmé, C. Di Primo, and S. Moreau, "Modulation of RNA function by oligonucleotides recognizing RNA structure," *Progress in Nucleic Acid Research and Molecular Biology*, vol. 69, pp. 1–46, 2001.

[21] J. J. Toulmé, F. Darfeuille, G. Kolb, S. Chabas, and C. Staedel, "Modulating viral gene expression by aptamers to RNA structures," *Biology of the Cell*, vol. 95, no. 3-4, pp. 229–238, 2003.

[22] G. Kolb, S. Reigadas, D. Castanotto et al., "Endogenous expression of an anti-TAR aptamer reduces HIV-1 replication," *RNA Biology*, vol. 3, no. 4, pp. 150–156, 2006.

[23] G. Upert, M. Mehiri, A. D. Giorgio, R. Condom, and N. Patino, "Solid-phase synthesis and thermal denaturation study of cyclic PNAs targeting the HIV-1 TAR RNA loop," *Bioorganic and Medicinal Chemistry Letters*, vol. 17, no. 21, pp. 6026–6030, 2007.

[24] P. E. Nielsen, "Gene targeting and expression modulation by peptide nucleic acids (PNA)," *Current Pharmaceutical Design*, vol. 16, no. 28, pp. 3118–3123, 2010.

[25] V. N. Pandey, A. Upadhyay, and B. Chaubey, "Prospects for antisense peptide nucleic acid (PNA) therapies for HIV," *Expert Opinion on Biological Therapy*, vol. 9, no. 8, pp. 975–989, 2009.

[26] R. Abes, A. A. Arzumanov, H. M. Moulton et al., "Cell-penetrating-peptide-based delivery of oligonucleotides: an overview," *Biochemical Society Transactions*, vol. 35, no. 4, pp. 775–779, 2007.

[27] Z. V. Zhilina, A. J. Ziemba, and S. W. Ebbinghaus, "Peptide nucleic acid conjugates: synthesis, properties and applications," *Current Topics in Medicinal Chemistry*, vol. 5, no. 12, pp. 1119–1131, 2005.

[28] A. Muratovska, R. N. Lightowlers, R. W. Taylor et al., "Targeting peptide nucleic acid (PNA) oligomers to mitochondria within cells by conjugation to lipophilic cations: implications for mitochondrial DNA replication, expression and disease," *Nucleic Acids Research*, vol. 29, no. 9, pp. 1852–1863, 2001.

[29] A. Filipovska, M. R. Eccles, R. A. J. Smith, and M. P. Murphy, "Delivery of antisense peptide nucleic acids (PNAs) to the cytosol by disulphide conjugation to a lipophilic cation," *FEBS Letters*, vol. 556, no. 1–3, pp. 180–186, 2004.

[30] M. Mehiri, G. Upert, S. Tripathi et al., "An efficient biodelivery system for antisense polyamide nucleic acid (PNA)," *Oligonucleotides*, vol. 18, no. 3, pp. 245–255, 2008.

[31] B. Armitage, D. Ly, T. Koch, H. Frydenlund, H. Ørum, and G. B. Schuster, "Hairpin-forming peptide nucleic acid oligomers," *Biochemistry*, vol. 37, no. 26, pp. 9417–9425, 1998.

[32] V. Planelles, F. Bachelerie, J. B. M. Jowett et al., "Fate of the human immunodeficiency virus type 1 provirus in infected cells: a role for vpr," *Journal of Virology*, vol. 69, no. 9, pp. 5883–5889, 1995.

[33] E. Riguet, S. Tripathi, B. Chaubey, J. Désiré, V. N. Pandey, and J. L. Décout, "A peptide nucleic acid-neamine conjugate that targets and cleaves HIV-1 TAR RNA inhibits viral replication," *Journal of Medicinal Chemistry*, vol. 47, no. 20, pp. 4806–4809, 2004.

[34] F. Kashanchi, R. Shibata, E. K. Ross, J. N. Brady, and M. A. Martin, "Second-site long terminal repeat (LTR) revertants of replication-defective human immunodeficiency virus: effects of revertant TATA box motifs on virus infectivity, LTR-directed expression, in vitro RNA synthesis, and binding of basal transcription factors TFIID and TFIIA," *Journal of Virology*, vol. 68, no. 5, pp. 3298–3307, 1994.

[35] T. C. Chou, "Relationships between inhibition constants and fractional inhibition in enzyme catalyzed reactions with different numbers of reactants, different reaction mechanisms, and different types and mechanisms of inhibition," *Molecular Pharmacology*, vol. 10, no. 2, pp. 235–247, 1974.

[36] T. C. Chou, "On the determination of availability of ligand binding sites in steady state systems," *Journal of Theoretical Biology*, vol. 65, no. 2, pp. 345–356, 1977.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 295719, 9 pages
doi:10.1155/2012/295719

Review Article

Practical Tips for Construction of Custom Peptide Libraries and Affinity Selection by Using Commercially Available Phage Display Cloning Systems

Keisuke Fukunaga and Masumi Taki

Bioscience and Technology Program, Department of Engineering Science, The Graduate School of Informatics and Engineering, The University of Electro-Communications (UEC), 7-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan

Correspondence should be addressed to Masumi Taki, taki@pc.uec.ac.jp

Received 4 June 2012; Revised 25 July 2012; Accepted 1 August 2012

Academic Editor: Hiroshi Murakami

Phage display technology is undoubtedly a powerful tool for affinity selection of target-specific peptide. Commercially available premade phage libraries allow us to take screening in the easiest way. On the other hand, construction of a custom phage library seems to be inaccessible, because several practical tips are absent in instructions. This paper focuses on what should be born in mind for beginners using commercially available cloning kits (Ph.D. with type 3 vector and T7Select systems for M13 and T7 phage, respectively). In the M13 system, Pro or a basic amino acid (especially, Arg) should be avoided at the N-terminus of peptide fused to gp3. In both systems, peptides containing odd number(s) of Cys should be designed with caution. Also, DNA sequencing of a constructed library before biopanning is highly recommended for finding unexpected bias.

1. Introduction

Phage display technology was born in 1985 when George Smith reported that foreign peptide could be displayed on the surface of filamentous bacteriophage [3]. Today, the phage display is a versatile tool for finding specific interactions between randomized library peptides/proteins on phage and target proteins, peptides, or other molecules. For example, it is applicable for generation of therapeutic peptides against cancer [4], microbe [5], novel functional protein [6], or fully humanized monoclonal antibody [7]. The advantages of the phage display technology over other selection methods are as follows. (1) Cost of a routine is cheap. (2) Time required for selection/amplification is fast. (3) Extreme care for handling, such as RNA isolation/selection, is not necessary. The phage is a DNA-containing virus that infects bacteria and makes many copies of the library within a very short time [8].

A phage that specifically binds a target can be selected from mixtures of billions of phages, propagated by *in vivo* amplification, and then subjected to additional rounds of affinity selection (Figure 1). This whole process is so-called "biopanning" [9]. After multiple rounds of the biopanning, enrichment of target-binding phage can be assessed by phage titering and enzyme-linked immunosorbent assay (ELISA). Finally, the peptide displayed on the phage can be analyzed by DNA sequencing.

1.1. Categorization of Phage Display Systems. Based on vector systems, the phage display systems can be categorized into two classes. One is a true phage vector system. The phage vector is often derived from genes encoding all phage proteins [10]. The library is to be cloned as a fusion with a component gene, which originally exists in the phage genome. Alternatively, some libraries are to be inserted in the same vector as an additional fusion gene encoding a displaying peptide and a phage protein [11].

Another is a phagemid vector system. The phagemid is a plasmid containing both a phage-derived replication origin and a plasmid-derived one [12]. A phage containing the

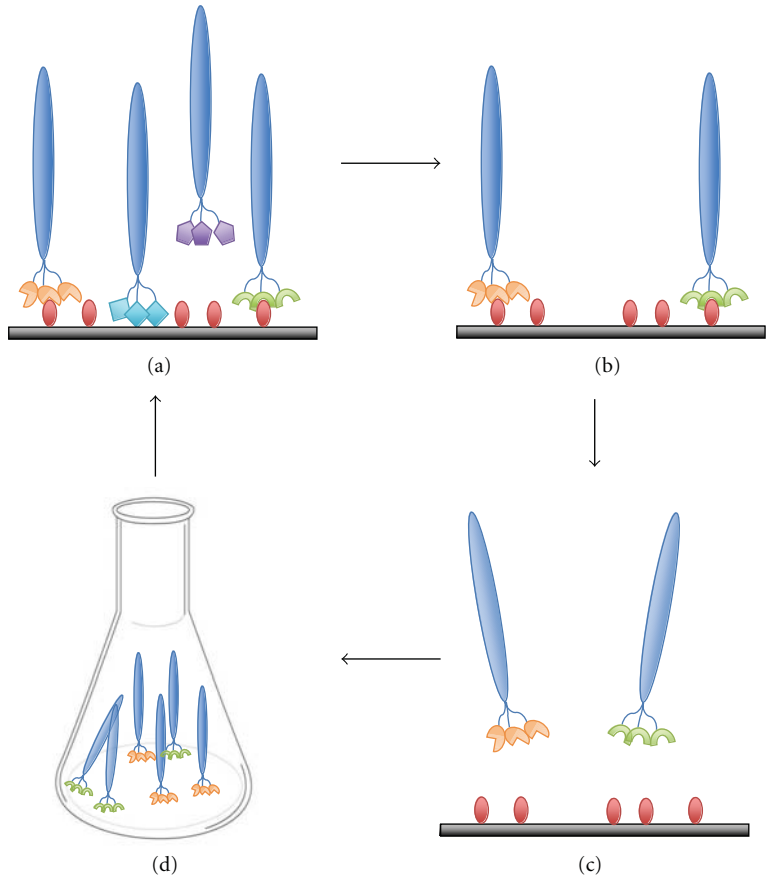


Figure 1: A typical procedure of the biopanning. (a) Incubation of phage library with an immobilized target. (b) Washing of unbound phage. (c) Elution of target-bound phage. (d) Amplification of the eluted phage for subsequent rounds of the biopanning.

phagemid can be generated only when phage components are secreted from bacterial host carrying a helper phage. In this system, two types of phages could be theoretically produced carrying either phagemid genome or helper-phage one. Practically, a helper phage with defective replication origin is used for the generation of phage proteins; production of the helper phage itself will be suppressed. This system yields a phage with the wild-type protein and library-fused one on the same virion, encoded by the helper phage and phagemid vector, respectively. Thus, numbers of the displaying peptides per virion from the phagemid system are less than those from the true vector system. This allows us to display not only small peptides but also large proteins [13], which is beyond the scope of this paper.

Among many different kinds of phages, M13 (filamentous bacteriophage) and T7 (lytic one) are exclusively used for the phage display. The M13 phage is composed of a circular single-stranded DNA genome and thousands copies of major capsid proteins (gp8) and capped by five copies of gp3 + gp6 on one end and five copies of gp7 + gp9 on the opposite (Figure 2). The most widely used M13 system is type 3. In this system, the peptide library is fused to the N-terminus of all five copies of the gp3. Other systems (e.g., type 33, type 8, etc.) are categorized by a peptide-displaying protein on the M13 phage and numbers of peptides per virion (Table 1) [14, 15].

The T7 phage is an icosahedral-shaped phage with a capsid shell that is composed of 415 copies of gp10, linear double-stranded DNA, and other proteins (Figure 2) [16]. The gp10 is made in two forms, gp10A (344 amino acids, aa) and its frameshifting product, gp10B (397 aa) [17]. In the T7 phage display systems, peptide library is always fused to the C-terminus of the gp10B. Numbers of peptides per virion and maximal size of the peptide are determined by the vector system (Table 1) [18].

1.2. Using Premade Phage Libraries. For screening, using a pre-made phage library is the most convenient way. Three types of M13 phage libraries, consisting of random linear/cyclic heptapeptides (Ph.D.-7/Ph.D.-C7C) and linear dodecapeptides (Ph.D.-12), are commercially distributable from New England Biolabs Inc. (NEB). In the C7C system, the randomized peptide is flanked by a pair of Cys, which are oxidized during the phage assembly to form an intramolecular disulfide bond. Several companies have constructed in-house pre-made peptide libraries; they provide screening services by using their phage libraries, instead of distributing

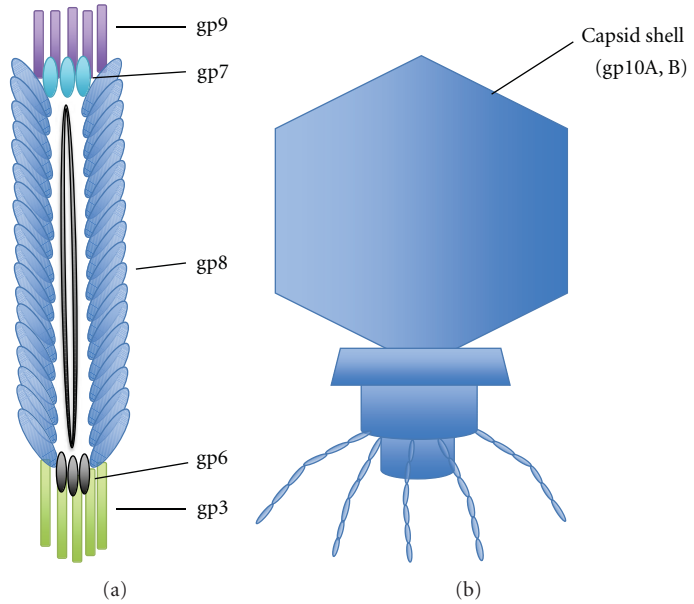


Figure 2: Structures of (a) a filamentous M13 bacteriophage and (b) a lytic T7 bacteriophage.

Table 1: Features of various systems of M13 and T7 phages.

	System	Size limit	Numbers of peptides per virion	Presentation region
M13	3	Unknown	5	N-terminus to gp3
	33	No limit	<1	
	3 + 3			
	8	Short	>2,700	N-terminus to gp8
	88	Unknown	<300	
	8 + 8	No limit	100–1000	
	8 + 8	Unknown		C-terminus to gp8
	6 + 6	No limit	<1	N-terminus to gp6
	6 + 6			C-terminus to gp6
	9 + 9			N-terminus to gp9
T7	T7Select1-1	1200aa	<1	C-terminus to gp10B
	T7Select1-2	900aa		
	T7Select10-3	1200aa	5–15	
	T7Select415-1	50aa	415	

Table 2: Consignment services of phage display with in-house libraries.

Company name	Peptide design	Peptide structure
Creative Biolabs	X_{10}, X_{16}, or X_{20}*1	linear
Dyax	$X_aCX_bCX_c$*1	cyclic
Bicycle Therapeutics	$X_aCX_bCX_cCX_d$*2	cyclic with a non-natural linker

X stands for any randomized amino acid.
*1The library was built by varying 19 aa at the randomized positions; the codon encoding Cys is excluded.
*2Bicyclic peptide library was made via thioether linkages [1].

ones. The chemical structures and features of the libraries are summarized in Table 2. Creative Biolabs Inc even accepts a service contract from a commercial pre-made library (e.g., Ph.D.-C7C system), a custom-constructed one in the company, or a hand-made one.

1.3. Construction of Custom Phage Library. Because of the limited kinds of resources, constructions of custom phage libraries are often performed by using kits available from NEB (Ph.D. Cloning System for M13 phage) or Merck Millipore (T7Select Cloning Kit for T7 one) [8]. Although

these instructions are well described, several practical tips are missing in both of them, which may lead beginners to pitfalls such as obtaining severe inherent bias of amino acid sequence in the randomized region. This paper focuses on instant tips for the construction of peptide libraries and affinity selection by using the commercial resources.

2. Ph.D. Cloning System

Ph.D. cloning system is based on a type 3 vector of M13 phage encoding N-terminal library peptide fused to a minor coat protein, gp3 [19]. Because gp3 plays a critical role for phage infection and randomized peptides are fused in all five copies of the gp3, infectivity of the M13 phage can be significantly affected by a sequence of the displaying peptide. Moreover, secretion of the M13 phage from *E. coli* closely depends on charges, hydrophilicity, and folding states of the displaying peptide [20, 21]. An amplification efficiency of the individual M13 phage clone is determined by a combination of the above infection and secretion rates. To avoid negative effects on the infection/secretion, one should be aware of the following in an insert DNA construction.

2.1. Signal Peptidase Cleavage. Positively charged basic amino acids, Lys and Arg, near the signal peptidase cleavage site inhibit the secretion of phages [22]; the cationic residue blocks translocation across the inner membrane of *E. coli* [23]. If the N-terminus of the displaying peptide should be positively charged, Lys has to be evidently chosen; 6 out of 99 arbitrarily chosen clones of the commercial 12 mer library (Ph.D.-12) contained Lys at the terminus, whereas N-terminal Arg was never found in the same 99 clones [24]. If the N-terminal Arg is inevitable, using noncommercial prlA suppressor strains such as ARI180 or ARI182 may help to avoid the secY-dependent secretion problem [22].

Pro at the terminus is also cumbersome. When a Pro is located next to the cleavage site, it inhibits the signal peptidase cleavage [25, 26]. Only one N-terminal Pro out of the 99 clones was found in the Ph.D.-12 library [24].

If it is necessary to encode a specific amino acid sequence just after the signal peptidase cleavage site, prediction of the position-specific cleavage is recommended to avoid risks of inappropriate or insufficient cleavage. For example, an Internet server, SignalP [27], instantly does this, and we usually use 0.3 for the threshold D-cutoff value in the gram-negative bacteria mode.

If one does not have any favorites of particular N-terminal sequence just after the cleavage site, "Ala-Glu" or simple "Ala" should be the first choice. There is an overabundance of negatively charged amino acids (Glu and Asp) at +1 and +2 and Ala at +1, in gram-negative signal peptidase cleavage sites (Figure 3) [24].

2.2. Unpaired Cys in a Displaying Peptide. If one generates a custom phage library displaying a disulfide-constrained peptide, an insert DNA encoding even number(s) of Cys, but not odd number(s), should be designed. This is because an intramolecular disulfide (S-S) bond could be formed between an unpaired Cys in a displaying peptide and an intrinsic Cys in the gp3 [28]. Phage assembly, infection, and/or secretion could be prevented by this unfavorable disulfide bond [24, 29]. It has been stated that an almost complete absence of odd number(s) of Cys was observed in the displaying peptide [28, 30], which is also identical to our experience. For example, when we sequenced 10 independent M13 phage clones encoding Cys-X_7-Cys where the X stands for any randomized amino acid, no Cys was observed in the X_7 region; only the designated Cys at both ends seemed to form an intramolecular disulfide bond (unpublished results). Given the difficulty, if one still tries to generate a phage library containing odd number(s) of Cys, M13 phages constructed by disulfide-free gp3 [1, 31] might be useful without using the Ph.D. system.

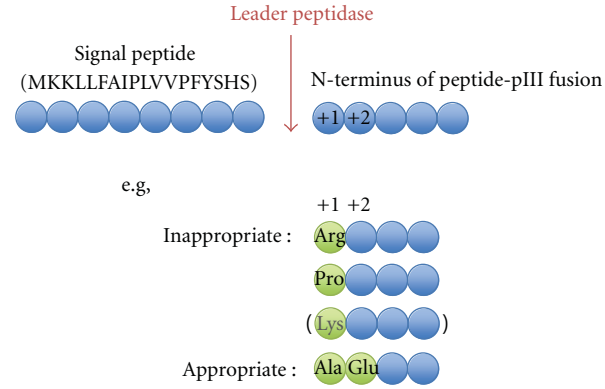


Figure 3: Sequence preference of the N-terminus of a peptide-pIII fusion in the M13 system.

3. T7Select Cloning System

Unlike the filamentous M13 system, T7 capsid shell displaying peptide library is not involved in phage infection and/or secretion. Indeed, it has been proven that libraries of the T7 phages exhibit less sequence bias than those of the M13 ones [29]. This is a great advantage for library construction, because it is less necessary to pay attention to the amino acid sequences described above. The T7 system is also good at displaying a rigid motif with a hydrophobic domain, namely, Trp cage [32]. This peptide motif was never displayed on the M13 system, presumably because the hydrophobic domain was anchored to the inner membrane of the *E. coli* prior to the phage assembly [32].

3.1. Codon Usage. To the best of our knowledge, there is no description of a relationship between codon usage and bias against translation for the T7 system in *E. coli*; in the M13KE system, it is reported that rare codons of *E. coli* seldom affect the bias of peptide libraries [24]. To avoid potential risks that minor codons could stress the translation system [33, 34], we simply use major codons (Table 3) for a nonrandomized region of a synthetic DNA insert.

Table 3: Codon usage in *E. coli* K-12 strain.

Amino acid	Codon	Codon frequency (%)
Phe	UUU	1.97
	UUC	1.50
Leu	UUA	1.52
	UUG	1.19
	CUU	1.19
	CUC	1.05
	CUA	0.53
	CUG	4.69
Ile	AUU	3.05
	AUC	1.82
	AUA	**0.37**
Met	AUG	2.48
Val	GUU	1.68
	GUC	1.17
	GUA	1.15
	GUG	2.64
Ser	UCU	0.57
	UCC	0.55
	UCA	0.78
	UCG	0.80
	AGU	0.72
	AGC	1.66
Pro	CCU	0.84
	CCC	0.64
	CCA	0.66
	CCG	2.67
Thr	ACU	0.80
	ACC	2.28
	ACA	0.64
	ACG	1.15
Ala	GCU	1.07
	GCC	3.16
	GCA	2.11
	GCG	3.85
Tyr	UAU	1.68
	UAC	1.46
His	CAU	1.58
	CAC	1.31
Gln	CAA	1.21
	CAG	2.77
Asn	AAU	2.19
	AAC	2.44
Lys	AAA	3.32
	AAG	1.21
Asp	GAU	3.79
	GAC	2.05
Glu	GAA	4.37
	GAG	1.84
Cys	UGU	0.59
	UGC	0.80
Stop	UAA	0.18
	UAG	0.00
	UGA	0.10
Trp	UGG	1.07
Arg	CGU	2.11
	CGC	2.60
	CGA	**0.43**
	CGG	**0.41**
	AGA	**0.14**
	AGG	**0.16**
Gly	GGU	2.13
	GGC	3.34
	GGA	0.92
	GGG	0.86

Codon frequency (%) is defined as the percent frequency of each codon which matches in whole open-reading frame of the *E. coli* K-12 genome. Minor codons (bold letters; below 0.5%) could be avoided for insert DNA construction. This table was cited from codon usage database (http://www.kazusa.or.jp/codon/) with some modifications.

3.2. Unpaired Cys in a Displaying Peptide. In our experiment, when a T7Select415-1b vector was used for the T7 packaging, the T7 phage failed to display a designated unpaired Cys (unpublished results). In this case, the library insert DNA was constructed using the genetic code of $(NNK)_6$-TGC-$(NNK)_6$, which encodes X_6-Cys-X_6. DNA sequencing of 8 independent phage clones revealed that peptides were truncated by the appearance of a TAG stop codon before the designated Cys that was supposed to be translated (Figure 4(a)).

The capsid shell used for randomized peptide display is composed of 415 copies of gp10 [35]. A structural study of T7 procapsid shell suggested that the gp10 might play an important role in the interaction between capsid shell and scaffolding proteins [36]. The designated Cys in the library peptide fused to the gp10 might form an intermolecular disulfide bond with the same kind of unpaired Cys in a neighboring library peptide. It also might form an intramolecular one with an intrinsic Cys in the gp10. Too many unpaired Cys may inhibit proper/efficient assembly of the capsid shell proteins. Although we do not have direct evidence for this hypothesis, Rosenberg et al. also speculated that some peptide sequences might be unfavorable for the T7Select415 system [18].

3.3. Paired Cys in a Displaying Peptide. Phages displaying the cyclic peptide by an intramolecular disulfide bond tend to exhibit higher target-binding ability, because their rigid structures minimize conformational entropy loss associated with the binding [37, 38]. Therefore, this kind of phage library is dominantly used for screening on the basis of not only M13 systems (e.g., Ph.D.-C7C library from NEB

Library design:	...DPNSGGSH **XXXXXX** C **XXXXXX**
Isolated clones:	...DPNSGGSH DDSL (stop)
	...DPNSGGSH GG (stop)
	...DPNSGGSH G (stop)
	...DPNSGGSH VS (stop)
	...DPNSGGSH KAEW (stop)
	...DPNSGGSH KTAL (stop)
	...DPNSGGSH MRSCR (stop)
	...DPNSGGSH AR (stop)

(a)

Library design:	AC **XXXX** HHHHHH **XXXX**C
Isolated clones:	AC FSRS HHHHHH HLLYC (**7**/12)
	AC TYGS HH**R**HHH PIFPC (1/12)
	AC LSQE HH**R**HHH SNFHC (1/12)
	AC HPFP HHHHHH NCHTC (1/12)
	AC NVCS HHHHHH LWCLC (1/12)
	AC RSPG HHHHHH RMIMC (1/12)

(b)

Figure 4: Unexpectedly isolated clones with high bias after library constructions. Bold "X" indicates any amino acids. (a) Stop codon appearance before the designated Cys. A combination of T7Select 415 vector and *E. coli* BL21 strain was used for *in vitro* packaging of the T7 phage, and individual clones were subjected to DNA sequencing. (b) Enrichment of a specific sequence and a mutation of the designated His (underlined). Randomly selected 12 individual clones of the M13 phage library were subjected to DNA sequencing. Parentheses indicate numbers of obtained clones.

[37, 39]) but also T7 ones [40, 41]. Disulfide constrained library of the T7 phage is most frequently constructed by using T7Select10-3b [29, 42] or 415-1b vector [2, 43–45] (Table 2).

For generation of the disulfide constrained (S-S) library using the T7Select415 system, it is recommended in the manual (Merck Millipore) to use *E. coli* Origami B or Rosetta-gami B strains, which tends to enhance disulfide bond formation in the cytoplasm. However, these strains may not be required for the library constructions. By using *E. coli* BLT5615 strain included in the T7 kit with the T7Select10-3b [29] or 415-1b [2] vector, the constrained library peptides were successfully displayed on the T7 phage, and high-affinity cyclic peptides were obtained.

3.4. Features of the T7 System. One of the features of the T7 phage, which grows much faster than the M13 one, is that it decreases the time for phage titering and amplification. After infection, clear plaques of T7 phages will usually appear within 2-3 hours on LB plate with no additives. Liquid amplification of the T7 phage after affinity selection can also be conducted within the same time.

It is also attractive for beginners that the T7 system does not require any special instruments like an electroporator for the library construction. Contrary to the kit instructions, ultracentrifugation of the T7 phage with CsCl is not necessary for all purification processes of ELISA assay and DNA sequencing. General procedure using polyethylene glycol (PEG)/NaCl with a conventional rotator is enough for the T7 phage purification, in the same way as the M13 system.

The T7 system can be useful for direct recovery of the highest-affinity phage with a very slow off-rate from a target-linked solid support. It has been reported that a target-bound lambda phage can be directly amplified by the addition of *E. coli* in midlog phase [46]. In a similar way, a library peptide displayed on the capsid shell does not interfere with the infectivity of the T7 phage. Indeed, we have experienced that a streptavidin-binding peptide containing the consensus sequence (HPQ [47]) was successfully obtained by this direct method (unpublished results). In the M13 system, phages may also be eluted by the addition of the host bacterial cells; however the elution of the highest-affinity binders may be hindered.

A minor drawback of the T7 system is that it is relatively expensive to construct a library with a high diversity. In a typical case, six whole tubes of T7 packaging extracts in a T7Select packaging kit (ca. $410) are required to obtain a diversity of 4.1×10^8 pfu [2].

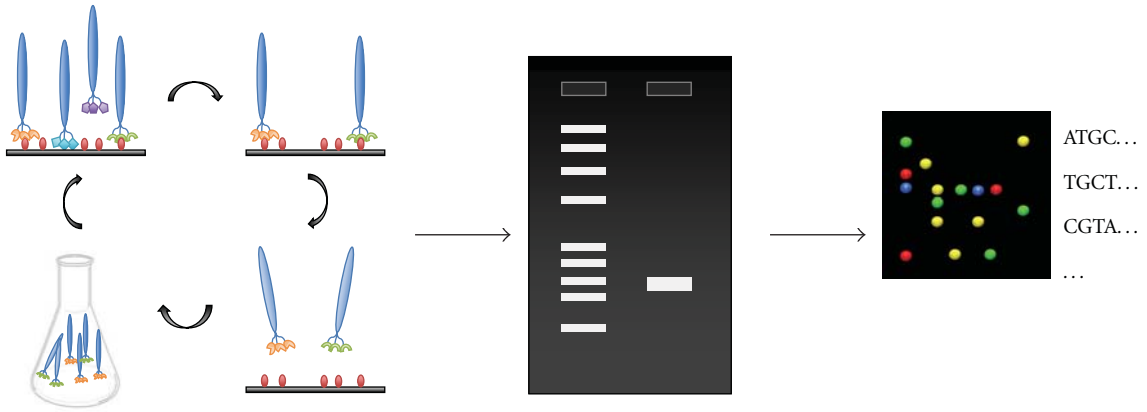
3.5. Handling Precautions. It should be emphasized that *in vitro* packaging has to be performed with extreme care. One must keep a stringent condition of the temperature and mixing. Only "fresh" T7 packaging extract will make a high quality library; freezing and thawing of the extract will result in apparent reduction of the packaging efficiency.

Diluted T7 phages with a buffer or water tend not to be infective. It should be diluted with a buffer containing a protectant such as gelatin or a growth media such as TB or LB.

4. Importance of DNA Sequencing for Finding Unfavorable Bias and False Positives at an Early Stage of the Affinity Selection

After the electroporation for the Ph.D. system or the packaging for the T7Select system, a qualitative assessment of the phage library should be performed by DNA sequencing prior to the biopanning. We always confirm it by a conventional DNA sequencer with at least 10 independent phage clones. For example, we obtained highly biased sequences when the random library encodes constraints with a His_6-tag (Ala-Cys-X_4-His_6-X_4-Cys) (unpublished results; Figure 4(b)). In this case, a specific sequence was predominantly enriched (7 out of 12 arbitrarily chosen clones). In addition, one of the designated His at the 3rd position of the His_6-tag was mutated to Arg accompanied with a codon replacement from CAC to CAT. Nature seems to exclude the constrained His-tag in the M13 system, and such a library should not be used for the biopanning.

4.1. Advantage of High-Throughput DNA Sequencing. A next-generation sequencer (NGS) makes it possible to sequence millions of inserts in parallel. If the NGS is available, one million reads of the library clones would be ideal for finding target-binding sequences even after first round of the biopanning (Figure 5) [48]. If false positive sequences such as target-unrelated (e.g., plastic or BSA)



(a) A few cycles of biopanning (b) DNA amplification and purification (c) DNA sequencing by NGS

Figure 5: Phage display screening with next-generation sequencing. (a) Biopanning with one or two cycle(s). (b) Randomized region of phage DNA is amplified with Polymerase Chain Reaction (PCR). The products are subjected to gel electrophoresis followed by further DNA purification. (c) Purified DNA is analyzed by a next-generation sequencer.

Table 4: Comparison of the M13 and T7 phage in library construction and affinity selection.

	M13 phage	T7 Phage
Cost	Routinely cheap, requires electroporator and cuvettes	Routinely expensive, no additional instruments required
Site of library peptide	N-terminus of gp3	C-terminus of gp10B
Library size (per μg DNA)	$\sim 10^9$	$\sim 10^8$ [2]
Peptide sequence bias	Highly biased	Less biased
Time required for phage tittering/amplification	Long	Short

binders or propagation accelerating peptide (e.g., HAIYPRH [49]) are predominantly enriched at an early stage, further biopanning will be useless. These meaningless false-positive sequences are well described and summarized in a recently published review [50] and can be found easily with online databases (SAROTUP [51], http://immunet.cn/sarotup/; PepBank [52], http://pepbank.mgh.harvard.edu/). Once candidate clones are selected after several rounds of biopanning, the false-positive sequences should be excluded in the same manner.

4.2. Precautions for Conventional DNA Sequencing. If the DNA sequencing is performed by a conventional sequencer but not by the NGS, one should be aware that the DNA sequencing of 50 randomly chosen clones after first or second rounds of the biopanning would be completely uninformative for finding target binders, because the population will be lacking [48]; it should be performed at a later round.

5. Conclusions

We summarized merits and demerits of the M13 and T7 systems in Table 4. It seems the T7 system is easier to handle for beginners, because there are several engineering tolerances in it. Additionally, the T7 phage is stable to detergents and denaturants, such as 1% sodium dodecyl sulfate (SDS), urea (up to 4M), and guanidine-HCl (up to 2M), for eliminating nonspecific binders during the biopanning. Although the T7 phage is robust against not only the chemicals but also an alkaline condition (pH 10), it is fragile at acidic conditions below pH 4. If an elution from target-linked solid support under the lower pH is necessary, the M13 system should be the first choice.

In both systems, the DNA sequencing of a constructed phage library before biopanning is highly recommended for finding unexpected bias.

Acknowledgments

This work was supported by development projects of the Industrial Technology Research Grant Program in 2009 from New Energy and Industrial Technology Development Organization (NEDO) of Japan. The authors also thank Dr. Laura Nelson for careful reading of this paper. The authors are grateful to Mr. Masaki Kanenobu and Ms. Akemi Fujiwara for their technical assistances of unpublished data. They also thank Professor Yuji Ito, Mr. Takaaki Hatanaka, and Dr. Aoi Shiraishi for their kind discussions.

References

[1] C. Heinis, T. Rutherford, S. Freund, and G. Winter, "Phage-encoded combinatorial chemical libraries based on bicyclic peptides," *Nature Chemical Biology*, vol. 5, no. 7, pp. 502–507, 2009.

[2] X. Fan, R. Venegas, R. Fey et al., "An in vivo approach to structure activity relationship analysis of peptide ligands," *Pharmaceutical Research*, vol. 24, no. 5, pp. 868–879, 2007.

[3] G. P. Smith, "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface," *Science*, vol. 228, no. 4705, pp. 1315–1317, 1985, Features of M13 phage, as well as basic strategies for insert DNA construction, are well described therein.

[4] R. Brissette, J. K. A. Prendergast, and N. I. Goldstein, "Identification of cancer targets and therapeutics using phage display," *Current Opinion in Drug Discovery and Development*, vol. 9, no. 3, pp. 363–369, 2006.

[5] T. K. Lu and M. S. Koeris, "The next generation of bacteriophage therapy," *Current Opinion in Microbiology*, vol. 14, no. 5, pp. 524–531, 2011.

[6] A. Skerra, "Alternative binding proteins: anticalins—Harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities," *FEBS Journal*, vol. 275, no. 11, pp. 2677–2683, 2008.

[7] N. Lonberg, "Fully human antibodies from transgenic mouse and phage display platforms," *Current Opinion in Immunology*, vol. 20, no. 4, pp. 450–459, 2008.

[8] L. R. H. Krumpe and T. Mori, "The use of phage-displayed peptide libraries to develop tumor-targeting drugs," *International Journal of Peptide Research and Therapeutics*, vol. 12, no. 1, pp. 79–91, 2006.

[9] E. Koivunen, W. Arap, D. Rajotte, J. Lahdenranta, and R. Pasqualini, "Identification of receptor ligands with phage display peptide libraries," *Journal of Nuclear Medicine*, vol. 40, no. 5, pp. 883–888, 1999.

[10] M. Russel, "Moving through the membrane with filamentous phages," *Trends in Microbiology*, vol. 3, no. 6, pp. 223–228, 1995.

[11] T. Haaparanta and W. D. Huse, "A combinatorial method for constructing libraries of long peptides displayed by filamentous phage," *Molecular Diversity*, vol. 1, no. 1, pp. 39–52, 1995.

[12] T. Bratkovič, "Progress in phage display: evolution of the technique and its applications," *Cellular and Molecular Life Sciences*, vol. 67, no. 5, pp. 749–767, 2010.

[13] M. Russel, H. B. Lowman, and T. Clackson, "Introduction to phage biology and phage display," in *Phage Display—A Practical Approach*, vol. 266, pp. 1–26, Oxford University Press, 2004.

[14] G. P. Smith and V. A. Petrenko, "Phage display," *Chemical Reviews*, vol. 97, no. 2, pp. 391–410, 1997.

[15] V. A. Petrenko and G. P. Smith, "Vectors and modes of display," in *Phage Display in Biotechnology and Drug Discovery*, pp. 63–98, CRC Press, 2005.

[16] X. Agirrezabala, J. Martín-Benito, J. R. Castón, R. Miranda, J. M. Valpuesta, and J. L. Carrascosa, "Maturation of phage T7 involves structural modification of both shell and inner core components," *EMBO Journal*, vol. 24, no. 21, pp. 3820–3829, 2005.

[17] B. G. Condron, J. F. Atkins, and R. F. Gesteland, "Frameshifting in gene 10 of bacteriophage T7," *Journal of Bacteriology*, vol. 173, no. 21, pp. 6998–7003, 1991.

[18] A. Rosenberg, K. Griffin, F.W. Studier et al., "T7 Select Phage Display System: a powerful new protein display system based on bacteriophage T7," *inNovations*, vol. 6, pp. 1–6, 1998.

[19] K. A. Noren and C. J. Noren, "Construction of high-complexity combinatorial phage display peptide libraries," *Methods*, vol. 23, no. 2, pp. 169–178, 2001.

[20] G. A. Barkocy-Gallagher, J. G. Cannon, and P. J. Bassford Jr., "β-Turn formation in the processing region is important for efficient maturation of *Escherichia coli* maltose-binding protein by signal peptidase I in vivo," *Journal of Biological Chemistry*, vol. 269, no. 18, pp. 13609–13613, 1994.

[21] E. H. Manting and A. J. M. Driessen, "*Escherichia coli* translocase: the unravelling of a molecular machine," *Molecular Microbiology*, vol. 37, no. 2, pp. 226–238, 2000.

[22] E. A. Peters, P. J. Schatz, S. S. Johnson, and W. J. Dower, "Membrane insertion defects caused by positive charges in the early mature region of protein pIII of filamentous phage fd can be corrected by prlA suppressors," *Journal of Bacteriology*, vol. 176, no. 14, pp. 4296–4305, 1994.

[23] H. Andersson and G. Von Heijne, "A 30-residue-long "export initiation domain" adjacent to the signal sequence is critical for protein translocation across the inner membrane of *Escherichia coli*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 21, pp. 9751–9754, 1991.

[24] D. J. Rodi, A. S. Soares, and L. Makowski, "Quantitative assessment of peptide sequence diversity in M13 combinatorial peptide phage display libraries," *Journal of Molecular Biology*, vol. 322, no. 5, pp. 1039–1052, 2002.

[25] I. Nilsson and G. Von Heijne, "A signal peptide with a proline next to the cleavage site inhibits leader peptidase when present in a sec-independent protein," *FEBS Letters*, vol. 299, no. 3, pp. 243–246, 1992.

[26] A. Plückthun and J. R. Knowles, "The consequences of stepwise deletions from the signal-processing site of beta-lactamase," *Journal of Biological Chemistry*, vol. 262, no. 9, pp. 3951–3957, 1987.

[27] T. N. Petersen, S. Brunak, G. Von Heijne, and H. Nielsen, "SignalP 4.0: discriminating signal peptides from transmembrane regions," *Nature Methods*, vol. 8, no. 10, pp. 785–786, 2011.

[28] S. J. McConnell, A. J. Uveges, D. M. Fowlkes, and D. G. Spinella, "Construction and screening of M13 phage libraries displaying long random peptides," *Molecular Diversity*, vol. 1, no. 3, pp. 165–176, 1996.

[29] L. R. H. Krumpe, A. J. Atkinson, G. W. Smythers et al., "T7 lytic phage-displayed peptide libraries exhibit less sequence bias than M13 filamentous phage-displayed peptide libraries," *Proteomics*, vol. 6, no. 15, pp. 4210–4222, 2006.

[30] B. K. Kay, N. B. Adey, Y. S. He, J. P. Manfredi, A. H. Mataragnon, and D. M. Fowlkes, "An M13 phage library displaying random 38-amino-acid peptides as a source of novel sequences with affinity to selected targets," *Gene*, vol. 128, no. 1, pp. 59–65, 1993.

[31] I. Kather, C. A. Bippes, and F. X. Schmid, "A stable disulfide-free gene-3-protein of phage fd generated by in vitro evolution," *Journal of Molecular Biology*, vol. 354, no. 3, pp. 666–678, 2005.

[32] R. E. Herman, D. Badders, M. Fuller et al., "The Trp cage motif as a scaffold for the display of a randomized peptide library on bacteriophage T7," *Journal of Biological Chemistry*, vol. 282, no. 13, pp. 9813–9824, 2007.

[33] X. Wu, H. Jörnvall, K. D. Berndt, and U. Oppermann, "Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and

secondary structure but not tRNA abundance," *Biochemical and Biophysical Research Communications*, vol. 313, no. 1, pp. 89–96, 2004.

[34] A. Hilterbrand, J. Saelens, and C. Putonti, "CBDB: the codon bias database," *BMC Bioinformatics*, vol. 13, no. 1, p. 62, 2012.

[35] J. J. Steven and B. L. Trus, "The structure of bacteriophage T7," *Electron Microscopy of Proteins*, vol. 5, pp. 1–35, 1991.

[36] X. Agirrezabala, J. A. Velázquez-Muriel, P. Gómez-Puertas, S. H. W. Scheres, J. M. Carazo, and J. L. Carrascosa, "Quasi-atomic model of bacteriophage t7 procapsid shell: insights into the structure and evolution of a basic fold," *Structure*, vol. 15, no. 4, pp. 461–472, 2007.

[37] M. A. McLafferty, R. B. Kent, R. C. Ladner, and W. Markland, "M13 bacteriophage displaying disulfide-constrained micro-proteins," *Gene*, vol. 128, no. 1, pp. 29–36, 1993.

[38] R. C. Ladner, "Constrained peptides as binding entities," *Trends in Biotechnology*, vol. 13, no. 10, pp. 426–430, 1995.

[39] K. Tanaka, M. Nishimura, Y. Yamaguchi et al., "A mimotope peptide of Aβ42 fibril-specific antibodies with Aβ42 fibrillation inhibitory activity induces anti-Aβ42 conformer antibody response by a displayed form on an M13 phage in mice," *Journal of Neuroimmunology*, vol. 236, no. 1-2, pp. 27–38, 2011.

[40] H. B. Lowman, "Bacteriophage display and discovery of peptide leads for drug development," *Annual Review of Biophysics and Biomolecular Structure*, vol. 26, pp. 401–424, 1997.

[41] F. Uchiyama, Y. Tanaka, Y. Minari, and N. Tokui, "Designing scaffolds of peptides for phage display libraries," *Journal of Bioscience and Bioengineering*, vol. 99, no. 5, pp. 448–456, 2005.

[42] K. Sakamoto, Y. Ito, T. Hatanaka, P. B. Soni, T. Mori, and K. Sugimura, "Discovery and characterization of a peptide motif that specifically recognizes a non-native conformation of human IgG induced by acidic pH conditions," *Journal of Biological Chemistry*, vol. 284, no. 15, pp. 9986–9993, 2009.

[43] P. Laakkonen, K. Porkka, J. A. Hoffman, and E. Ruoslahti, "A tumor-homing peptide with a targeting specificity related to lymphatic vessels," *Nature Medicine*, vol. 8, no. 7, pp. 751–755, 2002.

[44] M. Essler and E. Ruoslahti, "Molecular specialization of breast vasculature: a breast-homing phage-displayed peptide binds to aminopeptidase P in breast vasculature," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 4, pp. 2252–2257, 2002.

[45] H. Witt, K. Hajdin, K. Iljin et al., "Identification of a rhabdomyosarcoma targeting peptide by phage display with sequence similarities to the tumour lymphatic-homing peptide LyP-1," *International Journal of Cancer*, vol. 124, no. 9, pp. 2026–2032, 2009.

[46] Y. G. Mikawa, I. N. Maruyama, and S. Brenner, "Surface display of proteins on bacteriophage λ heads," *Journal of Molecular Biology*, vol. 262, no. 1, pp. 21–30, 1996.

[47] B. A. Katz, "Streptavidin-binding and -dimerizing ligands discovered by phage display, topochemistry, and structure-based design," *Biomolecular Engineering*, vol. 16, no. 1–4, pp. 57–65, 1999.

[48] P. A. C. 't Hoen, S. M. G. Jirka, B. R. ten Broeke et al., "Phage display screening without repetitious selection rounds," *Analytical Biochemistry*, vol. 421, no. 2, pp. 622–631, 2012.

[49] L. A. Brammer, B. Bolduc, J. L. Kass, K. M. Felice, C. J. Noren, and M. F. Hall, "A target-unrelated peptide in an M13 phage display library traced to an advantageous mutation in the gene II ribosome-binding site," *Analytical Biochemistry*, vol. 373, no. 1, pp. 88–98, 2008.

[50] M. Vodnik, U. Zager, B. Strukelj, and M. Lunder, "Phage display: selecting straws instead of a needle from a haystack," *Molecules*, vol. 16, no. 1, pp. 790–817, 2011.

[51] J. Huang, B. Ru, S. Li, H. Lin, and F. B. Guo, "SAROTUP: scanner and reporter of target-unrelated peptides," *Journal of Biomedicine and Biotechnology*, vol. 2010, Article ID 101932, 7 pages, 2010.

[52] T. Duchrow, T. Shtatland, D. Guettler, M. Pivovarov, S. Kramer, and R. Weissleder, "Enhancing navigation in biomedical databases by community voting and database-driven text classification," *BMC Bioinformatics*, vol. 10, article 1471, p. 317, 2009.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 518162, 8 pages
doi:10.1155/2012/518162

Review Article

Recent Advances in Chemical Modification of Peptide Nucleic Acids

Eriks Rozners

Department of Chemistry, Binghamton University, State University of New York, 4400 Vestal Parkway East, Binghamton, NY 13902, USA

Correspondence should be addressed to Eriks Rozners, erozners@binghamton.edu

Received 2 June 2012; Revised 12 July 2012; Accepted 20 July 2012

Academic Editor: Masayasu Kuwahara

Peptide nucleic acid (PNA) has become an extremely powerful tool in chemistry and biology. Although PNA recognizes single-stranded nucleic acids with exceptionally high affinity and sequence selectivity, there is considerable ongoing effort to further improve properties of PNA for both fundamental science and practical applications. The present paper discusses selected recent studies that improve on cellular uptake and binding of PNA to double-stranded DNA and RNA. The focus is on chemical modifications of PNA's backbone and heterocyclic nucleobases. The paper selects representative recent studies and does not attempt to provide comprehensive coverage of the broad and vibrant field of PNA modification.

1. Introduction

Peptide nucleic acid (PNA) is a DNA analogue that has the entire sugar-phosphodiester backbone replaced by a pseudopeptide linkage built of 2-aminoethylglycine residues (Figure 1) [1]. PNA is highly stable chemically and, because of the unnatural backbone, highly resistant to enzymatic degradation, which makes it an excellent candidate for in vivo applications as an oligonucleotide analogue. The neutral pseudopeptide backbone eliminates electrostatic repulsion (a factor that negatively affects oligonucleotide binding) and PNA binds to DNA and RNA with excellent affinity. PNA binds to double helical DNA via two competing binding modes, triple helix (PNA : DNA, 1 : 1), and strand invasion, where PNA displaces one of the DNA strands, typically followed by a triplex formation (PNA : DNA, 2 : 1) [1]. PNA also forms exceptionally strong and sequence-specific Watson-Crick duplexes with single-stranded DNA and RNA [2]. Interestingly, the sequence specificity of duplexes involving PNA is substantially higher than that of unmodified nucleic acids. Because of these superior qualities, PNA has become a powerful tool in chemical biology and biotechnology [3–5]. The main applications of PNA are as hybridization probes and molecular diagnostics of high affinity and selectivity for single-stranded DNA and RNA. PNA also holds a promise of becoming a novel gene therapy agent for targeting specific RNA molecules [3, 4].

Although PNA binds single-stranded DNA and RNA with superior affinity and selectivity, there are other properties of PNA that can be further improved. Most importantly, in vivo applications of unmodified PNA are hindered by poor cellular uptake and endosomal entrapment [6]. Current methods to enhance the cellular uptake of PNA, such as conjugation with cell penetrating peptides (CPP) [7, 8], are complicated and require high PNA-peptide concentrations that may cause off-target binding and toxicity in vivo. Another problem is the limited sequence scope of double-stranded nucleic acids that can be recognized by PNA. While PNA can bind any sequence of single-stranded DNA and RNA with high affinity and selectivity, recognition of double helical DNA has been limited to polypurine tracts and binding to double helical RNA has been little explored. The present paper focuses on most recent developments in chemical modification of PNA to enhance cellular uptake and recognition of double helical nucleic acids. Several comprehensive reviews have recently discussed modification of PNA backbone [9, 10] and nucleobases [11] in a broader context.

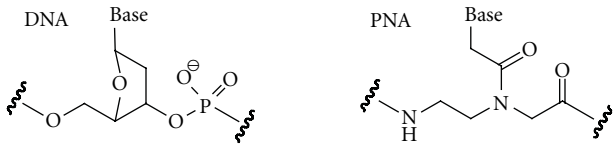


Figure 1: Structures of DNA and PNA repeating units.

2. Conjugation of PNA with Cationic Peptides to Improve the Cellular Uptake

Inefficient crossing of cellular membrane of mammalian cells by unmodified PNA has been a major problem for practical in vivo applications of PNA. Because of the neutral backbone, PNA does not associate with delivery vehicles based on cationic lipids. To use such standard oligonucleotide transfectants as Lipofectamine, PNA needs to be hybridized to complementary oligodeoxynucleotide (ODN) that aids the electrostatic complexation with the positively charged lipids [12]. Recently, a new approach to PNA delivery was developed by Wooley, Taylor and coworkers [13] who used cationic shell-cross-linked knedel-like nanoparticles (cSCKs) to deliver either PNA-ODN hybrid or PNA covalently attached to cSCKs nanoparticles through a biodegradable disulfide linkage. cSCKs nanoparticles have a hydrophobic core and a positively charged cross-linked shell. The latter is highly functionalizable and mediates the cellular delivery through, most likely, an endocytotic mechanism. An elegant extension of this technology is reported in this special issue by Taylor and coworkers [14].

Perhaps, the most popular approach to enhance cellular delivery has been conjugation of PNA with cell penetrating peptides that deliver the conjugate through the endocytosis pathway [7, 8]. However, the low ability of PNA-CPP conjugates to escape from endosomes has been the bottleneck of this approach. Various endosomolytic compounds have been explored; unfortunately, most are too toxic for in vivo applications [7]. Conjugates with arginine-rich peptides have shown promising activity in HeLa cells in the absence of endosomolytic agents [15]. However, even in the most promising cases large amount of conjugates remained in endosomes, leaving plenty of room for further improvement [15]. The relatively high concentrations of PNA-CCP, which are required for efficient delivery, may cause off-target binding and toxicity in vivo. Moreover, CPPs are relatively large peptides, which complicate the preparation and use of PNA-CPP conjugates. Recently, several groups have demonstrated that relatively simple cationic modifications in PNA can substantially improve their cellular uptake and produce effect similar to that of longer and more complex CPPs.

The groups of Corey [16, 17] and Gait [15, 18, 19] showed that conjugation of PNA with short oligolysine (Figure 2, **1** and **2**, resp.) enabled efficient delivery in fibroblast and various cancer cell lines (T47D, MCF-7, Huh7, and HeLa). As few as four lysine residues achieved similar efficiency as R6-Penetratin, a CPP previously optimized for cellular delivery of PNA [15]. Using short oligolysine instead of longer CPP significantly reduced the complexity and effort required for PNA use in cell culture. Lysine conjugates have also been used to deliver PNA in mice [20, 21]. Most recently, Gait and coworkers showed that introduction of a terminal Cys residue further increased the cellular uptake of Cys-Lys-PNA-Lys_3 conjugate [22]. While some studies showed that conjugates built of the unnatural D-lysine were more effective [17], presumably due to higher biostability, other studies found little difference between the L and D series [22]. In a similar study, Fabbri et al. [23] demonstrated that PNA conjugated at the carboxyl terminus with octaarginine was efficiently taken up in human leukemic K562 cells and inhibited activity of the target microRNA-210.

Nielsen and coworkers have recently reported on conjugates of PNA with cationic ligands that showed improved cellular delivery and activity [24, 25]. In one study, addition of a lipid domain to the cationic peptides increased the activity of PNA conjugate by two orders of magnitude [24]. The lypophilic fatty acid contributed by promoting both endosomal uptake and endosomal escape of PNA. In another study, conjugation of PNA with polyethylenimine showed significantly higher antisense activity than PNA-octaarginine conjugates [25]. Polyethylenimine conjugates had lower toxicity than PNA-octaarginine conjugates. The polyethylenimine conjugate activity did not depend on the presence of lysosomolytic agents, which suggested that these conjugates are able to escape endosomes efficiently. These studies suggest that chemical approaches can be used to tailor cationic modifications that will improve cellular uptake and avoid the problem of endosomal entrapment.

Conjugation of PNA with a lipophilic triphenylphosphonium cation has been shown to increase the cellular delivery [26, 27]. In this special issue, Pandey, Patino and coworkers [28] report on cyclic and hairpin PNAs conjugated to the triphenylphosphonium cation via a disulfide linkage. The conjugates inhibit HIV replication by targeting the HIV-1 TAR RNA loop. Most recently, Shiraishi and Nielsen [29] reported on cellular uptake and antisense activity of PNA conjugated with cholesterol and cholic acid in HeLa pLuc705 cells. Although the conjugates alone were inactive, the delivery was dramatically improved by addition of Lipofectamine leading to nanomolar antisense activity.

As the numerous recent studies reviewed above suggest, design and optimization of CPP and other cationic ligands for cellular delivery of PNA is still a vigorous and important area of research. The focus has shifted to addressing endosomal escape, improving the end point activity and potential in vivo applications.

3. Cationic Backbone Modifications to Improve the Cellular Uptake of PNA

An alternative approach to conjugation of PNA has been direct modification of PNA's backbone. Several groups have explored cationic modifications of PNA [30–32]. Ly and coworkers introduced guanidine groups at α- [31] and γ-positions [32] of PNA's backbone by custom synthesis of monomers starting from diaminoethane and L or D arginine instead of glycine (Figure 3, L series shown). The α-guanidine-modified PNA (GPNA) derived from the

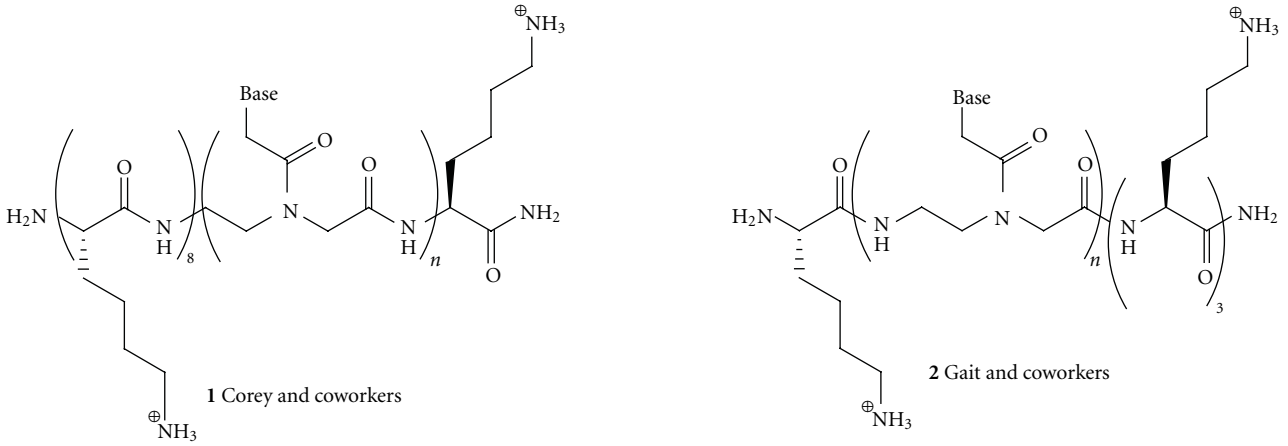


Figure 2: Conjugation of PNA with short oligolysines improves cellular uptake.

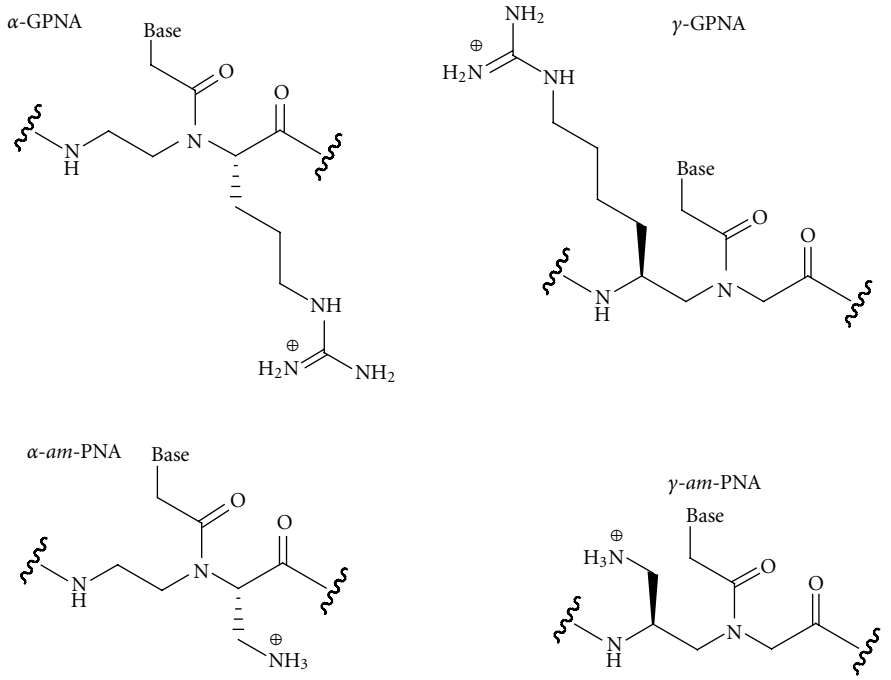


Figure 3: Cationic backbone modifications of PNA.

unnatural D-arginine had higher affinity for complementary DNA [33] and RNA [34]; good sequence selectivity was maintained. GPNA was readily taken up by several cell lines (HCT116, human ES, and HeLa), which was attributed to the cationic guanidine modifications. GPNA was less toxic to cells than a PNA-polyarginine conjugate and induced potent antisense inhibition of E-cadherin in A549 cells [35]. Our laboratory recently studied the triple helix formation between double helical RNA and α-GPNA. We found that the α-guanidine modification decreased RNA binding affinity and sequence selectivity of α-GPNA compared to unmodified PNA [36].

The γ-guanidine-modified PNA had higher affinity for complementary DNA and RNA than α-guanidine-modified PNA, presumably due to favorable preorganization of the γ-modified backbone into a right-handed helix [32]. In contrast to α-modified PNA, Englund and Appella found that the *S*-isomer of γ-modified PNA (derived from the natural L-lysine) had higher affinity for complementary DNA than the *R*-isomer [30]. Most recently, Manicardi et al. [37] used both α- and γ-modified GPNA 15-mers to inhibit microRNA-210 in K562 cells. Both isomers showed promising though not complete inhibition with the PNAs having eight consecutive γ-modification at the carboxyl terminus performing slightly better than other modification patterns [37].

Mitra and Ganesh reported similar results on DNA binding and cellular uptake of α- and γ-aminomethylene PNA (*am*-PNA, Figure 3) [38, 39]. The aminomethylene

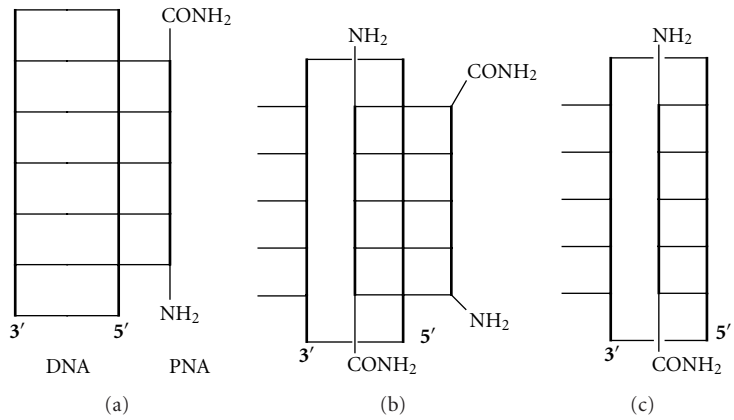


Figure 4: Binding modes for recognition of double-stranded DNA: triple helix (a), strand-displacement triplex (b), and strand-displacement duplex (c).

modification increased PNA binding to DNA, with γ-(*S*)*am*-PNA being significantly better than α-(*R*)*am*-PNA, which, in turn, was better than α-(*S*)*am*-PNA [39]. The cellular uptake was enhanced by these modifications in roughly the same order, with γ-(*S*)*am*-PNA giving the most promising results.

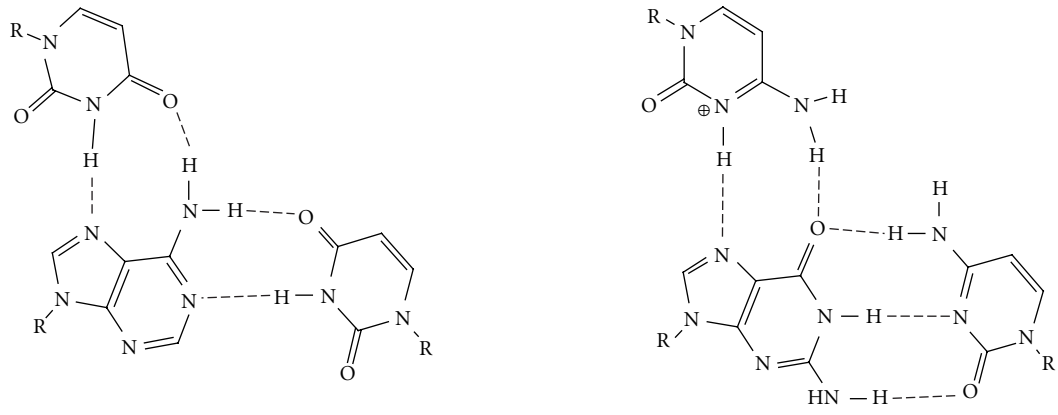
4. PNA Modifications to Expand the Recognition of Double-stranded Nucleic Acids

Recognition of single-stranded DNA and RNA following the Watson-Crick base pairing rules is relatively straightforward. Recognition of double-stranded nucleic acids is substantially more challenging because the Watson-Crick faces of nucleobases are already engaged in hydrogen bonding. PNA, as well as other oligonucleotide analogues, can recognize double-stranded nucleic acids by forming either a parallel triple helix (Figure 4(a), the amino end of PNA aligned with the 5′ end of DNA) or a strand-invasion complex, where PNA displaces one of the DNA strands. The strand-invasion is typically a competing mode for triplex (PNA : DNA, 1 : 1) and usually results in a strand-displacement triplex (PNA : DNA, 2 : 1). The PNA strand that is replacing the DNA strand aligns antiparallel with the DNA strand (Figure 4(b), the carboxyl end of PNA aligned with the 5′ end of DNA). Both binding modes are limited to nucleic acid duplexes featuring so-called polypurine tracts where one strand is built of purines, while the other strand consists of pyrimidines. This is because the standard Hoogsteen triplets (U*A-U and C+*G-C) recognize only purine bases (Figure 5(a)).

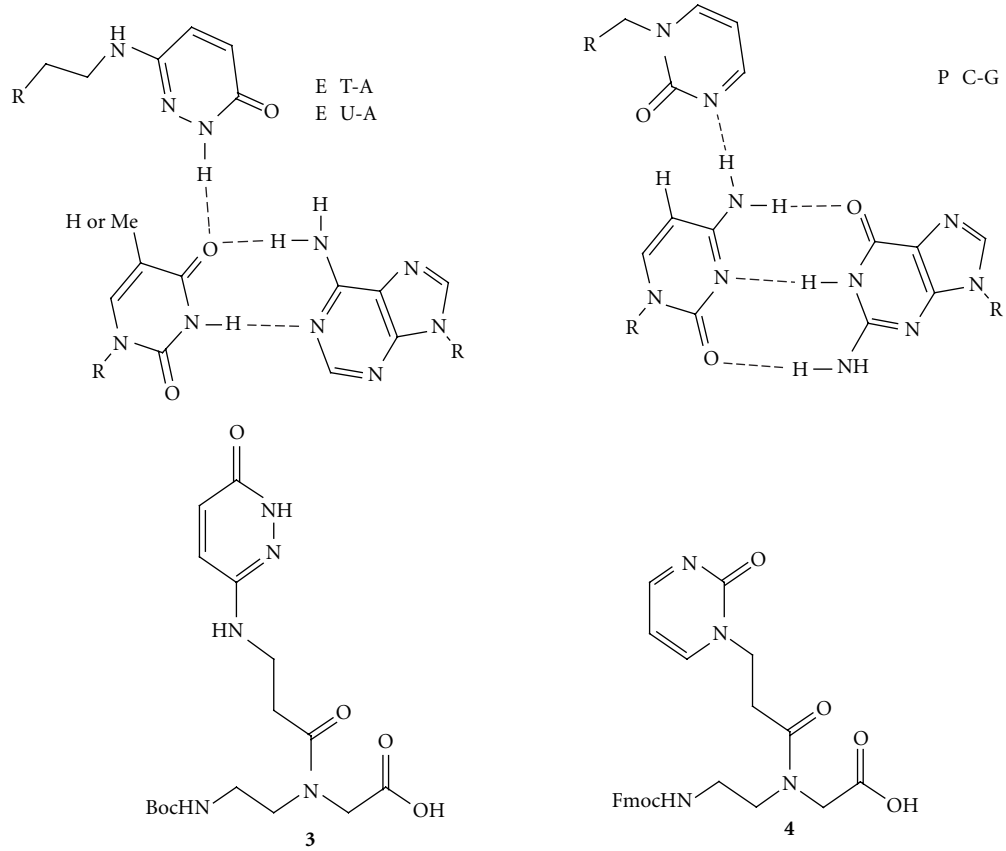
The strand-displacement triplex approach (Figure 4(b)) typically uses PNA clamps that have the two PNA strands connected by a short linker, which enhances the binding affinity and favors strand invasion. To expand the repertoire of sequences that can be recognized by the triplex forming part of PNA, Dahl and Nielsen designed 3-oxo-2,3-dihydropyridazine nucleobase (**E**, Figure 5(b)) to recognize thymidine in T-A base pairs of DNA [40]. This modification substantially increased the thermal stability of a PNA clamp targeting 10-nucleotide long DNA stretch that had two thymidines interrupting the purine-rich strand [40]. Despite the promising preliminary results, this modification has not been widely applied either in strand-displacement triplex or in triple helical approaches.

A PNA : DNA 1 : 1 strand-displacement duplex (Figure 4(c)) would be a highly desired binding mode because, in principle, any sequence of DNA could be recognized without the need for the presence of a purine-rich strand. However, this recognition mode is complicated by the fact that duplex forming PNA does not have enough thermodynamic advantage to displace a DNA stand from a duplex. Ly and coworkers recently showed that γ-methylation (Figure 6, **5**) preorganized PNA into right-handed helix and enhanced its ability to form strand-displacement complex with mixed sequence DNA [41]. The properties of invading γ-modified PNAs were further improved by incorporation of G-clamp nucleobases [42] and replacement of the methyl group with MiniPEG (Figure 6, **6**) [43]. The latter modification was critical to optimize water solubility and minimize PNA aggregation and enabled PNA built of monomers **6** invade essentially any sequence of double-stranded DNA in a highly sequence-specific manner [43].

The triple helical recognition of double-stranded DNA using PNA has received less attention than the strand-displacement approaches. However, in a recent and comprehensive study Nielsen and coworkers showed that this is a promising and perhaps underutilized approach [44]. Compared to DNA, molecular recognition of double-stranded RNA has been even less studied. This is perhaps because for a long time RNA was believed to be only a passive messenger in the transfer of genetic information from DNA to proteins. However, since the discovery that RNA can catalyze chemical reactions, the number and variety of noncoding RNAs and the important roles they play in biology have been growing steadily. While less than 2% of DNA encodes for functional proteins, almost 70% is transcribed into RNA. Today, the functional importance of most RNA transcripts is still unknown and it is fairly safe to predict that we will discover many more regulatory RNAs in the near future. The ability to



(a) Hoogsteen triplets



(b) Modified triplets for pyrimidine recognition

Figure 5: Recognition of purines and pyrimidines using Hoogsteen (a) and modified (b) triplets, respectively.

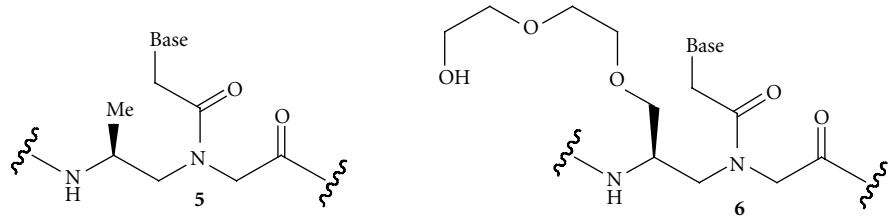


Figure 6: γ-Alkyl PNA for recognition of double-stranded DNA via strand-displacement duplex.

selectively recognize, detect, and inhibit the function of such RNAs will be highly useful for both fundamental biology and practical applications in biotechnology and medicine.

Recently, our laboratory started studies on triple helical recognition of double-stranded RNA using PNA [45, 46]. Before this effort, triple helices between RNA and PNA were virtually unknown; there was only one study by Toulme and coworkers that suggested that PNA may not be forming stable triple helix with RNA [47]. In contrast, we found that PNA formed a highly stable and sequence-selective triple helix with double-stranded RNA [45]. Interestingly, the RNA-PNA triplexes were at least an order of magnitude more stable than the DNA-PNA triplexes suggesting that PNA may be a significantly better ligand for the deep and narrow major groove of RNA than for the wider major groove of DNA [45]. To expand the sequence scope of RNA that can be recognized, we adopted monomer **E** for recognition of uridine in U-A base pair and designed a novel monomer, 2-pyrimidone **P** for recognition of cytidine in C-G base pair (Figure 5(b), **3** and **4**, resp.) [46]. Our design of **P** was inspired by the work of Leumann and coworkers [48, 49] who used 4-methyl-2-pyrimidone as an oligonucleotide modification for triple helical recognition of cytidine in C-G base pairs of DNA. Heterocycle **P** had not been used in PNA before our study. Incorporation of **E** and **P** in short PNA sequences allowed recognition of nine-nucleotide long polypurine tracts of double helical RNA containing single pyrimidine inversion. The selectivity was good and affinity matched that of the standard Hoogsteen triple helices (Figure 5(a)) [46]. Our results also showed that the extended linkers connecting **E** and **P** heterocycles to the PNA backbone were important design elements that optimized the binding affinity [46].

5. Conclusions

Since invention of PNA, synthetic chemists have been extensively modifying its structure [9–11]. Most of the work on backbone modifications of PNA has attempted, with mixed success, to improve the affinity and selectivity of Watson-Crick recognition of DNA and RNA. Reviewed herein are selected recent studies focused on improving cellular uptake of PNA and developing novel modes of binding, such as strand-invasion of mixed sequence double-stranded DNA and triple helical recognition of RNA. The preliminary results are very encouraging, and it is likely that more improvements and new discoveries will be made in the near future.

Acknowledgments

The PNA work in Rozners' laboratory has been supported by Binghamton University and NIH (R01 GM071461).

References

[1] P. E. Nielsen, M. Egholm, R. H. Berg, and O. Buchardt, "Sequence-selective recognition of DNA by strand displacement with thymine-substituted polyamide," *Science*, vol. 254, no. 5037, pp. 1497–1500, 1991.

[2] M. Egholm, O. Buchardt, L. Christensen et al., "PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules," *Nature*, vol. 365, no. 6446, pp. 566–568, 1993.

[3] P. E. Nielsen, "Peptide Nucleic Acids (PNA) in chemical biology and drug discovery," *Chemistry and Biodiversity*, vol. 7, no. 4, pp. 786–804, 2010.

[4] P. E. Nielsen, "Sequence-selective targeting of duplex DNA by peptide nucleic acids," *Current Opinion in Molecular Therapeutics*, vol. 12, no. 2, pp. 184–191, 2010.

[5] C. Achim, B. A. Armitage, D. H. Ly, and J. W. Schneider, "Peptide nucleic acids (PNAs)," *Wiley Encyclopedia of Chemical Biology*, vol. 3, pp. 588–597, 2009.

[6] P. E. Nielsen, "Addressing the challenges of cellular delivery and bioavailability of peptide nucleic acids (PNA)," *Quarterly Reviews of Biophysics*, vol. 38, no. 4, pp. 345–350, 2005.

[7] T. Shiraishi and P. E. Nielsen, "Enhanced delivery of cell-penetrating peptide-peptide nucleic acid conjugates by endosomal disruption," *Nature Protocols*, vol. 1, no. 2, pp. 633–636, 2006.

[8] F. S. Hassane, A. F. Saleh, R. Abes, M. J. Gait, and B. Lebleu, "Cell penetrating peptides: overview and applications to the delivery of oligonucleotides," *Cellular and Molecular Life Sciences*, vol. 67, no. 5, pp. 715–726, 2010.

[9] R. Corradini, S. Sforza, T. Tedeschi, F. Totsingan, A. Manicardi, and R. Marchelli, "Peptide nucleic acids with a structurally biased backbone. Updated review and emerging challenges," *Current Topics in Medicinal Chemistry*, vol. 11, no. 12, pp. 1535–1554, 2011.

[10] R. Corradini, S. Sforza, T. Tedeschi, F. Totsingan, and R. Marchelli, "Peptide nucleic acids with a structurally biased backbone: effects of conformational constraints and stereochemistry," *Current Topics in Medicinal Chemistry*, vol. 7, no. 7, pp. 681–694, 2007.

[11] F. Wojciechowski and R. H. E. Hudson, "Nucleobase modifications in peptide nucleic acids," *Current Topics in Medicinal Chemistry*, vol. 7, no. 7, pp. 667–679, 2007.

[12] D. A. Braasch and D. R. Corey, "Lipid-mediated introduction of peptide nucleic acids into cells.," *Methods in Molecular Biology*, vol. 208, pp. 211–223, 2002.

[13] H. Fang, K. Zhang, G. Shen, K. L. Wooley, and J. S. A. Taylor, "Cationic shell-cross-linked knedel-Like (cSCK) nanoparticles for highly efficient PNA delivery," *Molecular Pharmaceutics*, vol. 6, no. 2, pp. 615–626, 2009.

[14] Z. Wang, K. Zhang, K. L. Wooley, and J. S. Taylor, "Imaging mRNA expression in live cells via peptide nucleic acid (PNA) strand-displacement activated probes," *Journal of Nucleic Acids*, vol. 12, Article ID 962652, 2012.

[15] S. Abes, J. J. Turner, G. D. Ivanova et al., "Efficient splicing correction by PNA conjugation to an R6 -Penetratin delivery peptide," *Nucleic Acids Research*, vol. 35, no. 13, pp. 4495–4502, 2007.

[16] J. Hu, M. Matsui, K. T. Gagnon et al., "Allele-specific silencing of mutant huntingtin and ataxin-3 genes by targeting expanded CAG repeats in mRNAs," *Nature Biotechnology*, vol. 27, no. 5, pp. 478–484, 2009.

[17] J. Hu and D. R. Corey, "Inhibiting gene expression with peptide nucleic acid (PNA)-peptide conjugates that target chromosomal DNA," *Biochemistry*, vol. 46, no. 25, pp. 7581–7589, 2007.

[18] M. M. Fabani and M. J. Gait, "miR-122 targeting with LNA/2′-O-methyl oligonucleotide mixmers, peptide nucleic acids (PNA), and PNA-peptide conjugates," *RNA*, vol. 14, no. 2, pp. 336–346, 2008.

[19] J. J. Turner, G. D. Ivanova, B. Verbeure et al., "Cell-penetrating peptide conjugates of peptide nucleic acids (PNA) as inhibitors of HIV-1 Tat-dependent trans-activation in cells," *Nucleic Acids Research*, vol. 33, no. 21, pp. 6837–6849, 2005.

[20] E. V. Wancewicz, M. A. Maier, A. M. Siwkowski et al., "Peptide nucleic acids conjugated to short basic peptides show improved pharmacokinetics and antisense activity in adipose tissue," *Journal of Medicinal Chemistry*, vol. 53, no. 10, pp. 3919–3926, 2010.

[21] M. M. Fabani, C. Abreu-Goodger, D. Williams et al., "Efficient inhibition of miR-155 function in vivo by peptide nucleic acids," *Nucleic Acids Research*, vol. 38, no. 13, Article ID gkq160, pp. 4466–4475, 2010.

[22] A. G. Torres, M. M. Fabani, E. Vigorito et al., "Chemical structure requirements and cellular targeting of microRNA-122 by peptide nucleic acids anti-miRs," *Nucleic Acids Research*, vol. 40, no. 5, pp. 2152–2167, 2012.

[23] E. Fabbri, A. Manicardi, T. Tedeschi et al., "Modulation of the biological activity of microRNA-210 with peptide nucleic acids (PNAs)," *ChemMedChem*, vol. 6, no. 12, pp. 2192–2202, 2011.

[24] U. Koppelhus, T. Shiraishi, V. Zachar, S. Pankratova, and P. E. Nielsen, "Improved cellular activity of antisense peptide nucleic acids by conjugation to a cationic peptide-lipid (CatLip) domain," *Bioconjugate Chemistry*, vol. 19, no. 8, pp. 1526–1534, 2008.

[25] P. R. Berthold, T. Shiraishi, and P. E. Nielsen, "Cellular delivery and antisense effects of peptide nucleic acid conjugated to polyethyleneimine via disulfide linkers," *Bioconjugate Chemistry*, vol. 21, no. 10, pp. 1933–1938, 2010.

[26] A. Muratovska, R. N. Lightowlers, R. W. Taylor et al., "Targeting peptide nucleic acid (PNA) oligomers to mitochondria within cells by conjugation to lipophilic cations: implications for mitochondrial DNA replication, expression and disease," *Nucleic Acids Research*, vol. 29, no. 9, pp. 1852–1863, 2001.

[27] M. Mehiri, G. Upert, S. Tripathi et al., "An efficient biodelivery system for antisense polyamide nucleic acid (PNA)," *Oligonucleotides*, vol. 18, no. 3, pp. 245–255, 2008.

[28] G. Upert, A. Di Giorgio, A. Upadhyay et al., "Inhibition of HIV replication by cyclic and hairpin PNAs targeting the HIV-1 TAR RNA loop," *Journal of Nucleic Acids*, vol. 12, Article ID 591025, 2012.

[29] T. Shiraishi and P. E. Nielsen, "Nanomolar cellular antisense activity of peptide nucleic acid (PNA) cholic acid ("Umbrella") and cholesterol conjugates delivered by cationic lipids," *Bioconjugate Chemistry*, vol. 23, no. 2, pp. 196–202, 2012.

[30] E. A. Englund and D. H. Appella, "γ-substituted peptide nucleic acids constructed from L-lysine are a versatile scaffold for multifunctional display," *Angewandte Chemie, International Edition*, vol. 46, no. 9, pp. 1414–1418, 2007.

[31] P. Zhou, M. Wang, L. Du, G. W. Fisher, A. Waggoner, and D. H. Ly, "Novel binding and efficient cellular uptake of guanidine-based peptide nucleic acids (GPNA)," *Journal of the American Chemical Society*, vol. 125, no. 23, pp. 6878–6879, 2003.

[32] B. Sahu, V. Chenna, K. L. Lathrop et al., "Synthesis of conformationaly preorganized and cell-permeable guanidine-based γ-peptide nucleic acids (γGPNAs)," *Journal of Organic Chemistry*, vol. 74, no. 4, pp. 1509–1516, 2009.

[33] P. Zhou, A. Dragulescu-Andrasi, B. Bhattacharya et al., "Synthesis of cell-permeable peptide nucleic acids and characterization of their hybridization and uptake properties," *Bioorganic and Medicinal Chemistry Letters*, vol. 16, no. 18, pp. 4931–4935, 2006.

[34] A. Dragulescu-Andrasi, P. Zhou, G. He, and D. H. Ly, "Cell-permeable GPNA with appropriate backbone stereochemistry and spacing binds sequence-specifically to RNA," *Chemical Communications*, no. 2, pp. 244–246, 2005.

[35] A. Dragulescu-Andrasi, S. Rapireddy, G. He et al., "Cell-permeable peptide nucleic acid designed to bind to the 5′-untranslated region of E-cadherin transcript induces potent and sequence-specific antisense effects," *Journal of the American Chemical Society*, vol. 128, no. 50, pp. 16104–16112, 2006.

[36] P. Gupta, O. Muse, and E. Rozners, "Recognition of double stranded RNA by guanidine-modified peptide nucleic acids (GPNA)," *Biochemistry*, vol. 51, no. 1, pp. 63–73, 2012.

[37] A. Manicardi, E. Fabbri, T. Tedeschi et al., "Cellular uptakes, biostabilities and anti-miR-210 activities of chiral arginine-PNAs in leukaemic K562 cells," *ChemBioChem*, vol. 13, no. 9, pp. 1327–1337, 2012.

[38] R. Mitra and K. N. Ganesh, "PNAs grafted with (α/γ, R/S)-aminomethylene pendants: regio and stereo specific effects on DNA binding and improved cell uptake," *Chemical Communications*, vol. 47, no. 4, pp. 1198–1200, 2011.

[39] R. Mitra and K. N. Ganesh, "Aminomethylene peptide nucleic acid (am-PNA): synthesis, regio-/stereospecific DNA binding, and differential cell uptake of (α/γ, R/S)am-PNA analogues," *Journal of Organic Chemistry*, vol. 77, no. 13, pp. 5696–5704, 2012.

[40] A. B. Eldrup, O. Dahl, and P. E. Nielsen, "A novel peptide nucleic acid monomer for recognition of thymine in triple-helix structures," *Journal of the American Chemical Society*, vol. 119, no. 45, pp. 11116–11117, 1997.

[41] A. Dragulescu-Andrasi, S. Rapireddy, B. M. Frezza, C. Gayathri, R. R. Gil, and D. H. Ly, "A simple γ-backbone modification preorganizes peptide nucleic acid into a helical structure," *Journal of the American Chemical Society*, vol. 128, no. 31, pp. 10258–10267, 2006.

[42] S. Rapireddy, R. Bahal, and D. H. Ly, "Strand invasion of mixed-sequence, double-helical B-DNA by γ-peptide nucleic acids containing g-clamp nucleobases under physiological conditions," *Biochemistry*, vol. 50, no. 19, pp. 3913–3918, 2011.

[43] B. Sahu, I. Sacui, S. Rapireddy et al., "Synthesis and characterization of conformationally preorganized, (R)-diethylene glycol-containing γ-peptide nucleic acids with superior hybridization properties and water solubility," *Journal of Organic Chemistry*, vol. 76, no. 14, pp. 5614–5627, 2011.

[44] M. E. Hansen, T. Bentin, and P. E. Nielsen, "High-affinity triplex targeting of double stranded DNA using chemically modified peptide nucleic acid oligomers," *Nucleic Acids Research*, vol. 37, no. 13, pp. 4498–4507, 2009.

[45] M. Li, T. Zengeya, and E. Rozners, "Short peptide nucleic acids bind strongly to homopurine tract of double helical RNA at pH 5.5," *Journal of the American Chemical Society*, vol. 132, no. 25, pp. 8676–8681, 2010.

[46] P. Gupta, T. Zengeya, and E. Rozners, "Triple helical recognition of pyrimidine inversions in polypurine tracts of RNA by nucleobase-modified PNA," *Chemical Communications*, vol. 47, no. 39, pp. 11125–11127, 2011.

[47] K. Aupeix, R. Le Tinévez, and J. J. Toulmé, "Binding of oligopyrimidines to the RNA hairpin responsible for the ribosome gag-pol frameshift in HIV-1," *FEBS Letters*, vol. 449, no. 2-3, pp. 169–174, 1999.

[48] I. Prévot-Halter and C. J. Leumann, "Selective recognition of a C-G base-pair in the parallel DNA triple-helical binding motif," *Bioorganic and Medicinal Chemistry Letters*, vol. 9, no. 18, pp. 2657–2660, 1999.

[49] S. Buchini and C. J. Leumann, "Stable and selective recognition of three base pairs in the parallel triple-helical DNA binding motif," *Angewandte Chemie, International Edition*, vol. 43, no. 30, pp. 3925–3928, 2004.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 971581, 7 pages
doi:10.1155/2012/971581

Research Article

Combinatorial Synthesis, Screening, and Binding Studies of Highly Functionalized Polyamino-amido Oligomers for Binding to Folded RNA

Jonathan K. Pokorski and Daniel H. Appella

Laboratory of Bioorganic Chemistry, NIH, NIDDK, Bethesda, MD 20892, USA

Correspondence should be addressed to Daniel H. Appella, appellad@niddk.nih.gov

Received 1 June 2012; Revised 28 June 2012; Accepted 29 June 2012

Academic Editor: Eriks Rozners

Folded RNA molecules have recently emerged as critical regulatory elements in biological pathways, serving not just as carriers of genetic information but also as key components in enzymatic assemblies. In particular, the transactivation response element (TAR) of the HIV genome regulates transcriptional elongation by interacting specifically with the Tat protein, initiating the recruitment of the elongation complex. Preventing this interaction from occurring *in vivo* halts HIV replication, thus making RNA-binding molecules an intriguing pharmaceutical target. Using α-amino acids as starting materials, we have designed and synthesized a new class of polyamino-amido oligomers, called PAAs, specifically for binding to folded RNA structures. The PAA monomers were readily incorporated into a 125-member combinatorial library of PAA trimers. In order to rapidly assess RNA binding, a quantum dot-based fluorescent screen was developed to visualize RNA binding on-resin. The binding affinities of hits were quantified using a terbium footprinting assay, allowing us to identify a ligand (SFF) with low micromolar affinity ($k_d = 14\,\mu$M) for TAR RNA. The work presented herein represents the development of a flexible scaffold that can be easily synthesized, screened, and subsequently modified to provide ligands specific for binding to folded RNAs.

1. Introduction

Elucidation of the role played by small RNA molecules in the context of viral life cycles has led to the identification of new drug targets that fall outside the realm of traditional drug design [1]. In particular, the transactivation response element (TAR) of HIV is critical in regulating transcriptional elongation of the HIV genome [2]. TAR is a highly conserved stem-loop RNA located at the 5′-end of viral transcripts. For elongation to occur, the HIV transactivator protein (Tat) must first bind to TAR [2, 3]. The ensuing Tat-TAR complex is then responsible for recruitment of the positive transcription elongation factor complex, pTEFb, resulting in processive elongation [4]. The Tat-TAR complex is a particularly attractive drug target because it is highly conserved and difficult for the virus to develop resistance, which is a problem that hinders the effectiveness of conventional HIV therapies [5].

The most effective strategies to inhibit Tat-TAR formation rely upon mimicking the binding region of the Tat protein. Initial attempts to design TAR-binding molecules focused on dissecting the Tat peptide and synthesizing peptides corresponding to the highly basic, arginine rich region of Tat that binds to the trinucleotide bulge of TAR [6]. The truncated Tat peptides are able to compete with Tat *in vitro* for binding to TAR, however peptides are typically not viable drug candidates due to low bioavailability and stability [7]. Further studies have sought to mimic the Tat peptide with nonnatural oligomers. For instance, a nonamer peptoid was developed that had significant potency against the Tat-TAR interaction [8, 9]. Alternatively, cyclic peptides, and peptides comprised of D-amino acids had similar effects [10–12]. All of these molecules share a common feature in that they display highly cationic sidechains. We felt that the high degree of cationic charge associated with these molecules would lead to promiscuous RNA binders. Thus, we sought to

create molecules that could retain similar binding properties while minimizing cationic charge about the periphery of the scaffold.

Initial results from our lab showed that highly functionalized polyamines were able to serve as TAR-binding molecules specific for the trinucleotide bulge [13]. Our design strategy served to relegate cationic charge to the interior of the polyamines to decrease nonspecific charge-charge interactions, while projecting sidechains out from the backbone to direct specificity. The first generation of polyamines showed promising results, yielding a polyamine that bound TAR specifically and with a K_D of ~6 μM at the bulge. These polyamines, however, suffered from an inefficient oligomer synthesis that was not possible to develop into a combinatorial library. Due to the limited structural information available on TAR-binding molecules, we felt that expanding our strategy to allow for combinatorial library synthesis and subsequent screening was more likely to yield positive results. As such, we designed a new class of molecules, polyamino-amido oligomers (PAAs), that could be synthesized through modified solid phase peptide synthesis. The design principle of these molecules is illustrated in Figure 1. The secondary amines in the backbone, which should be protonated at physiological pH, are spaced at intervals corresponding to the spacing of the phosphate groups of the RNA backbone in order to form ammonium-phosphate salt bridges. From this general backbone scheme, amino acid derived sidechains can be introduced to direct binding specificity. While molecular modeling is often used as a tool for drug discovery, we felt that diversity-oriented approaches were more likely to yield positive results due to the dynamic nature of TAR RNA and the limited rational-based information available for design of small-molecule RNA binders. Furthermore, we wanted to reduce the number of highly exposed cationic charges that are prevalent in most RNA-binding small molecules which convey high binding affinity but also afford low specificity. In this paper, we describe the synthesis of the monomeric building blocks for PAAs as well as a parallel library synthesis. A quantum dot-based screen was implemented for an on-bead screening of the library for TAR binding. The binding affinity of hits was quantified using RNA footprinting assays.

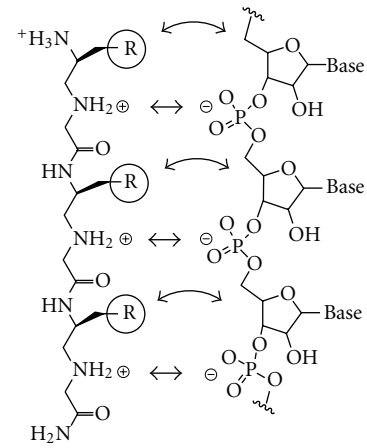


Figure 1: Structural comparison between PAA and RNA showing potential salt-bridges between ammonium groups in the PAA and the phosphodiester backbone of the RNA (straight double-headed arrows) and potential hydrophobic or hydrogen-bonding interactions between the sidechains of the PAA (R) and the RNA (curved double-headed arrow).

2. Materials and Methods

2.1. General Procedure A: Weinreb Amide Synthesis. An Fmoc-protected amino acid (10 mmol) and DIEA (1.6 mL) were dissolved in dry CH_2Cl_2 (50 mL) and cooled to 0°C. EDC (2.4 g, 12 mmol) and HOBt (1.8 g, 12 mmol) were added to the reaction. The reaction mixture was stirred at 0°C for 10 minutes to pre-activate the carboxylic acid. After 10 minutes, N,O-dimethylhydroxylamine hydrochloride (1.2 g, 12 mmol) and DIEA (2.0 mL) were added to the reaction mixture. The reaction was allowed to slowly warm to room temperature and was stirred for a total of 16 hours. The reaction was then transferred to a separatory funnel with CH_2Cl_2 (100 mL) and washed with 2 M HCl (3x, 50 mL), saturated aqueous $NaHCO_3$ (2x, 50 mL), and brine (2x, 50 mL). The organic layer was dried over sodium sulfate and solvent was removed under reduced pressure to yield pure Weinreb amides as solid white foams.

2.2. General Procedure B: Reduction of Weinreb Amides to α-Amino Aldehydes. Dry THF (50 mL) and Weinreb amide (5 mmol) were added to an oven dried 250 mL round bottom flask, placed under N_2, and cooled to 0°C in an ice bath. Lithium aluminum hydride (250 mg, 6.25 mmol) was slowly added to the solution over a period of approximately 30 seconds. The reaction was stirred vigorously at 0°C for 60 minutes and then slowly quenched with a 1 M solution of $NaHSO_4$ (50 mL). The biphasic mixture continued to stir at 0°C for 10 minutes at which point it was transferred to a separatory funnel using EtOAc (50 mL) and brine (50 mL). The aqueous layer was extracted with EtOAc (75 mL) and the combined organic layers were washed with 1.5 M HCl (2x, 50 mL), saturated aqueous $NaHCO_3$ (2x, 50 mL), brine (2x, 50 mL). The organic layer was dried over sodium sulfate and concentrated under reduced pressure. The resulting amino aldehydes were taken forward without additional purification.

2.3. General Procedure C: Reductive Amination to Form the PAA Backbone. An Fmoc-protected α-amino aldehyde (4.5 mmol) was dissolved in dry CH_2Cl_2 (30 mL) at room temperature. To the stirring solution was added a benzyl protected amino acid hydrochloride (4.9 mmol) and DIEA (0.85 mL). $NaBH(OAc)_3$ (6.3 mmol) was immediately added to the reaction mixture and stirred vigorously for 75 minutes. The reaction was quenched with a mixture of saturated aqueous K_2CO_3 (10 mL) and saturated aqueous $NaHCO_3$ (30 mL) and allowed to stir for an additional 10 minutes. The biphasic mixture was transferred to a separatory funnel and extracted with CH_2Cl_2 (3x, 30 mL). The organic layers were combined and dried over sodium sulfate, filtered, and concentrated under reduced pressure. The crude product was

purified using a biotage flash chromatography system (40+M column, 60 : 40 Hexanes: EtOAc).

2.4. General Procedure D: Boc Protection of Secondary Amine in the PAA Backbone. The PAA secondary amine backbone (2.0 mmol) was dissolved in dry CH_2Cl_2 (40 mL) and stirred at room temperature. To the solution was added di-tert-butyl dicarbonate (0.9 g, 4 mmol) and DIEA (0.6 mL, 2 mmol). The reaction was allowed to stir for 48 hours and was then transferred to a separatory funnel with CH_2Cl_2 (50 mL). The reaction mixture was washed with 1 M HCl (2x, 50 mL), saturated aqueous $NaHCO_3$ (2x, 50 mL), and brine (2x, 50 mL). The organic layer was dried over sodium sulfate and concentrated under reduced pressure. The crude product was purified using a biotage flash chromatography system (40+M column, 80 : 20 Hexanes: EtOAc).

2.5. General Procedure E: Hydrogenolysis of Benzyl Ester to Yield PAA Monomers. A Parr flask was purged with N_2 and then charged with 10% palladium on carbon. The palladium catalyst was wetted with a minimum amount of methanol (~5 mL) while under an N_2 atmosphere. The PAA-monomer ester (1.5 mmol) was dissolved in a minimal amount of methanol (typically ~20 mL) and added to the Parr flask. The flask was placed under an H_2 atmosphere (40 psi) and shaken on a Parr shaker for 2 hours. The reaction mixture was then filtered through a bed of celite to remove the palladium from the mixture. The resulting solution was concentrated under reduced pressure to yield an off-white solid as the crude product. The crude product was purified using a biotage flash chromatography system (40+M column, 0%–5% MeOH gradient in CH_2Cl_2).

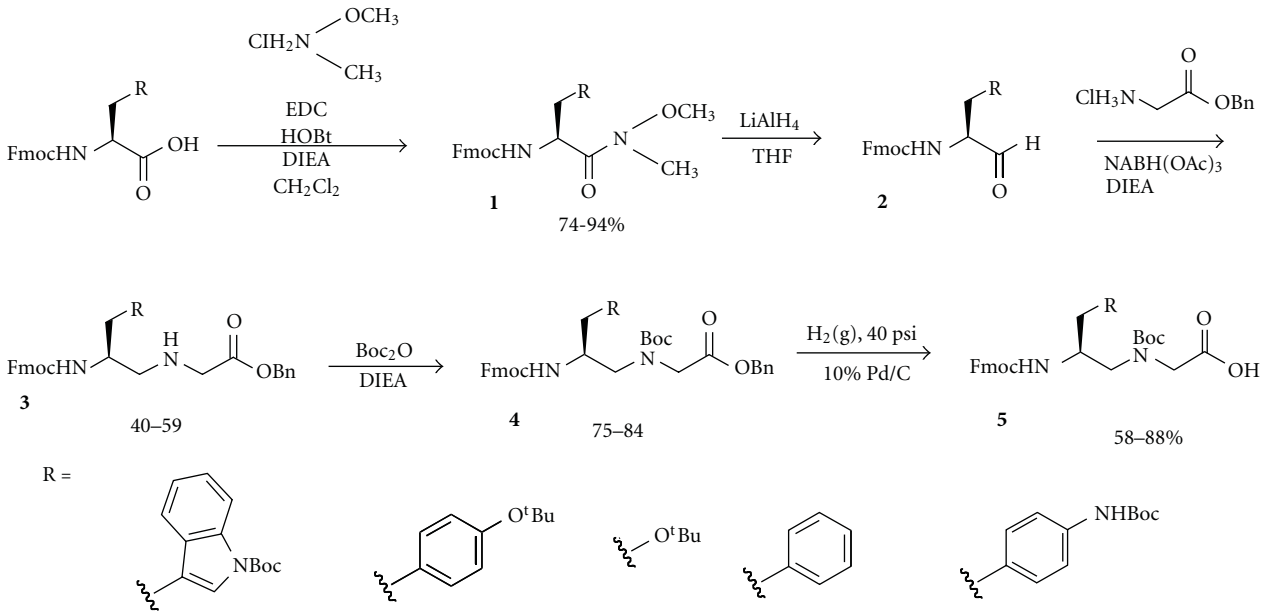
2.6. Screen for TAR RNA-Binding to PAA Library Members. Several beads from each well of the PAA library were transferred to a 384-well filter plate, keeping their spatial separation and orientation intact. The beads were first washed with water (5x, 50 μL), then 1x TK buffer (50 mM Tris, 20 mM KCl, 0.1% Triton X-100, pH 7.4; 4x–50 μL). To each well, BSA (0.1 mg/mL) was added in 1x TK buffer (20 μL) and agitated with mechanical shaking for 60 minutes at room temperature. The microplate was drained under vacuum and washed with 1x TK buffer (3x, 50 μL). Following BSA blocking, bulge mutant TAR in 1x TK buffer (2.5 μM, 20 μL/well) was added to each individual well. The library was incubated with bulge mutant TAR for 24 hours at 4°C before being drained under vacuum. Immediately following solvent removal, a mixture of bulge mutant TAR (2.5 μM) and 5′-biotin labeled TAR (Dharmacon, 250 nM) in 1x TK buffer (20 μL/well) were introduced to the library. (Note: RNA was snap-cooled by heating at 95°C for 5 minutes followed by an immediate transfer to dry ice for 5 minutes to promote hairpin formation.) The library was incubated with this solution for 2.5 days at 4°C, drained, and washed with water. To each well a solution of Qdot605 (50 nM, 15 μL/well) in 1x TK buffer was added and agitated at room temperature for 3 hours. The solution was drained and each well was washed with 1x TK (3x–50 μL), followed by a 2 hour wash with 1x TK buffer and drainage under vacuum. The library was then visualized using a fluorescent microscope equipped with a triple bandpass filter. Beads that appeared red or orange under the microscope were selected for further characterization while those that were green were disregarded.

3. Results and Discussion

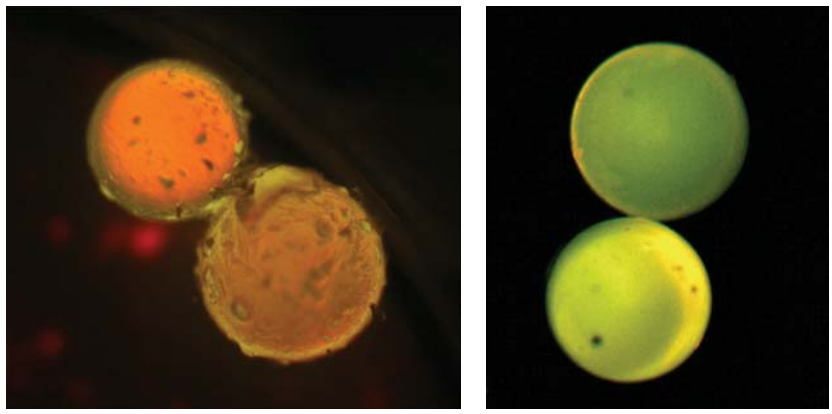
3.1. Synthesis of Polyamines. Polyamine monomers were synthesized starting from commercially available Fmoc-protected amino acids utilizing a solution-phase reductive amination strategy. Initially we sought to synthesize five polyamine monomers starting from orthogonally protected serine, tryptophan, tyrosine, 4-amino-phenylalanine, and phenylalanine (Scheme 1). The monomers were selected based on their likelihood to interact with folded RNA, thus amino acids capable of π-stacking, hydrophobic interactions, and hydrogen bonding were selected. We chose to exclude certain high-affinity moieties, such as guanidinium groups, due to their strongly cationic nature and tendency to interact non-specifically. This choice aligns with our design strategy, in which positive charge is sequestered to the backbone to reduce non-specific charge-charge interactions with the RNA backbone. The Fmoc-protected amino acid was first converted to the Weinreb amide ***(1)*** in high yield under EDC-mediated amide bond forming conditions. Subsequently, ***1*** was reduced to the corresponding aldehyde ***(2)*** using lithium aluminum hydride. The aldehyde product was taken forward without additional purification to the reductive amination step where ***2*** was condensed with the hydrochloride salt of benzyl glycinate. After imine formation, reduction with sodium triacetoxy borohydride provided ***3***. Following chromatographic purification, the secondary amine was protected with a Boc group to afford the PAA monomer ester. Hydrogenolysis of the benzyl ester yielded the PAA monomer ***(5)***.

From these five monomers, a library of 125 PAA trimers was synthesized in parallel on solid support. The synthesis was designed to be compatible with the development of an on-bead screen for RNA binding. In order to create a library suitable for aqueous screening conditions, Tentagel-NH_2 resin was chosen as a synthetic platform due to its unique ability to swell in both aqueous and organic solvents. In addition, the library synthesis was performed in 96-well filter plates to provide a physical separation between distinct PAAs. The strategy outlined provided an accessible synthetic platform that negated the need for molecular deconvolution during the screening process. The synthesis began by functionalizing the resin with an Fmoc-β-alanine spacer. The spacer was deprotected and PAA trimers were synthesized through HATU-mediated solid phase peptide synthesis (see supplementary material available online at doi:10.1155/2012/971581). Upon completion of the trimers, a global deprotection of the backbone and sidechain protecting groups afforded a 125-member library of resin-bound PAA trimers.

3.2. On Bead Screen for TAR Binding. Initially we aimed to develop a simple fluorescent screening procedure for



Scheme 1: Synthesis of PAA monomers.



(a) (b)

Figure 2: Representative pictures from screen of PAA library. The left hand picture represents a positive hit, while the picture on the right is a negative result within our screen.

assessing TAR binding to our resin-bound PAAs. To this end, we adapted protocols developed by the Rana and Kodadek laboratories to our system [12, 14–16]. Since on-bead binding and high activity in solution are not necessarily correlated, our assay design was carefully chosen based on validated literature results. The incubation conditions mimicked those reported by Rana [12], whose bead-based assays showed strong positive correlation to appropriate off-bead activity measurements. The PAA-containing beads were first incubated with a solution of BSA to block any nonspecific binding to the bead surface. Subsequently, the library was incubated with a bulge mutant TAR construct containing a single base bulge, rather than the wildtype trinucleotide bulge. The bulge mutant acted as a competitive inhibitor to exclude compounds that were not specific for the bulge region of TAR. Next, 5′-biotinylated wildtype TAR was added to the resin in the presence of an excess of bulge mutant TAR. Binding events were visualized using a fluorescent microscope equipped with a triple bandpass filter after the addition of streptavidin-coated quantum dots (Qdot605). This method allowed for the visualization of positive beads as red in color where beads containing nonbinding polyamines were seen as green under the microscope (Figure 2). From this initial screen the six brightest beads were selected, as determined by visual inspection and verified independently by a second researcher, and resynthesized on a larger scale to allow determination of binding constants. To describe the PAAs, the standard one-letter amino acid abbreviations are used to describe the sequence of sidechains with the understanding that in this paper the backbone is represented

by the chemical structure in Figure 1 for a PAA backbone. The six PAAs selected based on the screen were SFF, YFF, FFF, SYS, YSF, and FYY.

3.3. Terbium Footprinting Experiments. Previous work in our lab found RNA footprinting studies using terbium (III) ions as an RNA cleavage agent to be a reliable method for quantification of polyamine binding to TAR RNA [13, 17]. The six selected PAAs were synthesized, purified by reversed phase HPLC, and quantified by UV absorbance. The PAAs were then titrated into buffered solutions containing TAR up to 1 mM concentrations and effects on RNA cleavage patterns were assessed as a function of PAA concentration via denaturing gel electrophoresis. The results of these experiments identified the sequence XFF as a binding motif specific for the bulge region of TAR. Three of the six ligands selected showed appreciable binding affinity for the bulge region, with SFF, YFF, and FFF exhibiting binding constants in the low micromolar range (Figure 3). The best ligand derived from our initial screen was SFF (Figure 4), exhibiting a K_D of 14 μM for the bulge region. Additionally, no binding was observed for the loop region of the TAR RNA, suggesting that our screen effectively selected for molecules that specifically interacted with the desired bulge region rather than nonspecific RNA binders (see supplementary material).

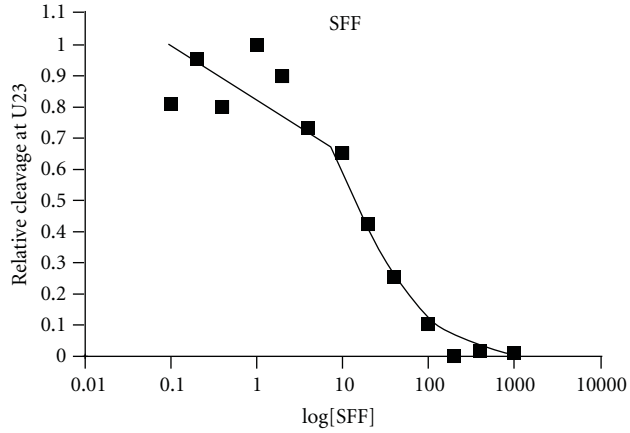
3.4. Alanine Scan. We next sought to probe the importance of sidechain interactions in an effort to define a minimal binding motif for the bulge region. The most direct way for us to probe this question was to substitute each sidechain in our best ligand (SFF) with a moiety deemed to be unlikely to interact with the RNA. In analogy to alanine scans in peptides, we synthesized the alanine-derived PAA monomer and iteratively replaced each monomer in the SFF PAA with one bearing a methyl sidechain [18]. The three PAAs were synthesized and subjected to terbium footprinting assays. We found that all three sidechain replacements yielded compounds with very low affinity for the TAR target, thus leading us to conclude that all three sidechains were critical to maintain binding affinity.

3.5. Secondary Library Synthesis and Screen. The PAA scaffold was intentionally designed such that modifications to the backbone could be easily installed in order to determine structural activity relationships (SARs). Therefore, modification to the backbone of the SFF core motif could lead to enhanced binding affinity for TAR. In this vein, we aimed to synthesize a library of compounds based on the SFF core but bearing sidechains in place of the glycine subunit of the PAA monomers. Six new monomers were synthesized, where the key step involved condensation between serine or phenylalanine aldehydes and the hydrochloride salts of either tryptophan, tyrosine, or lysine (Figure 5). The new monomers and the original serine and phenylalanine monomers were incorporated into a 64-member PAA library, where all library members contained the previously identified SFF core. The library was again subjected to the Qdot based screen and the six brightest hits were chosen for further

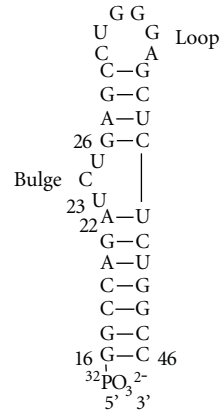
Sequence	K_D (μM)[a,b]
SFF	14.3
YFF	18.3
FFF	30.0
SYS	ND
YSF	ND
FYY	ND

[a]$K_D$was determined by assessing cleavage patterns at U23 in the TAR bulge.
[b]ND indicates a footprinting assay that relevant data could not be extracted from due to non-specific or very weak binding.

(a)



(b)



(c)

Figure 3: (a) Sequences of PAA "hits" from the initial screen and their corresponding dissociation constants for the bulge region of TAR. (b) Binding curve for SFF as determined by quantifying changes in cleavage at U23 as a function of SFF concentration (c) Structure of TAR RNA hairpin.

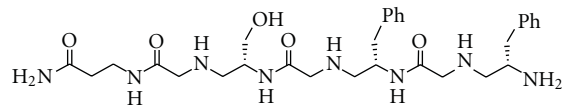


Figure 4: Structure of SFF ligand.

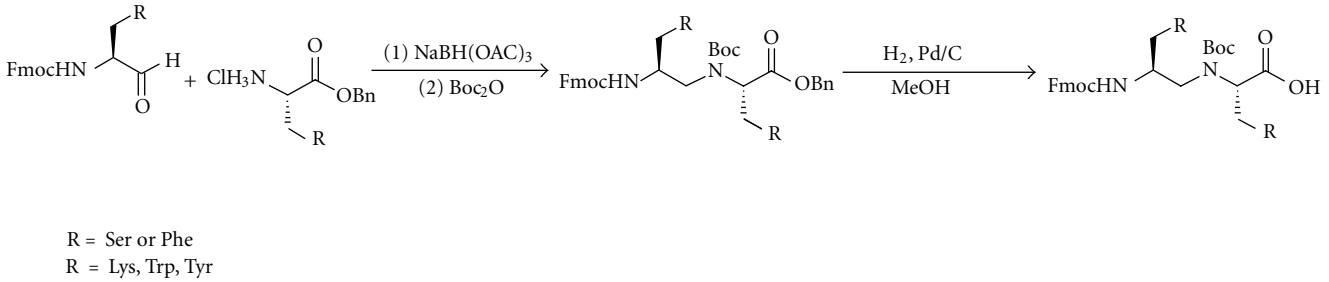


Figure 5: Synthesis of secondary library monomers.

characterization. Unfortunately, all PAA derived from the second screen showed either very weak affinity for TAR or no affinity at all in the footprinting assays (see supplementary material).

As biological knowledge has advanced, RNA has emerged as a viable pharmaceutical target. To date, there are few *de novo* designed molecules that are capable of specifically interacting with folded RNA. In this paper, we have presented a class of readily synthesized oligomers that can specifically bind to the bulge region of folded TAR RNA. In developing these molecules, we implemented a quantum dot based fluorescent screening protocol that readily identified ligands for RNA. Our screening protocol allowed for the determination of a minimal binding motif from the initial PAA library. Quantification of binding affinity via footprinting analysis yielded a low micromolar binder of TAR RNA specific for the bulge region. We anticipate that further understanding of RNA-mediated biological pathways will serve to validate RNA as a viable drug target. Consequently, the ability to rapidly develop drug candidates that are able to distinguish between subtle differences in RNA secondary structure will be of the utmost importance. We feel that the methods and results presented in this work represent a step toward developing a general class of RNA binding molecules that can suit this purpose (see supplementary material).

Acknowledgment

This work was supported by the National Institutes of Health, NIDDK intramural research program.

References

[1] J. R. Thomas and P. J. Hergenrother, "Targeting RNA with small molecules," *Chemical Reviews*, vol. 108, no. 4, pp. 1171–1224, 2008.

[2] B. Berkhout, R. H. Silverman, and K. T. Jeang, "Tat trans-activates the human immunodeficiency virus through a nascent RNA target," *Cell*, vol. 59, no. 2, pp. 273–282, 1989.

[3] B. Berkhout and K. T. Jeang, "Trans activation of human immunodeficiency virus type 1 is sequence specific for both the single-stranded bulge and loop of the trans-acting-responsive hairpin: a quantitative analysis," *Journal of Virology*, vol. 63, no. 12, pp. 5501–5504, 1989.

[4] P. Wei, M. E. Garber, S. M. Fang, W. H. Fischer, and K. A. Jones, "A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA," *Cell*, vol. 92, no. 4, pp. 451–462, 1998.

[5] B. Klaver and B. Berkhout, "Evolution of a disrupted TAR RNA hairpin structure in the HIV-1 virus," *EMBO Journal*, vol. 13, no. 11, pp. 2650–2659, 1994.

[6] F. Kashanchi, M. R. Sadaie, and J. N. Brady, "Inhibition of HIV-1 transcription and virus replication using soluble Tat peptide analogs," *Virology*, vol. 227, no. 2, pp. 431–438, 1997.

[7] D. K. Malik, S. Baboota, A. Ahuja, S. Hasan, and J. Ali, "Recent advances in protein and peptide drug delivery systems," *Current Drug Delivery*, vol. 4, no. 2, pp. 141–151, 2007.

[8] F. Hamy, E. R. Felder, G. Heizmann et al., "An inhibitor of the tat/TAR RNA interaction that effectively suppresses HIV-1 replication," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 8, pp. 3548–3553, 1997.

[9] T. Klimkait, E. R. Felder, G. Albrecht, and F. Hamy, "Rational optimization of a HIV-1 Tat inhibitor: rapid progress on combinatorial lead structures," *Biotechnology and Bioengineering*, vol. 61, no. 3, pp. 155–168, 1998.

[10] M. S. Lalonde, M. A. Lobritz, A. Ratcliff et al., "Inhibition of both HIV-1 reverse transcription and gene expression by a cyclic peptide that binds the Tat-Transactivating response element (TAR) RNA," *PLoS Pathogens*, vol. 7, no. 5, Article ID e1002038, 2011.

[11] Y. H. Niu, A. Jones, H. F. Wu, G. Varani, and J. F. Cai, "γ-AApeptides bind to RNA by mimicking RNA-binding proteins," *Organic & Biomolecular Chemistry*, vol. 9, pp. 6604–6609, 2011.

[12] S. Hwang, N. Tamilarasu, K. Ryan et al., "Inhibition of gene expression in human cells through small molecule-RNA interactions," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 23, pp. 12997–13002, 1999.

[13] G. R. Lawton and D. H. Appella, "Nonionic side chains modulate the affinity and specificity of binding between functionalized polyamines and structured RNA," *Journal of the American Chemical Society*, vol. 126, no. 40, pp. 12762–12763, 2004.

[14] P. G. Alluri, M. M. Reddy, K. Bachhawat-Sikder, H. J. Olivos, and T. Kodadek, "Isolation of protein ligands from large peptoid libraries," *Journal of the American Chemical Society*, vol. 125, no. 46, pp. 13995–14004, 2003.

[15] H. J. Olivos, K. Bachhawat-Sikder, and T. Kodadek, "Quantum dots as a visual aid for screening bead-bound combinatorial libraries," *ChemBioChem*, vol. 4, no. 11, pp. 1242–1245, 2003.

[16] T. Kodadek and K. Bachhawat-Sikder, "Optimized protocols for the isolation of specific protein-binding peptides or

peptoids from combinatorial libraries displayed on beads," *Molecular BioSystems*, vol. 2, no. 1, pp. 25–35, 2006.

[17] N. G. Walter, N. Yang, and J. M. Burke, "Probing non-selective cation binding in the hairpin ribozyme with Tb(III)," *Journal of Molecular Biology*, vol. 298, no. 3, pp. 539–555, 2000.

[18] G. Corzo, J. K. Sabo, F. Bosmans et al., "Solution structure and alanine scan of a spider toxin that affects the activation of mammalian voltage-gated sodium channels," *Journal of Biological Chemistry*, vol. 282, no. 7, pp. 4643–4652, 2007.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 923214, 11 pages
doi:10.1155/2012/923214

Review Article

Directed Evolution of Proteins through *In Vitro* Protein Synthesis in Liposomes

Takehiro Nishikawa,[1] Takeshi Sunami,[1,2] Tomoaki Matsuura,[1,3] and Tetsuya Yomo[1,2,4]

[1] *ERATO Japan Science and Technology (JST) and Yomo Dynamical Micro-Scale Reaction Environment Project, Graduate School of Information Sciences and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan*
[2] *Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan*
[3] *Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan*
[4] *Graduate School of Frontier Biosciences, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan*

Correspondence should be addressed to Tetsuya Yomo, yomo@ist.osaka-u.ac.jp

Received 8 June 2012; Accepted 10 July 2012

Academic Editor: Hiroshi Murakami

Directed evolution of proteins is a technique used to modify protein functions through "Darwinian selection." *In vitro* compartmentalization (IVC) is an *in vitro* gene screening system for directed evolution of proteins. IVC establishes the link between genetic information (genotype) and the protein translated from the information (phenotype), which is essential for all directed evolution methods, by encapsulating both in a nonliving microcompartment. Herein, we introduce a new liposome-based IVC system consisting of a liposome, the protein synthesis using recombinant elements (PURE) system and a fluorescence-activated cell sorter (FACS) used as a microcompartment, *in vitro* protein synthesis system, and high-throughput screen, respectively. Liposome-based IVC is characterized by *in vitro* protein synthesis from a single copy of a gene in a cell-sized unilamellar liposome and quantitative functional evaluation of the synthesized proteins. Examples of liposome-based IVC for screening proteins such as GFP and β-glucuronidase are described. We discuss the future directions for this method and its applications.

1. Introduction

Protein engineering is a technology that tailors a protein to function in a desired way. Rational design and directed evolution are two major approaches for introducing a change into the amino acid sequence of proteins. As a small change in the protein sequence can induce critical functional changes in proteins, altering the amino acid sequence is a crucial step in these approaches; the amino acid sequences are primarily altered by introducing mutations in the gene that encodes the protein of interest. In site-directed mutagenesis, specific mutations to the DNA sequence are introduced, which yields a desired function if the relationship between protein structure and function is clearly understood. However, directed evolution of proteins is based on Darwinian selection and thus does not necessarily require knowledge of the relationship between protein sequence and function [1, 2]. Using this method, mutations are generated through techniques such as random mutagenesis, recombination, or site-directed diversification [3]. Subsequently, the protein variants are synthesized from the mutated genes using living hosts (cells) or an *in vitro* transcription-translation system (IVTT), and they are screened for the desired function. Therefore, the methods used for directed evolution can be categorized as "*in vivo*" and "*in vitro*" approaches.

The difference between these two approaches (*in vivo* and *in vitro* approach) is the way that the genotype (genetic information encoding a protein) and a phenotype (the protein synthesized from the gene and its function) are linked for the genes of interest (Figure 1). Through an *in vivo* approach, the genotype-phenotype link is produced by using a living cell. For example, cell-surface display is an *in vivo* screening technique that uses a fusion protein to localize protein molecules to a cell membrane surface. Target proteins fused with a membrane protein are displayed on the cell membrane surface, screened for the desired function by

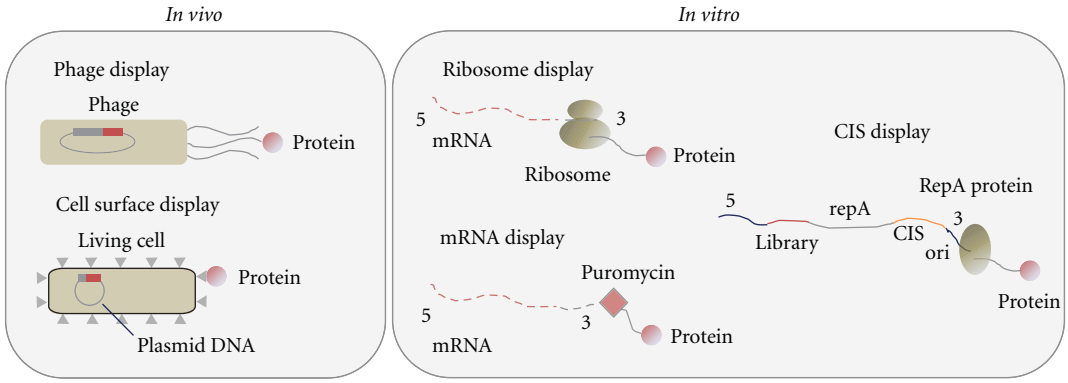


Figure 1: Genotype (genetic information)-phenotype (protein synthesized from the gene and its function) linkage and screening techniques for directed evolution of proteins. Screening techniques are categorized as *in vivo* and *in vitro* approaches.

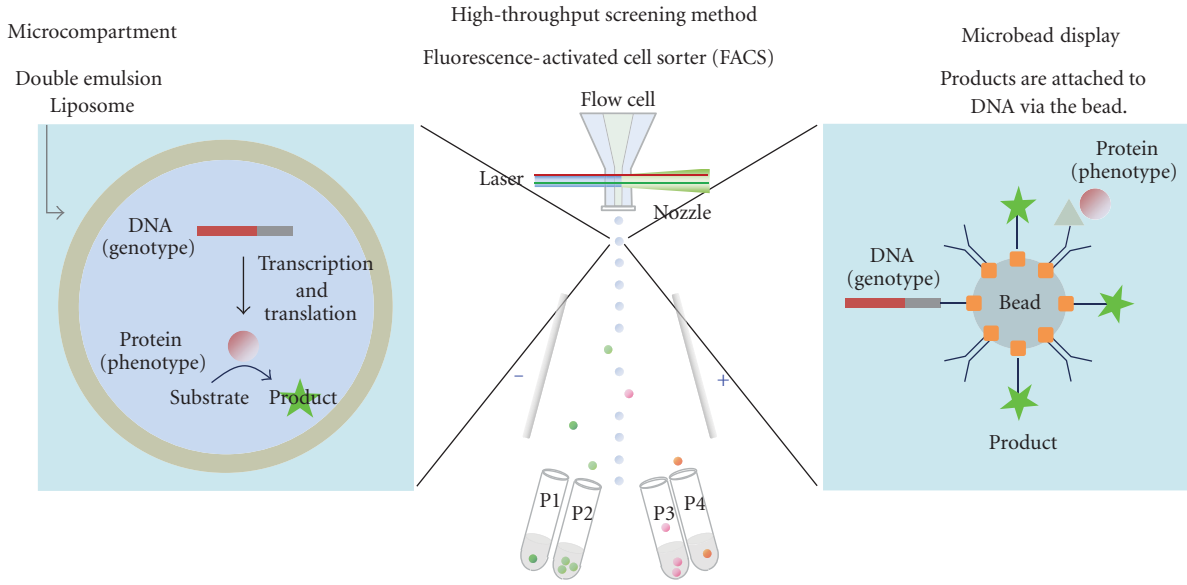


Figure 2: The underlying concept for *in vitro* compartmentalization (IVC) using double emulsion or liposome (left) and microbead (right). In both cases, a fluorescence-activated cell sorter (FACS) (center) is used for high throughput screening.

exposure to a colorimetric detection reagent, and selectively sorted using a fluorescence-activated cell sorter (FACS) [4, 5]. Phage display is another *in vivo* display technique that uses a phage for gene storage and protein display. In this technique, target proteins are fused with phage coat proteins (*g8p* or *g3p*) and displayed on a phage surface. These *in vivo* screening techniques have been applied to the directed evolution of proteins. However, these techniques are applicable to a limited number of proteins that are not toxic to growth of the host cell. Low transformation efficiency also limits genetic diversity (library size) by up to 10^8.

To overcome these technical drawbacks in *in vivo* techniques, *in vitro* display was proposed as a new display technique [6, 7]. In this technique, protein variants are synthesized from the gene using an IVTT, and the gene (genotype) is physically or covalently tethered to the translated protein (phenotype) via an adaptor or linker, such as ribosomes (ribosome display) [8], RepA (CIS display) [9], and puromycin (mRNA display) [10]. The proteins linked to the mutant gene are screened for the desired function. These *in vitro* display methods are suitable for improving protein equilibrium affinity, off rate, stability, and folding [8]. However, these display techniques are not suitable for improving the catalytic activity of enzymes because they rely on binding affinity between the displayed protein and an immobilized ligand for the screen [11]. *In vitro* compartmentalization (IVC) is a solution to direct screening for enzymatic reaction turnover entirely *in vitro*. The primary idea underlying IVC is that a DNA, an IVTT, and a fluorogenic detection reagent are encapsulated in a cell-like compartment to form a genotype-phenotype linkage (Figure 2, left). Proteins are translated from a single gene using an IVTT in each compartment, and they yield a fluorescent product that is screened directly for the catalytic activity of interest using an FACS [11]. We introduce herein the earlier studies on IVC-based directed evolution

of proteins, where water-in-oil (W/O) emulsions were used as microcompartments. We then introduce the IVC using cell-sized lipid vesicles, liposomes. Firstly, the technology underlying protein synthesis using an IVTT inside liposome is described. Then, construction of the liposome-based gene screening system using FACS and examples of the application of the liposome-based IVC to directed evolution of proteins are described. Finally, we remark on the future directions for liposome-based IVC in directed evolution that are impossible with other IVC techniques.

2. *In Vitro* Compartmentalization (IVC)

2.1. Emulsion-Based IVC. *In vitro* compartmentalization (IVC) is a technique for linking genotype to phenotype. Unlike other techniques used in conventional *in vitro* display, IVC does not connect directly the gene and encoded protein. IVC utilizes microcompartments for genotype-phenotype linkage. A single DNA and an IVTT are encapsulated in a microcompartment (Figure 2, left). Proteins encoded by the gene accumulate inside the microcompartment through *in vitro* protein synthesis. Colocalization of the gene and protein links the genotype and phenotype. W/O emulsion was first utilized for microcompartments in IVC-based genetic screening. With this technique, genes encoding the DNA methyltransferase *M. HaeIII* were enriched from a mixture containing 10^7-fold excess genes encoding dihydrofolate reductase [12]. Furthermore, toward high-throughput gene screening using an FACS, microbead display using IVC (Figure 2, right) was performed to screen catalytic activity of enzymes with a soluble non-DNA substrate [13]. This technique enables us to evaluate the catalytic activity of enzyme encapsulated in cell-size microcompartments under a variety of conditions that can inhibit the *in vitro* protein synthesis, because the evaluation of catalytic activity is separated from the protein synthesis. As a next advancement of IVC, water-in-oil-in-water emulsion (double emulsion) was adapted and enabled direct sorting of intact emulsion droplets. This double emulsion technique was first demonstrated through model selection of emulsion droplets encapsulating *FolA* genes from a droplet mixture with two separate W/O emulsions: a positive emulsion containing *FolA* genes and a fluorescent marker as well as a negative emulsion containing *M. HaeIII* genes and no fluorescent marker [14]. Reemulsification of W/O emulsion droplets in the aqueous phase creates double emulsion droplets, which can be directly analyzed and sorted using an FACS. Using the emulsion-based IVC and *in vitro* protein synthesis, Ebg, which is an *E. coli* protein of unknown function, was evolved into mutant proteins with β-galactosidase catalytic activity [15]. Single genes from the mutation library for *Ebg* as well as an IVTT and a fluorogenic substrate were compartmentalized in a W/O emulsion droplet. In an emulsion droplet, Ebg variants are translated from the mutant gene and yield fluorescent product if the variants express β-galactosidase catalytic activity. After reemulsification of the W/O emulsion in the aqueous phase, double emulsion droplets were screened directly for β-galactosidase activity (through the fluorescent signal from turnover reaction products).

2.2. Advantages and Limitations of Emulsion-Based IVC. Emulsion-based IVC is suitable for the quantitative screening of enzyme variants using an FACS because each emulsion droplet yields a fluorescent signal, which reflects the enzymatic activity of each variant. Other *in vitro* display techniques involve screening based on a binding event between a displayed protein and immobilized ligand and are not adapted for observing a catalytic turnover event. Although emulsion-based IVC has been useful and successful for directed evolution of enzymes, this method has two technical limitations. The first limitation concerns the stringency of the genotype-phenotype link (Figure 3, right). Double emulsion droplets containing multiple compartments are formed when W/O emulsion is reemulsified in an aqueous phase. During the reemulsification process, two types of microcompartments can be entrapped in a double emulsion droplet; one microcompartment can encapsulate the gene of interest and the other can encapsulate an unrelated gene. The genotype-phenotype link would be severed if two different mutant genes were in a double emulsion droplet for sorting using an FACS [16]. One approach to overcome the issues from multiple compartments is through a high-throughput screening platform using droplet-based microfluidics [16]. This screening platform comprises a droplet generation device (droplets for gene amplification), droplet fusion device (electrocoalescence between droplet pairs of a gene-containing droplet and an IVTT-containing droplet for the genotype-phenotype link), and sorting device (for recovery of the genes of interest). The second limitation of the emulsion-based IVC is a technical hurdle for its application using a variety of protein classes, such as membrane proteins (Figure 3, left), which cannot be overcome by the use of the aforementioned droplet-based microfluidics.

For the technical issue of single droplet containing substructures, and that preclude membrane protein use in directed evolution of proteins (Figure 3), cell-sized microcompartments with a phospholipid bilayer membrane are ideal solution for both issues [17–19]. We have been studying *in vitro* protein synthesis in liposomes [20–23] and constructed a high-throughput gene screening system using liposomes (liposome-based IVC) and an FACS [24, 25]. Our experimental system for protein synthesis in liposomes comprises a liposome as the bioreactor, chemical components for protein synthesis, and analytical tools for quantitation of the proteins produced. The following sections survey the liposomes used in preparation methods for cell-sized compartments (Section 3), *in vitro* protein synthesis using a PURE system (Section 4), high-throughput analysis using an FACS (Section 5), and liposome-based IVC for directed evolution of proteins (Section 6).

3. Liposomes as Cell-Sized Microcompartments

3.1. Liposomes. A phospholipid vesicle is a spherical hollow capsule that has an inner aqueous phase surrounded by a phospholipid bilayer membrane. The vesicular structure is formed spontaneously by dispersing phospholipids in an aqueous medium (Figure 4). Vesicle formation from egg lecithin was first reported by Bangham and Horne in

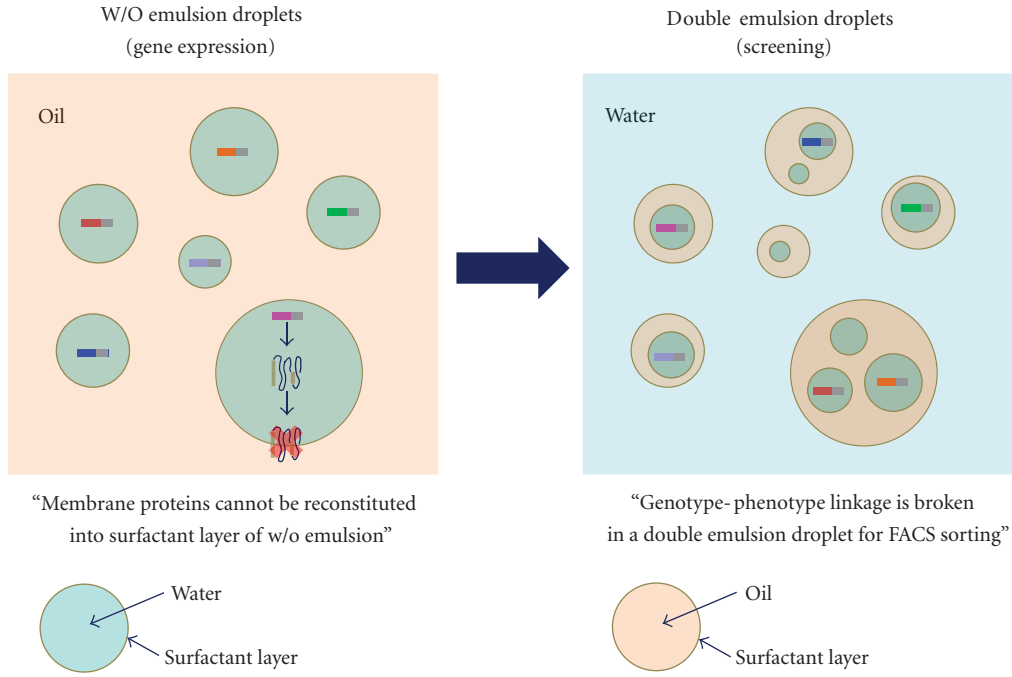


Figure 3: The technical limitations of emulsion-based IVC due to the "W/O emulsion" and "double emulsion" structures.

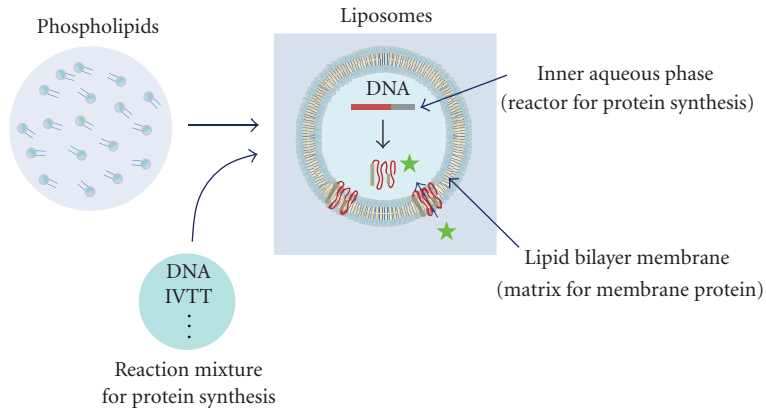


Figure 4: Liposomes as a platform for *in vitro* protein synthesis. Reaction mixture for protein synthesis constitutes an inner aqueous phase of liposome, in which protein is synthesized from a DNA. Membrane protein synthesized inside liposome can be embedded into a lipid bilayer membrane.

1964 [26]. They observed dried samples from an aqueous dispersion of lecithin by electron microscopy and discovered a spherical structure with a 4.4 nm thick lipid layer of lamellae. "Liposome" is a term for a phospholipid vesicle and was proposed by Sessa and Weissmann in 1968 [27]. This term is generally accepted. Since the first report by Bangham, liposomes have been utilized in various biophysical and biochemical studies, including model membranes, microreactors, supramolecular assemblies for biomimetic systems, and drug carriers for drug delivery systems [28]. Currently, liposome-related studies are motivated by a growing interest in "synthetic cells" [29] and the "origin of life" [29], both of which are intended to address how living things might emerge from nonliving matter [30]. The recent trend in liposome-related studies regards liposomes as a protocell model in which biochemical reactions inside a living cell are executed by filling liposomes with the required components. *In vitro* protein synthesis in liposomes and its application to genetic screening (liposome-based IVC) are examples of bioengineering as well as liposome-related studies.

3.2. Preparation Methods for Cell-Sized Liposomes. Liposomes are diverse in size (from several tens of nm to hundreds of μm in diameter), lamellarity (singly lamellar or multilamellar), and internal structure (single compartment or multiple compartments). This diverse structure depends

on the liposome preparation methods. However, not all sizes of liposomes are applicable as microcompartments for IVC-based gene screening due to detection limits (approximately 1 μm in diameter) in FACS measurements (see Section 5 for details). Therefore, the liposome size suitable for this experiment is similar to a cell size (larger than 1 μm in diameter), and such a cell-sized liposome is referred to as a "giant liposome" [31]. Giant liposomes are primarily generated using the "hydration of thin film method," "rehydration of freeze-dried empty liposome (FDEL) method," or "inverted-emulsion method" [32].

In liposome-based IVC, a single DNA and an IVTT are encapsulated in the same giant liposome to link genotype and phenotype (Figure 4). Feasibility of liposome preparation, encapsulation of the reactants for *in vitro* protein synthesis, and the internal structure of the liposome are significantly influenced by the preparation method for the giant liposome. For the hydration of thin film or rehydration of freeze-dried empty liposome methods of giant liposome preparation, the liposomes are formed by reconstituting dried lipid film with a reaction buffer for protein synthesis. The advantage of the hydration of thin film method is that liposomes can be prepared using various phospholipids irrespective of electrical charge. However, the disadvantage of the method is that a relatively large quantity of reaction mixture is necessary for swelling the dried thin film of the lipids during liposome preparation, and macromolecular compounds in the reaction mixture are difficult to trap in the liposomes [32]. When giant liposomes are prepared by the rehydration of freeze-dried empty liposomes, the liposome structure is sufficiently strong to withstand the osmotic pressure change in the outer solution and the sorting operation during the FACS screen [33]. However, the giant liposomes generated by this method are unsuitable for quantitative evaluation of in-liposome protein synthesis using an FACS because certain giant liposomes have multiple compartments and lamella; thus, the liposome size measured using an FACS does not represent the compartment size for protein synthesis [34]. Consequently, liposomes with a single compartment and lamella (e.g., giant unilamellar liposomes) are required for quantitative evaluation of in-liposome protein synthesis (liposome size and product quantity) by high-throughput analysis using an FACS.

The inverted-emulsion method is a preparation method for giant unilamellar liposomes [35]. This method comprises the following steps. The aqueous phase is emulsified in an oil phase containing phospholipids to prepare a water-in-oil emulsion. The emulsion is layered on an outer aqueous solution and centrifuged to sediment the emulsion droplets towards the oil-water interface where a lipid monolayer forms. The emulsion droplets generate a second lipid layer upon crossing the interface and transform into unilamellar liposomes in the outer aqueous solution. When this type of giant unilamellar liposome is applied to *in vitro* protein synthesis, the compartmentalized reaction mixture is separated from the outer aqueous phase until the liposomes are formed. Thus, researchers can control the composition of both inner and outer aqueous phases. The inverted-emulsion method is promising for construction of a suitable microcompartment to quantitatively evaluate in-liposome protein synthesis and reconstitution of membrane proteins.

4. Protein Synthesis in Liposomes

4.1. The PURE System for In Vitro Protein Synthesis. The primary components for *in vitro* protein synthesis are DNA encoding the protein of interest, an IVTT, and a detection reagent. These components should be encapsulated firmly in a liposome when *in vitro* protein synthesis is performed in liposomes (Figure 4). An IVTT is a multimolecular machine that facilitates protein synthesis from DNA in a test tube. Cell extracts from *E. coli*, wheat germ, rabbit reticulocytes, and insect cells have been used as IVTTs for *in vitro* protein synthesis [36]. However, cell extracts comprise certain unknown constituents. Furthermore, proteases, DNase, RNase, and intrinsic enzymes (e.g., β-galactosidase) remain in the cell extracts, and these remnants considerably decrease the production of protein and interfere with the detection of protein function. These problems are inevitable as long as cell extracts are used for *in vitro* protein synthesis. To overcome these problems, we have been using the IVTT developed by reassembling the individual components for protein synthesis, which were extracted from *E. coli* cells overexpressing the protein factors with a histidine tag and thoroughly purified. This new IVTT is referred to as a "protein synthesis using recombinant elements (PURE) system" [37]. *In vitro* protein synthesis is a coupled reaction system comprising transcription, aminoacylation of tRNA, translation, and energy source regeneration. The PURE system includes the entire reaction system and is prepared by reconstituting protein factors, ribosomes, tRNA mixture, and substrates (20 amino acids and four nucleoside triphosphates) in a buffer solution. The protein factors are T7 RNA polymerase, pyrophosphatase, 20 aminoacyl-tRNA synthetases, creatine kinase, myokinase, and nucleoside diphosphate kinase in addition to 10 translation factors (three initiation factors (IF), three elongation factors (EF), three release factors (RF), and a ribosome recycling factor (RRF)).

4.2. Protein Synthesis from a Single Gene in a Liposome. Using the experimental system with liposomes, DNA, and an IVTT, *in vitro* protein synthesis in liposomes has been studied by a number of groups [31]. The review article by Stano et al. [31] is a comprehensive survey of biomacromolecule synthesis in liposomes for the creation of semisynthetic minimal cells and provides the most recent and comprehensive list of publications on protein synthesis inside liposomes. Thus, our primary focus is on protein synthesis that begins with a single gene in a liposome, which is a crucial part of liposome-based IVC, because genotype and phenotype must be linked for the gene screening process. Our strategy, which links genotype and phenotype inside a liposome, includes DNA that encodes the protein of interest and is encapsulated at a single molecule level with the PURE system in liposomes. Liposomes in which green fluorescent protein (GFP) was translated from a single gene were successfully detected, analyzed, and sorted for a fluorescence signal from GFP using an FACS [24]. β-glucuronidase catalytic activity expressed

from a single gene inside the liposomes was also detected and quantitatively evaluated using a fluorogenic substrate and FACS [34].

5. High-Throughput Analysis of Liposomes Using an FACS

5.1. Application of an FACS to Liposome Measurement. In the liposome-based IVC, an extremely large number (more than 10^8) of liposomes are created for *in vitro* protein synthesis beginning with a single DNA. Liposomes that encapsulate a gene of interest are screened from the large population of liposomes for the desired protein function. The protein function expressed inside an individual liposome should be detected and quantitatively analyzed to identify the liposomes for sorting. Protein production and function inside the liposome are often measured and quantified using analytical tools such as a fluorescence spectrometer and fluorescence microscope [31]. A fluorescence spectrometer detects an averaged fluorescence signal from an ensemble of liposomes. A fluorescence microscope measures a fluorescence signal from an individual liposome where proteins are synthesized from genes. Microscopy measurements for liposomes provide data on the morphology (shape and size) and fluorescence intensity of a liposome (internal reaction). However, this technique is only effective for a much smaller liposome population than the population required for statistical analysis and gene screening in IVC-based directed evolution of proteins. FACS is a promising technique for observing a large population of liposomes because of its capacity for high-throughput analysis and simultaneous measurement of multiple characteristics.

An FACS is a powerful experimental apparatus for analyzing and sorting live cells simultaneously. The apparatus comprises a fluidics system for transporting one cell at a time, an interrogation system for detecting the cell by laser illumination, and a sorting system for collecting the cells of interest from one to millions of cells. Using this technique, cells exhibiting a specific biological characteristic are separated from a heterogeneous population of cells using fluorescence and light scattering from individual cells in the population. The FACS was invented in the late 1960s, commercialized in the early 1970s, and has been utilized since then for basic studies in cell biology as well as clinical applications such as diagnosis, disease classification, and *in vivo* therapies [38]. Recently, FACS measurements have not only been used for cell-oriented applications but also for molecular screening in directed evolution of proteins [15, 39]. In addition, the FACS has been utilized for measuring nonbiological particles such as submicron-size liposomes [40] and double emulsion droplets [41] for particle size and fluorescent marker entrapment. We have used FACS to characterize liposomes for structure and biochemical reactions.

We first successfully detected a GFP synthesis in liposomes using FACS based on fluorescence signals from the synthesized GFP [22]. For liposome structure, the internal aqueous volume and membrane volume of individual liposomes were quantitatively evaluated using light scattering intensity data from an FACS measurement [42]. The liposome population selected using these structural parameters was sorted using an FACS and observed by optical microscopy. The structural parameters generated using the FACS correlated with liposome structural heterogeneity, as demonstrated by microscopy observations. Population analysis of giant liposomes with an FACS was used to identify the subpopulation of unilamellar liposomes in a 2D contour map of surface area and internal aqueous volume generated for giant liposomes [43] (Figure 5). Furthermore, substructure of the multilamellar giant liposomes has been identified by encapsulating β-glucuronidase synthesis in liposome, and analyzed by an FACS [34].

5.2. Evaluation of an In-Liposome Reaction Using FACS. Analysis of biochemical reactions in liposomes using an FACS is based on a quantitative evaluation of liposome size and reaction products in liposomes. Liposome size is evaluated by measuring the fluorescence intensity of a fluorescent protein as a volume marker molecule, which is encapsulated in a liposome at a high concentration [24, 34]. The fluorescence intensity of the marker protein is converted to the number of marker molecules in a liposome and then to volume of the internal aqueous phase in the liposome. The reaction product is quantified by measuring the fluorescence intensity of new synthesized proteins or the fluorescent product of expression of a protein function [20, 25]. Through this analytical method using an FACS, a large population of liposomes can be measured for size and reaction as well as analyzed throughout a population or subpopulation that is defined by reactivity and a specific size [33]. For liposome-based IVC enzyme screening, a fluorescent volume marker and fluorogenic substrate are encapsulated in liposomes for a screening assay using an FACS (Figure 5, left). Details on this screening system are described in Section 6.

Using our system, GFP synthesis inside the liposomes was quantitatively evaluated, and the influence of lipid membrane composition on protein synthesis was discussed [44]. The study suggested that phospholipids and other liposomal membrane components for liposome preparation should neither inhibit nor impair the protein synthesis reaction steps. Furthermore, GFP synthesis inside the liposomes proceeds similarly to that in the test tube in spite that liposomes have very high surface-to-volume ratio in comparison to a test tube. This indicates that phospholipids and other liposomal membrane components for liposome preparation neither inhibit nor impair the protein synthesis reaction steps [20]. Consequently, liposome provides a reaction environment that is equally good as a test tube and provides an extremely large number (more than $10^{10}/100\,\mu$L reaction volume) of microcompartments.

6. Liposome-Based IVC for Directed Evolution of Proteins

6.1. Liposome-Based IVC. We constructed a novel gene screening system using a liposome-based IVC for directed evolution of proteins [24, 25]. Liposome-based IVC is a technique used to link genotype and phenotype. The idea

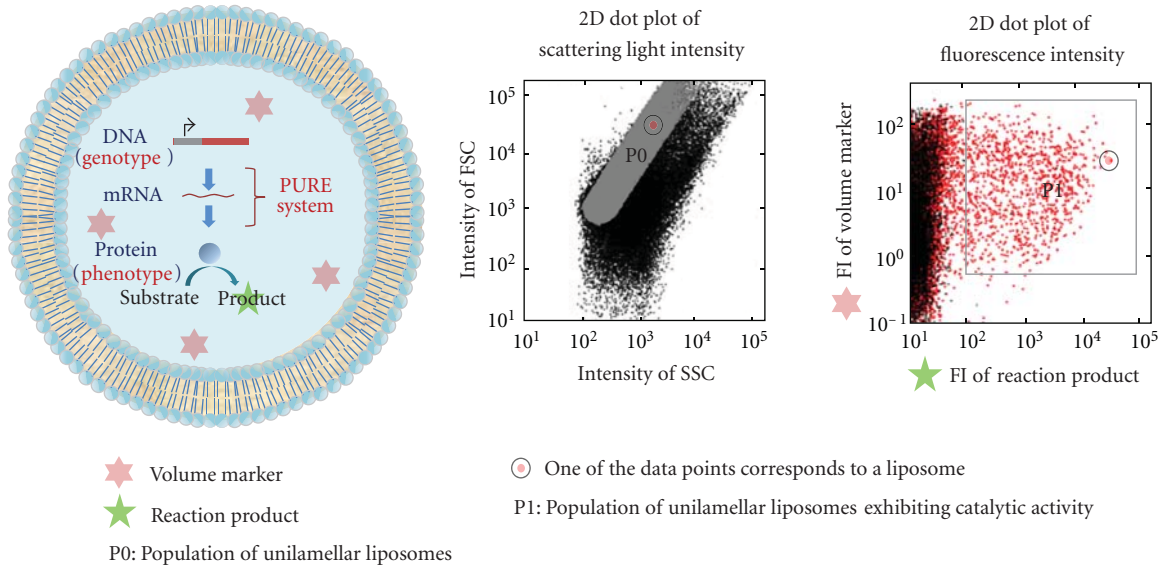


Figure 5: Characterization of in-liposome protein synthesis using an FACS. DNA, PURE system, fluorogenic substrate, and fluorescent volume marker are encapsulated in a giant unilamellar liposome (left). Subpopulation of unilamellar liposomes is represented by P0 in 2D dot plot of scattering light intensity (middle). Catalytic activity of enzymes expressed in liposomes is confirmed in P1 region of 2D dot plot of fluorescence intensity obtained by FACS measurement (right).

underlying this technique is that a mutant gene library is compartmentalized as a single molecule into cell-sized liposomes (giant liposomes) and a protein variant is synthesized from the encapsulated mutant gene through the PURE system in each liposome. Liposome-based IVC has two primary advantages over emulsion-based IVC. Unlike W/O emulsion droplets, liposomes are directly loaded onto an FACS apparatus when they are analyzed for gene screening (no reemulsification process is required for FACS analysis). The catalytic activity of an enzyme expressed in a liposome is quantitatively evaluated using an FACS when a giant unilamellar liposome is utilized for enzyme synthesis and catalytic activity expression. An FACS measurement collects signals from two or more different fluorescence colors (a fluorescent volume marker and fluorogenic substrate) from individual liposomes simultaneously and quantitates the liposome size and reaction product concentration, both of which are necessary for quantitative evaluation of catalytic activity. In addition, membrane proteins can be inserted into the phospholipid bilayer membrane of a liposome when giant unilamellar liposome is utilized for membrane protein synthesis. Membrane protein incorporated into lipid bilayer membrane is a prerequisite for quantitative evaluation of membrane protein function and subsequent genetic screening.

We first performed a pilot experiment for liposome-based IVC and demonstrated that the technique is promising for genetic screens [24]. Two GFP variants, GFPuv2 and GFPuv5, were used in the pilot experiment, and they were encoded in the pETG2tag and pETG5tag vectors, respectively. GFPuv5 emits a fluorescent signal eight times higher than GFPuv2 when excited at 488 nm. A mixture of the pETG2tag and pETG5tag DNA at molar ratio of 0.85 : 0.15 was compartmentalized into giant liposomes with the PURE system and a fluorescent volume marker. Giant liposomes were prepared by FDEL method (Section 3). After incubation for GFP synthesis, the liposomes were measured for fluorescent signals from the translated GFP as well as volume marker and sorted using the higher fluorescent intensity of GFP and a certain liposome size. The pETG5tag was enriched over 10-fold from the initial genetic mixture when the liposomes were collected from two liposome subpopulations; one subpopulation ranged from 1.4fL to 6.7fL and the other ranged from 6.7fL to 13fL. Therefore, the genotype- (GFP gene-) phenotype (GFP) link was securely constructed in individual liposomes, which encapsulated a single copy of DNA. Thus, the pilot experiment successfully showed that GFP genes encapsulated in a liposome can be screened for the fluorescence intensity from GFP emission.

However, we anticipated the following technical issue, which can be caused by multiple compartments and lamella in giant liposomes prepared by the FDEL method [34]. The issue is underestimation of catalytic activity where a gene is expressed only in a subset of the multiple compartments in a giant liposome. This yields an inaccurate evaluation of catalytic activity in individual liposomes and lower enrichment in the gene of interest. In addition, translocation of a membrane protein into the membrane of a multilamellar liposome is a technical hurdle for detection of a functional membrane protein. To solve these problems, we used a giant unilamellar liposome for liposome-based IVC. A giant unilamellar liposome was prepared using the inverted-emulsion method (Section 3.2). We constructed a genetic screening system composed of *in vitro* protein synthesis encapsulated within a giant unilamellar liposome and an FACS (Figure 6). A mock genetic library for β-glucuronidase

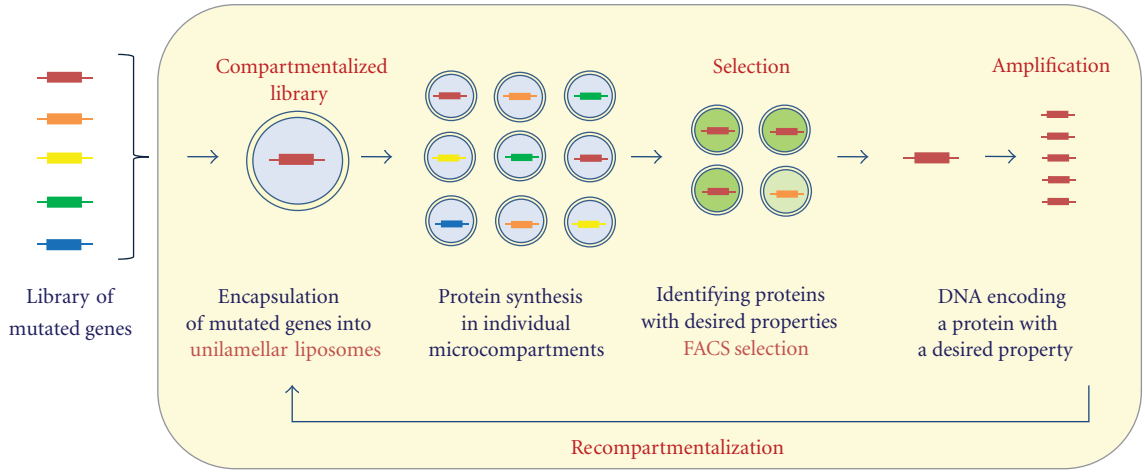


Figure 6: Flow chart for protein screening using liposome-based IVC. Library of mutated genes are compartmentalized with PURE system and other reagents into liposome. Each protein variant is synthesized from a single copy of gene in each liposome. Liposomes encapsulating the gene of interest are screened using an FACS. DNA extracted from the liposomes is amplified to be transferred to the next round of gene screening.

(GUS) was compartmentalized into liposomes as a single molecule. The liposomes that exhibited green fluorescence from hydrolysis of the fluorogenic substrate through the synthesized GUS were sorted from the subpopulation of giant unilamellar liposomes using an FACS. More than a 10-fold enrichment of the GUS gene with a higher catalytic activity was generated when a single copy of the GUS gene was encapsulated in each liposome. Quantitative analysis of the enrichment factors and their liposome size dependencies showed that the experimentally generated and theoretical values agreed. Using this method, the genes encoding active GUS were then enriched from a gene library of randomly mutated GUS genes. Only three rounds of screening were required, which was also consistent with our theoretical estimation. The consistency between the theoretical and experimental values generated using our screening system indicates that the screening system operates as expected.

6.2. Protein Evolution Directed by Compartment Size. Here, the directed evolution of protein is discussed through the effect of compartment size on protein function. Nature contains living prokaryotes with cell sizes that range from 0.02fL to 400fL [45]. The lower limit of the cell size is determined by the catalytic efficiency of enzymes, protein synthesis machinery, and machinery to cope with sudden environmental changes [45]. Under this theory, smaller cells could be generated if the enzyme catalytic efficiency was greater. Naturally occurring proteins have evolved in the cell through Darwinian selection. However, directed evolution of protein has never been discussed regarding compartment size because conventional microcompartments (emulsions) have been unsuitable for this purpose. Of the gene screening techniques for directed evolution of proteins, liposome-based IVC is the most promising technique for studying how compartment size influences protein evolution because the internal aqueous phase volume of the liposome is accurately evaluated by FACS measurement. A molecular evolution system using liposome-based IVC is an experimental approach for simulating the evolutionary process of protein function in a certain cell size.

We propose a molecular evolution system for evaluating the effect of compartment size on protein evolution. The system comprises a giant unilamellar liposome, GUS, and an FACS. Giant unilamellar liposomes are polydisperse in size ranging from 0.5fL to 250fL or larger, which includes the cell size discussed. GUS is a tetrameric enzyme, and GUS tetramer formation is a rate limiting step in catalytic activity expression [46]. Kinetic analysis of GUS tetramer formation in emulsion droplets showed that tetramer formation is susceptible to compartment size when GUS is synthesized from a single gene in a W/O emulsion droplet [47]. Monomeric GUS is prone to assemble in a smaller compartment because tetramer formation is the rate limiting step. In our molecular evolution system, a library of randomly mutated GUS genes and the PURE system are compartmentalized in giant unilamellar liposomes. GUS variants are synthesized in individual liposomes. Liposomes exhibiting GUS catalytic activity are sorted from the subpopulation defined by a certain liposome size (100fL) and green fluorescence intensity above threshold value. We predict that GUS variants prone to assemble in a larger compartment (100fL) will be generated after iterative rounds of genetic screening. Our genetic screening experiment is in progress and will continue until it generates a gene encoding active GUS variants, which are fit to a certain liposome size.

6.3. Adaptation of Membrane Protein Function to a Liposome Environment via Directed Evolution. Membrane proteins perform a variety of functions in cells, including material transport, signal transduction, and cell-cell contact. With recent progress in minimal cell research using liposomes and an IVTT, experimental methods for including membrane proteins are under development. Giant unilamellar liposomes are an ideal cell-mimetic environment because

the lipid composition can be optimized for reconstitution of membrane proteins. Although certain water soluble proteins have been synthesized using an IVTT inside liposomes (Section 4), thus far only a few membrane proteins have been synthesized inside liposomes and reconstituted into a lipid bilayer membrane. For synthesis of membrane proteins inside liposomes, a few groups have succeeded in synthesizing and characterizing α-hemolysin for membrane permeation of nutrient molecules in giant unilamellar liposomes [18] as well as sn-glycerol-3-phosphate acyltransferase (GPAT) and lysophosphatidic acid acyltransferase (LPAAT) for lipid synthesis in liposomes [48]. We believe that through liposome-based IVC development, *in vitro* molecular evolution of membrane proteins become possible. Advantages of using liposome-based IVC on the molecular evolution of membrane proteins are expected as follows. (1) Various kinds of membrane proteins can be engineered irrespective of their toxicity threatening cells' lives, (2) functions of membrane proteins can be evaluated under various reaction conditions and also in membranes with various lipid compositions. Thus, research on membrane proteins is entering a new stage for applications aimed at complex molecular machines, such as biosensors for a monitoring device, biochips for diagnosis, and biointerfaces for computing.

7. Conclusions

In this paper, we reviewed a novel *in vitro* genetic screening system comprising liposome-based IVC and an FACS. Liposome-based IVC is a new technique developed to link genotype and phenotype. This technique utilizes a giant unilamellar liposome and a PURE system. A library of mutant genes and a PURE system are compartmentalized into giant unilamellar liposomes for *in vitro* protein synthesis. A protein variant (phenotype) translated from a single DNA is colocalized with the DNA (genotype) inside a liposome. Using an FACS for high-throughput screening, liposomes encapsulating the gene of interest are sorted from a large population of liposomes using fluorescent signals generated from expression of a protein function. The genes of interest are enriched through iterative rounds of genetic screening.

With the gene screening system, genetic diversity at approximately 10^7 can be screened in a day [25]. The diversity size is sufficiently large for directed evolution of proteins. The system can screen various proteins including enzymes and membrane proteins. A large population of liposomes of different sizes (from 0.5fL to 250fL) facilitates the search for a protein function that has adapted to a certain compartment size. In addition, the semipermeable character of the liposomal membrane facilitates external feeding of a liposome compartment using additional solutes. If the protein function of interest is coupled to the external solute, then the protein screen is controlled by the timing of feeding and/or solute quantity.

Liposome-based IVC was successfully proven effective for screening a protein function under simple conditions. However, in nature, proteins must have evolved to adapt to more complex and dynamic environments where many biochemical reactions are coupled and organized to control cell behavior. This suggests that a reaction system comprising many proteins and enzymes can evolve to perform more efficiently and more productively. Liposome-based IVC will be a useful method for simulating versatile conditions by assembling the necessary components into a liposome reactor. It is expected that such liposome reactors containing a coupled reaction system have high potential as biochemical sensors for monitoring chemicals (e.g., carcinogens, toxins, and environmental hormones) and microreactors to produce biologically active substances for daily use (e.g., anticancer drugs and antibiotics) with high efficiency and selectivity.

References

[1] H. Leemhuis, R. M. Kelly, and L. Dijkhuizen, "Directed evolution of enzymes: library screening strategies," *IUBMB Life*, vol. 61, no. 3, pp. 222–228, 2009.

[2] F. H. Arnold, L. Giver, A. Gershenson, H. Zhao, and K. Miyazaki, "Directed evolution of mesophilic enzymes into their thermophilic counterparts," *Annals of the New York Academy of Sciences*, vol. 870, pp. 400–403, 1999.

[3] D. Lipovsek, M. Mena, S. M. Lippow, S. Basu, and B. M. Baynes, "Library construction for protein engineering," *Protein Engineering and Design*, pp. 83–108, 2010.

[4] S. Becker, H. Hübenreich, A. Vogel et al., "Single-cell high-throughput screening to identify enantioselective hydrolytic enzymes," *Angewandte Chemie*, vol. 47, no. 27, pp. 5085–5088, 2008.

[5] D. Lipovšek, E. Antipov, K. A. Armstrong et al., "Selection of horseradish peroxidase variants with enhanced enantioselectivity by yeast surface display," *Chemistry and Biology*, vol. 14, no. 10, pp. 1176–1185, 2007.

[6] P. Amstutz, P. Forrer, C. Zahnd, and A. Plückthun, "*In vitro* display technologies: novel developments and applications," *Current Opinion in Biotechnology*, vol. 12, no. 4, pp. 400–405, 2001.

[7] H. Leemhuis, V. Stein, A. D. Griffiths, and F. Hollfelder, "New genotype-phenotype linkages for directed evolution of functional proteins," *Current Opinion in Structural Biology*, vol. 15, no. 4, pp. 472–478, 2005.

[8] J. Hanes and A. Plückthun, "*In vitro* selection and evolution of functional proteins by using ribosome display," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 10, pp. 4937–4942, 1997.

[9] R. Odegrip, D. Coomber, B. Eldridge et al., "CIS display: *in vitro* selection of peptides from libraries of protein-DNA complexes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 9, pp. 2806–2810, 2004.

[10] R. W. Roberts and J. W. Szostak, "RNA-peptide fusions for the *in vitro* selection of peptides and proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 23, pp. 12297–12302, 1997.

[11] V. Taly, B. T. Kelly, and A. D. Griffiths, "Droplets as microreactors for high-throughput biology," *ChemBioChem*, vol. 8, no. 3, pp. 263–272, 2007.

[12] D. S. Tawfik and A. D. Griffiths, "Man-made cell-like compartments for molecular evolution," *Nature Biotechnology*, vol. 16, no. 7, pp. 652–656, 1998.

[13] A. D. Griffiths and D. S. Tawfik, "Directed evolution of an extremely fast phosphotriesterase by *in vitro* compartmentalization," *EMBO Journal*, vol. 22, no. 1, pp. 24–35, 2003.

[14] K. Bernath, M. Hai, E. Mastrobattista, A. D. Griffiths, S. Magdassi, and D. S. Tawfik, "*In vitro* compartmentalization

by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting," *Analytical Biochemistry*, vol. 325, no. 1, pp. 151–157, 2004.

[15] E. Mastrobattista, V. Taly, E. Chanudet, P. Treacy, B. T. Kelly, and A. D. Griffiths, "High-throughput screening of enzyme libraries: *in vitro* evolution of a β-galactosidase by fluorescence-activated sorting of double emulsions," *Chemistry and Biology*, vol. 12, no. 12, pp. 1291–1300, 2005.

[16] A. Fallah-Araghi, J. C. Baret, M. Ryckelynck, and A. D. Griffiths, "A completely *in vitro* ultrahigh-throughput droplet-based microfluidic screening system for protein engineering and directed evolution," *Lab on a Chip*, vol. 12, no. 5, pp. 882–891, 2012.

[17] P. Girard, J. Pécréaux, G. Lenoir, P. Falson, J.-L. Rigaud, and P. Bassereau, "A new method for the reconstitution of membrane proteins into giant unilamellar vesicles," *Biophysical Journal*, vol. 87, no. 1, pp. 419–429, 2004.

[18] V. Noireaux and A. Libchaber, "A vesicle bioreactor as a step toward an artificial cell assembly," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 51, pp. 17669–17674, 2004.

[19] M. Kaneda, S.-I. M. Nomura, S. Ichinose et al., "Direct formation of proteo-liposomes by *in vitro* synthesis and cellular cytosolic delivery with connexin-expressing liposomes," *Biomaterials*, vol. 30, no. 23-24, pp. 3971–3977, 2009.

[20] K. Nishimura, T. Matsuura, T. Sunami, H. Suzuki, and T. Yomo, "Cell-free protein synthesis inside gant unilamellar vesicles analyzed by flow cytometry," *Langmuir*, vol. 28, no. 22, pp. 8426–8432, 2012.

[21] K. Ishikawa, K. Sato, Y. Shima, I. Urabe, and T. Yomo, "Expression of a cascading genetic network within liposomes," *FEBS Letters*, vol. 576, no. 3, pp. 387–390, 2004.

[22] W. Yu, K. Sato, M. Wakabayashi et al., "Synthesis of functional protein in liposome," *Journal of Bioscience and Bioengineering*, vol. 92, no. 6, pp. 590–593, 2001.

[23] T. Sunami, H. Kita, K. Hosoda, T. Matsuura, H. Suzuki, and T. Yomo, "Detection and analysis of protein synthesis and RNA replication in giant liposomes," *Methods in Enzymology*, vol. 464, pp. 19–30, 2009.

[24] T. Sunami, K. Sato, T. Matsuura, K. Tsukada, I. Urabe, and T. Yomo, "Femtoliter compartment in liposomes for *in vitro* selection of proteins," *Analytical Biochemistry*, vol. 357, no. 1, pp. 128–136, 2006.

[25] T. Nishikawa, T. Sunami, T. Matsuura, N. Ichihashi, and T. Yomo, "Construction of a gene screening system using giant unilamellar liposomes and a fluorescence-activated cell sorter," *Analytical Chemistry*, vol. 84, no. 11, pp. 5017–5024, 2012.

[26] A. D. Bangham and R. W. Horne, "Negative staining of phospholipids and their structural modification by surface-active agents as observed in the electron microscope," *Journal of molecular biology*, vol. 8, pp. 660–668, 1964.

[27] G. Sessa and G. Weissmann, "Phospholipid spherules (liposomes) as a model for biological membranes," *Journal of Lipid Research*, vol. 9, no. 3, pp. 310–318, 1968.

[28] H. Ringsdorf, B. Schlarb, and J. Venzmer, "Molecular architecture and function of polymeric oriented systems: models for the study of organization, surface recognition, and dynamics of biomembrnaes," *Angewandte Chemie*, vol. 27, no. 1, pp. 113–158, 1988.

[29] P. L. Luisi, F. Ferri, and P. Stano, "Approaches to semi-synthetic minimal cells: a review," *Naturwissenschaften*, vol. 93, no. 1, pp. 1–13, 2006.

[30] S. Mann, "Systems of creation: the emergence of life from nonliving matter," *Accounts of Chemical Research*. In press.

[31] P. Stano, P. Carrara, Y. Kuruma, T. P. de Souza, and P. L. Luisi, "Compartmentalized reactions as a case of soft-matter biotechnology: synthesis of proteins and nucleic acids inside lipid vesicles," *Journal of Materials Chemistry*, vol. 21, no. 47, pp. 18887–18902, 2011.

[32] P. Walde, K. Cosentino, H. Engel, and P. Stano, "Giant vesicles: preparations and applications," *ChemBioChem*, vol. 11, no. 7, pp. 848–865, 2010.

[33] T. Sunami, T. Matsuura, H. Suzuki, and T. Yomo, "Synthesis of functional proteins within liposomes," *Methods in Molecular Biology*, vol. 607, pp. 243–256, 2010.

[34] K. Hosoda, T. Sunami, Y. Kazuta, T. Matsuura, H. Suzuki, and T. Yomo, "Quantitative study of the structure of multilamellar giant liposomes as a container of protein synthesis reaction," *Langmuir*, vol. 24, no. 23, pp. 13540–13548, 2008.

[35] S. Pautot, B. J. Frisken, and D. A. Weitz, "Production of unilamellar vesicles using an inverted emulsion," *Langmuir*, vol. 19, no. 7, pp. 2870–2879, 2003.

[36] L. Jermutus, L. A. Ryabova, and A. Plückthun, "Recent advances in producing and selecting functional proteins by using cell-free translation," *Current Opinion in Biotechnology*, vol. 9, no. 5, pp. 534–548, 1998.

[37] Y. Shimizu, A. Inoue, Y. Tomari et al., "Cell-free translation reconstituted with purified components," *Nature Biotechnology*, vol. 19, no. 8, pp. 751–755, 2001.

[38] L. A. Herzenberg, D. Parks, B. Sahaf, O. Perez, M. Roederer, and L. A. Herzenberg, "The history and future of the fluorescence activated cell sorter and flow cytometry: a view from Stanford," *Clinical Chemistry*, vol. 48, no. 10, pp. 1819–1827, 2002.

[39] B. P. Tracy, S. M. Gaida, and E. T. Papoutsakis, "Flow cytometry for bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes," *Current Opinion in Biotechnology*, vol. 21, no. 1, pp. 85–99, 2010.

[40] R. R. Fuller and J. V. Sweedler, "Characterizing submicron vesicles with wavelength-resolved fluorescence in flow cytometry," *Cytometry*, vol. 25, no. 2, pp. 144–155, 1996.

[41] M. Hai, K. Bernath, D. Tawfik, and S. Magdassi, "Flow cytometry: a new method to investigate the properties of water-in-oil-in-water emulsions," *Langmuir*, vol. 20, no. 6, pp. 2081–2085, 2004.

[42] K. Sato, K. Obinata, T. Sugawara, I. Urabe, and T. Yomo, "Quantification of structural properties of cell-sized individual liposomes by flow cytometry," *Journal of Bioscience and Bioengineering*, vol. 102, no. 3, pp. 171–178, 2006.

[43] K. Nishimura, T. Hosoi, T. Sunami et al., "Population analysis of structural properties of giant liposomes by flow cytometry," *Langmuir*, vol. 25, no. 18, pp. 10439–10443, 2009.

[44] T. Sunami, K. Hosoda, H. Suzuki, T. Matsuura, and T. Yomo, "Cellular compartment model for exploring the effect of the lipidic membrane on the kinetics of encapsulated biochemical reactions," *Langmuir*, vol. 26, no. 11, pp. 8544–8551, 2010.

[45] A. L. Koch, "What size should a bacterium be? A question of scale," *Annual Review of Microbiology*, vol. 50, pp. 317–348, 1996.

[46] T. Matsuura, K. Hosoda, N. Ichihashi, Y. Kazuta, and T. Yomo, "Kinetic analysis of β-galactosidase and β-glucuronidase tetramerization coupled with protein translation," *Journal of Biological Chemistry*, vol. 286, no. 25, pp. 22028–22034, 2011.

[47] T. Matsuura, K. Hosoda, N. Ichihashi, Y. Kazuta, H. Suzuki, and T. Yomo, "Effects of compartment size on the kinetics of intracompartmental multimeric protein synthesis," *ACS SyntheticBiology*. In press.

[48] Y. Kuruma, P. Stano, T. Ueda, and P. L. Luisi, "A synthetic biology approach to the construction of membrane proteins in semi-synthetic minimal cells," *Biochimica et Biophysica Acta*, vol. 1788, no. 2, pp. 567–574, 2009.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 538129, 7 pages
doi:10.1155/2012/538129

Research Article

A Concept for Selection of Codon-Suppressor tRNAs Based on Read-Through Ribosome Display in an *In Vitro* Compartmentalized Cell-Free Translation System

Atsushi Ogawa,[1,2] Masayoshi Hayami,[1] Shinsuke Sando,[1,3] and Yasuhiro Aoyama[1,4]

[1] *Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Katsura, Nishikyo-ku, Kyoto 615-8510, Japan*
[2] *Ehime University, 3 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan*
[3] *INAMORI Frontier Research Center, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan*
[4] *Department of Molecular Chemistry and Biochemistry, Faculty of Science and Engineering, Doshisha University, 1-3 Tatara-Miyakodani, Kyotanabe, Kyoto 610-0321, Japan*

Correspondence should be addressed to Atsushi Ogawa, a-ogawa@ccr.ehime-u.ac.jp
and Shinsuke Sando, ssando@ifrc.kyushu-u.ac.jp

Received 30 May 2012; Accepted 29 June 2012

Academic Editor: Masayasu Kuwahara

Here is presented a concept for *in vitro* selection of suppressor tRNAs. It uses a pool of dsDNA templates in compartmentalized water-in-oil micelles. The template contains a transcription/translation trigger, an amber stop codon, and another transcription trigger for the anticodon- or anticodon loop-randomized gene for $tRNA^{Ser}$. Upon transcription are generated two types of RNAs, a tRNA and a translatable mRNA (mRNA-tRNA). When the tRNA suppresses the stop codon (UAG) of the mRNA, the full-length protein obtained upon translation remains attached to the mRNA (read-through ribosome display) that contains the sequence of the tRNA. In this way, the active suppressor tRNAs can be selected (amplified) and their sequences read out. The enriched anticodon (CUA) was complementary to the UAG stop codon and the enriched anticodon-loop was the same as that in the natural $tRNA^{Ser}$.

1. Introduction

Selection/amplification is a general tool for directed evolution of nucleic acids and proteins [1], which is much more complicated for reaction promoters (biocatalysts) than for simple binders [2]. Selection and identification for the former require some sort of catalyst-product pairing in an isolated compartment. *In vivo* selection using living cells has been typical choice for such a purpose. Meanwhile, *in vitro* selection, as opposed to *in vivo* selection, is simple and convenient to carry out, is free from cytotoxicity problems, and allows for starting with a library of great diversity.

Griffiths and Tawfik proposed an *in vitro* compartmentalization (IVC) technique for the *in vitro* evolution of biomolecules including biocatalysts, where biocatalysts are transcribed/translated in a compartmentalized water-in-oil emulsion to allow catalyst-product pairing [2]. In practice, the IVC technique has been applied successfully for evolving or improving biocatalysts such as ribozymes and enzymes [3–12].

Here, we applied the IVC technique for evolution of tRNAs (one of catalysts in protein translation systems) [13–16] and present a promising concept for *in vitro* selection of suppressor tRNAs by the combination of read-through ribosome display (Rt-RD, *vide infra*) [17, 18].

2. Materials and Methods

2.1. General Design, Transcription/Translation System, and Analysis. Biological reagents and solvents were purchased from standard suppliers and used without further purification. Binding of suppressor tRNA to an amber stop codon is in competition with that of release factor 1 (RF1) to

terminate the translation. The amber codon is also often misread by the exogenous tRNA for Gln. To maximize the suppression efficiency and minimize incorporation of Gln at the amber codon, we used a reconstituted prokaryotic cell-free translation system (PURESYSTEM Classic) [19], in which RF1, and Gln and Gln-tRNA synthetase had not been added unless otherwise stated. The T7-promoted translation is generally more efficient for longer templates. To keep a "balance" in the amounts of the fused mRNA (mRNA-tRNA) (longer) and the tRNA (shorter) transcribed, we put GCC immediately downstream of the first T7-promoter so as to lower the translation efficiency for the fused mRNA (It is known that a G-less sequence, immediately downstream of the T7 promoter, suppresses the transcription efficiency). Gel electrophoresis and blotting were carried out on a BE-250 electrophoresis apparatus (BIO CRAFT) and a Trans-Blot SD system (Bio-Rad), respectively.

2.2. Preparation of Template DNAs for Selection. pDHFR, encoding *E. coli* DHFR, was a gift from Dr. Y. Shimizu. The first PCR was carried out in 20 μL of a reaction mixture containing 4 pmol of a forward primer containing a FLAG domain (bold) and a TAG amber stop codon (italic) 5′-d(AA GGA GAT ATA CCA ATG **GAC TAC AAG GAT GAC GAT GAC AAG** *TAG* ATC AGT CTG ATT GCG GCG TTA G)-3′, 4 pmol of a reverse primer with the lower T7 promoter (underlined) 5′-d(GTT CAG CCG CTC CGG CAT CTC TCC TAT AGT GAG TCG TAT TAC CGG GTG ACT GCT GAG GA)-3′ for the tRNA-fused template or 5′-d(TGG CGG AGA GAG GGG GAT TTG AAC CGG GTG ACT GCT GAG GA)-3′ for nonfused template, 20 ng of pDHFR, 1.25 U of *Pfu Ultra* HF DNA polymerase (Stratagene), 4 nmol each of dNTPs (TOYOBO), and 2 μL of 10 ×*Pfu Ultra* HF reaction buffer. After the PCR reaction, the product was purified by agarose gel electrophoresis. The second PCR was carried out in 20 μL of a reaction mixture containing 4 pmol of a forward primer with the upper T7 promoter (underlined) and the following sequence after modification (vide supra) of the widely-used one 5′-d(T AAT ACG ACT CAC TAT AGC CCG GCC ACA ACG GCT GGG CTC TAG AAA TAA TTT TGT TTA ACT TTA AGA AGG AGA TAT ACC A)-3′, 4 pmol of a reverse primer with the sequence for *E. coli* tRNA for Ser (SerU) 5′-d(TGG CGG AGA GAG GGG GAT TTG AAC CCC CGG TAG AGT TGC CCC TAC TCC GGT L_1L_2X YZL_3 L_4AC CGG TCC GTT CAG CCG CTC CGG CAT C)-3′ ($L_{1\sim4}$ and XYZ represent the anticodon-loop and anticodon bases, respectively), ∼20 fmol of the purified first PCR product, 1.25 U of *Pfu Ultra* HF DNA polymerase, 4 nmol each of dNTPs, and 2 μL of 10 ×*Pfu Ultra* HF reaction buffer.

2.3. In Vitro Selection of Suppressor tRNAs. In vitro coupled transcription/translation of template DNAs in the reverse-phase micelle was carried out as follows. A 50 μL portion of a cell-free translation system containing 50 pM of a DNA template (2.5 fmol) was added gradually to 950 μL of mineral oil (Sigma) containing detergents Span 85 (Nacalai Tesque) (4.5% v/v) and Tween 20 (Sigma) (0.6% v/v) under stirring on ice. The diameter (d) of the micelles (water droplets) in the resulting emulsion was about 2 μm, indicating that the number of micelles was $N = V/v = 1.2 \times 10^{10}$, where V is the total volume of the water phase (50 μL) and v is the volume of a micelle with $d = 2$ μm. The dsDNA used (2.5 fmol) contains 1.5×10^9 template molecules. This number is one-order of magnitude smaller than that of the micelles. Thus, each micelle is expected to encapsulate maximally one template.

After incubation of the mixture at 37°C for 1 h, the emulsion was spun at 2000 ×g for 10 sec. A 200 μL portion of the supernatant was carefully taken from the upper part and mixed with 200 μL of an ice-cold selection buffer (Phosphate-K, pH 7.3) containing 92.2 mM of K^+, 300 mM of Na^+, 50 mM of Mg^{2+}, 0.05% of Tween 20 (WAKO), 2% of Block Ace (Dainippon Pharmaceutical Co.), and 1 mL of ice-cold water-saturated ether. The mixture was inverted twenty times and centrifuged at 16100 ×g for 10 min at 4°C, and then the organic ether phase was removed. The water phase was washed with 1 mL of ether and mixed with 200 μL of ice-cold selection buffer. The mixture was applied on a column packed with prewashed anti-FLAG M2 agarose (Sigma) and gently inverted for 2 h at 4°C. After washing the gel retaining the ribosome-protein-mRNA (PRM) complex with 200 μL of selection buffer five times, the mRNAs were eluted upon collapse of the RPM complexes with 200 μL of an elution buffer (Phosphate-K, pH 7.2) containing 92.2 mM of K^+, 300 mM of Na^+, 30 mM of EDTA, 0.05% of Tween 20, and 2% of Block Ace. The mRNAs eluted were purified with an RNeasy MinElute Cleanup Kit (QIAGEN) and amplified with a QIAGEN One-Step RT-PCR Kit according to the manufacture's protocol using a forward primer with T7 promoter (underlined) and a FLAG domain (bold) followed by a TAG amber stop codon (italic) 5′-d(T AAT ACG ACT CAC TAT AGC CCG GCC ACA ACG GCT GGG CTC TAG AAA TAA TTT TGT TTA ACT TTA AGA AGG AGA TAT ACC A ATG **GAC TAC AAG GAT GAC GAT GAC AAG** *TAG*)-3′ and a reverse primer 5′-d(TGG CGG AGA GAG GGG GAT TTG AAC CCC CGG TAG AGT TGC C)-3′. The resulting RT-PCR products were used for the next round of selection and, after three (for anticodon-randomized tRNA) or five (for anticodon-loop-randomized tRNA) rounds, were monocloned using a PCR Cloning Kit (QIAGEN) and sequenced. Single-round coupled transcription/translation under normal (noncompartmentalized) or compartmentalized conditions (referring to Figure 1(b), lanes 1–4 or 5, resp.) was carried out using 400 pM of a template in 25 μL of a cell-free translation system under otherwise identical conditions.

2.4. Preparation of Nonfused Template mRNAs for Western Blotting. The first PCR was carried out in 20 μL of a reaction mixture containing 4 pmol of a forward primer with a FLAG domain (bold) with or without a TAG amber stop codon (italic) 5′-d(AA GGA GAT ATA CCA ATG **GAC TAC AAG GAT GAC GAT GAC AAG** [*TAG*] ATC AGT CTG ATT GCG GCG TTA G)-3′, 4 pmol of a reverse primer with an ochre stop codon (italic) 5′-d(TAT TCA *TTA* CCG CCG CTC CAG AAT CT)-3', 20 ng of pDHFR, 1.25 U of *Pfu Ultra* HF DNA polymerase (Stratagene), 4 nmol each of

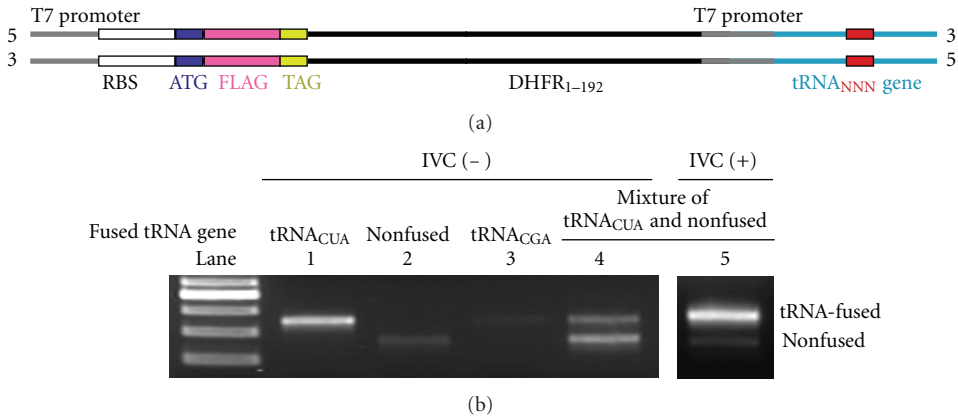


Figure 1: (a) Schematic sequence of the tRNA-fused template with two T7 promoters. The triple N in red represents the anticodon- or anticodon-loop-randomized region. (b) Agarose gel electrophoretic assay of the template DNAs recovered from $tRNA_{CUA}$-fused template (lane 1), nonfused template lacking the tRNA domain (lane 2), a 1 : 1 mixture thereof (lanes 4, 5), or $tRNA_{CGA}$-fused template (lane 3), after coupled transcription/translation under normal (noncompartmentalized) (lanes 1–4) or compartmentalized (lane 5) conditions.

dNTPs (TOYOBO), and 2 μL of 10 ×*Pfu Ultra* HF reaction buffer. After the PCR reaction, the product was purified by agarose gel electrophoresis. The second PCR was carried out in 20 μL of a reaction mixture containing 4 pmol of a forward primer with T7 promoter (underlined) 5′-d(GAA ATT AAT ACG ACT CAC TAT AGG GAG ACC ACA ACG GTT TCC CTC TAG AAA TAA TTT TGT TTA ACT TTA AGA AGG AGA TAT ACC A)-3′, 4 pmol of a reverse primer 5′-d(TAT TCA TTA CCG CCG CTC CAG AAT CT)-3′, ~20 fmol of the purified first PCR product, 1.25 U of *Pfu Ultra* HF DNA polymerase, 4 nmol each of dNTPs, and 2 μL of 10 ×*Pfu Ultra* HF reaction buffer. Template mRNAs were obtained by run-off transcription of the 5′-FLAG-TAG-DHFR or 5′-FLAG-DHFR template DNA obtained using a T7-MEGAshortscript Kit (Ambion). Thus, a T7-transcription mixture (10 μL) containing 3 μL of the PCR solution of template DNA was incubated at 37°C for 2 h. After addition of 1 U of DNase I, the mixture was incubated for additional 15 min. mRNAs were purified with an RNeasy MinElute Cleanup Kit (QIAGEN) and concentrations of the purified specimens were determined by the absorbance at 260 nm.

2.5. Preparation of $tRNA^{SerU}_{L1L2XYZL3L4}$. *E. coli* $tRNA^{SerU}_{L1L2XYZL3L4}$ was prepared by run-off *in vitro* transcription. Template DNA was prepared by PCR in 20 μL of a reaction mixture containing 20 pmol of a forward primer with T7 promoter (underlined) 5′-d(GTAA TAC GAC TCA CTA TA GGA GAG ATG CCG GAG CGG CTG AAC)-3′, 20 pmol of a reverse primer 5′-d(TGG CGG AGA GAG GGG GAT TTG AAC CCC CGG TAG AGT TGC C)-3′, 100 fmol of a template DNA (initial, RT-PCR amplified, or monocloned), 1.25 U of *Pfu Ultra* HF DNA polymerase (Stratagene), 10 nmol each of dNTPs (TOYOBO), and 2 μL of 10 ×*Pfu Ultra* HF reaction buffer. The resulting PCR solution was used for transcription using a T7-MEGAshortscript Kit (Ambion) for 37°C for 20 h. The transcribed tRNAs were purified by denaturing PAGE (8%), followed by ethanol precipitation and, after dissolution in 500 μL of water, further by passing successively through Microcon YM-30 (Millipore) and G-25 Microspin Columns (Amersham). Concentrations of the purified tRNAs were determined by the absorbance at 260 nm.

2.6. Translation of mRNA and Western Blotting Analysis with Exogenous tRNAs. Translation of a template mRNA (5′-FLAG-UAG-DHFR or 5′-FLAG-DHFR as a UAG(−) positive control) (2 μg) was carried out at 37°C for 1 h in the presence of an exogenous tRNA (2 μg) in 10 μL of a reconstituted cell-free translation system, in which RF1 had not been added, but Gln and Gln-tRNA synthetase had. To the reaction mixture were added 165 μL of water and 175 μL of a sample-loading buffer (125 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 20% (w/v) glycerol, 0.002% (w/v) bromophenol blue, and 10% (v/v) 2-mercaptoethanol). The resulting solution was incubated at 95°C for 5 min and applied on 15% SDS-PAGE. Western blotting was performed on a PVDF membrane (Hybond-P, Amersham). FLAG-tagged proteins were visualized with anti-FLAG-HRP conjugate (Sigma) and ECL Plus Western Blotting Detection Reagent (Amersham). The suppression efficiencies were evaluated by comparing the band intensities of the FLAG-TAG-DHFR protein, determined by using the Image J software (NIH), with those of serially diluted (10, 20, 40, 60, 80, 100%) solutions of the stop-free FLAG-DHFR reference protein translated under otherwise identical conditions.

3. Results and Discussion

We previously introduced Rt-RD [18], in which expressed protein could be fully displayed upon suppression of the stop codon(s) downstream of the open reading frame by appropriate suppressor tRNAs. In this paper, we conversely apply the Rt-RD technique for the selection of suppressor tRNAs that are coded at the 3′-terminus of the very template that contains the stop codon to be suppressed. We prepared

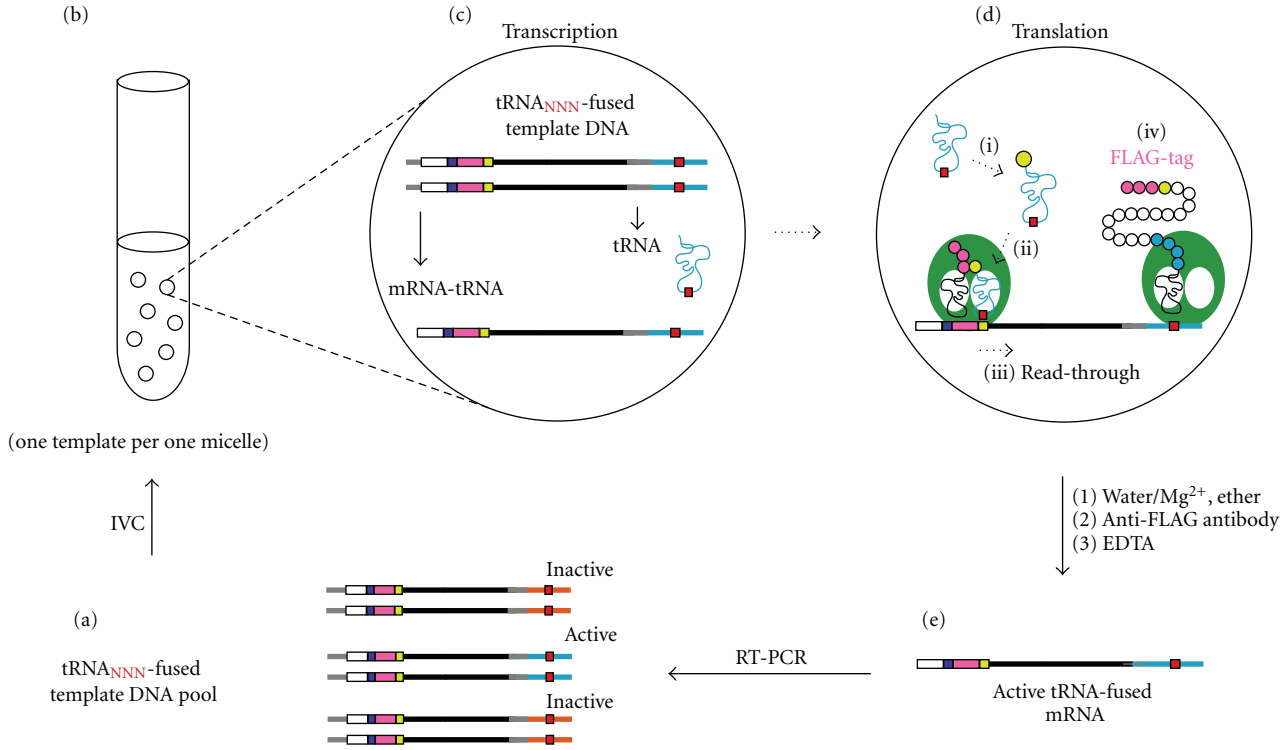


Figure 2: Rt-RD/IVC-based selection/amplification cycle. See the text for explanation.

a dsDNA template containing, under the control of a T7 promoter, the RBS (ribosome binding site), the ATG start codon, a FLAG domain, the TAG (amber) stop codon, and a spacer ($DHFR_{1–192}$) derived from the *E. coli* DHFR gene covering the amino acids 1–64, followed by the tRNA sequence together with its own T7 promoter (Figure 1(a)). There are two T7 promoters in this template, transcription of which thus generates two types of RNAs; the one that contains both an RBS and an AUG serves as an mRNA, and the other, which lacks them, serves only as a tRNA. When the latter (tRNA) happens to suppress (read through) the amber (UAG) stop codon in the transcribed mRNA, a fused protein containing the FLAG, $DHFR_{1–192}$, and tRNA regions would result upon translation and remain attached to the mRNA with the $DHFR_{1–192}$-tRNA portion, serving as a spacer to be anchored in the ribosome tunnel to squeeze the FLAG-tag peptide for full display (Rt-RD; Figure 2).

We first confirmed that the right amber suppressor ($tRNA^{SerU}_{CUA}$) generated in this way worked as such. The anticodon-adjusted (CUA) $tRNA^{SerU}$ is known to be aminoacylated with Ser by endogenous Ser-tRNA synthetase and hence suppress (read through) the amber codon with concomitant incorporation of Ser at that position [18, 20]. Coupled transcription/translation for 1 h of the template fused with the gene for $tRNA^{SerU}_{CUA}$ (Figure 1(a), NNN = CUA) in a reconstituted prokaryotic (*E. coli*) cell-free translation system [19] containing T7 RNA polymerase was followed by affinity selection (4°C and $[Mg^{2+}]$ = 50 mM) of the FLAG-tag peptide domain displayed (Rt-RD) in the protein-ribosome-mRNA (PRM) complex. The mRNA template coding the $tRNA^{SerU}_{CUA}$ sequence was recovered upon disruption of the ternary complex with EDTA, RT-PCR amplified, and identified as such (Figure 1(b), lane 1). However, the template was by no means recovered efficiently when it was lacking (nonfused) the tRNA domain (lane 2) or fused with anticodon-mismatched (CGA) natural tRNA for Ser (NNN = CGA) (lane 3). (In respect to the weak spots seen in lanes 2 and 3 (Figure 1(b)) , it is known that the stop codons can be suppressed to some extent (up to several %) in an RF1-minus translation system by misreading even in the absence of a suppressor tRNA.) In these cases, there is no generation of the correct tRNA for amber suppression and hence translation stops at the amber (UAG) codon, giving rise only to the FLAG peptide with no linkage to the genotype (mRNA). Conversely, when equal amounts of nonfused (suppression irrelevant) and $tRNA^{SerU}_{CUA}$-fused (relevant) templates were used, both templates were recovered (lane 4) (In respect to the apparently stronger spot for the nonfused template in lane 4 (Figure 1(b)) , it is generally true that shorter templates are more easily amplified by PCR.) This is because suppressor $tRNA^{SerU}_{CUA}$ generated from the latter template can suppress the amber codon of the former. To avoid such a crossover, we needed to compartmentalize the reactions of each template using a water-in-oil emulsion system [3–12]. With this technique (see below, Figure 2), we could selectively recover the $tRNA^{SerU}_{CUA}$-fused template coding the active suppressor tRNA (lane 5) from the above mixture.

We then moved on to the selection of suppressor tRNAs. The selection cycle is shown in Figure 2. An initial pool

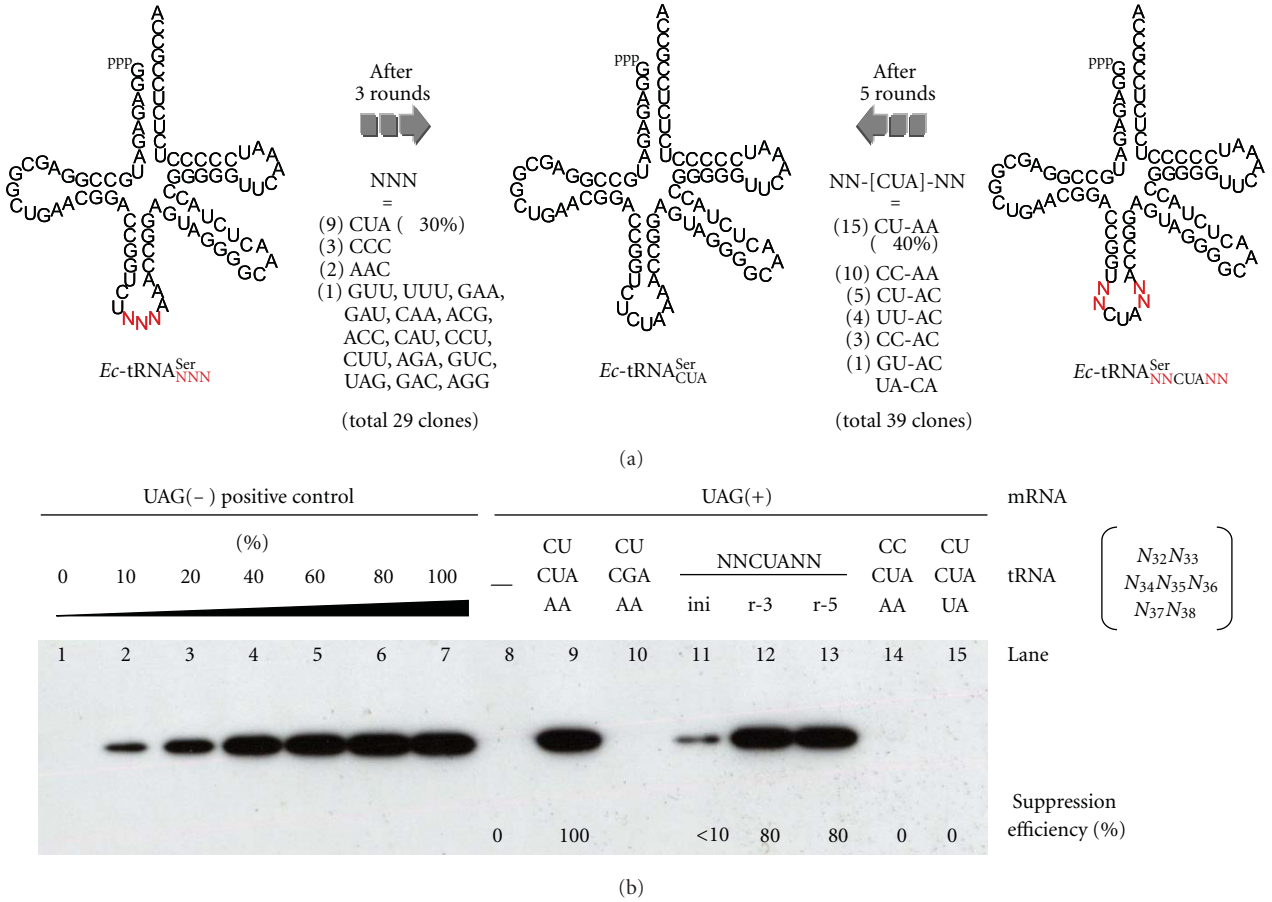


Figure 3: (a) Summary of selection/amplification of anticodon-randomized or anticodon-matched/anticodon-loop-randomized tRNAs for Ser. The numbers of clones that possess the sequence shown are indicated in parentheses. (b) Western blotting analysis of the translation of nonfused (tRNA-lacking) mRNA templates under normal (noncompartmentalized) conditions with (lanes 9–15) or without (lane 8) tRNA; $tRNA^{SerU}_{CUA}$ (lane 9), $tRNA^{SerU}_{CGA}$ (natural tRNA for Ser, lane 10), anticodon-loop-randomized $tRNA^{SerU}$ after 0, 3, or 5 rounds of selection/amplification (lanes 11, 12, or 13, respectively), or anticodon-matched but singly loop-mutated $tRNA^{SerU}$ (lanes 14 and 15). Lane 7 represents a control translation using amber-free mRNA with no use of tRNA as a positive control and lanes 1–7 are a calibration set.

(a) of fused templates (Figure 1(a), NNN = NNN with a diversity of $4^3 = 64$) with genes for anticodon-randomized $tRNA^{SerU}$ (Figure 3(a), left) was subjected to coupled transcription/translation in a water-in-oil emulsion (b) [3–12]. Consequently, each compartment (~2 μm) generates and contains a pair of template-related sister RNAs, a fused mRNA (shown as mRNA-tRNA) and a tRNA (c). The whole protein, susceptible to Rt-RD in the form of a PRM complex (step iv in d), can be translated (d) only when the tRNA serves as an amber suppressor (steps i–iii in d). The ternary complex thus formed was recovered by affinity selection of the displayed FLAG tag after breaking the emulsion and treated with EDTA to afford the active tRNA-fused template mRNAs (e), which were purified and amplified by RT-PCR back to the dsDNA templates (a), for use in another selection cycle. The DNAs obtained after three such cycles were monocloned and sequenced. The anticodon of approximately one third of the 29 monocloned tRNAs was CUA (Figure 3(a), center), which was the one expected based on codon-anticodon complementarity. These results indicate that the sensitive PRM complex survives the IVC and workup conditions to allow the Rt-RD/IVC method to select suppressor tRNAs.

Finally, we applied the Rt-RD/IVC method to the engineering of the two-base loop regions adjacent to the anticodon. These regions are variant in various tRNAs but are believed to be important in stabilizing codon-anticodon interactions [21]. Selection of fused templates with genes of anticodon-matched (CUA) and anticodon-loop-randomized (NN-CUA-NN with a diversity of $4^4 = 256$) $tRNA^{SerU}$ (Figure 3(a), right) was carried out as above. DNA templates recovered after five cycles were monocloned and sequenced. Interestingly, ~40% of the 39 monocloned tRNAs had the same loop sequence (CU-CUA-AA) as the natural tRNA for Ser (CU-CGA-AA) even in the base modification-free conditions (Figure 3(a)). These results indicate that the anticodon-loop sequence itself has played a *preserved* or *sophisticated* role in the evolution of tRNAs that now have many modified bases (Since neither anticodon nor anticodon-loop region is recognized by the serenyl-tRNA

synthetase [22], the selected loop sequence may play a role in stabilizing the interaction between the amber stop codon and the suppressor anticodon.) [22, 23]. Suppression efficiencies were evaluated by western blotting analysis using the nonfused template lacking the tRNA domain in an RF1-minus translation system under noncompartmentalized conditions (Figure 3(b)). Enrichment of active tRNA became notably pronounced after three rounds (lanes 11–13) (The reason for the high suppression efficiency (~80%) despite the low percentage amounts of the active $tRNA^{SerU}_{CUA}$ (10% in round 3 and 38% in round 5) is that excess amounts of $tRNA^{SerU}_{CUA}$ were used in the experiments for western blotting (Figure 3(b))) and the suppression efficiency of the monocloned $tRNA_{CU-CUA-AA}$ showed ~100% activity (lane 9) compared with the suppression-free translation using a UAG(–) reference template (lane 7). Interestingly, a single mutation in the anticodon-loop domain led to a dramatic loss of activity (lanes 14 and 15). This is in accord with the above argument.

In summary, a concept for *in vitro* selection of suppressor tRNAs is presented. An essential aspect of it is that tRNA is a part of mRNA (mRNA-tRNA fusion), being located downstream of the particular codon to be suppressed (the amber stop codon in this study). Therefore, only active tRNAs that can *self-suppress* the codon in a water-in-oil compartment are susceptible to amplification by the read-through ribosome-display technique.

Although the concept was well demonstrated, the present method must be further optimized especially regarding the selection efficiencies. Three or five rounds of selection were required for enrichment from the small library sizes ($n = 64$ and 256). One possible reason for this low selection efficiency is a misreading of stop codon in an RF1-minus translation system, which would result in the recovery of false positive tRNA-fused mRNAs. The misreading might be reduced by coexisting small amounts of RF-1. Another reason is the instability of RPM complex against extraction conditions. Optimization of extraction conditions or the use of mRNA display technique [24, 25] might overcome the instability. Of course, compartmentalization conditions or mRNA/tRNA ratios are also the factor to be checked.

Furthermore, it is not easy to understand that 10 clones out of 39 (26%) had the CC-CUA-AA anticodon-loop sequence after 5 rounds (Figure 3(a)), which turned out to be completely inactive in suppression (Figure 3(b), lane 14). Taking into account the fact that the CC-CUA-AA occupies 25% of pool even in round 3 in contrast to the less percentage amounts of the active CU-CUA-AA sequence (10%), it is feasible to speculate that the inactive CC-CUA-AA sequence might be already abundant in the initial pool and/or has been more easily amplified regardless of the selection process.

After sufficient improvements, the method may provide a promising *in vitro* tool for tRNA evolutions.

Acknowledgments

The authors thank Dr. N. Doi of Keio University and Dr. Y. Shimizu of RIKEN for advising as to IVC experiments and for providing the plasmid pDHFR, respectively. A. Ogawa acknowledges a fellowship from the JSPS. This work was partly supported by The Sumitomo Foundation and partially supported by "Special Coordination Funds for Promoting Science and Technology" from the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government.

References

[1] D. S. Wilson and J. W. Szostak, "*In vitro* selection of functional nucleic acids," *Annual Review of Biochemistry*, vol. 68, pp. 611–647, 1999.

[2] A. D. Griffiths and D. S. Tawfik, "Man-made enzymes—from design to *in vitro* compartmentalisation," *Current Opinion in Biotechnology*, vol. 11, no. 4, pp. 338–353, 2000.

[3] D. S. Tawfik and A. D. Griffiths, "Man-made cell-like compartments for molecular evolution," *Nature Biotechnology*, vol. 16, no. 7, pp. 652–656, 1998.

[4] N. Doi and H. Yanagawa, "STABLE: protein-DNA fusion system for screening of combinatorial protein libraries *in vitro*," *FEBS Letters*, vol. 457, no. 2, pp. 227–230, 1999.

[5] M. Levy, K. E. Griswold, and A. D. Ellington, "Direct selection of trans-acting ligase ribozymes by *in vitro* compartmentalization," *RNA*, vol. 11, no. 10, pp. 1555–1562, 2005.

[6] H. S. Zaher and P. J. Unrau, "Selection of an improved RNA polymerase ribozyme with superior extension and fidelity," *RNA*, vol. 13, no. 7, pp. 1017–1026, 2007.

[7] Y. Zheng and R. J. Roberts, "Selection of restriction endonucleases using artificial cells," *Nucleic Acids Research*, vol. 35, no. 11, article e83, 2007.

[8] B. T. Kelly and A. D. Griffiths, "Selective gene amplification," *Protein Engineering, Design and Selection*, vol. 20, no. 12, pp. 577–581, 2007.

[9] M. Levy and A. D. Ellington, "Directed evolution of streptavidin variants using *in vitro* compartmentalization," *Chemistry and Biology*, vol. 15, no. 9, pp. 979–989, 2008.

[10] Y. Chen, J. Mandic, and G. Varani, "Cell-free selection of RNA-binding proteins using *in vitro* compartmentalization," *Nucleic Acids Research*, vol. 36, no. 19, article e128, 2008.

[11] T. Sumida, N. Doi, and H. Yanagawa, "Bicistronic DNA display for *in vitro* selection of Fab fragments," *Nucleic Acids Research*, vol. 37, no. 22, Article ID gkp776, p. e147, 2009.

[12] Y. Tay, C. Ho, P. Droge, and F. J. Ghadessy, "Selection of bacteriophage lambda integrases with altered recombination specificity by *in vitro* compartmentalization," *Nucleic Acids Research*, vol. 38, no. 4, article e25, 2010.

[13] L. Wang and P. G. Schultz, "A general approach for the generation of orthogonal tRNAs," *Chemistry and Biology*, vol. 8, no. 9, pp. 883–890, 2001.

[14] N. Kotlova, T. M. Ishii, E. I. Zagryadskaya, and S. V. Steinberg, "Active suppressor tRNAs with a double helix between the D- and T-loops," *Journal of Molecular Biology*, vol. 373, no. 2, pp. 462–475, 2007.

[15] A. Frankel and R. W. Roberts, "*In vitro* selection for sense codon suppression," *RNA*, vol. 9, no. 7, pp. 780–786, 2003.

[16] H. Taira, T. Hohsaka, and M. Sisido, "*In vitro* selection of tRNAs for efficient four-base decoding to incorporate non-natural amino acids into proteins in an *Escherichia coli* cell-free translation system," *Nucleic Acids Research*, vol. 34, no. 5, article e44, 2006.

[17] D. Lipovsek and A. Plückthun, "*In-vitro* protein evolution by ribosome display and mRNA display," *Journal of Immunological Methods*, vol. 290, no. 1-2, pp. 51–67, 2004.

[18] A. Ogawa, S. Sando, and Y. Aoyama, "Termination-free prokaryotic protein translation by using anticodon-adjusted *E. coli* $tRNA^{Ser}$ as unified suppressors of the UAA/UGA/UAG stop codons. Read-through ribosome display of full-length DHFR with translated UTR as a buried spacer arm," *ChemBioChem*, vol. 7, no. 2, pp. 249–252, 2006.

[19] Y. Shimizu, A. Inoue, Y. Tomari et al., "Cell-free translation reconstituted with purified components," *Nature Biotechnology*, vol. 19, no. 8, pp. 751–755, 2001.

[20] J. Normanly, T. Ollick, and J. Abelson, "Eight base changes are sufficient to convert a leucine-inserting tRNA into a serine-inserting tRNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 12, pp. 5680–5684, 1992.

[21] G. Eggertsson and D. Söll, "Transfer ribonucleic acid-mediated suppression of termination codons in *Escherichia coli*," *Microbiological Reviews*, vol. 52, no. 3, pp. 354–374, 1988.

[22] J. R. Sampson and M. E. Saks, "Contributions of discrete $tRNA^{Ser}$ domains to aminoacylation by *E.coli* seryl-tRNA synthetase: a kinetic analysis using model RNA substrates," *Nucleic Acids Research*, vol. 21, no. 19, pp. 4467–4475, 1993.

[23] H. Murakami, A. Ohta, and H. Suga, "Bases in the anticodon loop of tRNAGGCAla prevent misreading," *Nature Structural and Molecular Biology*, vol. 16, no. 4, pp. 353–358, 2009.

[24] R. W. Roberts and J. W. Szostak, "RNA-peptide fusions for the *in vitro* selection of peptides and proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 23, pp. 12297–12302, 1997.

[25] N. Nemoto, E. Miyamoto-Sato, Y. Husimi, and H. Yanagawa, "*In vitro* virus: bonding of mRNA bearing puromycin at the 3′-terminal end to the C-terminal end of its encoded protein on the ribosome *in vitro*," *FEBS Letters*, vol. 414, no. 2, pp. 405–408, 1997.

Hindawi Publishing Corporation
Journal of Nucleic Acids
Volume 2012, Article ID 392039, 8 pages
doi:10.1155/2012/392039

Research Article

Structural and Functional Characterization of RecG Helicase under Dilute and Molecular Crowding Conditions

Sarika Saxena,[1,2] Satoru Nagatoishi,[1] Daisuke Miyoshi,[1,3] and Naoki Sugimoto[1,3]

[1] *Frontier Institute for Biomolecular Engineering Research (FIBER), Konan University, 7-1-20 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan*
[2] *Amity Institute of Biotechnology, Amity University Uttar Pradesh, Sector-125, Expressway Highway, Noida 201303, India*
[3] *Department of Nanobiochemistry, Faculty of Frontiers of Innovative Research in Science and Technology (FIRST), Konan University, 7-1-20 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan*

Correspondence should be addressed to Naoki Sugimoto, sugimoto@konan-u.ac.jp

Received 4 April 2012; Accepted 30 May 2012

Academic Editor: Masayasu Kuwahara

In an ATP-dependent reaction, the *Escherichia coli* RecG helicase unwinds DNA junctions *in vitro*. We present evidence of a unique protein conformational change in the RecG helicase from an α-helix to a β-strand upon an ATP binding under dilute conditions using circular dichroism (CD) spectroscopy. In contrast, under molecular crowding conditions, the α-helical conformation was stable even upon an ATP binding. These distinct conformational behaviors were observed to be independent of Na^+ and Mg^{2+}. Interestingly, CD measurements demonstrated that the spectra of a frayed duplex decreased with increasing of the RecG concentration both under dilute and molecular crowding conditions in the presence of ATP, suggesting that RecG unwound the frayed duplex. Our findings raise the possibility that the α-helix and β-strand forms of RecG are a preactive and an active structure with the helicase activity, respectively.

1. Introduction

The double-stranded conformation of genomic DNA must be unwound to provide single-stranded DNA (ssDNA) intermediates required for DNA replication, recombination, and repair. The ssDNA intermediates can adopt various structures like junctions, G-quadruplex, and intramolecular triplex [1–3]. In cells, the unwinding of double-stranded DNA (dsDNA) is catalyzed by a class of ubiquitous enzymes termed DNA helicases [4]. Helicases disrupt one or more base pairs within the duplex DNA and then translocate vectorially to the next duplex region to repeat the process [5–9]. The helicase activity is cycled by the binding and hydrolysis of an NTP through a number of energetic (conformational) states that have different affinities for ssDNA and dsDNA [10].

The structure of helicases plays a critical role in their catalytic functions. Previously, it was reported that almost all helicases appear to function as oligomers (usually dimers or hexamers) [10]. Oligomerization provides multiple binding sites necessary for DNA or RNA target recognition, interaction with accessory proteins, and ATP binding [5, 6]. RecG is a well-characterized helicase from *Escherichia coli* that unwinds DNA junctions *in vitro*. Biochemical studies revealed that RecG is active as a monomer [11]. It catalyzes the interconversion of forks and junctions [1, 12, 13]. It is necessary in cellular processes such as DNA replication, recombination, and repair [5, 6]. The conversion of a replication fork into a Holliday junction requires the simultaneous unwinding of the leading and lagging strands followed by the reannealing of the two parental strands and the annealing of the two nascent strands. There is no information available on the relationship between the structural states of RecG and its function. Any study to determine the mechanism of RecG action must, therefore, be addressing the differences between structural states of inactive and active forms of RecG.

Nucleic acids possess an intrinsic structural polymorphism critical in nucleic acid-nucleic acid, nucleic acid-protein, and nucleic acid-drug interactions. The polymorphic properties are influenced not only by sequence but

also by surrounding conditions. The structures adopted are especially influenced by ionic properties like ion concentration, charge, and size [14–16]. Moreover, living cells contain soluble and insoluble molecules such as proteins, nucleic acids, saccharides, lipids, and metabolites that can alter the stabilities of canonical and noncanonical nucleic acid structures [17]. The total concentration of biomolecules reaches 400 g L^{-1} in cells, leading to what is referred to as molecular crowding [17, 18]. In the crowded intracellular environment, water activity decreases and hydration is unfavorable. These crowded conditions stabilize the noncanonical DNA structures such as triplexes and G-quadruplexes, whereas they destabilize the duplex form [19, 20]. Elucidation of the interactions between helicases and DNA substrates with various structures is a very important step in understanding the mechanism through which helicases bind and unwind dsDNA. In this study, the structure and function of RecG were investigated under diluted and crowded conditions shedding light on how the structural properties of RecG correlate with activity.

2. Materials and Methods

2.1. DNA Sequences. DNA oligonucleotides of high performance liquid chromatography (HPLC) purification grade were purchased from Hokkaido System Science (see Table S1 in Supplementary Materials available online at doi:10.1155/2012/392039). Single-strand concentrations of the DNA sequences were determined by measuring absorbance at 260 nm at a high temperature using a Shimadzu 1700 spectrophotometer connected to a thermoprogrammer. Single-strand extinction coefficients were calculated from mononucleotide and dinucleotide data using the nearest neighbor approximation [21, 22].

2.2. Preparation of RecG. The gene encoding RecG was amplified using KOD-Plus DNA polymerase (Toyobo) with *E. coli* BL 21 genomic DNA as the template and the following primers: primer-S (5′-ggaattccatatgaaaggtcgcctgttagatg-3′) and primer-AS (5′-cccgctcgagtcatgcgttggagtaacgttc-3′). Restriction enzyme sites for digestion and ligation are underlined. The PCR products were inserted into the *Nde*I and *Xho*I sites of the pET-26b vector (Merck). RecG was expressed with a hexahistidine tag attached at the N-terminus.

RecG was expressed in *E. coli* strain Rosetta2 (DE3) at 28°C in LB medium supplemented with 30 mg L^{-1} kanamycin and 34 mg L^{-1} chloramphenicol. To induce RecG expression, isopropyl β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM when the optical density of the cells reached approximately 0.6 at 600 nm. The culture was shaken overnight at 28°C. The procedure for purification RecG was as follows: cells were suspended in 20 mM Tris-HCl (pH 8.5), 100 mM NaCl, and membranes were disrupted with a sonicator. The soluble fraction was loaded onto a HiTrap HP column (GE Healthcare). RecG was purified over a HisTrap HP column using buffer A (20 mM Tris (pH 8.0) containing 100 mM NaCl) and buffer B (20 mM Tris (pH 8.0) containing 2 M NaCl) and then a HiLoad 26/60 Superdex 200-pg column (GE Healthcare) using buffer A (20 mM Tris (pH 8.0) containing 100 mM NaCl and 5 mM imidazole) and buffer B (20 mM Tris (pH 8.0) containing 500 mM NaCl and 1 M imidazole). Protein purity was confirmed by SDS-PAGE.

2.3. Circular Dichroism Measurements. Circular dichroism (CD) experiments were performed on a J-820 spectropolarimeter (Jasco) at 4°C and 37°C in a 0.1-cm path length cuvette. Samples of 5 μM RecG were prepared in 30 mM MES (pH 7.0) and 0.5 mM Na_2EDTA containing 100 mM NaCl or 100 mM NaCl and 1 mM $MgCl_2$ or 5 mM $MgCl_2$, with and without 1 mM ATP and 0 wt% or 40 wt% poly(ethylene glycol) with an average molecular mass of 200 (PEG 200). The CD spectra shown are the average of at least three scans from 200 to 350 nm. The temperature of the cell holder was regulated by a temperature controller (PTC-348, Jasco), and the cuvette-holding chamber was flushed with a constant stream of dry N_2 gas to avoid condensation of water on the cuvette exterior. CD melting curves of RecG were recorded at 222 nm in the presence of 30 mM MES (pH 7.0) and 0.5 mM Na_2EDTA containing 100 mM NaCl or 100 mM NaCl and 1 mM $MgCl_2$ or 5 mM $MgCl_2$ and 0 wt% or 40 wt% PEG 200. To analyze the functional activity of RecG, 1 μM frayed duplex DNA was titrated with successive additions of 50 nM RecG and 0.1 mM ATP. Samples were prepared in the presence of 30 mM MES (pH 7.0) and 0.5 mM Na_2EDTA containing 100 mM NaCl or 100 mM NaCl and 1 mM $MgCl_2$ or 5 mM $MgCl_2$. Before measurement, the frayed duplex DNA was heated to 95°C, gently cooled at a rate of 0.5°C min^{-1}, and incubated at 4°C overnight.

2.4. UV Melting Analysis. UV absorbance was measured with the Shimadzu spectrophotometer equipped with the temperature controller. Melting curves of DNA structures were obtained by measuring the UV absorbance at 260 nm. Samples were prepared in 30 mM MES (pH 7.0) and 0.5 mM Na_2EDTA containing 100 mM NaCl or 100 mM NaCl and 1 mM $MgCl_2$ or 5 mM $MgCl_2$. Before measurement, the samples were heated to 95°C, gently cooled at a rate of 0.5°C min^{-1}, and incubated at 4°C overnight. Measurement was performed using a 1-cm path length cuvette. The melting temperature (T_m) values for DNA structures were obtained from the UV melting curves as described previously [21, 22]. The heating rate was 0.5°C min^{-1}.

3. Results and Discussion

3.1. Structures of RecG under Dilute Conditions with and without ATP. Firstly, we used CD spectroscopy to study the structural states of RecG with and without ATP in the presence of different cations (Na^+, Mg^{2+}, or both Na^+ and Mg^{2+}) at 0 wt% and 40 wt% PEG 200, a neutral cosolute to study systematically the effects of molecular crowding. RecG was expressed in *E. coli* and purified by using ion exchange chromatography and affinity chromatography followed by

dialysis. Figure 1(a) shows CD spectra at 4°C or 37°C of 5 μM RecG in the presence of 100 mM Na^+ with or without ATP at 0 wt% PEG 200. The CD spectra without ATP displayed a positive peak at 198 nm and negative peaks at 208 nm and 220 nm, which are characteristic of an α-helix [23]. In the presence of 1 mM ATP, the CD spectra of RecG had a positive peak at 195 nm and a negative peak at 225 nm. These CD signatures are in agreement with formation of a β-strand structure [23].

The crystal structure of *Thermotoga maritime* RecG reveals three structural domains [24] (Figure S1). The largest domain (domain 1) is at N-terminus, which forms a long α-helix and a β-strand [25]. Domains 2 and 3 are referred to as helicase domains. The C-terminal residues of the protein extend from the end of domain 3 and cross-back to contact domain 1, forming a hook that wraps around the extended α-helix, which provides a nucleotide binding site [24]. Recently, the structure of *E. coli* RecG was modeled based on the coordinates of *Thermotoga maritime* RecG usingSwiss-Pdb Viewer [26]. Apart from the missing N-terminal sequences that form a separate fold in the *Thermotoga maritime* RecG, the *E. coli* RecG structure is essentially identical [27]. Morikawa suggested that the insertion before the helicase core in the RecG sequence compared to the *Thermotoga maritime* RecG contributes to the specific recognition of the branched DNA structure by *E. coli* RecG [28]. The β-strand in the insertion domain recognizes the junction through a stacking interaction with several aromatic residues in RecG [24] (Figure S1). Thus, the α-helix and the β-strand structures observed by the CD measurements likely correspond to the ATP binding domain and the DNA junction recognition structure, respectively.

We also measured the CD spectra of RecG as a function of ATP concentration from 0 mM to 1 mM under physiologically relevant ionic conditions (100 mM NaCl and 1 mM $MgCl_2$) at 37°C (Figure S2a). The CD spectra of RecG without ATP displayed a positive peak at 196 nm and negative peaks at 210 nm and 222 nm that remained unchanged at ATP concentrations below 0.17 mM. At ATP concentrations from 0.34 mM to 1 mM, the CD spectra gradually shifted to a positive peak at 216 nm and a negative peak at 228 nm. Figures S2b and S1c show the plots of molar ellipticity at 210 nm and 225 nm, respectively, obtained from CD spectra versus increasing ATP concentration. The appearance of biphasic curves at 210 nm and 225 nm further confirms the structural transition. The difference of the transition points at 210 nm and at 225 nm might indicate that the dependency on ATP is different between the α-helix and the β-strand conformations. These CD results clearly indicate that the α-helix to β-strand transition in RecG depends on the ATP concentration.

3.2. Structures of RecG under Molecular Crowding Conditions with and without ATP. We also explored the effect of ATP binding on RecG under similar ionic conditions and a molecular crowding condition of 40 wt% PEG 200 (Figure 1(b)). CD spectra displayed positive peaks at 190 nm and 198 nm and negative peaks at 208 nm and 220 nm in the absence of ATP. In contrast to the spectra under the diluted conditions, no significant differences were observed in the CD spectra in the presence of 40% PEG 200 even after the addition of ATP, indicating that α-helix structure was dominant. In 100 mM NaCl and 1 mM $MgCl_2$ (Figures S3a and S3b) or 5 mM $MgCl_2$ (Figures S3c and S3d) at 0 wt% or 40 wt% PEG 200, we observed that RecG without ATP folded into α-helical conformation, whereas it was converted into the β-strand conformation after the addition of ATP. This result is identical to the result obtained with 100 mM Na^+ only, indicating that under dilute conditions the binding of ATP to RecG regulates the structural transition from an α-helix to β-strand independent on the nature of cation. On the other hand, the cell-mimicking molecular crowding condition favored the folding of RecG into the α-helical conformation even in the presence of ATP. Previous studies have shown that molecular crowders can stabilize the native state of a protein [29, 30], promote a oligomerization [31, 32], shift an open-closed equilibrium toward a closed form as a substrate-binding state [17], and affect a folding rate of protein [33–35]. Protein-folding variants are proposed to have key roles in a number of pathophysiological processes [36–40]. In view of published results that support the hypothesis that molecular crowding conditions are more relevant to the conditions in cells than dilute conditions, the ATP-independence of the RecG folding may have biological significance.

3.3. Thermal Stability of RecG under Dilute Conditions with and without ATP. To investigate in detail the structural changes induced in RecG upon ATP binding, we explored the thermal stability of RecG using CD melting at 222 nm. Normalized CD melting curves of 5 μM RecG in the presence of 100 mM Na^+ (Figure S4a), 100 mM Na^+ and 1 mM Mg^{2+} (Figure S4b), or 5 mM Mg^{2+} (Figure S4c) in the absence and presence of 1 mM ATP were recorded. A single melting transition was observed in all the conditions. The estimated values of melting temperature $T_{1/2}$ (the temperature at which 50% of a protein sample is denatured) are given in Table S2. Addition of ATP stabilized the RecG by 3°C, 3.5°C and 1°C in the presence of 100 mM Na^+, 100 mM Na^+ and 1 mM Mg^{2+}, or 5 mM Mg^{2+}, respectively. Close internal packing of the backbone atoms in β-strand structures of RecG optimizes van der Waals interactions and minimizes energetically unfavorable hydrophobic interactions between nonpolar protein groups and water molecules in the environment [41]. Collectively, these factors help to reduce the net free energy of the β-strand and thereby increase its stability. Our results suggest that RecG adopts its active functional structure after ATP binding to facilitate the process of unwinding of nucleic acid substrate, because ATP bindings play an important role in the functional activity of various proteins [42–45].

3.4. Thermal Stability of RecG under Molecular Crowding Conditions with and without ATP. In the absence of ATP, RecG was stabilized by 2°C, 1.0°C, and 2.5°C in the presence of 100 mM Na^+, 100 mM Na^+ and 1 mM Mg^{2+}, or 5 mM Mg^{2+} respectively at 40 wt% PEG 200 in comparison to 0 wt% PEG 200. Furthermore, addition of ATP thermally stabilized the RecG at 40 wt% PEG 200, and RecG was maximally

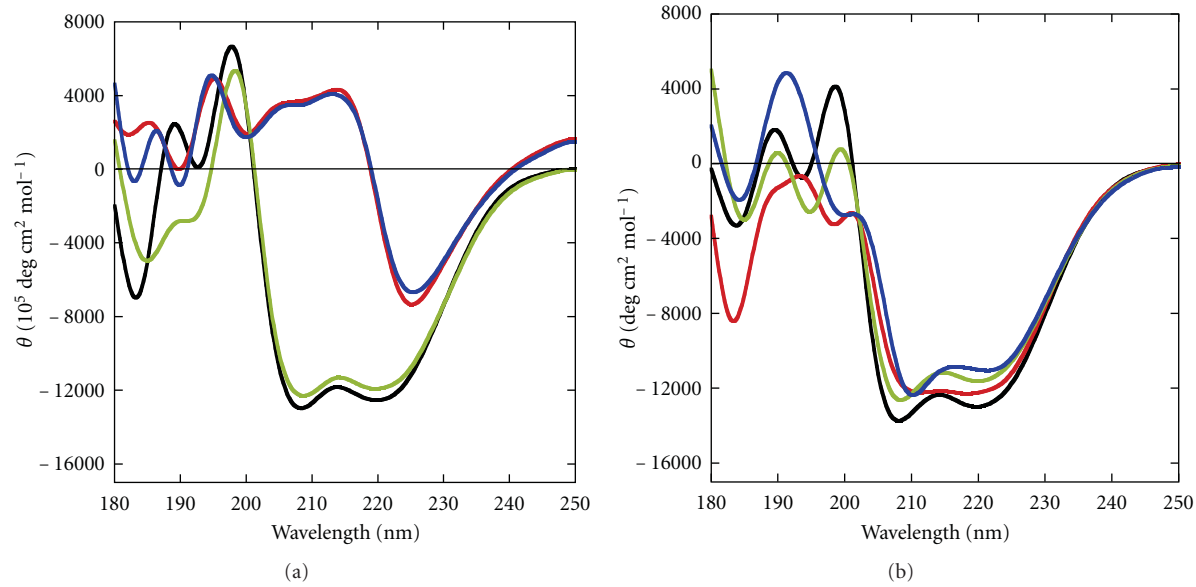


Figure 1: CD spectra of 5 μM RecG. Measurements were carried out in 30 mM MES buffer (pH 7.0) containing 100 mM Na^+ and 0.5 mM Na_2EDTA (a) at 0 wt% PEG 200 without ATP at 4°C (black) and 37°C (green), or with 1 mM ATP at 4°C (red) and 37°C (blue), and (b) at 40 wt% PEG 200 without ATP at 4°C (black) and 37°C (green), or with 1 mM ATP at 4°C (red) and 37°C (blue), respectively.

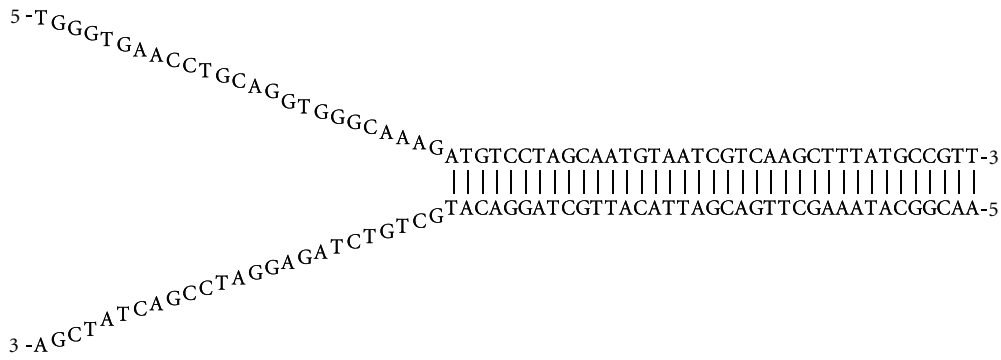


Figure 2: A nucleotide target sequence of frayed duplex.

stable ($T_{1/2}$ = 50°C) at 40 wt% PEG 200 in the presence of ATP. Folded proteins usually have hydrophobic cores and charged or polar side-chains occupy the solvent-exposed surface. Minimizing the number of hydrophobic side chains exposed to water is an important driving force behind the folding process [46]. Formation of intramolecular hydrogen bonds provides another important contribution to protein stability [47]. The strength of hydrogen bonds depends on their environment, thus hydrogen bonds enveloped in a hydrophobic core contribute more than hydrogen-bonds exposed to the aqueous environment [48]. Under molecular crowding conditions, low water activity may induce the inter-domain rearrangement in RecG to allow formation of the α-helix containing a hydrophobic core. The α-helix form was stabilized by 2°C ($T_{1/2}$ = 50°C) at 40 wt% PEG 200 than at 0 wt% PEG 200 ($T_{1/2}$ = 48°C) in the presence of 5 mM Mg^{2+}.

3.5. Thermal Stability of Frayed Duplex under Dilute and Molecular Crowding Conditions. We designed a 62-mer DNA oligonucleotide to form a frayed DNA duplex containing noncomplementary arms and a stem region (Figure 2). This arrangement of bases within the target duplex mimics that of a replication fork. The single-stranded non-complementary arms should facilitate the binding of RecG. The enzyme should then translocate to the dsDNA to allow the strand separation with a 3′–5′ polarity. The ssDNA extension was introduced at the 5′ end to stimulate the activity of RecG on the strand opposite to that used for the primary contact of RecG.

We first measured the stability of frayed duplex by recording the UV melting of 1 μM DNA duplex in the presence of 100 mM Na^+ (Figure S5a), 100 mM Na^+ and 1 mM Mg^{2+} (Figure S5b), and 5 mM Mg^{2+} (Figure S5c) at

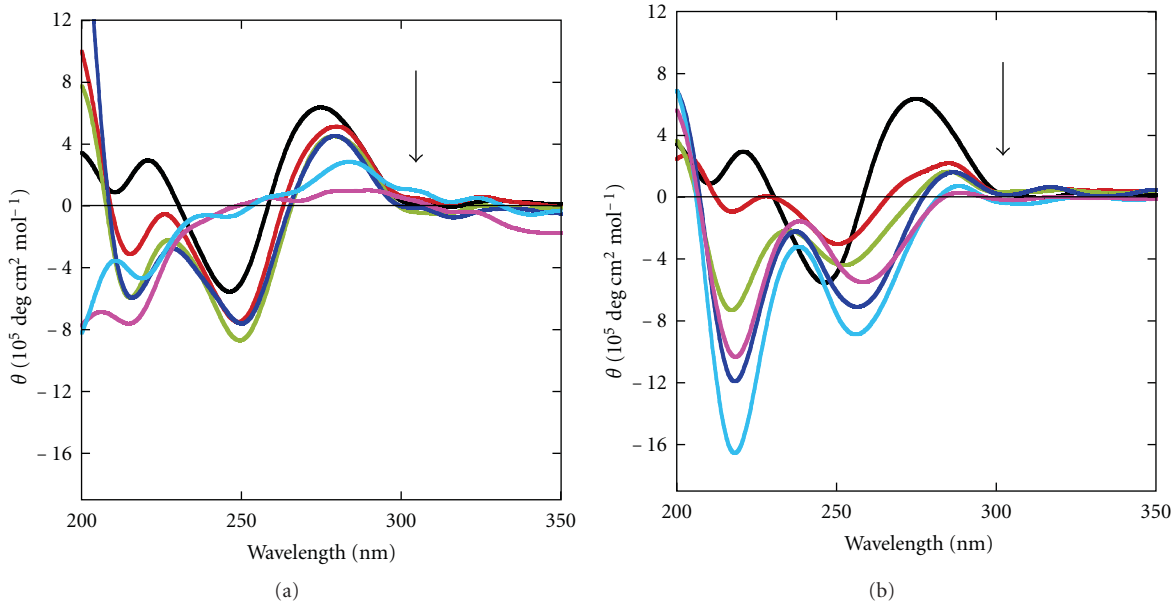


Figure 3: CD spectra of 1 μM DNA frayed duplex. Measurements were carried out at 37°C in 30 mM MES buffer (pH 7.0) containing 100 mM Na^+, 0.5 mM Na_2EDTA, RecG (0 nM) and ATP (0 mM) (black), RecG (50 nM) and ATP (0.1 mM) (red), RecG (100 nM) and ATP (0.2 mM) (green), RecG (150 nM) and ATP (0.3 mM, blue), RecG (200 nM) and ATP (0.4 mM, cyan), and RecG (250 nM) and ATP (0.5 mM) (pink), at (a) 0 wt% PEG 200 and (b) 40 wt% PEG 200.

0 wt% or 40 wt% PEG 200. We observed two transitions for the thermal melting of the frayed duplex. The lower temperature transition could be due to the intramolecular interaction of the bases constituting the arm region of the frayed duplex. The higher-temperature transition could be due to the melting of the stem region. In the absence of PEG 200, the mid-points of the higher-temperature transitions were 70°C, 71°C, and 72°C in the presence of 100 mM Na^+, 100 mM Na^+ and 1 mM Mg^{2+}, and 5 mM Mg^{2+}, respectively. In 40 wt% PEG 200, the higher-temperature transitions in the presence of 100 mM Na^+, 100 mM Na^+ and 1 mM Mg^{2+}, and 5 mM Mg^{2+} were 65°C, 66°C, and 66°C, respectively. To confirm the higher melting transition due to stem region, we have also recorded the stability of stem region of frayed duplex by recording the UV melting of 1 μM DNA duplex in the presence of 100 mM Na^+ (Figure S5a), 100 mM Na^+ and 1 mM Mg^{2+} (Figure S5b), and 5 mM Mg^{2+} (Figure S5c) at 0 wt% or 40 wt% PEG 200. We observed single transition for the thermal melting of the stem region of frayed duplex which confirms that upper transition in frayed duplex is due to the intermolecular structure (Figure S6). In the absence of PEG 200, the mid-points of the higher-temperature transitions were 74°C, 74.5°C, and 74.5°C in the presence of 100 mM Na^+, 100 mM Na^+ and 1 mM Mg^{2+}, and 5 mM Mg^{2+}, respectively. In 40 wt% PEG 200, the higher-temperature transitions in the presence of 100 mM Na^+, 100 mM Na^+ and 1 mM Mg^{2+}, and 5 mM Mg^{2+} were 74°C, 67°C and 66°C, respectively. The decrease in the $T_{1/2}$ values under molecular crowding conditions is in good agreement with previous reports of DNA duplex stabilities in molecular crowders [49].

3.6. Functional Activity of RecG under Dilute Conditions. The CD spectrum of the DNA in the presence of 100 mM Na^+ without PEG 200 was characterized by a positive peak at 275 nm and negative peaks at 211 nm and 247 nm, a spectrum typical of the B-form conformation [50] (Figure 3). We then added RecG and ATP to the DNA. At 50 nM RecG and 0.1 mM ATP, we observed a decrease in intensity of the positive peak at 275 nm and a red shift such that negative peaks were located at 215 nm and 250 nm (Figure 3(a)). There was no signal from RecG due to the low concentration. Spectra were recorded after successive additions of RecG and ATP, and after each addition we observed a decrease in intensity of the positive peak. At 250 nM RecG and 0.5 mM ATP, there was complete loss in original B-form conformation of the DNA, indicating unwinding of the duplex. Previous report described the role of RecG in the processing of stalled replication forks, and acted by reversing the fork past the damage to create a four-way junction that allows template switching and lesion bypass [24]. It has also been reported that RecG unwinds both the leading and lagging strand duplex arms of a three-way junction and the unwinding of these arms was found to be coordinated [24]. In our study, as frayed duplex contains only single-stranded long overhangs at both the terminus, therefore, we propose that RecG should bind to the frayed duplex (as it mimics the replication fork) and should convert the duplex into single strands. Therefore, to confirm the same we have recorded the CD spectra of single stands constituting the frayed duplex under dilute and molecular crowding conditions (Figure S7). The CD spectrum of each single strand of frayed duplex in the presence of 100 mM Na^+, 100 mM Na^+ and

1 mM Mg^{2+}, and 5 mM Mg^{2+} with and without PEG 200 was characterized by a positive peak at 274 nm and negative peaks at 244 nm. These CD signatures indicate that each single strand of the frayed duplex folds into typical of the B-form conformation due to the formation of intramolecular structure. To better understand the structures of frayed duplex after unwinding by RecG, secondary structures of single strands were predicted using M-fold [51, 52]. The monomers of frayed duplex showed 3 short complementary stems with a varying number of bases in loops along with short dangling ends at opposite terminus in each strand (Figure S8). When CD spectra were recorded after successive additions of RecG and ATP, and after each addition, we observed a decrease in intensity of the positive peak (Figure 3 and Figure S9). At 250 nM RecG and 0.5 mM ATP, there was complete loss in original B-form conformation of the DNA, indicating unwinding of the duplex into single strands solely due to the enzymatic activity of RecG in the presence of ATP. Earlier studies indicated that the substrate specificity of RecG was critically dependent on the concentrations of ATP and $MgCl_2$, and under certain conditions, RecG preferentially unwound three strand junctions of DNA [1, 11]. Here, we observed functional activity of RecG in the presence of Na^+ and absence of Mg^{2+}.

Next, we recorded the CD spectra in the presence of 100 mM Na^+ and 1 mM Mg^{2+} (Figure S9a) and in 100 mM Na^+ and 5 mM Mg^{2+} (Figure S9b) without PEG 200. We observed an overall decrease in CD ellipticity at 273 nm that shifted to 283 nm after the successive additions of RecG and ATP, although the extent of decrease of the positive peak was observed to be less than that observed without Mg^{2+}. These results clearly indicate that the functional activity of RecG was not dependent on the nature of the cations. A recent paper showed that the conformational change of the DNA-binding domain of hel308 helicase, which is a member of the same superfamily-2 helicase family as RecG, is coordinated by the ATP binding [53]. As we observed a structural transition from an α-helix to β-strand upon ATP binding under dilute conditions (Figure 1(a)) and functional activity under the same conditions, we propose that the β-strand structure is the conformation of RecG that unwinds our DNA substrate under dilute conditions. The β-strand structure could function as a "helix opener" by actively disrupting base pairs of a frayed DNA duplex, and this may be the basis of recognition of DNA junctions.

3.7. Functional Activity of RecG under Molecular Crowding Conditions. To evaluate the effect of molecular crowding, we repeated our analysis of the activity of RecG in solution containing 40 wt% PEG 200. The CD spectra of the DNA were measured in 100 mM Na^+ (Figure 3(b)), 100 mM Na^+ and 1 mM Mg^{2+} (Figure S9c), and 5 mM Mg^{2+} (Figure S9d) in 40 wt% PEG 200. Of note, the conformation of the frayed duplex is different in 40 wt% PEG 200 than in 0 wt% PEG 200, as mentioned above stability of the frayed duplex was found to be lower under molecular crowding conditions than under dilute conditions. In 100 mM Na^+ in the presence of the crowding agent, the positive peak at 275 nm shifted to 288 nm after the successive addition of 50 nM RecG and 0.1 mM ATP. This indicates the catalytic activity of RecG under molecular crowding conditions is different from that under the dilute conditions.

Similar decreases in intensities of the positive peak at 275 nm and the negative peaks at 221 nm and 247 nm were observed when RecG and ATP were added to the frayed duplex in the presence of 100 mM Na^+ and either 1 mM Mg^{2+} or 5 mM Mg^{2+} and 40 wt% PEG 200 (Figures S9c and S9d). These CD results clearly indicate that RecG unwound the frayed duplex both under dilute and molecular crowding condition independent on the nature of cations, although it is quite possible that the extent of unwinding of frayed duplex and catalytic efficiency of RecG may depend on the nature of cation.

3.8. Possible Relationship between the Structure and the Function of RecG. In the absence of crowder and in the presence of ATP, RecG was converted from an α-helix form to a β-strand form. Under the crowding conditions, on the other hand, RecG remained in α-helix form even when ATP was added. The ATP binding site of RecG is in domains 2 and 3 which both have α-helical conformations [25]. It has been reported that the SecA helicase, which has homologous structure and function to RecG undergoes a transition to an α-helix form upon ATP binding [54]. Therefore, the α-helix form of RecG is likely essential for ATP binding. The conformational change to β-strand form observed in this study may suggest a transition to the active DNA-binding form of RecG. Based on our results, we propose that the β-strand form observed under diluted conditions is the active structure for binding and unwinding of DNA, whereas the α-helix form observed under the crowded conditions is the preactive structure that stably binds ATP. The different unwinding activities observed may be due to the interdomain flexibility of RecG, which allows substantial conformational changes imparting the ability to recognize more than one DNA structure. Furthermore, our results can provide significant information to explore and design the small ligands regulating the RecG helicase activity with a molecular crowding condition.

4. Conclusion

The data presented in this study emphasize some important points. Firstly, the secondary structures of RecG were different with and without a molecular crowding agent, PEG 200. The structural transition from α-helix to β-strand observed under the dilute conditions depended solely on the binding of ATP and not on the nature of cations present in solution. The molecular crowding conditions favored the folding of RecG bound to ATP into an α-helical structure. This may be a consequence of hydration. Secondly, the frayed duplex was unwound by RecG under dilute and molecular crowding conditions, indicating that the β-strand conformation may be the active structure of RecG, with the α-helix form the pre-active structure of RecG. Molecular crowding conditions played a critical role in the functional structure of RecG, and our data suggest the reason for

the observed differences in the unwinding activity *in vivo* and *in vitro*. We are currently working to determine the binding stoichiometry of ATP with RecG and to quantify functional activity of RecG under dilute and molecular crowding conditions on various DNA substrates.

Acknowledgments

This work was supported in part by Grants-in-Aid for Scientific Research, the "Core research" project (2009–2014) from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and the Hirao Taro Foundation of the Konan University Association for Academic Research and the research grant of "Department of Science and Technology (DST)," Govt of India, (SERC/LS-0594/2010).

References

[1] P. McGlynn and R. G. Lloyd, "RecG helicase activity at three- and four-strand DNA structures," *Nucleic Acids Research*, vol. 27, no. 15, pp. 3049–3056, 1999.

[2] R. Giraldo, M. Suzuki, L. Chapman, and D. Rhodes, "Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 16, pp. 7658–7662, 1994.

[3] R. M. Brosh, A. Majumdar, S. Desai, I. D. Hickson, V. A. Bohr, and M. M. Seidman, "Unwinding of a DNA triple helix by the werner and bloom syndrome helicases," *Journal of Biological Chemistry*, vol. 276, no. 5, pp. 3024–3030, 2001.

[4] S. W. Matson and K. A. Kaiser-Rogers, "DNA helicases," *Annual Review of Biochemistry*, vol. 59, pp. 289–329, 1990.

[5] T. M. Lohman, "*Escherichia coli* DNA helicases: mechanisms of DNA unwinding," *Molecular Microbiology*, vol. 6, no. 1, pp. 5–14, 1992.

[6] T. M. Lohman, "Helicase-catalyzed DNA unwinding," *Journal of Biological Chemistry*, vol. 268, no. 4, pp. 2269–2272, 1993.

[7] G. T. Yarranton and M. L. Gefter, "Enzyme-catalyzed DNA unwinding: studies on *Escherichia coli* Rep protein," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 76, no. 4, pp. 1658–1662, 1979.

[8] N. Arai, K. I. Arai, and A. Kornberg, "Complexes of Rep protein with ATP and DNA as a basis for helicase action," *Journal of Biological Chemistry*, vol. 256, no. 10, pp. 5287–5293, 1981.

[9] I. Wong and T. M. Lohman, "Allosteric effects of nucleotide cofactors on *Escherichia coli* Rep helicase-DNA binding," *Science*, vol. 256, no. 5055, pp. 350–355, 1992.

[10] K. J. M. Moore and T. M. Lohman, "Helicase-catalyzed DNA unwinding: energy coupling by DNA motor proteins," *Biophysical Journal*, vol. 68, supplement 4, pp. 180s–185s, 1995.

[11] P. McGlynn, A. A. Mahdi, and R. G. Lloyd, "Characterisation of the catalytically active form of RecG helicase," *Nucleic Acids Research*, vol. 28, no. 12, pp. 2324–2332, 2000.

[12] P. McGlynn and R. G. Lloyd, "Modulation of RNA polymerase by (p)ppGpp reveals a RecG-dependent mechanism for replication fork progression," *Cell*, vol. 101, no. 1, pp. 35–45, 2000.

[13] P. McGlynn, R. G. Lloyd, and K. J. Marians, "Formation of Holliday junctions by regression of nascent DNA in intermediates containing stalled replication forks: RecG stimulates regression even when the DNA is negatively supercoiled," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 15, pp. 8235–8240, 2001.

[14] G. S. Manning, "On the application of polyelectrolyte "limiting laws" to the helix-coil transition of DNA. I. Excess univalent cations," *Biopolymers*, vol. 11, no. 5, pp. 937–949, 1972.

[15] M. T. Record Jr., C. F. Anderson, and T. M. Lohman, "Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity," *Quarterly Reviews of Biophysics*, vol. 11, no. 2, pp. 103–178, 1978.

[16] C. F. Anderson and M. T. Record, "Ion distributions around DNA and other cylindrical polyions: theoretical descriptions and physical implications," *Annual Review of Biophysics and Biophysical Chemistry*, vol. 19, pp. 423–465, 1990.

[17] H. Dong, S. Qin, and H. X. Zhou, "Effects of macromolecular crowding on protein conformational changes," *PLoS computational biology*, vol. 6, p. e1000833, 2010.

[18] D. Miyoshi and N. Sugimoto, "Molecular crowding effects on structure and stability of DNA," *Biochimie*, vol. 90, no. 7, pp. 1040–1051, 2008.

[19] D. Miyoshi, K. Nakamura, K. Tateishi, T. Ohmichi, and N. Sugimoto, "Hydration of watson crick base pairs and dehydration of hoogsteen base pairs inducing structural polymorphism under molecular crowding conditions," *Journal of the American Chemical Society*, vol. 131, no. 10, pp. 3522–3531, 2009.

[20] H. X. Zhou, G. Rivas, and A. P. Minton, "Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences," *Annual Review of Biophysics*, vol. 37, pp. 375–397, 2008.

[21] N. Sugimoto, S. I. Nakano, M. Katoh et al., "Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes," *Biochemistry*, vol. 34, no. 35, pp. 11211–11216, 1995.

[22] S. I. Nakano, T. Kanzaki, and N. Sugimoto, "Influences of ribonucleotide on a duplex conformation and its thermal stability: study with the chimeric RNA-DNA strands," *Journal of the American Chemical Society*, vol. 126, no. 4, pp. 1088–1095, 2004.

[23] M. K. Sharon, J. J. Thomas, and C. P. Nicholas, "How to study proteins by circular dichroism," *Biochimica et Biophysica Acta*, vol. 1751, no. 2, pp. 119–139, 2005.

[24] M. R. Singleton, S. Scaife, and D. B. Wigley, "Structural analysis of DNA replication fork reversal by RecG," *Cell*, vol. 107, no. 1, pp. 79–89, 2001.

[25] A. A. Mahdi, P. McGlynn, S. D. Levett, and R. G. Lloyd, "DNA binding and helicase domains of the *Escherichia coli* recombination protein RecG," *Nucleic Acids Research*, vol. 25, no. 19, pp. 3875–3880, 1997.

[26] N. Guex and M. C. Peitsch, "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling," *Electrophoresis*, vol. 18, no. 15, pp. 2714–2723, 1997.

[27] G. S. Briggs, A. A. Mahdi, G. R. Weller, Q. Wen, and R. G. Lloyd, "Interplay between DNA replication, recombination and repair based on the structure of RecG helicase," *Philosophical Transactions of the Royal Society B*, vol. 359, no. 1441, pp. 49–59, 2004.

[28] T. Nishino, K. Komori, D. Tsuchiya, Y. Ishino, and K. Morikawa, "Crystal structure and functional implications of *Pyrococcus furiosus* Hef helicase domain involved in branched DNA processing," *Structure*, vol. 13, no. 1, pp. 143–153, 2005.

[29] K. Sasahara, P. McPhie, and A. P. Minton, "Effect of dextran on protein stability and conformation attributed to

macromolecular crowding," *Journal of Molecular Biology*, vol. 326, no. 4, pp. 1227–1237, 2003.

[30] D. S. Spencer, K. Xu, T. M. Logan, and H. X. Zhou, "Effects of pH, salt, and macromolecular crowding on the stability of FK506-binding protein: an integrated experimental and theoretical study," *Journal of Molecular Biology*, vol. 351, no. 1, pp. 219–232, 2005.

[31] S. B. Qin and H.-X. Zhou, "Atomistic modeling of macromolecular crowding predicts modest increases in protein folding and binding stability," *Biophysical Journal*, vol. 97, no. 1, pp. 12–19, 2009.

[32] J. Batra, K. Xu, S. B. Qin, and H. X. Zhou, "Effect of macromolecular crowding on protein binding stability: modest stabilization and significant biological consequences," *Biophysical Journal*, vol. 97, no. 3, pp. 906–911, 2009.

[33] B. van den Berg, R. J. Ellis, and C. M. Dobson, "Effects of macromolecular crowding on protein folding and aggregation," *The EMBO Journal*, vol. 18, no. 24, pp. 6927–6933, 1999.

[34] R. J. Ellis, "Macromolecular crowding: obvious but underappreciated," *Trends in Biochemical Sciences*, vol. 26, no. 10, pp. 597–604, 2001.

[35] X. Ai, Z. Zhou, Y. Bai, and W. Y. Choy, "15N NMR spin relaxation dispersion study of the molecular crowding effects on protein folding under native conditions," *Journal of the American Chemical Society*, vol. 128, no. 12, pp. 3916–3917, 2006.

[36] J. Martin and F. U. Hartl, "Chaperone-assisted protein folding," *Current Opinion in Structural Biology*, vol. 7, no. 1, pp. 41–52, 1997.

[37] F. G. van der Goot, J. M. Gonzalez-Manas, J. H. Lakey, and F. Pattus, "A "molten-globule" membrane-insertion intermediate of the pore-forming domain of colicin A," *Nature*, vol. 354, no. 6352, pp. 408–410, 1991.

[38] J. Safar, H. Wille, V. Itri et al., "Eight prion strains have PrP(Sc) molecules with different conformations," *Nature Medicine*, vol. 4, no. 10, pp. 1157–1165, 1998.

[39] J. S. Dennis, "Folding proteins in fatal ways," *Nature*, vol. 426, no. 6968, pp. 900–904, 2003.

[40] F. Chiti and C. M. Dobson, "Protein misfolding, functional amyloid, and human disease," *Annual Review of Biochemistry*, vol. 75, pp. 333–366, 2006.

[41] V. Smirnovas, R. Winter, T. Funck, and W. Dzwolak, "Thermodynamic properties underlying the α-helix-to-β-sheet transition, aggregation, and amyloidogenesis of polylysine as probed by calorimetry, densimetry, and ultrasound velocimetry," *Journal of Physical Chemistry B*, vol. 109, no. 41, pp. 19043–19045, 2005.

[42] M. Saraste, P. R. Sibbald, and A. Wittinghofer, "The P-loop - a common motif in ATP- and GTP-binding proteins," *Trends in Biochemical Sciences*, vol. 15, no. 11, pp. 430–434, 1990.

[43] T. W. Traut, "The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites," *European Journal of Biochemistry*, vol. 222, no. 1, pp. 9–19, 1994.

[44] S. Klumpp, D. Kriha, G. Bechmann et al., "Phosphorylation of the growth factors bFGF, NGF and BDNF: a prerequisite for their biological activity," *Neurochemistry International*, vol. 48, no. 2, pp. 131–137, 2006.

[45] K. Rose, R. E. Gast, A. Seeger, J. Krieglstein, and S. Klumpp, "ATP-dependent stabilization and protection of fibroblast growth factor 2," *Journal of Biotechnology*, vol. 145, no. 1, pp. 54–59, 2010.

[46] C. N. Pace, B. A. Shirley, M. Mcnutt, and K. Gajiwala, "Forces contributing to the conformational stability of proteins," *The FASEB Journal*, vol. 10, no. 1, pp. 75–83, 1996.

[47] G. D. Rose, P. J. Fleming, J. R. Banavar, and A. Maritan, "A backbone-based theory of protein folding," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 45, pp. 16623–16633, 2006.

[48] S. Deechongkit, H. Nguyen, P. E. Dawson, M. Gruebele, J. W. Kelly, and E. T. Powers, "Context-dependent contributions of backbone hydrogen bonding to β-sheet folding energetics," *Nature*, vol. 430, no. 6995, pp. 101–105, 2004.

[49] D. Miyoshi, S. Matsumura, S. I. Nakano, and N. Sugimoto, "Duplex dissociation of telomere DNAs induced by molecular crowding," *Journal of the American Chemical Society*, vol. 126, no. 1, pp. 165–169, 2004.

[50] W. C. Johnson, "CD of nucleic acids," in *Circular Dichroism: Principles and Applications*, K. Nakanishi, N. Berova, and R. W. Woody, Eds., pp. 523–540, Wiley-VCH, New York, NY, USA, 1994.

[51] M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3406–3415, 2003.

[52] D. H. Mathews, M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, and D. H. Turner, "Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 19, pp. 7287–7292, 2004.

[53] K. Büttner, S. Nehring, and K.-P. Hopfner, "Structural basis for DNA duplex separation by a superfamily-2 helicase," *Nature Structural & Molecular Biology*, vol. 14, no. 7, pp. 647–652, 2007.

[54] D. Keramisanou, N. Biris, I. Gelis et al., "Disorder-order folding transitions underlie catalysis in the helicase motor of SecA," *Nature Structural & Molecular Biology*, vol. 13, no. 7, pp. 594–602, 2006.