Review

Creating hybrid genes by homologous recombination

Peter L. Wang

Centre for Protein Engineering and Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England, UK

Tel.: +44 1223 402104; Fax: +44 1223 402140;

E-mail: plw@mrc-lmb.cam.ac.uk

Recombination of homologous genes is a powerful mechanism for generating sequence diversity, and can be applied to protein analysis and directed evolution. *In vitro* recombination methods such as DNA shuffling are very flexible and can give hybrid genes with multiple crossovers; they have been used extensively to evolve proteins with improved and novel properties. *In vivo* recombination in both *E. coli* and yeast is greatly enhanced by double-strand breaks; for *E. coli*, mutant strains are often necessary to obtain high efficiency. Intra- and inter-molecular recombination *in vivo* have distinct features; both give hybrids with one or two crossovers, and have been used to study structure-function relationships of many proteins. Recently *in vivo* recombination has been used to generate diversity for directed evolution, creating a large phage display antibody library. Recombination methods will become increasingly useful in light of the explosion in genomic sequence data and potential for engineered proteins.

1. Introduction

Hybrids created by recombination of homologous genes benefit from the accumulated experience of evolution. Although homologs have diverged in sequence and often in function, their differences must be compatible with the same protein fold. Thus hybrid genes allow a large jump in "sequence space" that is much more likely to give a correctly folded protein than a similar jump due to random mutations. On a less radical scale, once individual random mutations have been vetted by selection, recombination of these is more likely to give functional proteins than new random mutations. Evolution can proceed faster with than without recombina-

tion (here meaning homologous recombination, unless qualified).

These principles apply equally to genetic change in nature or directed by man. Recombination is pervasive throughout nature and for sexual organisms is obligatory at each generation. However, creation of hybrid genes is less common than recombination between alleles; in the vast majority of cases this is likely to be deleterious [1], even though functional proteins may result. Examples from human genetics include recombination between δ and β globin genes to give hemoglobin Lepore [2], and recombination between red and green visual pigment genes resulting in color-blindness [3].

When rapid change is at a premium, as in the competition between host and parasite, recombination can be adapted to produce hybrid genes at very high rates. Chickens (and also rabbits, pigs, and cows) generate antibody diversity by gene conversion, a recombination process that copies bits of sequence from multiple pseudogenes into the single active antibody locus to create many different antibody genes [4]. A strikingly similar process is used by the bacterium *Borrelia burgdorferi* during Lyme disease infection: mosaic versions of the antigenic surface lipoprotein *vlsE* are generated by gene conversion of the active *vlsE* locus with an array of silent *vls* cassettes [5,6]; switching between antigenic variants allows evasion of the host immune response. Recombination is also used for antigenic variation by other pathogens including *Borrelia helmsii*, *Neisseria*, and *trypanosomes* [7].

Just as nature finds recombination a useful tool, so does the molecular biologist. This review is about creating hybrid genes using *in vitro* recombination and *in vivo* recombination in *Escherichia coli* and the yeast *Saccharomyces cerevisiae*, those organisms in which recombination is best understood. There are two main reasons for creating hybrid genes. The first is analysis: hybrids are useful for studying protein function because they access genetic differences yet are likely to maintain three-dimensional structure. The second is directed

evolution: recombination is an efficient way to create genetic variants that can be selected for improved or novel functions. Different recombination methods have advantages and disadvantages for these two types of applications.

2. In vitro recombination

Recombination *in vitro* was noted early in the history of PCR and was attributed to incomplete extension and re-annealing [8]. The percentage of PCR recombinants could be increased by shortening the extension time per cycle, and it was suggested that this might be useful for producing populations of hybrid DNAs [9]. DNA with breaks and other damage is particularly prone to PCR-induced recombination [10]. Clearly, the ability to undergo recombination is a direct consequence of the complementary structure of DNA.

Stemmer demonstrated the usefulness of *in vitro* recombination for directed evolution with his technique of “DNA shuffling” [11,12], also dubbed “sexual PCR” [13]. A gene is cut into small fragments with DNase I, and then re-assembled in a PCR-like process that also tends to introduce point mutations (Fig. 1(a)). The resulting products are subjected to selection or screening, and the survivors undergo another round of DNA shuffling. The first round is no different from error-prone PCR, but in subsequent rounds the evolutionary speed-up expected from recombination comes into play. A further important advance was to begin the shuffling process with a set of homologous genes (“family shuffling”) rather than a single gene, tapping into sequence diversity derived from natural evolution [14].

Related methods not involving DNA degradation have been described. Random-priming *in vitro* recombination (RPR) creates the short fragments needed for re-assembly by primer extension of random oligonucleotides (Fig. 1(b)) [15]. Staggered extension process (StEP) comprises cycles of denaturation and polymerase extension where the amount of extension per cycle is deliberately kept short by short times and lower temperature [16,17]. The gradually extending chains have an opportunity to switch templates with each cycle (Fig. 1(c)). Certain features are common to all these *in vitro* methods. Mutations are likely due to the extensive use of *in vitro* polymerization. The usual protocols are intended to produce many crossovers; single crossovers can be favored by altering parameters (such as fragment size, cycling), but this will also increase

the percentage of non-recombinants. Similarly, as one recombinates increasingly divergent genes, re-assembly of identical sequences will become strongly favored and the percentage of non-recombinants will increase.

In vitro recombination has been applied to the evolution of a wide range of proteins [18,19]. The majority of studies have shuffled only point mutations; the potential of family shuffling is just beginning to be explored. Shuffling of four cephalosporinases gave greatly increased activity against moxalactam [14]; shuffling of two biphenyl dioxygenases gave novel substrate ranges [20]; shuffling of two herpes thymidine kinases gave increased activity on zidovudine (AZT) [21]; shuffling of >20 human interferons gave better activity towards mouse cells [22]; and shuffling of 26 different subtilisin proteases gave combinations of improved properties: thermostability, solvent tolerance, and activity at acid or basic pH [23].

In vitro recombination has clearly been extremely useful for directed evolution. It can generate hybrids with many crossovers and it is very flexible. It should be very compatible with completely *in vitro* selection technologies such as ribosome display [24], RNA display [25], and man-made compartments [26]. On the other hand, for analysis of protein function the high mosaicism and mutations generated by *in vitro* methods complicate interpretation of results and increase the number of clones that need to be tested. They will probably most useful when a selection or rapid screen exists for the function of interest.

3. Mismatch repair

An interesting partially *in vivo* approach relies not on recombination machinery but mismatch repair. DNAs are constructed that are heteroduplex for the genes of interest; after transformation, mismatches are corrected *in vivo* (Fig. 2). If either there is no repair or if repair is concerted, only parental genes will be recovered. If only some mismatches are repaired or if not all repairs use the same strand as template (“independent repair”), then hybrid genes will result. The predominant mismatch repair pathway in *E. coli* does work by a concerted mechanism: one strand is targeted for replacement synthesis, and these replacement tracts can be several kilobases long [27]. Despite this, hybrid genes have been observed after transformation of multiply mismatched heteroduplexes, usually at low frequencies [28–31]. Heteroduplexes have typically been made by denaturing and re-annealing two plasmids cut

Fig. 1. *In vitro* recombination. Variations on PCR-like assembly of hybrid genes. Both DNA strands are pictured; existing DNA is a solid color, newly synthesized DNA is striped. Oligonucleotide primers are depicted as open rectangles, and point mutations by small open ellipses. For all methods, a smear is usually generated; a discrete-sized product is obtained by further cycling under conventional PCR conditions with specific primers. a) DNA shuffling: fragmentation is usually done with DNase I, and the desired size range obtained by preparative gel electrophoresis. Thermal cycling is similar to standard PCR. b) Random-primed recombination (RPR): random hexamer primers are extended with Klenow polymerase; the size of the extension products is inversely related to the hexamer concentration used. Template and primers are removed by ultrafiltration. c) StEP: Specific primers are used, and polymerization is kept short by the cycle conditions: for example, 30 seconds at 94°C then 5 seconds at 55°C, repeated 80 times.

at different sites; heteroduplexes re-form circles having nicks or single-strand gaps. Volkov et al. have tested several heteroduplex methods for producing hybrid genes [32]. Their most effective method was to anneal a gene heteroduplex from two asymmetric PCRs and ligate this into a cut vector; 30% of transformants had the desired crossover between two mutations 463 bp apart.

4. *In vivo* recombination: Background

Some background about recombination in *E. coli* and yeast is helpful in understanding its application to making hybrid genes. The primary reason recombination evolved was probably not to generate sequence diversity, but to repair DNA damage. Recombination is usually greatly enhanced by the presence of DNA damage,

Fig. 2. Mismatch repair. Different outcomes of a heteroduplex. A molecule is generated *in vitro* that has mismatches (pictured as spikes) only in the genes of interest (thick lines), while the plasmid backbone (thin lines) is perfectly matched. Depending on how the molecule is assembled, there may also be nicks or single-strand gaps, which are not shown. The pictured products are those occurring after any mismatch repair and one round of replication.

especially double-strand breaks and crosslinks. The initiation and outcome of recombination depend greatly on the precise structure of the DNA substrates – the presence of gaps or breaks, the lengths (chromosome vs plasmid), positioning of the homologous segments; different substrates may have different genetic requirements for effective recombination. Recent models have emphasized the inter-relationships between DNA recombination and replication ([33] and rest of that issue.)

4.1. Recombination in *E. coli*

Genetic analysis has defined three pathways of recombination in *E. coli*: the RecBCD, RecE, and RecF pathways. The pathway distinctions are useful but simplistic; in reality recombination is accomplished by a network of genes, many shared between the “pathways”. The multifunctional RecA protein is critical for most recombination reactions; it coats single-stranded DNA and promotes its invasion into homologous duplex DNA and subsequent strand exchange. For reviews on recombination in *E. coli* see [34,35].

The RecBCD pathway is the major pathway for chromosomal recombination in wildtype cells. However, RecBCD is not normally involved in recombination of small plasmids. RecBCD is a potent helicase and

nuclease that facilitates RecA loading onto the single strands it creates. Interestingly, when only its nuclease activity is inactivated (as in *recD* strains), intra-plasmid recombination is increased [36,37]. The other two recombination pathways were revealed by suppressor mutations that restored chromosomal recombination to *recB recC* strains. Both alternate pathways are much more proficient at intra-molecular plasmid recombination than wildtype, especially if a double-strand break is present [38].

One suppressor genotype (*recBC sbcB*) activates the RecF pathway for chromosomal recombination. Even in wildtype cells, the RecF pathway mediates plasmid recombination as this is inhibited by *recF*, *recO*, and *recR* mutations. These genes also are important in wildtype cells for recombinational repair of DNA lesions encountered during replication [35]. The other suppressor genotype discovered was *recBC sbcA*, activating the RecE pathway. The *sbcA* mutation turns on expression of *recE* and *recT*, a 5'→3' exonuclease and homologous pairing protein, respectively [39]. These genes are functionally similar to the phage lambda recombination genes *Redα* and *Redβ* (also known as *exo* and *bet*), while the *Redγ* (or *gam*) gene inactivates RecBCD – thus recombination in a *recBC sbcA* cell is completely analogous to that in a cell infected by wild-type phage lambda; in fact, *recE* and *recT* are remnants

of a lambda-like phage named Rac that integrated into the *E. coli* genome long ago.

4.2. Recombination in yeast

Unlike in *E. coli* where mutant strains have advantages for recombination, in yeast there is little need for special genotypes. Wildtype yeast are highly proficient at recombination, particularly at double-strand breaks. Double-strand breaks are important for initiating recombination during meiosis, and for switching of mating-type which occurs by gene conversion. Over 10 genes have been implicated in a recombination epistasis group that has *RAD52* as its central member: all types of recombination appear to involve *RAD52*, while mutants in other members show only partial recombination defects. The group also includes *RAD51*, a homolog of bacterial *recA* that also catalyzes homologous strand invasion and exchange. These and other *RAD52* group genes have counterparts in mammalian cells, so basic mechanisms are likely to be conserved from yeast to man. For a review on yeast recombination see [40].

5. *In vivo* recombination: Applications

An important distinction for recombination methods is whether the substrates lie on the same or different DNA molecules (intra- vs inter-molecular recombination). Intra-molecular recombination is more efficient since homology search is aided by the substrates being physically linked; however it usually involves a specially constructed DNA. Inter-molecular recombination is much more flexible but generally less efficient.

5.1. Intra-molecular plasmid recombination in *E. coli*

Weber and Weissman demonstrated the importance of double-strand breaks for efficient recombination [41]. They first created two plasmids with human interferon- $\alpha 1$ and $\alpha 2$ genes ($\sim 80\%$ nucleotide identity) having different resistance markers. Fragments from each, sharing homology only within the interferon genes, were ligated into linear concatamers and transformed (Fig. 3(a)). Circularization occurred via intra-molecular recombination *in vivo*, giving rise to doubly-resistant colonies. DNA sequencing identified 11 different single-crossover products.

Subsequent workers have chosen to first construct a plasmid containing both homologs in head-to-tail orientation, sometimes with intervening markers. Note

Fig. 3. *In vivo* recombination: Intra-molecular. a) The original scheme of Weber and Weissman. Vectors with different antibiotic resistance markers are digested. The fragments are purified and ligated (the free ends have incompatible overhangs so a circle cannot be generated); this is transformed and re-circularization occurs by recombination. Inter-molecular recombination is also possible but less likely. In most cases an intermediate plasmid is constructed containing both genes of interest: b) recombination of circles is infrequent, so a counter-selectable marker is required ("X"). c) If a double-strand break is made between the genes, recombination can re-create a circle; most surviving plasmids will be hybrids.

that one needs to construct two plasmids (with the order of the genes reversed) to get hybrids of both polarities. This plasmid is then subjected to *in vivo* recombination, either as a circle or as linear DNA. If circular DNAs are used (Fig. 3(b)), it is important to have an intervening counter-selectable marker such as *galK* [42] or a unique restriction site [43] since the recombination frequency is usually low. More commonly, the plasmid is cut between the two homologs and transformed as a linear fragment (Fig. 3(c)) [44,45]; this provides double-strand ends that stimulate recombination and also selects for re-circularization since linear fragments cannot replicate. Abastado et al. studied various factors in this reaction [46]. Products were the result of single crossovers, which were much more frequent near the homology ends but distributed throughout the homo-

gous region including more than 2.5 kb away from the break. Heterologous sequence at one end had no deleterious effect, but heterology on both ends completely blocked recombination. However, others observed recombination even with heterology on both ends [44, 45]. Surprisingly, *recA* is not required or is less important when a double-strand break is present [46–49]. One group found it advantageous to use a *recBC sbcA* strain for creating hybrid genes by this approach [50].

Intra-molecular recombination in *E. coli* has been used to create hybrids in a wide variety of gene families, for example: interferons [41], amylases [43], lipases [51], adenylyl cyclases [52], insecticidal toxins [49], *recA* [50], adenovirus E1A [53], outer membrane proteins [44,45], major histocompatibility complex antigens [46], neurotransmitter transporters [54], cAMP receptors [55], and chemokine receptors [56]. The process has also been called “random chimeragenesis” [52]. The degree of identity ranged from 47–90%; however, with sufficient screening it was possible to identify a few in-frame hybrids between cytochrome P450 genes sharing only 25% amino acid identity [57].

5.2. Intra- and inter-molecular plasmid recombination in yeast

Pompon and Nicolas reported using both intra- and inter-molecular recombination in yeast to create hybrid cytochrome P450 genes [58]. Intra-molecular recombination (Fig. 3(c)) was more efficient than inter-molecular recombination (discussed below) [59]. As in *E. coli*, the majority of transformants were the result of single crossovers, which occurred most frequently near a double-strand end. Whereas *rad52* mutation completely eliminates inter-molecular recombination, it has much less effect on intra-molecular recombination [59]. However, intra-molecular recombination in yeast has been little used for engineering purposes, presumably since inter-molecular recombination is reasonably efficient and is much more flexible.

Linear DNA fragments will readily undergo inter-molecular recombination with another homologous DNA in yeast [60]. Recombination of a fragment with homologous ends and a chromosomal locus (Fig. 4(a)) makes it simple to knock-out genes in yeast. Recombination between a fragment and a cut plasmid (Fig. 4(b)) is useful for plasmid construction [61,62]; recently it was used to clone the entire complement of yeast coding sequences for two-hybrid and biochemical analyses [63,64]. A similar reaction can be used to map mutations in cloned genes [65,66]. Pompon and Nico-

Fig. 4. *In vivo* recombination: Inter-molecular. With two or more recombining DNAs, there are many possibilities. Genes are pictured as rectangles of various shades; identical vector sequences are thick lines. a) Recombination of a linear fragment into an intact DNA is an infrequent event, but can be selected for using a marker and can be designed to inactivate the target locus. The target is shown as a circular plasmid, but can also be a linear yeast chromosome. b) Plasmid construction by recombination (*in vivo* cloning). A gene flanked by sequences homologous to the vector (which may be appended by PCR) is co-transformed with linearized vector; recombination results in a circle. c) Three-part hybrids: a gene fragment recombines with a plasmid-borne homolog that is cut at an internal site. d) Two-part hybrids: a gene fragment together with vector sequence recombines with a homologous gapped plasmid. e) Multi-parent hybrids: multiple co-transformed mutant genes or homologs may first recombine with each other and then with the linearized vector.

las used it to make hybrid genes [58]. The gene on one plasmid was cut at an internal restriction site and co-transformed with a homologous gene on another fragment, resulting in three-part hybrids (Fig. 4(c)). Two-part hybrids can be made by using a fragment with vector sequence at one end (Fig. 4(d)) [67]. Recombination between different cytochrome P450 genes (71% nucleotide identity) was 13–18-fold lower than between identical genes. Recombination between the yeast ARG4 gene and its human homolog (52% nucleotide identity) produced very few and mostly aberrant plasmids, although two in-frame, functional recombinants were identified [68]. Crossovers are most frequent near the double-strand breaks; treatment of ends with BAL-31 exonuclease results in a more even

distribution of crossovers. More complex hybrids (five-part rather than three-part) can also occur [69]. Yeast recombination was recently used to shuffle peroxidase mutants for improved stability: when multiple donor fragments were co-transformed, these could recombine with each other in addition to the cut plasmid (Fig. 4(e)) [70]. There have been fewer examples of hybrid recombination in yeast than in *E. coli*; it has mainly been used when gene expression will also be in yeast.

5.3. Inter-molecular recombination in *E. coli*

Applications of inter-molecular recombination in *E. coli* have so far mainly been for plasmid construction. Recombination between two circular plasmids occurs at low frequencies ($\sim 10^{-4}$ in wildtype bacteria [71,72]), but hybrid genes can be identified if one has a suitable selection [73]. Analogous to the yeast reaction above (Fig. 4(b)), insert and vector fragments whose ends shared 30 bp identity could recombine *in vivo* to create a circular plasmid [74]; even three overlapping fragments can be recombined [75]. Several strains have been used, including *recA* strains, but the reaction was more efficient in *recBC sbcA* [76] or *recBC sbcB* strains [77, 78]. A linear fragment with short end identities can also recombine into an intact DNA (Fig. 4(a)), but only in *recBC sbcA* strains and at a low frequency; the linear fragment must confer an antibiotic resistance so that recombinants can be selected [79]. This allows targeted mutagenesis of large DNAs such as bacterial artificial chromosomes and the *E. coli* genome. Other strain backgrounds can be used if the recombination genes are provided on a plasmid [79,80]. Lambda Red genes and *recD* strains have also been used for recombination of linear DNAs with the genome [81,82].

Inter-molecular recombination in *E. coli* can be used for creating hybrid genes as in Fig. 4(c) and (d) [83]. We observed recombination only in *recBC sbcA* strains, and not in wildtype, *recD*, or *recBC sbcB* strains. The efficiency of recombination was high between identical sequences or ones with two small regions of mismatch, but dropped significantly for sequences sharing 82% nucleotide identity (human and mouse p53 DNA binding domains). On the other hand, correct hybrids were still obtained between sequences sharing only 56% identity (human p53 and p73). We generated a series of hybrids between mouse and chicken p53 that allowed us to map the epitopes of the two antibodies that specifically recognize the wildtype conformation of p53 and not cancer-derived p53 mutants. Because

the epitopes require a properly folded p53 domain, they cannot be analyzed using deletion mutants or peptide-scanning. No special plasmid constructs are needed, and both two-part and three-part hybrids can be directly created. In addition, recombination can easily be iterated to make hybrids of hybrids for more precise analysis.

5.4. Phage delivery of genes for *in vivo* recombination in *E. coli*

Though number of recombinants is not usually a limiting factor when using recombination for analysis, the need for DNA transformation imposes a practical limit that is significant for directed evolution experiments – the more variants sampled, the better the chance of finding a variant with the desired characteristics. Numerical limitations can be partly overcome by using a more efficient way of getting DNA into cells, namely by phage infection. Lambda phage have long been used to study recombination, and can be used for creating hybrid genes [84,85]. Filamentous phage (Ff) infection is particularly suitable because any plasmid with an Ff origin (a “phagemid”) can be packaged as infectious particles and because Ff phage display is a powerful selection method. Griffiths et al. created a very large antibody library by combining heavy- and light-chain sub-libraries using Ff infection and site-specific recombination [86]; recently a simpler format has been developed [87]. Site-specific recombination allows crossover at a fixed point, whereas homologous recombination allows crossover at many points and so can generate a wider range of diversity.

We have created a library by homologous recombination *in vivo* (PLW and G. Winter; manuscript in preparation). In order to provide the conditions needed for efficient inter-molecular recombination, we have made a *recBC sbcA* strain that expresses the *EcoRI* restriction system, allowing the creation of double-strand breaks *in vivo*. We have also designed two phagemid vectors for cloning antibody single-chain Fv (scFv) genes (Fig. 5). These are packaged as viral particles and co-infected into the recombination strain. Upon entry they are cut by *EcoRI*, releasing fragments that can recombine to generate a phage display vector. One crossover occurs within the antibody genes, the other within ampicillin resistance (Amp^R) gene fragments, resulting in a hybrid antibody gene and a functional Amp^R gene. Sequences of progeny from an *in vivo* recombination between two clones included a variety of hybrids, but

Fig. 5. *In vivo* double-strand break recombination: Hybrid antibodies. Vectors for creating recombinant antibodies *in vivo* carry elements for phage display (Ab-1, Ab-2, pIII) as well as overlapping fragments of an ampicillin-resistance gene (5'-Amp, 3'-Amp). DNAs are packaged as filamentous phage particles and efficiently delivered into bacteria by infection. Double-strand breaks are generated *in vivo* by action of a restriction endonuclease, EcoRI, expressed in the bacteria (irrelevant fragments that are also generated are not shown.) Recombination of the homologous ends gives a hybrid antibody (Ab-3) and a complete resistance gene (Amp), and re-creates a plasmid circle.

crossovers were most common near the double-strand ends.

Since chicken antibodies are natural products of homologous recombination, we made two small sub-libraries of chicken scFv ($<10^7$ clones) and then recombined them *in vivo* by co-infection into 1 liter of bacteria. This resulted in 8×10^{10} ampicillin-resistant recombinants, which were directly used to produce antibody-expressing phage. This library was selected for antibodies to several antigens including groEL, p53, Werner's syndrome protein, fluorescein, and Cy5. The antibodies work well in ELISAs and in western blots. Nucleotide polymorphisms allowed us to definitively

identify several of the antibody genes as hybrid products of recombination. This completely *in vivo* recombination system should be generally useful for directed evolution of antibodies and other genes for which selection methods are available.

6. Conclusion

There are several ways to make hybrid genes, and the best one will vary depending on the specific application. All methods will perform less well as the sequences to recombine become more divergent. Natural gene families often have alternating conserved and variable regions, so despite low overall similarity they may recombine readily in the conserved regions. Intra-molecular recombination may be more effective than inter-molecular recombination for generating hybrids between highly diverged homologs. But certainly the lower reaches of similarity identifiable by sequence comparison (20–25% amino acid identity) or three-dimensional structure comparison will not be amenable to homologous recombination methods; perhaps nonhomologous methods may be useful for these cases [88].

In vivo recombination can be done in either yeast or *E. coli*. Recombination in yeast is highly efficient, but is most practical when the hybrid genes will also be expressed in yeast. Recombination in mutant strains of *E. coli* seems to be as powerful as recombination in yeast and the hybrid genes created can either be analyzed directly in *E. coli* for which many techniques are available, or can be transfected into the species of interest. Both intra- and inter-molecular recombination can be used, with the latter being more flexible.

For analysis of gene function, both *in vitro* and *in vivo* recombination methods may be useful. *In vivo* methods give simple hybrids and do not introduce mutations; this means that an ordered strategy can be used, and fewer clones need to be tested and sequenced. If a functional selection or rapid screen exists, then *in vitro* recombination becomes more practical. And often analysis of function may shade into creating new function, that is, directed evolution.

For directed evolution, *in vitro* recombination methods like DNA shuffling have the major advantage that they can create hybrids with many crossovers. It is important to be able to bring together several beneficial mutations that may be anywhere in the sequence; if one is limited to one or two crossovers at a time this requires more than one round of recombination. *In vitro*

recombination should be the method of choice for totally *in vitro* selection methods. On the other hand, for selections or screens requiring an *in vivo* step such as phage display, two-hybrid, or gene complementation, there may be practical advantages for *in vivo* recombination methods; for example, we have used *in vivo* recombination to create a large antibody phage display library.

Recombinational cloning and mutagenesis are proving to be important tools for genome-scale studies of function ("functional genomics"). The explosion of new homologs uncovered by genome sequencing projects should also greatly benefit recombinational approaches to gene analysis and directed evolution. They may even tell us something about natural evolution by re-creating possible ancestors of current-day molecules [89].

Acknowledgements

This work was supported by a Hitchings-Elion fellowship from the Burroughs-Wellcome Fund and by the Medical Research Council. Thanks to Greg Winter for helpful discussion.

References

- [1] J.R. Lupski, Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits, *Trends Genet* **14** (1998), 417–422.
- [2] A.B. Metzberg, G. Wurzer, T.H. Huisman and O. Smithies, Homology requirements for unequal crossing over in humans, *Genetics* **128** (1991), 143–161.
- [3] J. Neitz, M. Neitz and G.H. Jacobs, Analysis of fusion gene and encoded photopigment of colour-blind humans, *Nature* **342** (1989), 679–682.
- [4] C.A. Reynaud, B. Bertocci, A. Dahan and J.C. Weill, Formation of the chicken B-cell repertoire: ontogenesis, regulation of Ig gene rearrangement, and diversification by gene conversion, *Adv Immunol* **57** (1994), 353–378.
- [5] J.R. Zhang, J.M. Hardham, A.G. Barbour and S.J. Norris, Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes, *Cell* **89** (1997), 275–285.
- [6] J.R. Zhang and S.J. Norris, Genetic variation of the *Borrelia burgdorferi* gene *vlsE* involves cassette-specific, segmental gene conversion, *Infect Immun* **66** (1998), 3698–3704.
- [7] K.W. Deitsch, E.R. Moxon and T.E. Wellems, Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections, *Microbiol Mol Biol Rev* **61** (1997), 281–293.
- [8] R.K. Saiki, D.H. Gelfand, S. Stoffel, S.J. Scharf and R. Higuchi et al., Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, *Science* **239** (1988), 487–491.
- [9] A. Meyerhans, J.P. Vartanian and S. Wain-Hobson, DNA recombination during PCR, *Nucleic Acids Res* **18** (1990), 1687–1691.
- [10] S. Pääbo, D.M. Irwin and A.C. Wilson, DNA damage promotes jumping between templates during enzymatic amplification, *J Biol Chem* **265** (1990), 4718–4721.
- [11] W.P. Stemmer, Rapid evolution of a protein *in vitro* by DNA shuffling, *Nature* **370** (1994), 389–391.
- [12] W.P. Stemmer, DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution, *Proc Natl Acad Sci USA* **91** (1994), 10747–10751.
- [13] G.P. Smith, Applied evolution. The progeny of sexual PCR, *Nature* **370** (1994), 324–325.
- [14] A. Cramer, S.A. Raillard, E. Bermudez and W.P. Stemmer, DNA shuffling of a family of genes from diverse species accelerates directed evolution, *Nature* **391** (1998), 288–291.
- [15] Z. Shao, H. Zhao, L. Giver and F.H. Arnold, Random-priming *in vitro* recombination: an effective tool for directed evolution, *Nucleic Acids Res* **26** (1998), 681–683.
- [16] M.S.B. Judo, A.B. Wedel and C. Wilson, Stimulation and suppression of PCR-mediated recombination, *Nucleic Acids Res* **26** (1998), 1819–1825.
- [17] H. Zhao, L. Giver, Z. Shao, J.A. Affholter and F.H. Arnold, Molecular evolution by staggered extension process (StEP) *in vitro* recombination, *Nat Biotechnol* **16** (1998), 258–261.
- [18] F.H. Arnold and A.A. Volkov, Directed evolution of biocatalysts, *Curr Opin Chem Biol* **3** (1999), 54–59.
- [19] J. Minshull and W.P. Stemmer, Protein evolution by molecular breeding, *Curr Opin Chem Biol* **3** (1999), 284–290.
- [20] T. Kumamaru, H. Suenaga, M. Mitsuoka, T. Watanabe and K. Furukawa, Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase, *Nat Biotechnol* **16** (1998), 663–666.
- [21] F.C. Christians, L. Scapozza, A. Cramer, G. Folkers and W.P. Stemmer, Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling, *Nat Biotechnol* **17** (1999), 259–264.
- [22] C.C. Chang, T.T. Chen, B.W. Cox, G.N. Dawes and W.P. Stemmer et al., Evolution of a cytokine using DNA family shuffling, *Nat Biotechnol* **17** (1999), 793–797.
- [23] J.E. Ness, M. Welch, L. Giver, M. Bueno and J.R. Cherry et al., DNA shuffling of subgenomic sequences of subtilisin, *Nat Biotechnol* **17** (1999), 893–896.
- [24] C. Schaffitzel, J. Hanes, L. Jermutus and A. Pluckthun, Ribosome display: an *in vitro* method for selection and evolution of antibodies from libraries, *J Immunol Methods* **231** (1999), 119–135.
- [25] R.W. Roberts, Totally *in vitro* protein selection using mRNA-protein fusions and ribosome display, *Curr Opin Chem Biol* **3** (1999), 268–273.
- [26] D.S. Tawfik and A.D. Griffiths, Man-made cell-like compartments for molecular evolution, *Nat Biotechnol* **16** (1998), 652–656.
- [27] R. Wagner, Jr. and M. Meselson, Repair tracts in mismatched DNA heteroduplexes, *Proc Natl Acad Sci USA* **73** (1976), 4135–4139.
- [28] J.P. Abastado, B. Cami, T.H. Dinh, J. Igolen and P. Kourilsky, Processing of complex heteroduplexes in *Escherichia coli* and Cos-1 monkey cells, *Proc Natl Acad Sci USA* **81** (1984), 5792–5796.
- [29] B. Cami, P. Chambon and P. Kourilsky, Correction of complex heteroduplexes made of mouse H-2 gene sequences in *Escherichia coli* K-12, *Proc Natl Acad Sci USA* **81** (1984), 503–507.

- [30] W.L. Sung, D.M. Zahab, F.L. Yao, R. Wu and S.A. Narang, Simultaneous synthesis of human-, mouse- and chimeric epidermal growth factor genes via 'hybrid gene synthesis' approach, *Nucleic Acids Res* **14** (1986), 6159–6168.
- [31] R. Fishel and R. Kolodner, Gene conversion in *Escherichia coli*: the recF pathway for resolution of heteroduplex DNA, *J Bacteriol* **171** (1989), 3046–3052.
- [32] A.A. Volkov, Z. Shao and F.H. Arnold, Recombination and chimera-genesis by *in vitro* heteroduplex formation and *in vivo* repair, *Nucleic Acids Res* **27** (1999), e18.
- [33] S.C. Kowalczykowski, Initiation of genetic recombination and recombination-dependent replication, *Trends Biochem Sci* **25** (2000), 156–165.
- [34] S.C. Kowalczykowski, D.A. Dixon, A.K. Eggleston, S.D. Lauder and W.M. Rehrauer, Biochemistry of homologous recombination in *Escherichia coli*, *Microbiol Rev* **58** (1994), 401–465.
- [35] A. Kuzminov, Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda, *Microbiol Mol Biol Rev* **63** (1999), 751–813.
- [36] D.P. Biek and S.N. Cohen, Identification and characterization of *recD*, a gene affecting plasmid maintenance and recombination in *Escherichia coli*, *J Bacteriol* **167** (1986), 594–603.
- [37] S.T. Lovett, C. Luisi-DeLuca and R.D. Kolodner, The genetic dependence of recombination in *recD* mutants of *Escherichia coli*, *Genetics* **120** (1988), 37–45.
- [38] C. Luisi-DeLuca, S.T. Lovett and R.D. Kolodner, Genetic and physical analysis of plasmid recombination in *recB recC sbcB* and *recB recC sbcA* *Escherichia coli* K-12 mutants, *Genetics* **122** (1989), 269–278.
- [39] R. Kolodner, S.D. Hall and C. Luisi-DeLuca, Homologous pairing proteins encoded by the *Escherichia coli* *recE* and *recT* genes, *Mol Microbiol* **11** (1994), 23–30.
- [40] F. Paques and J.E. Haber, Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*, *Microbiol Mol Biol Rev* **63** (1999), 349–404.
- [41] H. Weber and C. Weissmann, Formation of genes coding for hybrid proteins by recombination between related, cloned genes in *E. coli*, *Nucleic Acids Res* **11** (1983), 5661–5669.
- [42] G.F. Crouse, L. New and L.A. Stivaletta, Gene engineering by selectable intraplasmid recombination: construction of novel dihydrofolate reductase minigenes, *Gene* **84** (1989), 165–172.
- [43] G.L. Gray, S.E. Mainzer, M.W. Rey, M.H. Lamsa and K.L. Kindle et al., Structural genes encoding the thermophilic alpha-amylases of *Bacillus stearothermophilus* and *Bacillus licheniformis*, *J Bacteriol* **166** (1986), 635–643.
- [44] T. Nogami, T. Mizuno and S. Mizushima, Construction of a series of *ompF-ompC* chimeric genes by *in vivo* homologous recombination in *Escherichia coli* and characterization of the translational products, *J Bacteriol* **164** (1985), 797–801.
- [45] J. Tommassen, P. van der Ley, M. van Zeijl and M. Agterberg, Localization of functional domains in *E. coli* K-12 outer membrane porins, *EMBO J* **4** (1985), 1583–1587.
- [46] J.P. Abastado, S. Darche, F. Godeau, B. Cami and P. Kourilsky, Intramolecular recombination between partially homologous sequences in *Escherichia coli* and *Xenopus laevis* oocytes, *Proc Natl Acad Sci USA* **84** (1987), 6496–6500.
- [47] M.J. Doherty, P.T. Morrison and R. Kolodner, Genetic recombination of bacterial plasmid DNA. Physical and genetic analysis of the products of plasmid recombination in *Escherichia coli*, *J Mol Biol* **167** (1983), 539–560.
- [48] E.C. Conley and J.R. Saunders, Recombination-dependent recircularization of linearized pBR322 plasmid DNA following transformation of *Escherichia coli*, *Mol Gen Genet* **194** (1984), 211–218.
- [49] T. Caramori, A.M. Albertini and A. Galizzi, *In vivo* generation of hybrids between two *Bacillus thuringiensis* insect-toxin-encoding genes, *Gene* **98** (1991), 37–44.
- [50] T. Ogawa, A. Shinohara, H. Ogawa and J. Tomizawa, Functional structures of the RecA protein found by chimera analysis, *J Mol Biol* **226** (1992), 651–660.
- [51] M.D. van Kampen, N. Dekker, M.R. Egmond and H.M. Verheij, Substrate specificity of *Staphylococcus hyicus* lipase and *Staphylococcus aureus* lipase as studied by *in vivo* chimera-genesis, *Biochemistry* **37** (1998), 3459–3466.
- [52] L.R. Levin and R.R. Reed, Identification of functional domains of adenyllyl cyclase using *in vivo* chimeras, *J Biol Chem* **270** (1995), 7573–7579.
- [53] T. Jelinek and F.L. Graham, Recombinant human adenoviruses containing hybrid adenovirus type 5 (Ad5)/Ad12 E1A genes: characterization of hybrid E1A proteins and analysis of transforming activity and host range, *J Virol* **66** (1992), 4117–4125.
- [54] K.R. Moore and R.D. Blakely, Restriction site-independent formation of chimeras from homologous neurotransmitter-transporter cDNAs, *Biotechniques* **17** (1994), 130–135, 137.
- [55] J.Y. Kim and P.N. Devreotes, Random chimera-genesis of G-protein-coupled receptors. Mapping the affinity of the cAMP chemoattractant receptors in *Dictyostelium*, *J Biol Chem* **269** (1994), 28724–28731.
- [56] Z. Wang, B. Lee, J.L. Murray, F. Bonneau and Y. Sun et al., CCR5 HIV-1 coreceptor activity. Role of cooperativity between residues in N-terminal extracellular and intracellular domains, *J Biol Chem* **274** (1999), 28413–28419.
- [57] I.A. Pikuleva, I. Bjorkhem and M.R. Waterman, Studies of distant members of the P450 superfamily (P450scc and P450c27) by random chimera-genesis, *Arch Biochem Biophys* **334** (1996), 183–192.
- [58] D. Pompon and A. Nicolas, Protein engineering by cDNA recombination in yeasts: shuffling of mammalian cytochrome P-450 functions, *Gene* **83** (1989), 15–24.
- [59] C. Mezard and A. Nicolas, Homologous, homeologous, and illegitimate repair of double-strand breaks during transformation of a wild-type strain and a *rad52* mutant strain of *Saccharomyces cerevisiae*, *Mol Cell Biol* **14** (1994), 1278–1292.
- [60] T.L. Orr-Weaver, J.W. Szostak and R.J. Rothstein, Yeast transformation: a model system for the study of recombination, *Proc Natl Acad Sci USA* **78** (1981), 6354–6358.
- [61] S. Kunes, D. Botstein and M.S. Fox, Transformation of yeast with linearized plasmid DNA. Formation of inverted dimers and recombinant plasmid products, *J Mol Biol* **184** (1985), 375–387.
- [62] H. Ma, S. Kunes, P.J. Schatz and D. Botstein, Plasmid construction by homologous recombination in yeast, *Gene* **58** (1987), 201–216.
- [63] M.R. Martzen, S.M. McCraith, S.L. Spinelli, F.M. Torres and S. Fields et al., A biochemical genomics approach for identifying genes by the activity of their products, *Science* **286** (1999), 1153–1155.
- [64] P. Uetz, L. Giot, G. Cagney, T.A. Mansfield and R.S. Judson et al., A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*, *Nature* **403** (2000), 623–627.
- [65] S. Kunes, H. Ma, K. Overbye, M.S. Fox and D. Botstein, Fine structure recombinational analysis of cloned genes using yeast transformation, *Genetics* **115** (1987), 73–81.
- [66] D.L. Marykwas and S.E. Passmore, Mapping by multifragment cloning *in vivo*, *Proc Natl Acad Sci USA* **92** (1995), 11701–11705.

- [67] P. Supply, A. de Kerchove d'Exaerde, T. Roganti, A. Goffeau and F. Foury, In-frame recombination between the yeast H(+)-ATPase isogenes PMA1 and PMA2: insights into the mechanism of recombination initiated by a double-strand break, *Mol Cell Biol* **15** (1995), 5389–5395.
- [68] C. Mezard, D. Pompon and A. Nicolas, Recombination between similar but not identical DNA sequences during yeast transformation occurs within short stretches of identity, *Cell* **70** (1992), 659–670.
- [69] V. Goguel, A. Delahodde and C. Jacq, Connections between RNA splicing and DNA intron mobility in yeast mitochondria: RNA maturase and DNA endonuclease switching experiments, *Mol Cell Biol* **12** (1992), 696–705.
- [70] J.R. Cherry, M.H. Lamsa, P. Schneider, J. Vind and A. Svendsen et al., Directed evolution of a fungal peroxidase, *Nat Biotechnol* **17** (1999), 379–384.
- [71] A. Laban and A. Cohen, Interplasmidic and intraplasmidic recombination in *Escherichia coli* K-12, *Mol Gen Genet* **184** (1981), 200–207.
- [72] A.A. James, P.T. Morrison and R. Kolodner, Genetic recombination of bacterial plasmid DNA. Analysis of the effect of recombination-deficient mutations on plasmid recombination, *J Mol Biol* **160** (1982), 411–430.
- [73] W.P. Schneider, B.P. Nichols and C. Yanofsky, Procedure for production of hybrid genes and proteins and its use in assessing significance of amino acid differences in homologous tryptophan synthetase alpha polypeptides, *Proc Natl Acad Sci USA* **78** (1981), 2169–2173.
- [74] D.H. Jones and B.H. Howard, A rapid method for recombination and site-specific mutagenesis by placing homologous ends on DNA using polymerase chain reaction, *Biotechniques* **10** (1991), 62–66.
- [75] P. Bubeck, M. Winkler and W. Bautsch, Rapid cloning by homologous recombination in vivo, *Nucleic Acids Res* **21** (1993), 3601–3602.
- [76] J.D. Oliner, K.W. Kinzler and B. Vogelstein, In vivo cloning of PCR products in *E. coli*, *Nucleic Acids Res* **21** (1993), 5192–5197.
- [77] E. Degryse, B. Dumas, M. Dietrich, L. Laruelle and T. Achstetter, In vivo cloning by homologous recombination in yeast using a two-plasmid-based system, *Yeast* **11** (1995), 629–640.
- [78] E. Degryse, In vivo intermolecular recombination in *Escherichia coli*: application to plasmid constructions, *Gene* **170** (1996), 45–50.
- [79] Y. Zhang, F. Buchholz, J.P. Muirers and A.F. Stewart, A new logic for DNA engineering using recombination in *Escherichia coli*, *Nat Genet* **20** (1998), 123–128.
- [80] J.P. Muirers, Y. Zhang, G. Testa and A.F. Stewart, Rapid modification of bacterial artificial chromosomes by ET-recombination, *Nucleic Acids Res* **27** (1999), 1555–1557.
- [81] C.B. Russell, D.S. Thaler and F.W. Dahlquist, Chromosomal transformation of *Escherichia coli* recD strains with linearized plasmids, *J Bacteriol* **171** (1989), 2609–2613.
- [82] K.C. Murphy, Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*, *J Bacteriol* **180** (1998), 2063–2071.
- [83] P.L. Wang, F. Sait and G. Winter, The “wildtype” conformation of p53: epitope mapping using hybrid proteins, *Oncogene* **20** (2001) (in press).
- [84] G.F. Crouse, R.N. McEwan and M.L. Pearson, Expression and amplification of engineered mouse dihydrofolate reductase minigenes, *Mol Cell Biol* **3** (1983), 257–266.
- [85] D.K. Struck, D.S. Durica and R. Young, Gene fusion in vivo using transductional cointegrates, *Gene* **47** (1986), 221–230.
- [86] A.D. Griffiths, S.C. Williams, O. Hartley, I.M. Tomlinson and P. Waterhouse et al., Isolation of high affinity human antibodies directly from large synthetic repertoires, *EMBO J* **13** (1994), 3245–3260.
- [87] D. Sblattero and A. Bradbury, Exploiting recombination in single bacteria to make large phage antibody libraries, *Nat Biotechnol* **18** (2000), 75–80.
- [88] M. Ostermeier, J.H. Shim and S.J. Benkovic, A combinatorial approach to hybrid enzymes independent of DNA homology, *Nat Biotechnol* **17** (1999), 1205–1209.
- [89] B.S.W. Chang and M.J. Donoghue, Recreating ancestral proteins, *Trends in Ecology & Evolution* **15** (2000), 109–114.