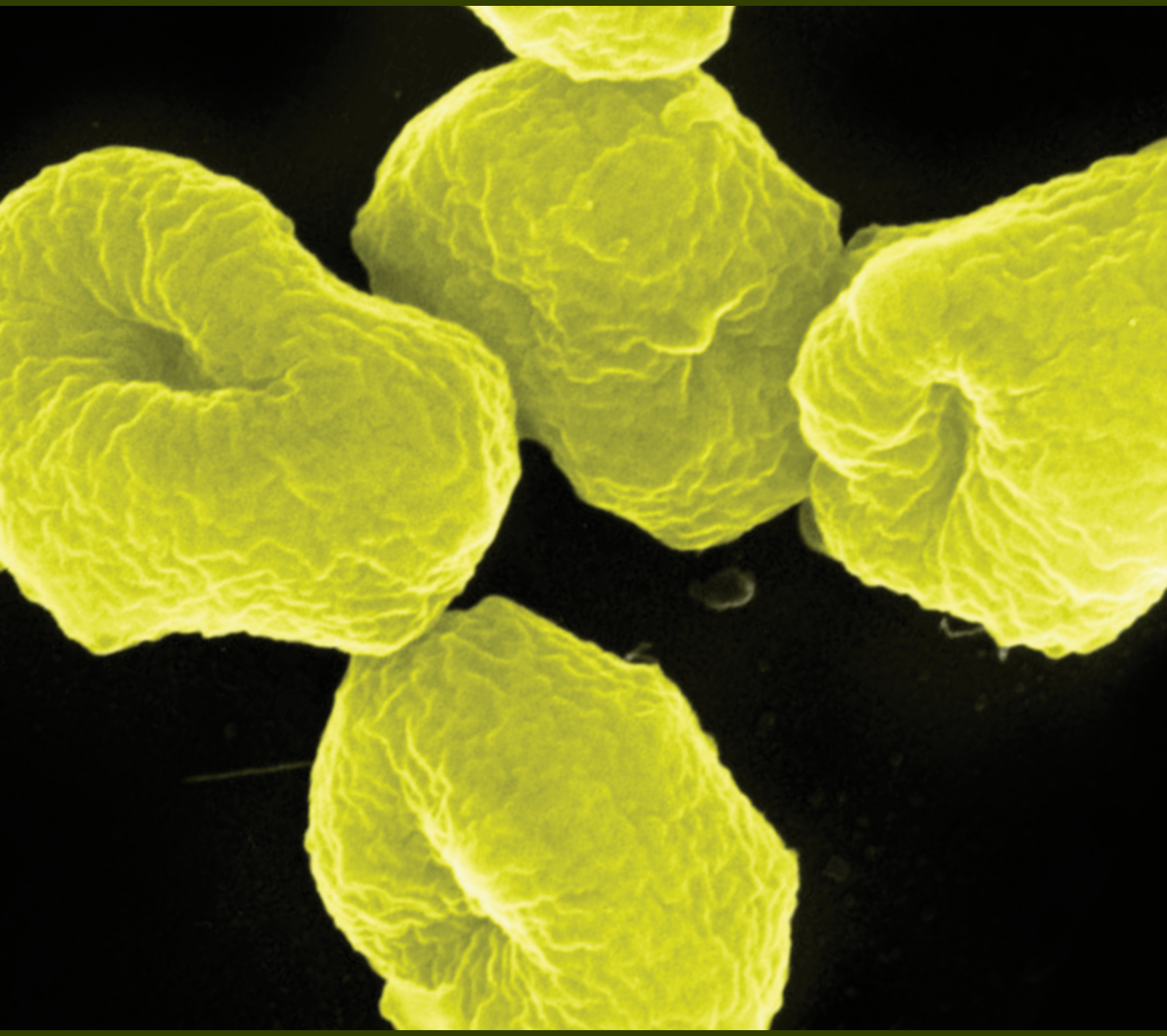
Archaea

Biotechnological Uses of Archaeal Proteins

Guest Editors: Frédéric Pecorari, Vickery L. Arcus, and Juergen Wiegel



Biotechnological Uses of Archaeal Proteins

Archaea

Biotechnological Uses of Archaeal Proteins

Guest Editors: Frédéric Pecorari, Vickery L. Arcus, and Juergen Wiegel

Editorial Board

Contents

Hindawi Publishing Corporation
Archaea
Volume 2015, Article ID 809758, 2 pages
http://dx.doi.org/10.1155/2015/809758

Editorial

Biotechnological Uses of Archaeal Proteins

Frédéric Pecorari,[1] Vickery L. Arcus,[2] and Juergen Wiegel[3]

[1]*Cancer Research Center of Nantes-Angers, INSERM UMR 892, CNRS UMR 6299, University of Nantes, 44007 Nantes, France*
[2]*School of Science, University of Waikato, Hamilton, New Zealand*
[3]*Departments of Microbiology and of Biochemistry & Molecular Biology, University of Georgia, Athens, GA 30605, USA*

Correspondence should be addressed to Frédéric Pecorari; frederic.pecorari@univ-nantes.fr

Received 21 September 2015; Accepted 21 September 2015

Many industrial/biotechnological processes take place under extreme conditions of temperature, pH, salinity, or pressure which are not suitable for activities of proteins from model eukaryotic or common neutrophilic, mesophilic, and prokaryotic microorganisms. In contrast, Archaea offer a large panel of extremophile organisms that express proteins that are able to remain properly folded and functional under the harshest biophysical conditions.

The study of this group of organisms has uncovered archaeal enzymes and proteins with unusual properties compared to their traditional homologues. In addition, with their ease of production and better-behaved samples for X-ray crystallography, for example, archaeal proteins are often more convenient for structural biology studies than their eukaryotic equivalents. The knowledge thus gained can open routes to commercial biotechnological applications. These last years, with the emergence of next generation sequencing techniques to decode whole genomes and metagenomes and the pressure to develop "greener" industrial processes, the rate of new archaeal proteins reported has significantly increased, thereby widening again their potential of applications. In this special issue of Archaea, we present selected papers dealing with the uses of archaeal proteins as tools for various fields of biotechnologies and research.

DNA and RNA ligases are essential enzymes in living cells and have applications in molecular biology. A review by M. Tanabe et al. discusses the uses of DNA ligases and recent progress in deciphering their catalytic mechanisms *via* structural studies, and they describe how protein engineering can improve ligation efficiency of an archaeal DNA ligase over a broad temperature range. In another paper on ligases, C. R. Chambers and W. M. Patrick present the current state of knowledge on archaeal nucleic acid ligases including RNA ligases, highlighting their remarkable properties relevant to biotechnologists, and they discuss the modifications of the activities of archaeal RNA ligases by directed mutagenesis to develop more efficient molecular biology protocols.

J. A. Littlechild reviews research regarding the discovery and potential applications of a range of thermophilic archaeal proteins, illustrating the power of archaeal enzymes for various industrial biocatalysis. Then, an article by V. M. Gumerov et al. describes the characterization of a novel thermostable and multifunctional β-glycosidase from *Acidilobus saccharovorans* that displays a high tolerance to glucose, a desired property for such enzymes used to process lignocellulose biomass. C.-H. Wu et al. present a review summarizing the strategies used in engineering and characterizing three different forms of soluble hydrogenase I from the hyperthermophile *Pyrococcus furiosus*, an enzyme which has been used *in vitro* for hydrogen production.

Archaea are not only interesting for catalysis applications. J. C. Charlesworth and B. P. Burns give a comprehensive overview of archaeal low-molecular weight compounds including peptides with antimicrobial properties which can be exploited for biotechnological purposes. Finally, in their review, J. M. Miller and E. J. Enemark exemplify with MCM helicases how crystallography of archaeal homologues can be helpful to decipher structure/function relationships of their eukaryotic versions which are more difficult to crystallize.

We hope this special issue will provide the reader with an up-to-date overview of some of the diverse applications of archaeal proteins and, perhaps, prompt further research in this largely untapped field with rich prospects for the future of many biotechnological applications.

Frédéric Pecorari
Vickery L. Arcus
Juergen Wiegel

Hindawi Publishing Corporation
Archaea
Volume 2015, Article ID 912582, 8 pages
http://dx.doi.org/10.1155/2015/912582

Review Article

Production and Application of a Soluble Hydrogenase from *Pyrococcus furiosus*

Chang-Hao Wu, Patrick M. McTernan, Mary E. Walter, and Michael W. W. Adams

Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA

Correspondence should be addressed to Michael W. W. Adams; adams@bmb.uga.edu

Received 6 May 2015; Accepted 22 July 2015

Academic Editor: Vickery L. Arcus

Hydrogen gas is a potential renewable alternative energy carrier that could be used in the future to help supplement humanity's growing energy needs. Unfortunately, current industrial methods for hydrogen production are expensive or environmentally unfriendly. In recent years research has focused on biological mechanisms for hydrogen production and specifically on hydrogenases, the enzyme responsible for catalyzing the reduction of protons to generate hydrogen. In particular, a better understanding of this enzyme might allow us to generate hydrogen that does not use expensive metals, such as platinum, as catalysts. The soluble hydrogenase I (SHI) from the hyperthermophile *Pyrococcus furiosus*, a member of the euryarchaeota, has been studied extensively and used in various biotechnological applications. This review summarizes the strategies used in engineering and characterizing three different forms of SHI and the properties of the recombinant enzymes. SHI has also been used in *in vitro* systems for hydrogen production and NADPH generation and these systems are also discussed.

1. Introduction

Hydrogen is a potential renewable and carbon neutral energy carrier. It has three times the energy content per unit mass of fossil fuels [1]. The concept of replacing current gasoline-based vehicles with hydrogen fuel cell vehicles (HFCVs) has gained a lot of attention recently [2]. A major advantage of HFCVs is that water is the only waste product, and hence they eliminate the harmful exhaust of current vehicles, thereby benefiting human health and the climate [2, 3]. With the introduction of commercially available HFCVs in many counties in 2015, the demand for hydrogen is anticipated to dramatically increase in the near future [3]. Unfortunately, current methods of producing hydrogen rely on fossil fuels and are expensive. They include steam reforming of natural gas, which produces greenhouse gases, and electrolysis to split water uses the expensive metal platinum as a catalyst [4]. New and renewable methods are obviously needed for the generation of hydrogen and biological-based systems have a great deal of potential.

The enzyme hydrogenase catalyzes the simplest chemical reaction in nature, the reversible interconversion of protons, electrons, and hydrogen gas: $2H^+ + 2e^- \leftrightarrow H_2$. Such enzymes are widespread in bacteria and Archaea and are even found in some Eukarya [5]. Hydrogenases enable organisms to remove excess reducing power generated during metabolism by evolving hydrogen, or they can oxidize hydrogen to generate reducing power for growth [6]. Hydrogenases can be classified into three types based on the metal content of their catalytic sites, and they are referred to as [NiFe] hydrogenases, [FeFe] hydrogenases, and mononuclear Fe hydrogenases [7]. The [NiFe] hydrogenases are the most ubiquitous and have been extensively studied [5]. They are further classified into four different types (groups 1–4) based on the peptide sequence used to bind the [NiFe]-containing active site [7]. Group 1 [NiFe] hydrogenases are the best studied among the four groups [5]. The assembly of the [NiFe] catalytic site of these hydrogenases requires eight maturation proteins, based on the mechanism elucidated for *Escherichia coli* hydrogenase 3 [8]. The [NiFe] hydrogenases are also reversibly inactivated in the presence of oxygen [9].

Herein we focus on the [NiFe] hydrogenases of *Pyrococcus furiosus*, a strictly anaerobic archaeon that grows optimally at 100°C. This organism utilizes carbohydrates as a carbon source for growth and generates acetate, carbon dioxide, and hydrogen gas as end products. *P. furiosus* contains three

different types of [NiFe] hydrogenase, a membrane-bound enzyme (MBH) and two soluble hydrogenases (SHI and SHII). MBH is the hydrogenase responsible for producing H_2 during its fermentative metabolism wherein it oxidizes the reduced ferredoxin generated during the oxidation of glucose to acetate [15, 16]. In contrast, SHI and SHII utilize NADP(H) and NAD(H) as electron carriers, respectively, and while their functions have not been established, it is assumed that they can recycle some of the H_2 produced by MBH under the appropriate growth conditions. All three hydrogenases have been purified and characterized [6, 17–19]. This review focuses on the engineering, properties, and applications of SHI.

2. Expression and Purification

P. furiosus SHI is a heterotetramer encoded by a four-gene operon (PF0891–0894). A structural model of SHI has been proposed based on sequence analyses of the four subunits [10]. As shown in Figure 1, PF0894 is the subunit harboring the Fe- and Ni-containing catalytic site wherein the Fe atom has three diatomic ligands, one -CO and two -CN. PF0892 contains the flavin and a [2] cluster and is the site of interaction with NADP(H). PF0891 and PF0893 contain two and three [4] clusters, respectively, for electron transfer between the flavin and the active site. SHI was first purified and characterized using four chromatographic steps, which yielded the intact heterotetramer [6]. The yield from this purification was very low and an improvement in yield was needed in order to generate the enzyme for detailed characterization studies.

In order to improve the yield of SHI, an attempt was made to heterologously express SHI in *E. coli*, with coexpression of the genes encoding the accessory proteins that are necessary for proper assembly of the [NiFe] active site [10]. This was also the first example of heterologously expressing a functional [NiFe] hydrogenase in *E. coli*, as well as demonstrating expression of a hydrogenase in a distantly related host. Unfortunately, the yield of this heterologous expressed SHI was lower than the natively purified SHI from *P. furiosus* [10]. Although the expression of SHI in a genetically tractable host, such as *E. coli*, was a significant achievement, this system was not suitable to produce large amounts of this enzyme.

Once a genetic system was established in *P. furiosus*, the host organism could be used to both overproduce and engineer SHI [20]. The genetic system was established by removing the *pyrF* gene in a genetically tractable strain of *P. furiosus* termed COM1. *pyrF* is essential for uracil biosynthesis and allows for selection and counter selection based on uracil biosynthesis in a minimal medium. Moreover, it was demonstrated that the genes encoding SHI could be deleted from *P. furiosus* without any apparent effect on cell growth [20]. This suggested that SHI was not an essential enzyme and could be engineered in various ways without affecting the metabolism of *P. furiosus*, and this proved to be the case.

In the first attempt to overexpress the four genes encoding SHI and to affinity-tag the enzyme [11], transcription of SHI was put under the control of a strong constitutive promoter, P_{slp}, which controls expression of the gene encoding the S-layer protein. In addition, PF0891 was engineered to include a Strep-II affinity tag (Figure 2). A Strep tag was chosen instead of a polyhistidine tag as the latter might interfere with the incorporation of nickel into the catalytic site of SHI. Using this approach and a one-step affinity purification, approximately seven times more SHI (per gram of cells) was purified from the cytoplasmic fraction of *P. furiosus* compared to the original purification of SHI [11]. Interestingly, expression of the genes encoding the [NiFe]-maturation proteins was at the same level in the recombinant strain as in the parent strain even though the genes encoding SHI (under the control of P_{slp}) were increased by about 10-fold. The native levels of the maturation proteins were therefore able to synthesize almost an order of magnitude more SHI and produce the active enzyme.

Since a functional SHI is not required for growth of *P. furiosus* [20], this allowed the engineering of nonfunctional forms that did not utilize H_2 and/or NADP(H) as substrates. A dimeric version of SHI that contained only PF0893 and PF0894 was successfully produced (Figure 1). This enzyme evolved H_2 from artificial electron donors but did not oxidize NADPH, as it lacked the NADPH-oxidizing subunit [13]. Engineering dimeric SHI also involved the development of another selectable marker, arginine decarboxylase (*pdaD*), to be used for genetic manipulations in *P. furiosus*. In addition, the purified dimeric SHI had a polyhistidine (9-His) tag at the N-terminus of PF0893. The results demonstrated that this type of tag does not interfere with the assembly of nickel into the catalytic site of SHI as the enzyme retained its H_2-production activity (using an artificial dye as the electron donor) after purification using the immobilized nickel-affinity chromatography step.

Based on the success in engineering and purifying 9x-His tag dimeric SHI, the tetrameric SHI was engineered to contain a polyhistidine affinity tag to determine if this would improve the efficiency of purification compared to the Strep-tag II tetramer [12]. The same strategy to overexpress SHI (tagging at the N-terminus of PF0891) was used, except that the Strep-tag II was replaced by a 9x-His tag (Figure 2). This resulted in an 8-fold improvement in the yield compared to the Strep-tag II and a 50-fold higher yield of SHI compared to the original native purification. A comparison of the yields for the different purification procedures is shown in Table 1 and the strains constructed for SHI expression are shown in Table 2.

During the affinity purification of the His-tagged tetrameric form of SHI, a trimeric form was observed eluting from the affinity column that lacked the large [NiFe]-containing subunit (PF0894, see Figure 1). This Ni-free trimeric SHI represented approximately 2% of total SHI [12]. This discovery supports the proposed maturation mechanism of [NiFe] hydrogenases in which the three subunits of the *P. furiosus* enzyme (PF0891–PF0893) form a trimeric complex before the catalytic subunit (PF0894) is assembled [8]. This complex then binds to the catalytic subunit to generate the active enzyme. Hence, in an overexpressed strain, the processing machinery may not be able to keep up with the production of the four protein subunits, such

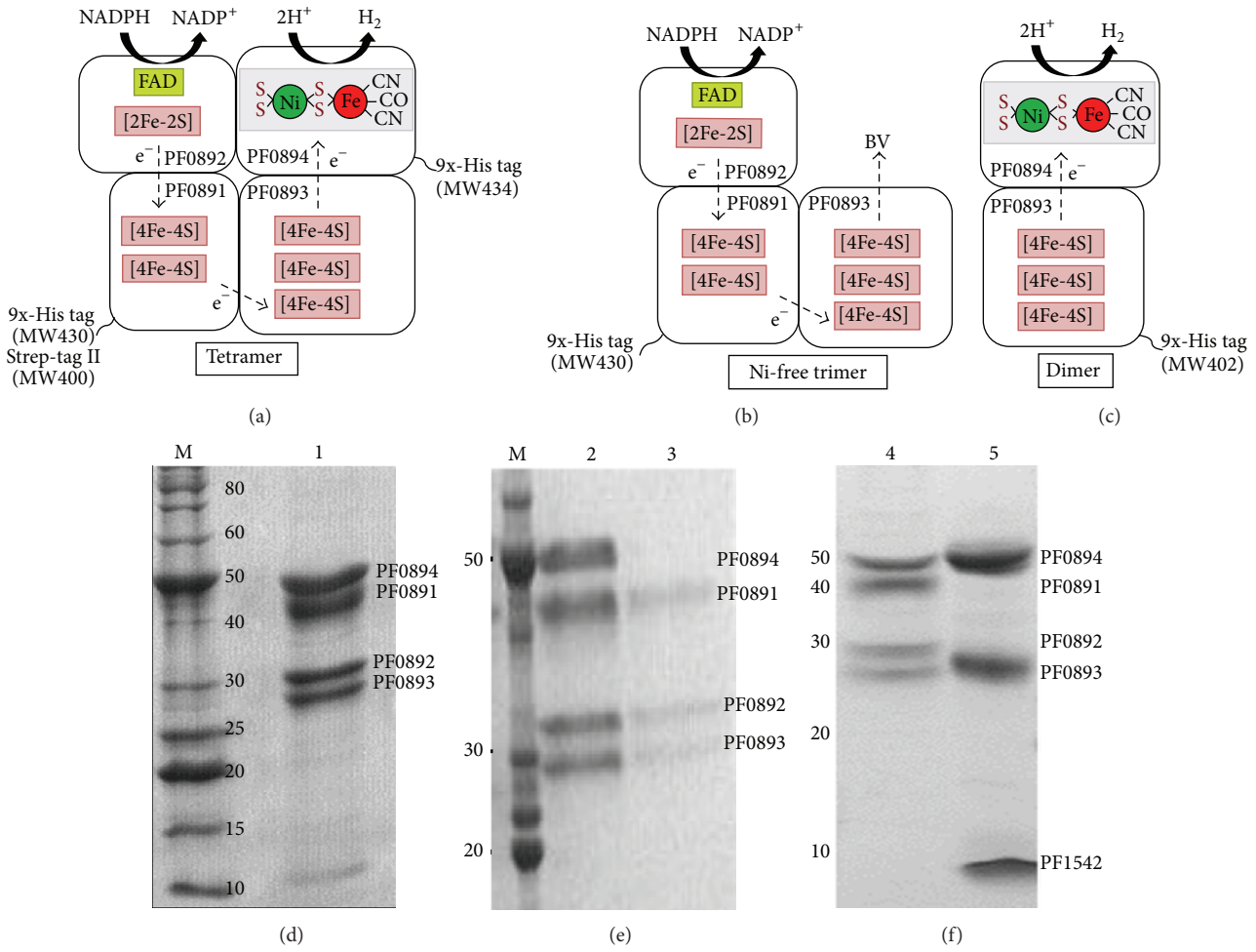


Figure 1: Models of tetrameric (a), Ni-free trimeric (b), and dimeric (c) forms of SHI. These are modified from [10] and are based on the cofactor and iron-sulfur cluster contents in sequence analysis. The proposed electron flow from NADPH oxidation to hydrogen evolution is also shown. Four different strains of *P. furiosus* were constructed to obtain the various forms of SHI. They are designated as MW400, MW430, MW434, and MW402 and their properties are listed in Table 2. These were used to prepare PF0891 Strep-tag II SHI [11], PF0891 9x-His tag SHI [12], PF0894 9x-His tag SHI [12], and PF0893 9x-His tag dimeric SHI [13], respectively. SDS PAGE gels show the purity of the different forms: (d) lane 1, Strep-tag II tetrameric SHI; (e) lane 2, 9x-His tag tetrameric SHI; lane 3, 9x-His tag Ni-free trimeric SHI; (f) lane 4, native SHI; lane 5, 9x-His tag dimeric SHI (PF1542 is an unrelated protein that is a persistent contaminant that copurified with dimeric SHI). The SDS PAGE gel data were modified from [11–13].

Table 1: Yields of SHI from different expression systems.

Protein	Expression host	Affinity tag	Purification steps	Protein yield (mg)[1]	Reference
Native SHI	*P. furiosus*	—	4	2.5	[6]
Recombinant SHI	*E. coli*	—	3	0.16	[10]
Strep-tag II SHI	*P. furiosus*	Strep-tag II	1	17	[11]
9x-His Dimeric SHI	*P. furiosus*	9x-His tag	2	16	[13]
9x-His SHI	*P. furiosus*	9x-His tag	1	135	[12]

[1]Protein yield from 100 g of cells (wet weight).

that there is a very slight excess of the Ni-free trimeric form. These results are also consistent with the notion that the catalytic subunit alone cannot be expressed and isolated in an active form and needs the other hydrogenase subunits to be processed. For example, when the catalytic subunit of a cytoplasmic hydrogenase of *Thermococcus kodakarensis* was expressed in *E. coli*, the purified subunit was inactive and contained Fe, Zn, and Ca atoms but not Ni, indicating that it was not properly assembled [21].

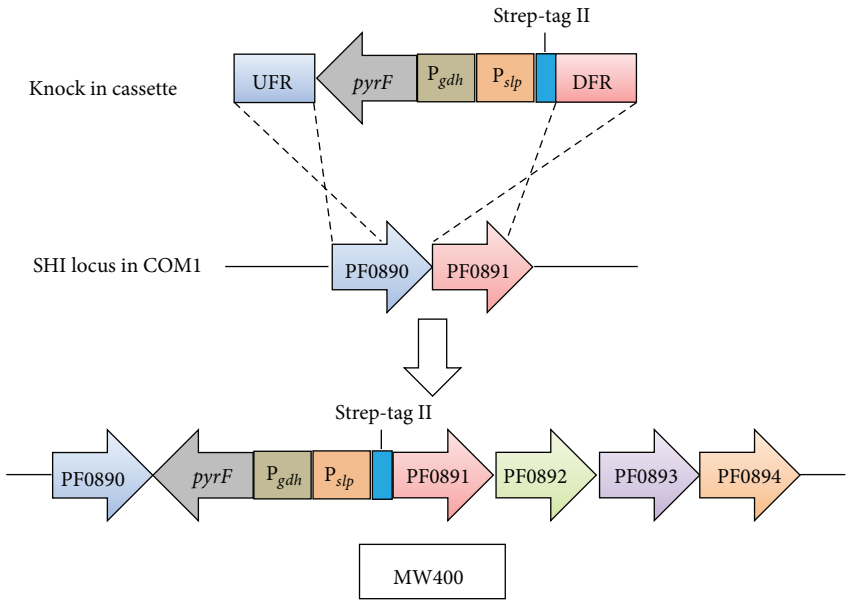


Figure 2: Genetic strategy for overexpression of SHI. MW400 is used as the example. The knock in cassette contains upstream flanking region (UFR) and downstream flanking region (DFR) homologous to PF0890 and PF0891, respectively. This cassette also contains a selectable marker *pyrF* with the promoter P_{gdh}, the promoter for the S-layer protein (P_{slp}) to drive expression of the SHI genes, and a Strep-tag II at the N-terminus of PF0891. By homologous recombination, this cassette was inserted into the SHI locus in *P. furiosus* COM1.

Table 2: Strains for SHI expression.

Strain designation	Genotype	Deleted or inserted ORF/elements	Reference
MW400	P_{slp} Strep-tag II-*shIβγδα*	P_{slp} Strep-tag II, P_{gdh}-*pyrF*	[11]
MW402	Δ*shIβγδα* P_{slp} 9x-His-*shIδα*	P_{slp} 9x-His-*shIδα*, P_{pdaD}-*pdaD*	[13]
MW430	P_{slp} 9x-His-*shIβγδα*	P_{slp}9x-His, P_{gdh}-*pyrF*	[12]
MW434	P_{slp} 9x-His-*shIβγδα*	P_{slp}9x-His, P_{gdh}-*pyrF*	[12]

3. Properties

SHI is classified as a group 3 bidirectional [NiFe] hydrogenase based on the amino acid sequence that surround the four cysteinyl residues that coordinate the [NiFe] catalytic site [7]. In *in vitro* assays, SHI oxidizes H_2 and reduces $NADP^+$ and can also reduce protons *in vitro* to evolve H_2 using NADPH as the electron donor [22]. Kinetic studies on SHI showed that it has ten times higher hydrogen consumption activity than hydrogen production and has a high affinity for hydrogen (K_m 20 μM) and $NADP^+$ (K_m 37 μM) [6, 23, 24]. This suggested that the physiological function of SHI is to regenerate NADPH from the hydrogen produced by MBH [24]. However, since the SHI deletion mutant strain did not have a phenotype under the growth conditions used in the laboratory, the true physiological function of SHI is still a mystery [20]. As shown in Figure 1, it is predicted that NADPH binds to the flavin-containing subunit PF0892 and the electrons are transferred through the iron-sulfur clusters of PF0892, PF0891, and PF0893 and finally to the catalytic site in PF0894 in order to evolve hydrogen. As expected, the dimeric form of SHI did not interact with NADP(H) but, interestingly, it accepted electrons directly from pyruvate ferredoxin oxidoreductase (POR) to produce hydrogen in the absence of an intermediate electron carrier. This two-enzyme system therefore directly oxidized pyruvate to hydrogen gas (and acetyl-CoA) [13]. Potentially, the POR subunit that would normally reduce ferredoxin is able to directly transfer electrons to iron-sulfur clusters in dimeric SHI that are exposed due to the lack of the other two subunits (Figure 1).

P. furiosus is a hyperthermophile that grows at 100°C so it would be expected that SHI is extremely stable at high temperature and this proved to be the case. The half-life of SHI at 90°C (as measured by its hydrogen evolution activity) is 14 hours for the native enzyme and 6 hours for the affinity tagged enzyme [12]. SHI is not a very oxygen-sensitive enzyme. The half-life in air at 25°C (as measured by loss of hydrogen evolution activity at 80°C) was 21 hours for native SHI and 25 hours for the affinity tagged version [11]. SHI is much less sensitive to inactivation by oxygen compared to the well-characterized group 1 [NiFe] hydrogenases, which are typically inactivated within an hour after exposure to oxygen [25]. Although it is regarded as a strictly anaerobic microorganism, *P. furiosus* is also resistant to oxygen and contains a mechanism of oxygen detoxification. This allows

it to grow even in the presence of 8% (v/v) oxygen. SHI does not contribute to the resistance mechanism as the SHI deletion strain behaved similarly to the parent strain [26]. The general resistance to oxygen of SHI is an attractive property for biofuel-related applications.

SHI has been characterized previously by electron paramagnetic resonance (EPR) and Fourier transform infrared (FTIR) spectroscopy [6, 27–29]. The EPR properties of the enzyme are consistent with the iron-sulfur clusters predicted in the model as shown in Figure 1 [6, 29]. EPR can also be used to follow the electronic state of the [NiFe] active site. In general, the Ni atom in [NiFe] hydrogenases typically exhibits three paramagnetic states referred to as Ni-A, Ni-B, and Ni-C. Ni-A is referred to as the inactive unready state and requires incubation under reducing conditions for hours to become active. Ni-B is referred to as the inactive ready state of the enzyme and this can be reactivated within minutes under reducing conditions [5]. These EPR-active states are further distinguished by the type of oxygen ligand bound in the active site. Ni-A is believed to harbor a peroxide ligand while Ni-B is thought to harbor a bound hydroxide ligand, which may explain the faster reactivation for Ni-B state as the hydroxide would be easier to remove upon reduction. Ni-C represents the active ready state of the enzyme, which is free of any oxygen species, and performs the catalytic reaction with hydrogen [5]. These EPR detectable states (Ni-A-like, Ni-B-like, Ni-C) have been observed within *P. furiosus* SHI although a heat treatment step was required to observe some of these signals [27].

The diamagnetic states of the Ni atom in [NiFe] hydrogenases can be observed by FTIR, which detects the vibration frequencies of the CO and CN ligands bound to the Fe atom in the active site. The FTIR signals on the CN ligand of SHI have been reported [28]. The frequencies at 1959, 1950, 1967, and 1954 cm^{-1} were assigned to the $Ni_{u/r}$-A/B, Ni_a-S, Ni_a-C, and Ni_a-SR states, respectively [5, 28]. These are in agreement with the data obtained from the extensively studied group 1 hydrogenases. The results from X-ray absorption spectroscopy also show that the nickel coordination geometry of SHI is identical to that of the active site of the group 1 hydrogenases [30]. Taken together, all of these data show that the [NiFe] site of SHI assumes the same redox states and similar architecture as the catalytic sites of the standard group 1 hydrogenases. Indeed, SHI was recently used as a model [NiFe] hydrogenase to investigate the catalytic mechanism using nanosecond transient infrared and visible absorbance spectroscopy [31]. This approach identified three new catalytic intermediates and established the first elementary mechanistic description of catalysis by a [NiFe] hydrogenase.

4. Biotechnological Applications

Like all hydrogenases, SHI catalyzes the reversible oxidation of hydrogen but it is extremely unusual in that it is one of the few hydrogenases that use NAD(P)H as an electron carrier. Hence, SHI can oxidize NADPH to produce hydrogen or can use hydrogen to reduce $NADP^+$, and applications exist that rely on both of these reactions. In general industrial applications, oxidoreductase-type enzymes have been used as biocatalysts in organic synthesis where they catalyze stereoselective reductions. The products from these syntheses include pharmaceuticals, artificial flavors, and agrochemicals [32]. However, such oxidoreductases require a cofactor, either NADH or NADPH, as a source of reductant, but these are too expensive to be used directly in industrial synthesis. SHI was the first hydrogenase reported to be used in an application to regenerate NADPH using hydrogen as the source of reducing power. It was used to regenerate NADPH in enantioselective reductive reactions *in vitro* catalyzed by *Thermoanaerobium* alcohol dehydrogenase [32]. With SHI and hydrogen, the yield of product was greatly improved compared to using NADPH alone. It was also reported that SHI was able to produce approximately 200 μmol of NADPH in 100 hours of repetitive batch reactions. SHI was used at 40°C, at which temperature it had a half-life of 208 hours [32]. These results demonstrate that SHI functions efficiently at temperatures far below its optimum (>90°C). Instead of batch reactions, SHI has been attached on graphite and glass beads in a continuous NADPH production system, and it was also demonstrated that SHI was electrochemically active in the presence of hydrogen in cyclic voltammetric experiments [33]. Electrochemical and spectrophotometric studies also supported the potential applications on biofuel cell and bioelectrocatalytic applications, where SHI immobilized electrodes are used to replace electrocatalysts, such as platinum [34, 35].

SHI has also been used in several different hydrogen production systems. In an *in vitro* hydrogen photoproduction system, SHI accepted electrons from a light activated semiconductor, titanium dioxide (TiO_2) [36, 37]. In this system, a mediator, methyl viologen, was used initially as the electron carrier, but it was found that SHI accepted electrons directly from the photoactivated TiO_2. Improvements on this system have also been reported [38, 39]. Instead of using a mixture of SHI and TiO_2, a two-compartment system was developed that was separated by a membrane. Anodized tubular TiO_2 electrode (ATTE) was placed at the anode and SHI immobilized on the ATTE was used at the cathode. ATTE on the anode splits water into protons and oxygen, and the electrons were conducted into a solar cell and then transferred to cathodic SHI immobilized ATTE for hydrogen production. This system avoided using the rate-limiting step of electron transfer from the photocatalyst to the enzyme thereby improving system performance [38]. It was also shown that SHI and ATTE could be chemically cross-linked, which had a higher hydrogen production yield compared to either a slurry of SHI-ATTE or the direct adsorption of SHI onto ATTE [39].

SHI is also essential for a hydrogen-producing *in vitro* system that has been developed whereby a variety of sugars are completely oxidized to hydrogen and carbon dioxide. This stemmed from an initial report showing that purified SHI could successfully be used in an *in vitro* hydrogen production system when combined with purified glucose dehydrogenase (GDH) from *Thermoplasma acidophilum* [40]. In this system, glucose was oxidized by GDH to generate NADPH and the NADPH was utilized by SHI for hydrogen production. $NADP^+$ produced by SHI was recycled by GDH

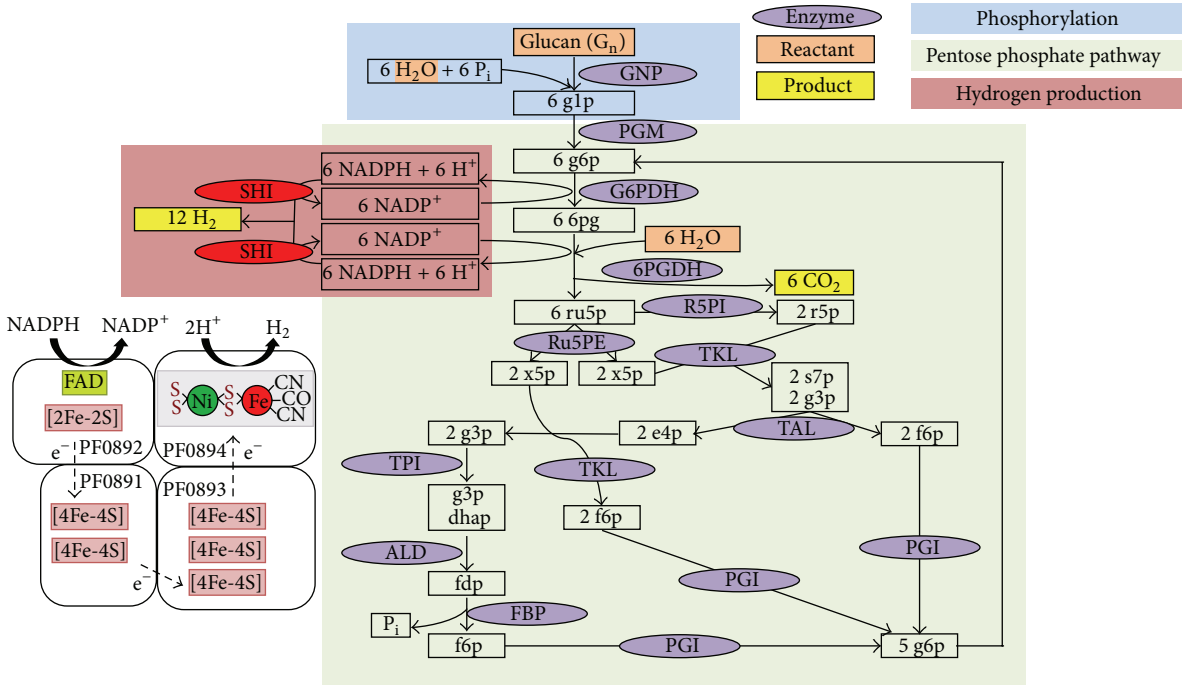


Figure 3: Biohydrogen production from glucan and water via SyPaB. SHI is colored in red and the model of tetrameric SHI (see Figure 1) shows how NADPH is oxidized to produce hydrogen. The abbreviations are GNP, glucan phosphorylase; PGM, phosphoglucomutase; G6PDH, G-6-P dehydrogenase; 6PGDH, 6-phosphogluconate dehydrogenase; R5PI, phosphoribose isomerase; Ru5PE, ribulose 5-phosphate epimerase; TKL, transketolase; TAL, transaldolase; TPI, triose phosphate isomerase; ALD, aldolase; FBP, fructose-1,6-bisphosphatase; PGI, phosphoglucose isomerase. g1p, glucose-1-phosphate; g6p, glucose-6-phosphate; 6pg, 6-phosphogluconate; ru5p, ribulose-5-phosphate; x5p, xylulose-5-phosphate; r5p, ribose-5-phosphate; s7p, sedoheptulose-7-phosphate; g3p, glyceraldehyde-3-phosphate; e4p, erythrose-4-phosphate; dhap, dihydroxyacetone phosphate; fdp, fructose-1,6-diphosphate; f6p, fructose-6-phosphate; and Pi, inorganic phosphate. Modified from [14].

for continual hydrogen production. In addition, this two-enzyme hydrogen production system also generated a high value by-product, gluconic acid, from glucose oxidation [40]. This system was further modified to use renewable resources for hydrogen production, including sucrose, lactose, cellulose, xylan, starch, and pretreated aspen wood, where corresponding enzymes were added to produce monosaccharides. These monosaccharides were further hydrolyzed by the appropriate enzymes for the hydrogen production [41, 42]. This has now been developed into a novel and highly efficient method for *in vitro* hydrogen production using a range of sugars and various mixtures of purified enzymes where, in all cases, SHI is the catalyst for hydrogen production [43]. Compared to biological fermentations (4 H_2/glucose), this cell-free synthetic pathway biotransformation (SyPaB) has a three-time higher theoretical yield (12 H_2/glucose). Different sugar sources can be used, including starch [43], cellulosic materials [44], xylose [45], and sucrose [46]. SyPaB has three basic modules for biohydrogen production as shown in Figure 3, for example, using starch as the sugar source, which is first phosphorylated using inorganic phosphate rather than ATP to produce glucose 6-phosphate (G6P). G6P is oxidized by the pentose phosphate pathway in the second module in order to generate NADPH and to regenerate G6P. In the third module, NADPH produced in the second module is used for hydrogen generation by SHI [43]. This coupled system of SHI and pentose phosphate pathway for biohydrogen production was first described by Woodward et al. [47]. The net reaction is given by the following equation: $C_6H_{10}O_5$ (l) + 7 H_2O (l) $\rightarrow$ 12 H_2 (g) + 6 CO_2 (g). The rate of hydrogen production from glucose was reported to be two times higher than the other substrate sources [48]. Moreover, this pathway exhibited one of the highest biohydrogen generation rates, 157 mmol/L/h, while using glucose 6-phosphate as the electron source [48]. The cost of biohydrogen production by SyPaB using carbohydrates as the source of reductant could be 40–75% lower than commodity prices [14]. Interestingly, a very recent study showed that glucose and xylose from plant biomass could be completely converted to hydrogen by the *in vitro* enzymatic pathway containing *P. furiosus* SHI [49], which bodes well for the future development of this synthetic approach.

5. Conclusions

P. furiosus SHI has been successfully overexpressed in the native host in affinity-tagged forms and can be purified from the cytoplasmic extract in high yield by a single chromatography step. The most efficient purification used a polyhistidine tag, and the overall yield was 50 times higher than that obtained in the original purification of native SHI, which used multistep column chromatography. SHI has

been characterized spectroscopically using EPR and FTIR. Although SHI is classified as a group 3 hydrogenase, the properties of its [NiFe] catalytic site appear to be very similar to those of the extensively characterized group 1 enzymes. SHI has a wide temperature spectrum of enzyme activity (30–95°C) and is less sensitive to oxygen inactivation than typical [NiFe] hydrogenases. It is one of the few hydrogenases that uses NADP(H) as an electron carrier and this has been taken advantage of in several biotechnological applications. These include using SHI to regenerate NADPH with hydrogen as the electron donor for NADPH-dependent oxidoreductase reactions and using SHI to produce hydrogen from NADPH in cell-free synthetic pathways that oxidize a variety of sugars, as well as those in plant biomass, completely to hydrogen gas and carbon dioxide. The availability of significant quantities of recombinant and affinity-tagged SHI should facilitate the further development of these applications, as well as enabling more fundamental structure-function studies of this fascinating enzyme.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work was supported by a Grant (DE-FG05-95ER20175 to Michael W. W. Adams) from the Division of Chemical Sciences, Geosciences and Biosciences, Office of Basic Energy Sciences of the Department of Energy.

References

[1] T. N. Vezirolu and F. Barbir, "Hydrogen: the wonder fuel," *International Journal of Hydrogen Energy*, vol. 17, no. 6, pp. 391–404, 1992.

[2] M. Z. Jacobson, W. G. Colella, and D. M. Golden, "Cleaning the air and improving health with hydrogen fuel-cell vehicles," *Science*, vol. 308, no. 5730, pp. 1901–1905, 2005.

[3] J. E. Kang, T. Brown, W. W. Recker, and G. S. Samuelsen, "Refueling hydrogen fuel cell vehicles with 68 proposed refueling stations in California: measuring deviations from daily travel patterns," *International Journal of Hydrogen Energy*, vol. 39, no. 7, pp. 3444–3449, 2014.

[4] H.-S. Lee, W. F. J. Vermaas, and B. E. Rittmann, "Biological hydrogen production: prospects and challenges," *Trends in Biotechnology*, vol. 28, no. 5, pp. 262–271, 2010.

[5] W. Lubitz, H. Ogata, O. Rüdiger, and E. Reijerse, "Hydrogenases," *Chemical Reviews*, vol. 114, no. 8, pp. 4081–4148, 2014.

[6] F. O. Bryant and M. W. Adams, "Characterization of hydrogenase from the hyperthermophilic archaebacterium, *Pyrococcus furiosus*," *The Journal of Biological Chemistry*, vol. 264, no. 9, pp. 5070–5079, 1989.

[7] P. M. Vignais and B. Billoud, "Occurrence, classification, and biological function of hydrogenases: an overview," *Chemical Reviews*, vol. 107, no. 10, pp. 4206–4272, 2007.

[8] A. Böck, P. W. King, M. Blokesch, and M. C. Posewitz, "Maturation of hydrogenases," *Advances in Microbial Physiology*, vol. 51, pp. 1–71, 2006.

[9] J. O. Eberly and R. L. Ely, "Thermotolerant hydrogenases: biological diversity, properties, and biotechnological applications," *Critical Reviews in Microbiology*, vol. 34, no. 3-4, pp. 117–130, 2008.

[10] J. Sun, R. C. Hopkins, F. E. Jenney Jr., P. M. McTernan, and M. W. W. Adams, "Heterologous expression and maturation of an NADP-Dependent [NiFe]-Hydrogenase: a key enzyme in biofuel production," *PLoS ONE*, vol. 5, no. 5, Article ID e10526, 2010.

[11] S. K. Chandrayan, P. M. McTernan, R. C. Hopkins, J. Sun, F. E. Jenney Jr., and M. W. W. Adams, "Engineering hyperthermophilic archaeon *Pyrococcus furiosus* to overproduce its cytoplasmic [NiFe]-hydrogenase," *The Journal of Biological Chemistry*, vol. 287, no. 5, pp. 3257–3264, 2012.

[12] S. K. Chandrayan, C.-H. Wu, P. M. McTernan, and M. W. W. Adams, "High yield purification of a tagged cytoplasmic [NiFe]-hydrogenase and a catalytically-active nickel-free intermediate form," *Protein Expression and Purification*, vol. 107, pp. 90–94, 2015.

[13] R. C. Hopkins, J. Sun, F. E. Jenney, S. K. Chandrayan, P. M. McTernan, and M. W. W. Adams, "Homologous expression of a subcomplex of *Pyrococcus furiosus* hydrogenase that interacts with pyruvate ferredoxin oxidoreductase," *PLoS ONE*, vol. 6, no. 10, Article ID e26569, 2011.

[14] Y.-H. P. Zhang, "A sweet out-of-the-box solution to the hydrogen economy: is the sugar-powered car science fiction?" *Energy and Environmental Science*, vol. 2, no. 3, pp. 272–282, 2009.

[15] P. J. Silva, E. C. D. van den Ban, H. Wassink et al., "Enzymes of hydrogen metabolism in *Pyrococcus furiosus*," *European Journal of Biochemistry*, vol. 267, no. 22, pp. 6541–6551, 2000.

[16] R. Sapra, K. Bagramyan, and M. W. W. Adams, "A simple energy-conserving system: proton reduction coupled to proton translocation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 13, pp. 7545–7550, 2003.

[17] K. Ma, R. Weiss, and M. W. W. Adams, "Characterization of hydrogenase II from the hyperthermophilic archaeon *Pyrococcus furiosus* and assessment of its role in sulfur reduction," *Journal of Bacteriology*, vol. 182, no. 7, pp. 1864–1871, 2000.

[18] P. M. McTernan, S. K. Chandrayan, C.-H. Wu et al., "Intact functional fourteen-subunit respiratory membrane-bound [NiFe]-hydrogenase complex of the hyperthermophilic archaeon *Pyrococcus furiosus*," *The Journal of Biological Chemistry*, vol. 289, no. 28, pp. 19364–19372, 2014.

[19] P. M. McTernan, S. K. Chandrayan, C. H. Wu, B. J. Vaccaro, W. A. Lancaster, and M. W. Adams, "Engineering the respiratory membrane-bound hydrogenase of the hyperthermophilic archaeon *Pyrococcus furiosus* and characterization of the catalytically active cytoplasmic subcomplex," *Protein Engineering, Design & Selection*, vol. 28, no. 1, pp. 1–8, 2015.

[20] G. L. Lipscomb, K. Stirrett, G. J. Schut et al., "Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases," *Applied and Environmental Microbiology*, vol. 77, no. 7, pp. 2232–2238, 2011.

[21] D. Sasaki, S. Watanabe, T. Kanai, H. Atomi, T. Imanaka, and K. Miki, "Characterization and *in vitro* interaction study of a [NiFe] hydrogenase large subunit from the hyperthermophilic archaeon *Thermococcus kodakarensis* KOD1," *Biochemical and Biophysical Research Communications*, vol. 417, no. 1, pp. 192–196, 2012.

[22] K. S. Ma, Z. H. Zhou, and M. W. W. Adams, "Hydrogen production from pyruvate by enzymes purified from the hyperthermophilic archaeon, *Pyrococcus furiosus*: a key role for NADPH," *FEMS Microbiology Letters*, vol. 122, no. 3, pp. 245–250, 1994.

[23] D. J. van Haaster, P.-L. Hagedoorn, J. A. Jongejan, and W. R. Hagen, "On the relationship between affinity for molecular hydrogen and the physiological directionality of hydrogenases," *Biochemical Society Transactions*, vol. 33, no. 1, pp. 12–14, 2005.

[24] D. J. van Haaster, P. J. Silva, P.-L. Hagedoorn, J. A. Jongejan, and W. R. Hagen, "Reinvestigation of the steady-state kinetics and physiological function of the soluble NiFe-hydrogenase I of *Pyrococcus furiosus*," *Journal of Bacteriology*, vol. 190, no. 5, pp. 1584–1587, 2008.

[25] H. M. Van Der Westen, S. G. Mayhew, and C. Veeger, "Effect of growth conditions on the content and O_2-stability of hydrogenase in the anaerobic bacterium *Desulfovibrio vulgaris* (Hildenborough)," *FEMS Microbiology Letters*, vol. 7, no. 1, pp. 35–39, 1980.

[26] M. P. Thorgersen, K. Stirrett, R. A. Scott, and M. W. W. Adams, "Mechanism of oxygen detoxification by the surprisingly oxygen-tolerant hyperthermophilic archaeon, *Pyrococcus furiosus*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 45, pp. 18547–18552, 2012.

[27] P. J. Silva, B. de Castro, and W. R. Hagen, "On the prosthetic groups of the NiFe sulfhydrogenase from *Pyrococcus furiosus*: topology, structure, and temperature-dependent redox chemistry," *Journal of Biological Inorganic Chemistry*, vol. 4, no. 3, pp. 284–291, 1999.

[28] H. X. Wang, C. Y. Ralston, D. S. Patil et al., "Nickel L-edge soft X-ray spectroscopy of nickel-iron hydrogenases and model compounds—evidence for high-spin nickel(II) in the active enzyme," *Journal of the American Chemical Society*, vol. 122, no. 43, pp. 10544–10552, 2000.

[29] A. F. Arendsen, P. T. M. Veenhuizen, and W. R. Hagen, "Redox properties of the sulfhydrogenase from *Pyrococcus furiosus*," *FEBS Letters*, vol. 368, no. 1, pp. 117–121, 1995.

[30] J. Van Elp, G. Peng, Z. H. Zhou et al., "Nickel L-edge X-ray absorption spectroscopy of *Pyrococcus furiosus* hydrogenase," *Inorganic Chemistry*, vol. 34, no. 10, pp. 2501–2504, 1995.

[31] B. L. Greene, C. Wu, P. M. McTernan, M. W. Adams, and R. B. Dyer, "Proton-coupled electron transfer dynamics in the catalytic mechanism of a [NiFe]-hydrogenase," *Journal of the American Chemical Society*, vol. 137, no. 13, pp. 4558–4566, 2015.

[32] R. Mertens, L. Greiner, E. C. D. van den Ban, H. B. C. M. Haaker, and A. Liese, "Practical applications of hydrogenase I from *Pyrococcus furiosus* for NADPH generation and regeneration," *Journal of Molecular Catalysis B: Enzymatic*, vol. 24-25, pp. 39–52, 2003.

[33] L. Greiner, I. Schröder, D. H. Müller, and A. Liese, "Utilization of adsorption effects for the continuous reduction of $NADP^+$ with molecular hydrogen by *Pyrococcus furiosus* hydrogenase," *Green Chemistry*, vol. 5, no. 6, pp. 697–700, 2003.

[34] W. Johnston, M. J. Cooney, B. Y. Liaw, R. Sapra, and M. W. W. Adams, "Design and characterization of redox enzyme electrodes: new perspectives on established techniques with application to an extremeophilic hydrogenase," *Enzyme and Microbial Technology*, vol. 36, no. 4, pp. 540–549, 2005.

[35] O. G. Voronin, D. J. Van Haaster, E. E. Karyakina, W. R. Hagen, and A. A. Karyakin, "Direct bioelectrocatalysis by NADP-reducing hydrogenase from *Pyrococcus furiosus*," *Electroanalysis*, vol. 19, no. 21, pp. 2264–2266, 2007.

[36] A. Selvaggi, C. Tosi, U. Barberini, E. Franchi, F. Rodriguez, and P. Pedroni, "In vitro hydrogen photoproduction using *Pyrococcus furiosus* sulfhydrogenase and TiO_2," *Journal of Photochemistry and Photobiology A: Chemistry*, vol. 125, no. 1–3, pp. 107–112, 1999.

[37] P. Pedroni, G. M. Mura, G. Galli, C. Pratesi, L. Serbolisca, and G. Grandi, "The hydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*: from basic research to possible future applications," *International Journal of Hydrogen Energy*, vol. 21, no. 10, pp. 853–858, 1996.

[38] S. Bae, E. Shim, J. Yoon, and H. Joo, "Enzymatic hydrogen production by light-sensitized anodized tubular TiO_2 photoanode," *Solar Energy Materials and Solar Cells*, vol. 92, no. 4, pp. 402–409, 2008.

[39] J. Yoon, S. Bae, E. Shim, and H. Joo, "*Pyrococcus furiosus*-immobilized anodized tubular titania cathode in a hydrogen production system," *Journal of Power Sources*, vol. 189, no. 2, pp. 1296–1301, 2009.

[40] J. Woodward, S. M. Mattingly, M. Danson, D. Hough, N. Ward, and M. Adams, "In vitro hydrogen production by glucose dehydrogenase and hydrogenase," *Nature Biotechnology*, vol. 14, no. 7, pp. 872–874, 1996.

[41] J. Woodward and M. Orr, "Enzymatic conversion of sucrose to hydrogen," *Biotechnology Progress*, vol. 14, no. 6, pp. 897–902, 1998.

[42] J. Woodward, K. A. Cordray, R. J. Edmonston, M. Blanco-Rivera, S. M. Mattingly, and B. R. Evans, "Enzymatic hydrogen production: conversion of renewable resources for energy production," *Energy and Fuels*, vol. 14, no. 1, pp. 197–201, 2000.

[43] Y.-H. P. Zhang, B. R. Evans, J. R. Mielenz, R. C. Hopkins, and M. W. W. Adams, "High-yield hydrogen production from starch and water by a synthetic enzymatic pathway," *PLoS ONE*, vol. 2, no. 5, article e456, 2007.

[44] X. Ye, Y. Wang, R. C. Hopkins et al., "Spontaneous high-yield production of hydrogen from cellulosic materials and water catalyzed by enzyme cocktails," *ChemSusChem*, vol. 2, no. 2, pp. 149–152, 2009.

[45] J. S. Martín del Campo, J. Rollin, S. Myung et al., "High-yield production of dihydrogen from xylose by using a synthetic enzyme cascade in a cell-free system," *Angewandte Chemie—International Edition*, vol. 52, no. 17, pp. 4587–4590, 2013.

[46] S. Myung, J. Rollin, C. You et al., "In vitro metabolic engineering of hydrogen production at theoretical yield from sucrose," *Metabolic Engineering*, vol. 24, pp. 70–77, 2014.

[47] J. Woodward, M. Orr, K. Cordray, and E. Greenbaum, "Enzymatic production of biohydrogen," *Nature*, vol. 405, no. 6790, pp. 1014–1015, 2000.

[48] J. A. Rollin, X. Ye, J. M. Del Campo, M. W. Adams, and Y. H. Zhang, "Novel hydrogen bioreactor and detection apparatus," *Advances in Biochemical Engineering/Biotechnology*, 2014.

[49] J. A. Rollin, J. Martin del Campo, S. Myung et al., "High-yield hydrogen production from biomass by in vitro metabolic engineering: mixed sugars coutilization and kinetic modeling," *Proceedings of the National Academy of Sciences*, vol. 112, no. 16, pp. 4964–4969, 2015.

Hindawi Publishing Corporation
Archaea
Volume 2015, Article ID 305497, 14 pages
http://dx.doi.org/10.1155/2015/305497

Review Article

Archaeal MCM Proteins as an Analog for the Eukaryotic Mcm2–7 Helicase to Reveal Essential Features of Structure and Function

Justin M. Miller and Eric J. Enemark

Department of Structural Biology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA

Correspondence should be addressed to Eric J. Enemark; eric.enemark@stjude.org

Received 27 January 2015; Accepted 5 April 2015

Academic Editor: Vickery L. Arcus

In eukaryotes, the replicative helicase is the large multisubunit CMG complex consisting of the Mcm2–7 hexameric ring, Cdc45, and the tetrameric GINS complex. The Mcm2–7 ring assembles from six different, related proteins and forms the core of this complex. In archaea, a homologous MCM hexameric ring functions as the replicative helicase at the replication fork. Archaeal MCM proteins form thermostable homohexamers, facilitating their use as models of the eukaryotic Mcm2–7 helicase. Here we review archaeal MCM helicase structure and function and how the archaeal findings relate to the eukaryotic Mcm2–7 ring.

1. Introduction

The process of cell division requires precise duplication of genetic material. This duplication is carried out by replisomes that coordinate multiple protein activities to separate parental DNA strands and to synthesize new strands of complementary DNA. DNA strand separation and also the progression of the replisome along DNA are fueled by a replicative helicase. In eukaryotes, the replicative helicase is a large multiprotein complex, termed the CMG complex, which encircles leading-strand DNA at the replication fork [1–3]. The CMG complex consists of Cdc45, the Mcm2–7 heterohexamer ring, and the GINS tetramer. The Mcm2–7 ring contains six unique gene products, which are loaded with Cdt1 at replication origins by the Origin Recognition Complex (ORC) and Cdc6 [4] to yield two Mcm2–7 rings that inactively encircle the DNA as a double-hexamer [4, 5] (Figure 1(a)). Helicase activation requires the Dbf4-dependent Cdc7 kinase (DDK) and cyclin-dependent kinases (CDKs) to phosphorylate Mcm2–7 and drive recruitment of Cdc45 and the GINS complex. Physical interaction of GINS and Cdc45 with phosphorylated Mcm2–7 supports formation of the active helicase complex [6–13]. Once activated, the two CMG complexes separate and translocate independently in opposite directions on opposing ssDNA strands in the $3' \rightarrow 5'$ direction to generate two active replication forks [3, 12, 14–18] (Figure 1(a)).

MCM proteins were first identified in a screen for genes essential for minichromosome maintenance in yeast [19]. The identified genes include 6 highly similar proteins with identifiable ATPase motifs [20, 21], Mcm2–7 (Figure 1(b)). The proposed function as the replicative helicase in eukaryotes was initially controversial because Mcm2–7 did not unwind DNA *in vitro*. However, this was resolved by observations that the larger CMG complex unwinds circular and forked double-stranded DNA substrates *in vitro* [2] and that the *Sc*Mcm2–7 complex can unwind DNA by itself when glutamate or acetate was included in the reaction buffer [1].

Electron microscopy studies have revealed general features of Mcm2–7 architecture, including a dynamic gap between the Mcm2 and Mcm5 subunits, an interface previously identified to serve as a "gate" in the ring [1, 22]. EM studies have also shown that Cdc45 and the GINS tetramer associate with the Mcm2–7 hexamer near the Mcm2/5 gate [23, 24], potentially closing the gate (Figure 1(c)). The CMG complex and the Mcm2–7 complex have not been visualized at atomic resolution since each complex has thus far resisted crystallization. A crystal to be used for structure determination must be well ordered, which requires all

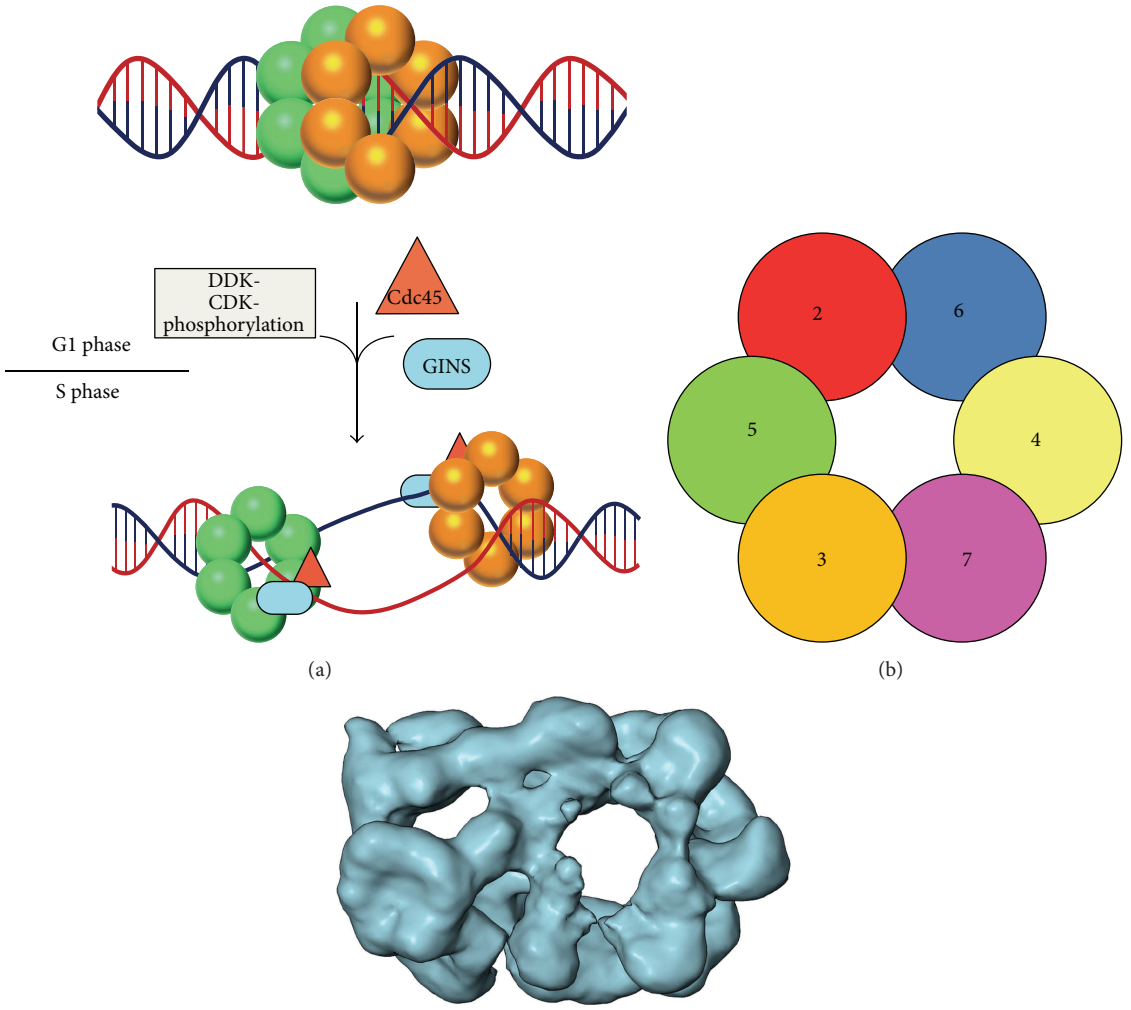


Figure 1: Mcm2–7 activation and organization. (a) Mcm2–7 rings are loaded as inactive double-hexamers at origins of replication by the Origin Recognition Complex (ORC), Cdc6, and Cdt1 (not shown). In a cell-cycle dependent fashion, the Dbf4-dependent Cdc7 kinase (DDK) and cyclin-dependent kinases (CDKs) drive the association of Cdc45 (red triangle) and the GINS complex (blue oval) with the phosphorylated Mcm2–7 ring to yield the active replicative helicase complex, termed the CMG complex (Cdc45-Mcm2–7-GINS). (b) Schematized view of the Mcm2–7 ring from the N-terminal face with subunits labeled to illustrate the order of subunits around the ring [65, 86]. The ring orientation has been selected to correspond with the orientation of the CMG complex shown in (c). (c) The three-dimensional electron microscopy reconstruction of the CMG complex illustrates basic architectural features. The CMG structure representation was prepared with the UCSF Chimera software package [87] and has been labeled with the Electron Microscopy Data Bank (EMDB) accession number.

related molecules to be consistently oriented throughout the crystal. A grossly symmetric ring, such as Mcm2–7, has special crystallization difficulties if the ring is able to adopt multiple orientations that are permuted. These will look similar at the macroscopic level but not at the atomic level. In contrast, archaeal MCM complexes often consist of six identical subunits, and thus all permutations are identical at the macroscopic and atomic level, potentially facilitating crystallization. Indeed, archaeal MCM complexes have been crystallized, which has led to atomic-level descriptions for major functions like DNA binding and ATP hydrolysis. Here we review archaeal MCM helicase structure and function with an emphasis on the advancements in our knowledge of the Mcm2–7 complex that have been derived from archaeal crystallographic studies.

2. Archaeal MCM as a Model for Eukaryotic Mcm2–7

As in eukaryotes, the replisomes of archaea centrally consist of a hexameric ring of MCM proteins that exhibits a $3' \rightarrow 5'$ polarity in *in vitro* double-stranded DNA

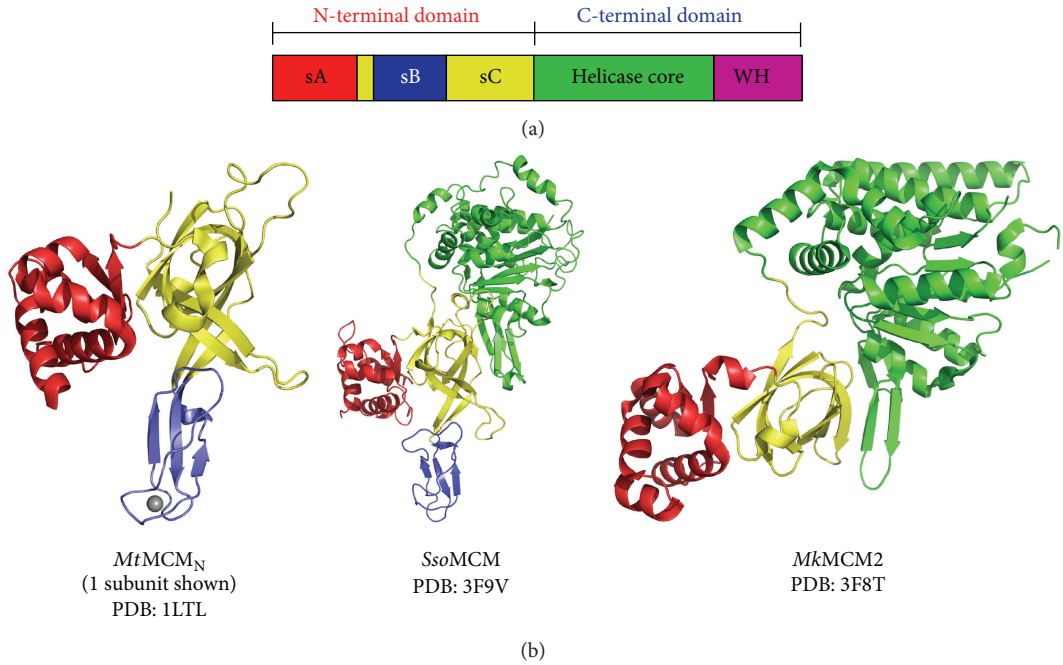


FIGURE 2: General architecture of MCM proteins. (a) MCM monomers can be subdivided into N- and C-terminal tiers. Each tier can be further subdivided, where the N-terminal half includes subdomains A (red), B (blue), and C (yellow) and the C-terminal domain contains the helicase core (green) and a winged helix domain (magenta). (b) The crystal structures of one subunit of the *Methanobacterium thermoautotrophicum* MCM N-terminal domain double-hexamer (*Mt*MCM$_N$, PDB: 1LTL), monomeric *Sulfolobus solfataricus* MCM (*Sso*MCM, PDB: 3F9V), and monomeric *Methanopyrus kandleri* MCM2 (*Mk*MCM2, PDB: 3F8T) each illustrate the general architecture with distinct subdomains A, B, C, and a AAA+ ATPase domain. Subdomains A, B, and C consist of a helical bundle (red), a zinc-binding domain (blue), and an oligonucleotide/oligosaccharide- (OB-) binding fold (yellow), respectively. The AAA+ domain architecture has five β-sheets flanked by α-helices. All structure representations of Figure 2 were prepared with the Pymol software package [88].

(dsDNA) unwinding experiments [17, 25, 26]. The amino acid sequences of eukaryotic and archaeal MCM proteins are highly conserved. Due to the strong functional and sequence conservation, archaeal MCM proteins have proven to be powerful tools for elucidating essential features of MCM function. The MCM complexes of many archaea form homohexamers from a single gene product [27]. As such, these archaeal MCM complexes represent simplified versions of the eukaryotic Mcm2–7 complex and can serve as a model, both structurally and biochemically. These models have played a critical role in deciphering essential features of MCM structure and function because exclusively archaeal MCM complexes have generated crystal structures thus far. The identified essential features are likely to be conserved in all MCM complexes, including eukaryotic Mcm2–7.

3. MCM Overall Architecture

Based on electron microscopy studies, MCM hexamers form a ring with distinct N-terminal and C-terminal tiers [25, 28–31]. Both tiers can independently bind DNA with the N-terminal tier showing a stronger affinity than the C-terminal tier [32]. The C-terminal tier contains the highly conserved AAA+ (ATPases associated with various cellular activities) ATPase/helicase core [33–36] and a winged helix (WH) domain [37] (Figure 2(a)). The C-terminal ATPase tier alone from either *Sso*MCM [32] or *Ape*MCM [38] is sufficient for *in vitro* unwinding of dsDNA. The N-terminal tier (MCM$_N$) shows three consistent subdomains in crystal structures, A (sA), B (sB), and C (sC) [39–44] (Figures 2(a)-2(b)), as defined from the crystal structure of *Mt*MCM$_N$ [39].

3.1. Subdomain A: Peripheral Helical Bundle. Subdomain A is a helical bundle located at the ring periphery and is directly connected to subdomain C by a short linker that may allow for dynamic interaction with the body of the protein [45–48] (Figure 2(b)). Dynamic ring association could play a role in regulating MCM activities, perhaps by restricting access to protein interaction surfaces [44, 46, 49]. Deletion of subdomain A significantly reduces both single-stranded DNA (ssDNA) and dsDNA binding by MCM [28]. The 5′-tail of Y-shaped DNA substrates has been proposed to wrap around the MCM exterior and fit into a binding pocket formed between solvent-exposed C-terminal residues and subdomain A [28, 50–52]. Hence, subdomain A orientation may impact protein:protein interactions, protein:DNA interactions, or both.

The *mcm5-BOB1* P83L mutation in yeast allows cells to bypass the requirement for the S-phase activator protein Cdc7 [53]. The crystal structure of *Mt*MCM$_N$ reveals that the corresponding residue sits at the center of the interface between subdomain A and subdomains B/C [39].

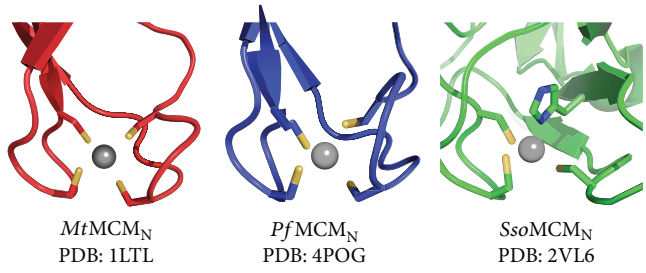
A "domain-push" model was proposed where mutation of the buried proline residue to an amino acid with a bulkier side-chain weakens the interaction between subdomain A and subdomains B/C. This hypothesis was tested by introducing P83L, P83W, or P83K Mcm5 mutations in yeast cells [39]. Consistent with the "domain-push" model, bulky side-chains (P83L, P83W, or P83K) allowed S-phase checkpoint bypass, but P83G Mcm5 mutant-containing cells behaved equivalent to wild-type cells. The N-terminal subdomain A has most often been observed in EM and crystal structures to pack against subdomains B/C [39, 40, 43, 45, 48], but subdomain A has also been observed in a different conformation that is extended away from the body of the protein [44, 45, 54]. The "domain-push" may expose a protein interaction surface that is normally only exposed following DDK phosphorylation. Thus, the observation that the *mcm5-BOB1* mutant is able to bypass the S-phase checkpoint may be a result of a shift in protein interaction partners.

3.2. Subdomain B: Zinc-Binding Domain. All X-ray crystal structures of MCM_N reveal a tetrahedrally coordinated zinc ion bound to subdomain B [39, 40, 43] (Figure 3(a)). Critical zinc-binding amino acid side-chains are located on three antiparallel β-strands [39–41, 43, 44]. MCM biochemical activities show extreme sensitivity to mutation of the zinc-binding residues, and a cysteine to serine point mutation in *Mt*MCM ablates dsDNA unwinding, DNA-dependent ATPase activity, and ssDNA binding [55]. The zinc-binding sequence motif is commonly $CX_2CX_nCX_2C$ (C_4 type) [39, 43], although a histidine residue is also possible, as observed in the crystal structure of $SsoMCM_N$ (Figures 3(a)-3(b)) [40] and in the sequence of *Ape*MCM (Figure 3(b)). Other than Mcm3, all of the subunits of eukaryotic Mcm2–7 possess four highly conserved cysteines that likely coordinate a zinc ion with a tetrahedral geometry. The spacing of the cysteines in these Zn-binding sequence motifs show family-specific patterns (Figure 3(b)). Mcm2 has a $CX_2CX_nCX_2C$ motif analogous to those of *Mt*MCM and *Pf*MCM. Moreover, a highly similar $CX_2CX_nCX_{2-4}C$ motif of Mcm4, Mcm6, and Mcm7 also likely binds a zinc ion [55, 56]. In contrast, the four conserved cysteines of Mcm5 have an unusually large spacing between the third and fourth cysteine. The family-specific conservation of the zinc-binding motif, including the lack of an obvious motif in Mcm3, suggests that the zinc-binding sites provide an important mechanism for regulation of eukaryotic Mcm2–7 activities.

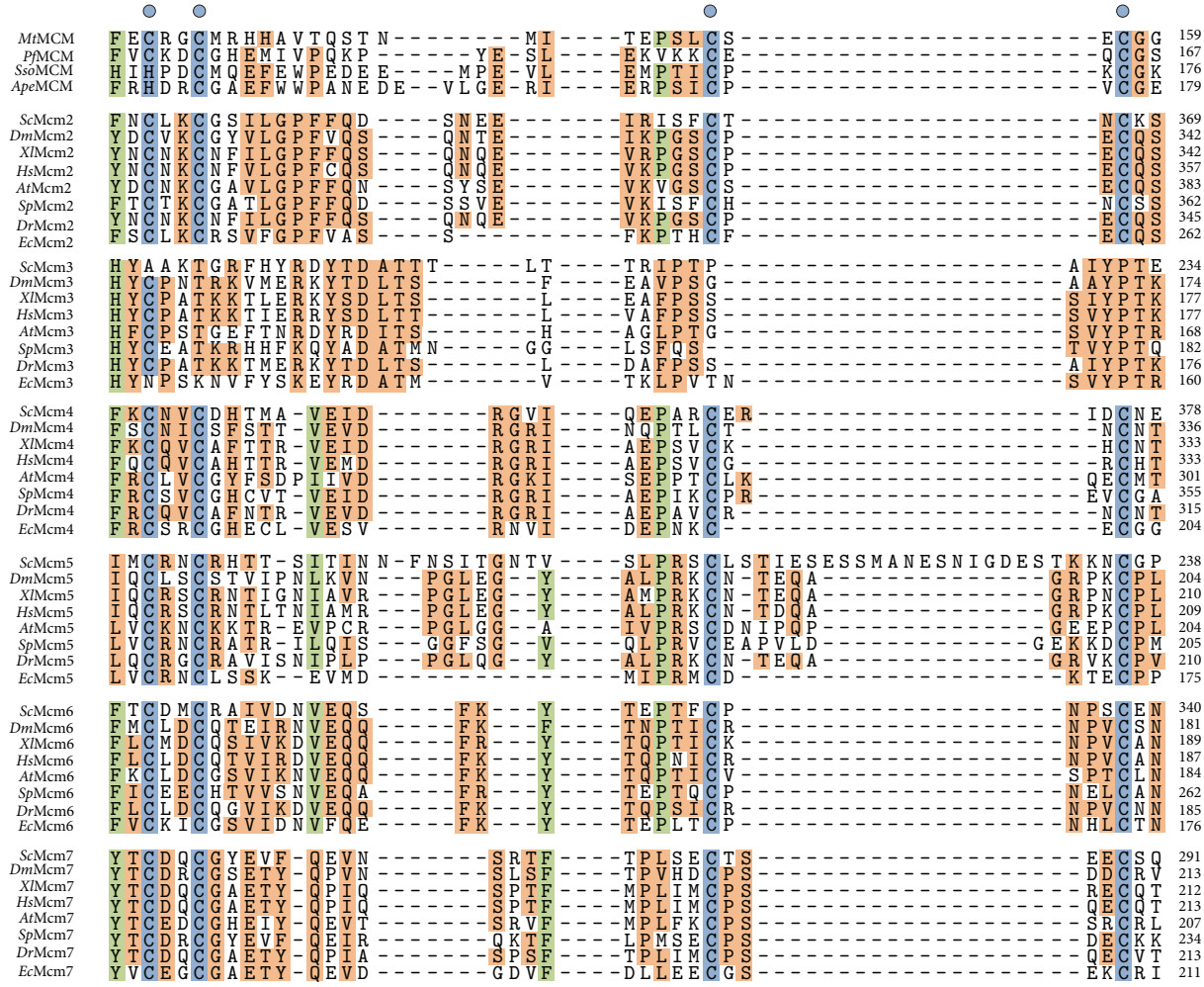
3.3. Subdomain C: OB-Fold. Subdomain C has an oligonucleotide/oligosaccharide fold (OB-fold), a fold associated with DNA binding [57] (Figure 2(b)). Multiple residues in subdomain C are critical for DNA binding and hexamerization [41, 43, 49, 58]. Further, β-loop motifs found in subdomain C appear to facilitate allosteric communication between the N- and C-terminal tiers [59, 60]. A phenylalanine to isoleucine mutation (F345I) in mouse Mcm4 subdomain C has been termed Chaos3 (chromosome aberrations occurring spontaneously 3) [61]. Homozygous $Mcm4^{Chaos3/Chaos3}$ mice develop mammary adenocarcinomas within an average of 12 months for >80% of cases, and mouse cells that possess only F345I mutant Mcm4 (either $Mcm4^{Chaos3/-}$ or $Mcm4^{Chaos3/Chaos3}$) exhibit hypersensitivity to chromosomal breakage [61]. Strains of *Saccharomyces cerevisiae* with the analogous mutation (*Sc*Mcm4 F391I) exhibit the minichromosome loss phenotype typically associated with a loss-of-function for Mcm2–7 [61]. The phenylalanine residue involved is highly conserved as an aromatic residue in all eukaryotic Mcm2–7 subunits and also in *Mt*MCM (*Mt*MCM F170) [61]. The highly conserved aromatic residue is at the center of a hydrophobic intersubunit interface in the structure of *Mt*MCM [39] and is therefore likely critical to the structural integrity of the MCM ring. Although the Chaos3 mutation may not completely prevent interaction of Mcm4 with other Mcm2–7 subunits, the mutation could alter the intersubunit orientations to disrupt DNA binding by the MCM single-stranded binding motif (MSSB, described further below) [43] to generate genomic instability, a general hallmark of cancer cells [62–64].

3.4. AAA+ ATPase Core Domain. Crystal structures of MCM proteins containing both N- and C-terminal tiers have revealed a canonical AAA+ ATP binding and hydrolysis site in the C-terminal tier. The MCM AAA+ domain has 5 parallel β-strands flanked on either side by α-helices [41, 42, 44] (Figure 2(b)). Consistent with other AAA+ family members, such as domain 2 of N-ethylmaleimide sensitive fusion protein, the strand order is $\beta5$–$\beta1$–$\beta4$–$\beta3$–$\beta2$, which strictly positions key catalytic residues to generate a competent ATPase site [34, 41, 42, 44]. The MCM AAA+ ATP binding and hydrolysis site contains canonical *cis*- and *trans*-acting elements like the *cis*-acting Walker A and Walker B motifs and the *trans*-acting "SRF" arginine finger motif [65, 66]. Mutation of nearly any residue of the MCM AAA+ active abolishes ATPase activity [32, 66, 67].

3.5. Winged Helix-Turn-Helix Domain. Amino acid residues found at the extreme C-terminus of archaeal MCM proteins [37], Mcm6 [68], and the replication factor Mcm10 [69, 70] are predicted to adopt a flexible winged helix-turn-helix fold (Figure 2(a)). For most archaeal MCM proteins, this represents the C-terminal 60–70 amino acids. The NMR structure of the Mcm6 winged helix (WH) domain indicates that the isolated domain is well ordered [68, 71]. However, it is connected to the core of the protein via a flexible linker that allows for mobility of the domain similar to subdomain A [48]. Consistent with this, all X-ray crystal structures containing the C-terminal tier reported to date have been with constructs that lack the WH domain, suggesting that the orientation of this domain is intrinsically flexible [41, 42, 44]. The deletion of the WH region in *Sso*MCM or *Mt*MCM results in a nearly 2-fold enhancement in ATPase activity [32, 72]. For *Sso*MCM, deletion of the WH domain also yields a 15-fold increase in dsDNA unwinding activity [32]. Taken together, the results suggest that the WH domain does not play an essential role during normal unwinding and may instead function during the initial assembly or activation of the helicase or to sense atypical DNA structures, such as damaged DNA, and to signal their presence. Consistent with



(a)



(b)

Figure 3: Subdomain B contains a zinc-binding motif. (a) All MCM$_N$ X-ray crystal structures contain a zinc atom bound with a tetrahedral geometry. Each structure is depicted in a cartoon representation with zinc-binding side-chains shown in stick. Cysteine sulfurs are colored in yellow, and histidine nitrogen atoms are colored in blue. The remainder of each protein is shown as red, blue, and green for *Mt*MCM$_N$ (PDB: 1LTL), *Pf*MCM$_N$ (PDB: 4POG), and *Sso*MCM$_N$ (PDB: 2VL6), respectively. Zinc atoms are shown as gray spheres. All structure representations of Figure 3 were prepared with the Pymol software package [88]. (b) A multiple sequence alignment shows the strong sequence conservation of the MCM zinc-binding domain. The spacing of the cysteine residues in MCM Zn-binding sequence motifs shows family-specific patterns, where residues displaying a high degree of sequence conservation across all MCM proteins or family-specific sequence conservation have been colored as either green or orange, respectively. Residues involved in zinc-binding are highlighted in blue with a blue dot for emphasis. Mcm3 does not possess an obvious zinc-binding motif, and Mcm5 has an abnormally large spacing between the third and fourth cysteine residues.

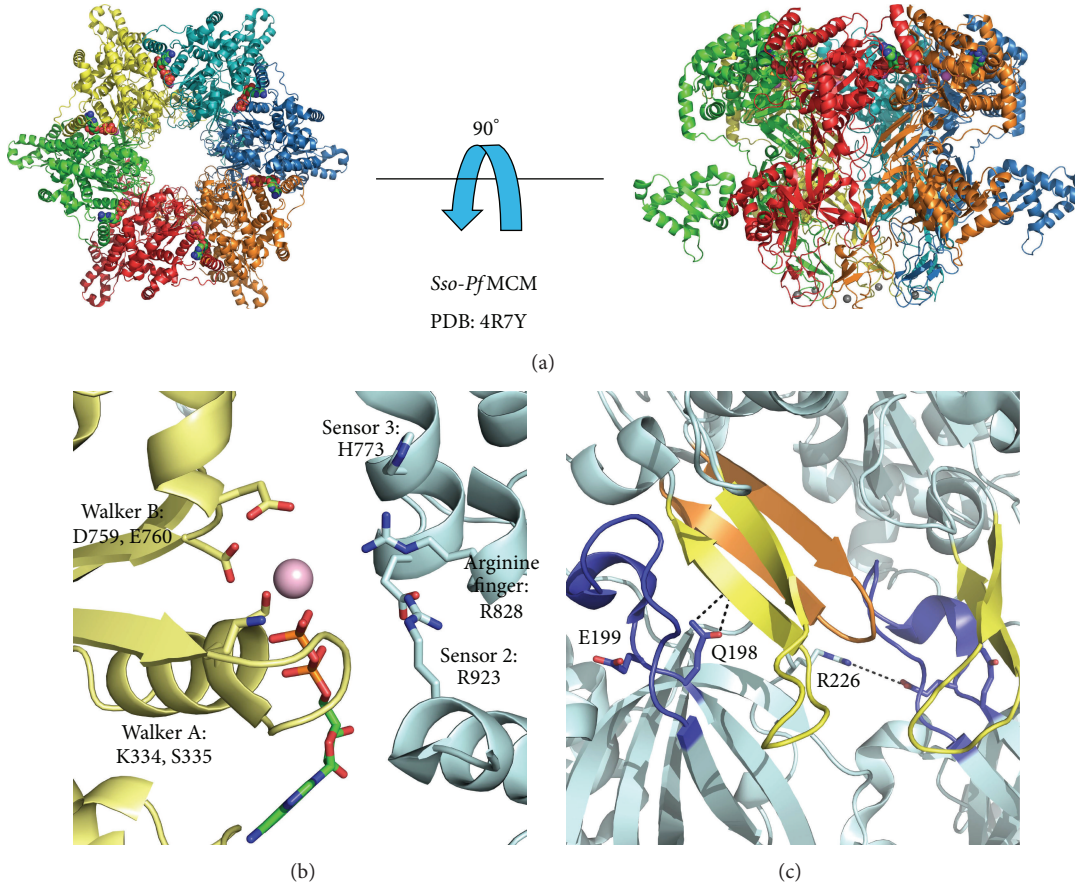


Figure 4: Crystal structure of a nearly full length MCM hexamer. An ADP-bound, nearly full length MCM protein was crystallized by creating a chimeric fusion protein of the N-terminal domain of *Sso*MCM and the C-terminal tier of *Pf*MCM. (a) Views of the *Sso-Pf*MCM hexamer crystal structure parallel and perpendicular to the central channel with each subunit uniquely colored. Magnesium ions are shown as magenta spheres, ADP molecules are shown in space-filling view, and zinc ions are shown as gray spheres. In the view parallel to the central channel, the ATPase domains are projected out of the page. In the perpendicular view, the ATPase domains are located at the top and the N-terminal domains are located on the bottom. (b) The ATPase site is formed at subunit interfaces by *cis*- and *trans*-acting residues shown in stick and labeled. The Walker A and Walker B residues of one subunit are positioned at the left side of the site (yellow), while three basic residues are located on the right side of the site (blue). The bound ADP molecule is shown in stick, and the magnesium ion is represented as a magenta sphere. (c) Residues of the allosteric communication loop (ACL) interact across subunit interfaces and also with main-chain atoms of the helix-2-insert (h2i) β-hairpin motif. The ACL is shown in blue and the h2i and presensor-1-β-hairpin (ps1β) are shown in yellow and orange, respectively. Dashed lines indicate the discussed molecular interactions. All structure representations of Figure 4 were prepared with the Pymol software package [88] and PDB accession code 4R7Y.

such a role, this domain is expected to sit directly proximal to the entering dsDNA to be unwound [73].

4. MCM Hexamer

Although archaeal MCM proteins are often robust hexamers in solution [32, 66, 73–75], they have proven to be highly resistant to crystallization in a hexameric form. This challenge was overcome by a chimeric fusion of the N-terminal domain of *Sso*MCM with the ATPase domain of *Pf*MCM (*Sso-Pf*MCM), which crystallized as a hexamer bound to Mg:ADP [44]. In the *Sso-Pf*MCM hexamer structure, the N- and C-terminal domains form distinct tiers stacked on top of one another (Figure 4(a)). The cocrystallized ADP revealed the general architecture of the AAA+ ATP binding site and the specific interactions between the protein and bound nucleotide (Figure 4(b)). A "sensor 3" residue was identified based on the observation of a histidine side-chain that projects into the active site similar to sensor 2 and the arginine finger. The structure also revealed atomic details of the intermolecular interactions between the subunits and also intramolecular interactions between the tiers (Figure 4(c)). In addition, the *Sso-Pf*MCM structure crystallized with subdomain A in an extended conformation, in contrast to the previous N-terminal MCM crystal structures with subdomain A packed against subdomains B/C, consistent with an ability for subdomain A to adopt multiple discrete conformations.

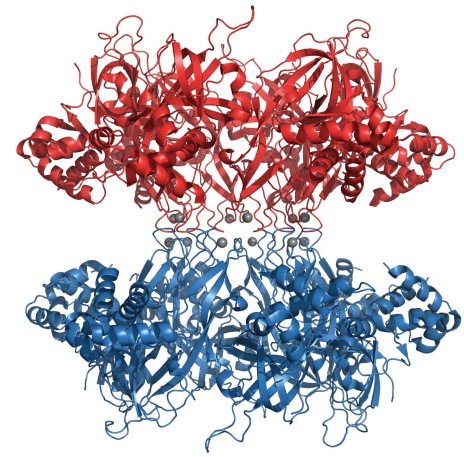
5. ATPase Site

Consistent with other AAA+ superfamily members, the MCM ATP binding and hydrolysis site is at the interface of two adjacent subunits with catalytic residues contributed both in *cis* and in *trans*. The site is competent to catalyze ATP hydrolysis only when all requisite residues are precisely arranged. In the ADP-bound *Sso-Pf*MCM crystal structure, Walker A residue K334 is observed to project into the ATP binding site, and residue S335 coordinates a magnesium cation (Figure 4(b)) [44]. Based on previously proposed AAA+ ATPase catalytic mechanisms, this magnesium cation likely positions the β- and γ-phosphates of the bound ATP molecule [76] and neutralizes an accumulated negative charge on the γ-phosphate during hydrolysis [34, 76]. Based on crystal structures of other AAA+ family members, the second conserved acidic residue of the Walker B motif (E760) is the catalytic base for a water molecule to perform a nucleophilic attack on the γ-phosphate of the bound ATP.

The *trans*-acting elements of AAA+ ATPase sites occupy different positions depending on whether ATP, ADP, or no nucleotide is bound to communicate the active site status to the rest of the protein [66, 77, 78]. The *trans*-acting residues must be able to reach the ATP molecule to generate a hydrolysis site with the adjacent subunit. The subunit interface is likely dynamic because the positions of the *trans*-acting residues in the ADP-bound hexameric *Sso-Pf*MCM structure are too distant for such interaction, and the subunit interface therefore needs to constrict in order for the *trans*-acting residues to interact with ATP [44, 77]. *Trans*-acting MCM residues include "sensor 2" [44, 66, 79], the "arginine finger" [44, 66], and "sensor 3" [44]. The arginine finger may function to polarize the γ-phosphate during ATP hydrolysis [76]. The *trans*-acting "sensor 2" element of MCM contrasts most AAA+ ATPases that typically use this motif as a *cis*-acting residue [42, 44, 66, 79].

6. MCM Double-Hexamer

In eukaryotes, two Mcm2–7 rings are initially loaded as an inactive double-hexamer at replication origins [4, 5]. The X-ray crystal structure of *Mt*MCM_N revealed an MCM double-hexamer with two hexamers arranged head to head via subdomain B domain interactions (Figure 5) [39], providing a model for the Mcm2–7 double-hexamer interaction. A single arginine to alanine mutation at the *Mt*MCM_N hexamer:hexamer interface disrupted the interaction in favor of single hexamers [80]. A double-hexamer structure had previously been observed for full length *Mt*MCM in scanning transmission electron microscopy images [25]. In contrast to the stable double-hexamer of *Mt*MCM, X-ray crystal structures of *Sso*MCM_N and *Pf*MCM_N consist of single hexamers [40, 43]. Although a double-hexamer structure has not yet been universally detected among archaeal MCM complexes, double-hexamer architecture is likely to be conserved when MCM rings are first loaded onto DNA prior to the onset of bidirectional replication [81].



MtMCM_N
PDB: 1LTL

Figure 5: *Mt*MCM_N double-hexamer crystal structure. The *Mt*MCM_N double-hexamer crystal structure is shown in cartoon representation with one hexamer colored red and the other blue. The two hexamers interact head to head via subdomain B interactions. Zinc atoms are shown as gray spheres. The figure was prepared with the Pymol software package [88] and PDB accession code 1LTL.

7. DNA Binding

The N-terminal and C-terminal domains of MCM are independently able to bind DNA with several modules that line the interior channel of the hexameric ring. These modules and their interaction with DNA are detailed further below.

7.1. MCM Single-Stranded Binding Motif (MSSB). The MCM N-terminal domain binds ssDNA as revealed by a recent X-ray crystal structure of *Pf*MCM_N bound to ssDNA [43]. In this structure, ssDNA binds in the plane of the ring, rather than perpendicular as would be expected if DNA were threaded through the central channel (Figure 6(a)) [43]. This DNA binding configuration contrasts the configurations previously seen for the nucleic acid complexes of the hexameric helicases E1 [77], DnaB [82], and Rho [83] where nucleic acid binds perpendicular to the plane of the hexamer. When viewed from the N-terminal side of the complex (Figure 6(a)), the ssDNA binds with $3'$ to $5'$ polarity in the clockwise direction around the ring. The MCM_N:ssDNA structure revealed that 4 ssDNA nucleotides are bound per subunit, which also contrasts with other hexameric helicases that bind either one (E1 [77] and Rho [83]) or two (DnaB [82]) nucleotides per subunit. Interestingly, the MCM_N:ssDNA structure reveals that ssDNA is bound across subunit interfaces but is not bound at all interfaces of a hexamer (Figure 6(b)) [43]. The intersubunit distance appears to dictate whether ssDNA can bind at a particular subunit interface with ssDNA observed only at tighter interfaces [43].

The *Pf*MCM_N:ssDNA cocrystal structure revealed specific amino acid residues of the OB-fold subdomain C that

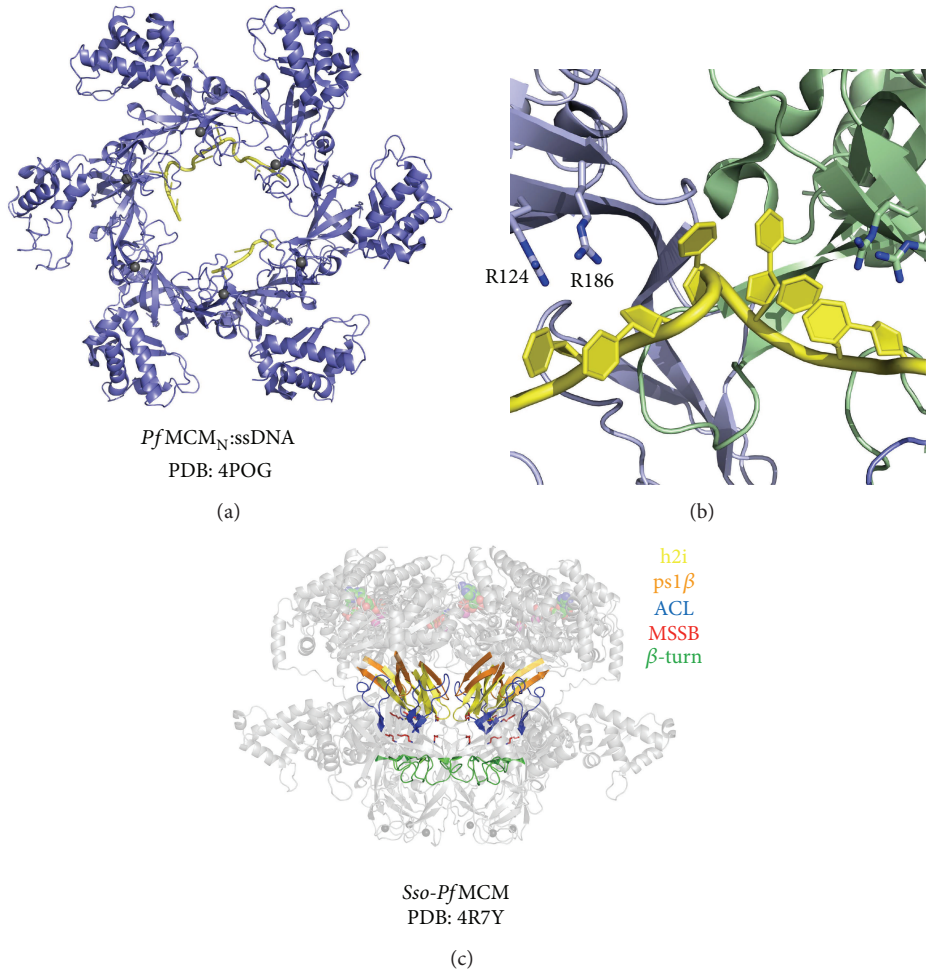


FIGURE 6: Features of DNA binding in MCM hexamers. (a) The cocrystal structure of *Pf*MCM$_N$ bound to ssDNA revealed that DNA binds in the plane of the hexamer. The DNA binds with a 3′ to 5′ polarity in the clockwise direction around the ring when viewed from the N-terminal side of the hexamer. DNA is shown in yellow and zinc ions are shown as grey spheres. (b) DNA binds across subunit interfaces through interactions between the DNA and MCM Single-Strand Binding (MSSB) residues R186 and R124 (*Pf*MCM$_N$:ssDNA structure, PDB: 4POG), shown in stick. (c) Important residues of the central channel in the MCM hexamer structure (PDB: 4R7Y) with other components transparent. The h2i, ps1β, ACL, MSSB, and N-terminal β-turn motifs are shown in yellow, orange, blue, red, and green, respectively. All structure representations of Figure 6 were prepared with the Pymol software package [88].

bind to ssDNA (Figure 6(b)). This motif is termed the MCM Single-Strand Binding (MSSB) motif (Figure 6(b)) and forms a positively charged DNA binding pocket near the interface of the N- and C-terminal tiers (Figure 6(c)). Residues R124 and R186 were demonstrated to most significantly impair DNA binding where the affinity of R124A and R186A point mutants for ssDNA decreased 7- and 6-fold, respectively [43]. The R124A/R186A double-mutant exhibits a 25-fold reduction in ssDNA binding affinity with much higher protein concentrations required for observation of binding relative to the wild-type protein. Mutation of similarly positioned residues in *Sso*MCM abolishes dsDNA unwinding activity [74].

In *Saccharomyces cerevisiae*, the *Pf*MCM R124 and R186 MSSB residues are conserved as either an arginine or lysine in Mcm4, Mcm6, and Mcm7 [43]. In Mcm2, Mcm3, and Mcm5, only one of the two positively charged residues is conserved but not both. Interestingly, complementation studies in yeast demonstrated that while single-subunit MSSB mutants are viable, cells with two-subunit MSSB mutations are lethal: Mcm4/Mcm6, Mcm4/Mcm7, and Mcm6/Mcm7 [43] (Figure 1(b)). In fact, any of these combinations leave the Mcm2–7 ring defective in helicase loading and severely defective for replication [43]. These data suggest a role for the MSSB during Mcm2–7 helicase loading and activation.

7.2. N-Terminal β-Turn. A second N-terminal motif implicated in DNA interaction is a β-turn that projects into the central channel of the hexamer (Figure 6(c)). In *Sso*MCM, this motif spans residues 241–251 (polypeptide sequence QDSPVKRGSRA) between β-strands β11 and β12 [40]. The β-turn is present in all archaeal MCM proteins and has been studied in *Sc*Mcm4 and *Sc*Mcm5 [39, 40, 43, 84]. The length and sequences of this loop vary among Mcm2–7 subunits. In X-ray crystal structures of archaeal MCM

hexamers, the N-terminal β-turn is the narrowest part of the central channel, varying from 17 Å to 23 Å [39, 40, 43]. Despite poor sequence conservation in this region, the presence of positively charged residues on the β-turn is required for DNA binding [39, 73]. Double-mutation of two positively charged residues (K246A/R247A) on the tip of the N-terminal β-turn in *Sso*MCM leads to an 8-fold reduction in DNA binding [73]. Similarly, DNA binding is abolished for the comparable alanine double-mutant of *Mt*MCM (R226A/228A) [39]. Based on these experiments, the DNA binding activity attributed to the N-terminal β-turn is proposed to play a critical role in DNA loading [73, 84]. Structure-guided yeast genetic approaches have suggested that the β-turn of *Sc*Mcm5 is important for binding to origins of replication and subsequent initiation of DNA replication [84]. Thus, the N-terminal β-turn likely has a major role in facilitating initial DNA binding.

7.3. Presensor-1 β-Hairpin (ps1β) and Helix-2-Insert (h2i). The ATPase domain of each MCM monomer contains two β-hairpin motifs that project into the central channel of the hexamer, the helix-2-insert (h2i), and the presensor-1-β-hairpin (ps1β) (Figures 7(a)-7(b)) [41, 42, 44, 72]. The h2i motif is defined by a β-α-β insertion in primary sequence between Walker A and Walker B motifs [35, 72, 79]. This feature is found as a unique insert within helix-2 of the conserved ASCE (additional strand catalytic glutamate) ATPase family [35, 72, 85]. Similarly, the ps1β feature represents an insertion between sensor 1 motif and the preceding helix [35]. Within the AAA+ superfamily, the ps1β superclade includes the SFIII helicases, HslU, ClpX, Lon, ClpA, MCM, dynein, midasin, YifB, and many others [35]. The ps1β superclade is subdivided into HslU/ClpX/Lon/ClpAB-C and h2i subclades, where MCM proteins belong to the h2i subclade since they contain both the ps1β and the h2i motifs [35].

MCM mutants lacking either the ps1β or h2i motifs are unable to catalyze dsDNA unwinding, despite still being competent for DNA binding [72, 73]. The ps1β-deletion mutant displays a weakened binding affinity for Y-shaped DNA of 343 nM versus 152 nM for wild-type protein [73]. Interestingly, this contrasts an h2i-deletion mutant, which displays a shift in binding affinity for a blunt dsDNA substrate from greater than 1 μM for the wild-type protein to ~8 nM for the h2i-deleted mutant [72]. The observation of tighter binding for an h2i-deleted MCM protein may suggest a role for the h2i in destabilizing DNA:protein interactions, which could facilitate the movement of DNA during unwinding. The primary role of the ps1β motif is most likely to bind DNA since deletion of the ps1β feature yields the weakened affinity expected for a motif intimately involved in DNA binding. The lack of detectable unwinding activity for either the ps1β- or h2i-deletion mutants indicates that both C-terminal β-hairpins are essential.

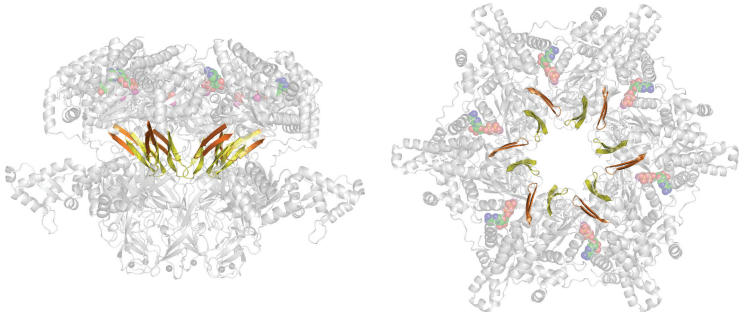
8. DNA Translocation

The conformation of the h2i and ps1β motifs has been proposed to depend on whether ATP or ADP is bound [41, 42, 72]. In the ADP-bound MCM hexamer structure, these loops are observed to be in a "down" position (Figure 7(a)) [44]. Thus, a view perpendicular to the central channel of the hexamer defines the C-terminal domain face as the "up" direction and the N-terminal face as the "down" direction. The orientation of the MCM hexamer shown in Figure 7(a) would result in DNA being translocated with a net movement from the top to the bottom face. Consequently, this motion implies that the h2i and ps1β motifs would be in the "up" position in the ATP-bound state. Following hydrolysis of ATP to ADP, the loops would move to the "down" position shown in Figure 7(a). This direction of DNA translocation is consistent with earlier FRET studies showing that MCM hexamers bind forked DNA with the C-terminus facing the leading direction [73].

The projection of each C-terminal β-hairpin into the central channel of the MCM hexamer is expected to position positively charged residues for interaction with DNA (Figure 7(b)). This is conceptually similar to other hexameric helicases such as E1 [77], DnaB [82], and Rho [83]. In the crystal structures of each of these hexameric helicases, positive residues on loops projected into the central channel interact with nucleic acid in a spiral-staircase-like arrangement (Figures 7(c)–7(h)). In this mode of binding, the central channel loops from adjacent subunits occupy different heights around the ring to form a "spiral-staircase" shape. The movement of DNA is then achieved by the concerted movement of the entire staircase of loops [77, 82, 83]. As an example, the adjacent subunits of the E1 structure adopt ATP-like, ADP-like, and apo forms around the ring that clearly demonstrate that the loop positions depend on whether ATP, ADP, or no nucleotide is bound at the associated ATP binding site [77]. Basic features of this mode of DNA binding and translocation are most likely shared by all the hexameric helicase superfamilies: SFIII (E1), SFIV (DnaB), SFV (Rho), and SFVI (MCM) [89].

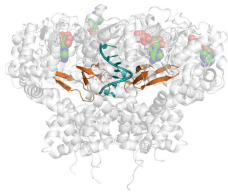
Although MCM proteins display high homology with SFIII helicases, the presence of both the h2i and ps1β features in MCM proteins suggests significantly different mechanisms for coupling the energy of ATP binding and hydrolysis to dsDNA unwinding. Deletion of the h2i β-hairpin in *Mt*MCM results in ~12-fold increase in dsDNA-stimulated ATPase activity relative to wild-type *Mt*MCM [72]. However, no detectable dsDNA unwinding is observed for the h2i-deletion mutant, despite the observation that the h2i-deleted mutant still binds both ssDNA and dsDNA [72]. Taken together, these observations suggest that the deletion of the h2i motif actually disrupts the coupling of ATP binding and hydrolysis to DNA unwinding such that the helicase can bind dsDNA and also hydrolyze ATP but cannot unwind dsDNA. This is conceptually analogous to an automobile engine shifted to neutral with deletion of the h2i motif disengaging the MCM motor from dsDNA unwinding.

8.1. Allosteric Communication Loop. While the C-terminal helicase core can catalyze unwinding of dsDNA in the absence of the MCM N-terminal tier, inclusion of the N-terminal tier results in an increase in the processivity of MCM catalyzed dsDNA unwinding [32]. Such increases in processivity are observed whether the two halves are expressed as

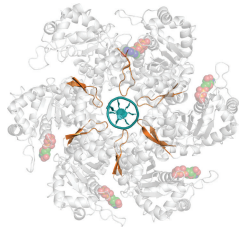


MCM (PDB: 4R7Y)
(a)

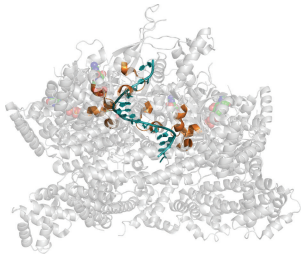
MCM (PDB: 4R7Y)
(b)



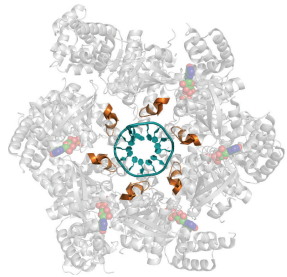
E1 (PDB: 2XGA)
(c)



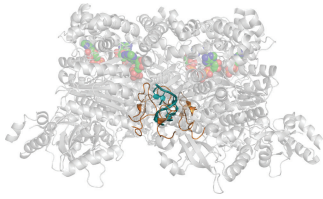
E1 (PDB: 2XGA)
(d)



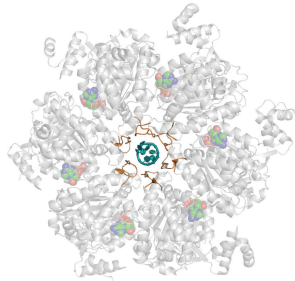
DnaB (PDB: 4ESV)
(e)



DnaB (PDB: 4ESV)
(f)



Rho (PDB: 3ICE)
(g)



Rho (PDB: 3ICE)
(h)

Figure 7: Comparison of MCM with representative members of other hexameric helicase superfamilies. The hexameric MCM structure viewed either perpendicular (a) or parallel (b) to the central channel axis. The MCM h2i and ps1β central channel modules are colored yellow and orange, respectively. (c–h) Comparison with representative members of other helicase superfamilies such as E1 ((c-d), SFIII, PDB: 2XGA), DnaB ((e-f), SFIV, PDB: 4ESV), and Rho ((g-h), SFV, PDB: 3ICE) suggests that MCM central channel modules will move "up" and "down" to translocate DNA. In the crystal structures of each of these hexameric helicases, basic residues on loops projected into the central channel interact with nucleic acid in a spiral-staircase-like arrangement. In this mode of binding, the central channel loops from adjacent subunits occupy different heights around the ring to form a "spiral-staircase" shape. The movement of DNA is then achieved by the movement of the central channel loops. DNA is colored in teal and DNA binding loops are colored orange. All structures are viewed either perpendicular (left column) or parallel (right column) to the central channel axis. All structure representations of Figure 7 were prepared with the Pymol software package [88].

a single gene product or when each half is expressed separately and incubated together during unwinding experiments [32]. An important interaction between each half is mediated through a conserved loop found at the interface of the N- and C-terminal tiers termed the "allosteric communication loop" (ACL, Figure 6(c)). Mutational studies have revealed that point mutations in this loop can have either stimulatory or inhibitory effects on the observed MCM dsDNA unwinding activity [60]. For the majority of point mutations introduced to this loop, the rate of *Mt*MCM catalyzed dsDNA unwinding is observed to decrease relative to the wild-type protein, but mutation of either Q181 or E185 to alanine actually results in an increase in unwinding rate. Deletion of the entire ACL causes a loss of unwinding activity. However, activity is restored by the additional deletion of the N-terminal β-turn, which suggests the position of the N-terminal β-turn could be controlled by the ATP hydrolysis cycle via the ACL [59]. In the *Sso-Pf*MCM hexamer crystal structure, a fully conserved glutamine residue (homologous to *Mt*MCM Q181) at the C-terminal end of the ACL interacts with the h2i β-hairpin of the C-terminal AAA+ domain (Figure 4(c)). This glutamine:h2i interaction and the enhanced unwinding rate observed for *Mt*MCM Q181A are mutually consistent with an interaction that restricts interdomain movement and thus regulates helicase activity [60]. When the interaction is removed by mutation of the glutamine to alanine, the helicase activity is no longer regulated and can occur at an increased rate.

Additional roles for the ACL in the regulation of MCM catalyzed helicase activity are found in the observation that *Mt*MCM E182R mutation results in a 7-fold decrease in unwinding rate relative to wild-type *Mt*MCM for unwinding of forked DNA substrates [60]. In the *Sso-Pf*MCM structure, the homologous glutamate residue (*Sso-Pf*MCM E199) interacts with an arginine residue of an adjacent subunit, *Sso-Pf*MCM R226 [44] (Figure 4(c)). Thus, the *Mt*MCM E182R mutation would disrupt a salt bridge between adjacent subunits and likely increase the distance between subunits by electrostatic repulsion. Furthermore, the ACL is observed to project towards the AAA+ ps1β motif of an adjacent subunit, which could be disrupted in the *Mt*MCM E182R mutant due to an altered intersubunit distance. The proximity of the ACL to the h2i of the same subunit and the ps1β of a neighboring subunit has previously been shown by Double Electron-Electron Resonance (DEER) spectroscopy, with the ACL:ps1β distance estimated to be approximately 30 Å [59]. While the role of the ACL:ps1β interaction is not clear, it may function like the ACL glutamine:h2i interaction to stabilize the ps1β and regulate MCM helicase activity. Thus, multiple roles have emerged for the ACL, including regulation of helicase activity, maintenance of intersubunit interfaces, and intrasubunit interactions.

9. Concluding Comments

Much like any machine, the MCM helicase requires specific events to occur in a precise order to unwind dsDNA. The identification of the essential events and their timing is complicated in the eukaryotic Mcm2–7 complex because the subunits do not equivalently interact, and they also interact with Cdc45 and the subunits of the GINS complex. For this reason, archaeal MCM proteins have emerged as useful models for elucidating the essential features of the complex interaction network present in MCM hexamers. A thorough understanding of the many interactions within the replisome is essential to understand the replication fork and DNA replication. Atomic resolution crystal structures will continue to reveal mechanistically important conformational states of the MCM complex, including how MCM hexamers specifically respond to binding different nucleotides, oligonucleotides, and protein interaction partners. As new information becomes available, both for the structural/biochemical details of MCM function and for disease-associated mutations, the translation of basic science information will undoubtedly aid in the treatment of primary disease states.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported in part by ALSAC and Grant R01GM098771 (to Eric J. Enemark) from NIGMS.

References

[1] M. L. Bochman and A. Schwacha, "The Mcm2-7 complex has in vitro helicase activity," *Molecular Cell*, vol. 31, no. 2, pp. 287–293, 2008.

[2] I. Ilves, T. Petojevic, J. J. Pesavento, and M. R. Botchan, "Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins," *Molecular Cell*, vol. 37, no. 2, pp. 247–258, 2010.

[3] Y. V. Fu, H. Yardimci, D. T. Long et al., "Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase," *Cell*, vol. 146, no. 6, pp. 931–941, 2011.

[4] D. Remus, F. Beuron, G. Tolun, J. D. Griffith, E. P. Morris, and J. F. X. Diffley, "Concerted loading of Mcm2–7 double hexamers around DNA during DNA replication origin licensing," *Cell*, vol. 139, no. 4, pp. 719–730, 2009.

[5] C. Evrin, P. Clarke, J. Zech et al., "A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 48, pp. 20240–20245, 2009.

[6] T. Masuda, S. Mimura, and H. Takisawa, "CDK- and Cdc45-dependent priming of the MCM complex on chromatin during S-phase in *Xenopus* egg extracts: possible activation of MCM helicase by association with Cdc45," *Genes to Cells*, vol. 8, no. 2, pp. 145–161, 2003.

[7] J. A. Tercero, K. Labib, and J. F. X. Diffley, "DNA synthesis at individual replication forks requires the essential initiation factor Cdc45p," *The EMBO Journal*, vol. 19, no. 9, pp. 2082–2093, 2000.

[8] B. Hopwood and S. Dalton, "Cdc45p assembles into a complex with Cdc46p/Mcm5p, is required for minichromosome maintenance, and is essential for chromosomal DNA replication," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 22, pp. 12309–12314, 1996.

[9] Y. Kubota, Y. Takase, Y. Komori et al., "A novel ring-like complex of *Xenopus* proteins essential for the initiation of DNA replication," *Genes and Development*, vol. 17, no. 9, pp. 1141–1152, 2003.

[10] A. Gambus, R. C. Jones, A. Sanchez-Diaz et al., "GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks," *Nature Cell Biology*, vol. 8, no. 4, pp. 358–366, 2006.

[11] K. Labib and A. Gambus, "A key role for the GINS complex at DNA replication forks," *Trends in Cell Biology*, vol. 17, no. 6, pp. 271–278, 2007.

[12] S. E. Moyer, P. W. Lewis, and M. R. Botchan, "Isolation of the Cdc45/Mcm2–7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 27, pp. 10236–10241, 2006.

[13] T. Yoshimochi, R. Fujikane, M. Kawanami, F. Matsunaga, and Y. Ishino, "The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity," *The Journal of Biological Chemistry*, vol. 283, no. 3, pp. 1601–1609, 2008.

[14] H. Yardimci, A. B. Loveland, S. Habuchi, A. M. van Oijen, and J. C. Walter, "Uncoupling of sister replisomes during eukaryotic DNA replication," *Molecular Cell*, vol. 40, no. 5, pp. 834–840, 2010.

[15] G. T. Haugland, C. R. Rollor, N.-K. Birkeland, and Z. Kelman, "Biochemical characterization of the minichromosome maintenance protein from the archaeon Thermoplasma acidophilum," *Extremophiles*, vol. 13, no. 1, pp. 81–88, 2009.

[16] Y. Ishimi, "A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex," *The Journal of Biological Chemistry*, vol. 272, no. 39, pp. 24508–24513, 1997.

[17] Z. Kelman, J.-K. Lee, and J. Hurwitz, "The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* ΔH contains DNA helicase activity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 26, pp. 14783–14788, 1999.

[18] J.-K. Lee and J. Hurwitz, "Isolation and characterization of various complexes of the minichromosome maintenance proteins of *Schizosaccharomyces pombe*," *The Journal of Biological Chemistry*, vol. 275, no. 25, pp. 18871–18878, 2000.

[19] G. T. Maine, P. Sinha, and B. W. Tye, "Mutants of *S. cerevisiae* defective in the maintenance of minichromosomes," *Genetics*, vol. 106, no. 3, pp. 365–385, 1984.

[20] E. V. Koonin, "A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication," *Nucleic Acids Research*, vol. 21, no. 11, pp. 2541–2547, 1993.

[21] M. Lutzmann, D. Maiorano, and M. Méchali, "Identification of full genes and proteins of MCM9, a novel, vertebrate-specific member of the MCM2–8 protein family," *Gene*, vol. 362, no. 1-2, pp. 51–56, 2005.

[22] M. L. Bochman and A. Schwacha, "Differences in the single-stranded DNA binding activities of MCM2-7 and MCM467: MCM2 and MCM5 define a slow ATP-dependent step," *The Journal of Biological Chemistry*, vol. 282, no. 46, pp. 33795–33804, 2007.

[23] A. Costa, I. Ilves, N. Tamberg et al., "The structural basis for MCM2-7 helicase activation by GINS and Cdc45," *Nature Structural and Molecular Biology*, vol. 18, no. 4, pp. 471–479, 2011.

[24] A. Costa, L. Renault, P. Swuec et al., "DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome," *Elife*, vol. 3, Article ID e03273, 2014.

[25] J. P. J. Chong, M. K. Hayashi, M. N. Simon, R.-M. Xu, and B. Stillman, "A double-hexamer archaeal minichromosome maintenance protein is an ATP- dependent DNA helicase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 4, pp. 1530–1535, 2000.

[26] D. F. Shechter, C. Y. Ying, and J. Gautier, "The intrinsic DNA helicase activity of *Methanobacterium thermoautotrophicum* delta H minichromosome maintenance protein," *The Journal of Biological Chemistry*, vol. 275, no. 20, pp. 15049–15059, 2000.

[27] F. Sarmiento, F. Long, I. Cann, and W. B. Whitman, "Diversity of the DNA replication system in the archaea domain," *Archaea*, vol. 2014, Article ID 675946, 15 pages, 2014.

[28] A. Costa, G. Van Duinen, B. Medagli et al., "Cryo-electron microscopy reveals a novel DNA-binding site on the MCM helicase," *The EMBO Journal*, vol. 27, no. 16, pp. 2250–2258, 2008.

[29] X. Yu, M. S. VanLoock, A. Poplawski et al., "The *Methanobacterium thermoautotrophicum* MCM protein can form heptameric rings," *EMBO Reports*, vol. 3, no. 8, pp. 792–797, 2002.

[30] T. Pape, H. Meka, S. Chen, G. Vicentini, M. van Heel, and S. Onesti, "Hexameric ring structure of the full-length archael MCM protein complex," *EMBO Reports*, vol. 4, no. 11, pp. 1079–1083, 2003.

[31] Y. Gómez-Llorente, R. J. Fletcher, X. S. Chen, J. M. Carazo, and C. S. Martin, "Polymorphism and double hexamer structure in the archaeal minichromosome maintenance (MCM) helicase from Methanobacterium thermoautotrophicum," *The Journal of Biological Chemistry*, vol. 280, no. 49, pp. 40909–40915, 2005.

[32] E. R. Barry, A. T. McGeoch, Z. Kelman, and S. D. Bell, "Archaeal MCM has separable processivity, substrate choice and helicase domains," *Nucleic Acids Research*, vol. 35, no. 3, pp. 988–998, 2007.

[33] A. F. Neuwald, L. Aravind, J. L. Spouge, and E. V. Koonin, "AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes," *Genome Research*, vol. 9, no. 1, pp. 27–43, 1999.

[34] T. Ogura and A. J. Wilkinson, "AAA^+ superfamily ATPases: common structure-diverse function," *Genes to Cells*, vol. 6, no. 7, pp. 575–597, 2001.

[35] L. M. Iyer, D. D. Leipe, E. V. Koonin, and L. Aravind, "Evolutionary history and higher order classification of AAA+ ATPases," *Journal of Structural Biology*, vol. 146, no. 1-2, pp. 11–31, 2004.

[36] P. I. Hanson and S. W. Whiteheart, "AAA+ proteins: have engine, will work," *Nature Reviews Molecular Cell Biology*, vol. 6, no. 7, pp. 519–529, 2005.

[37] L. Aravind and E. V. Koonin, "DNA-binding proteins and evolution of transcription regulation in the archaea," *Nucleic Acids Research*, vol. 27, no. 23, pp. 4658–4670, 1999.

[38] N. Atanassova and I. Grainge, "Biochemical characterization of the minichromosome maintenance (MCM) protein of the crenarchaeote *Aeropyrum pernix* and its interactions with the origin recognition complex (ORC) proteins," *Biochemistry*, vol. 47, no. 50, pp. 13362–13370, 2008.

[39] R. J. Fletcher, B. E. Bishop, R. P. Leon, R. A. Sclafani, C. M. Ogata, and X. S. Chen, "The structure and function of MCM from archaeal *M. thermoautotrophicum*," *Nature Structural Biology*, vol. 10, no. 3, pp. 160–167, 2003.

[40] W. Liu, B. Pucci, M. Rossi, F. M. Pisani, and R. Ladenstein, "Structural analysis of the *Sulfolobus solfataricus* MCM protein N-terminal domain," *Nucleic Acids Research*, vol. 36, no. 10, pp. 3235–3243, 2008.

[41] A. S. Brewster, G. Wang, X. Yu et al., "Crystal structure of a near-full-length archaeal MCM: functional insights for an AAA+ hexameric helicase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 51, pp. 20191–20196, 2008.

[42] B. Bae, Y. H. Chen, A. Costa et al., "Insights into the architecture of the replicative helicase from the structure of an archaeal MCM homolog," *Structure*, vol. 17, no. 2, pp. 211–222, 2009.

[43] C. A. Froelich, S. Kang, L. B. Epling, S. P. Bell, and E. J. Enemark, "A conserved MCM single-stranded DNA binding element is essential for replication initiation," *eLife*, vol. 2014, no. 3, Article ID e01993, 2014.

[44] J. M. Miller, B. T. Arachea, L. B. Epling, and E. J. Enemark, "Analysis of the crystal structure of an active MCM hexamer," *eLife*, vol. 3, Article ID e03433, 2014.

[45] Y.-J. Chen, X. Yu, R. Kasiviswanathan, J.-H. Shin, Z. Kelman, and E. H. Egelman, "Structural polymorphism of *Methanothermobacter thermautotrophicus* MCM," *Journal of Molecular Biology*, vol. 346, no. 2, pp. 389–394, 2005.

[46] R. J. Fletcher and X. S. Chen, "Biochemical activities of the BOB1 mutant in Methanobacterium thermoautotrophicum MCM," *Biochemistry*, vol. 45, no. 2, pp. 462–467, 2006.

[47] S. Krueger, J.-H. Shin, S. Raghunandan, J. E. Curtis, and Z. Kelman, "Atomistic ensemble modeling and small-angle neutron scattering of intrinsically disordered protein complexes: applied to minichromosome maintenance protein," *Biophysical Journal*, vol. 101, no. 12, pp. 2999–3007, 2011.

[48] S. Krueger, J.-H. Shin, J. E. Curtis, K. A. Rubinson, and Z. Kelman, "The solution structure of full-length dodecameric MCM by SANS and molecular modeling," *Proteins*, vol. 82, no. 10, pp. 2364–2374, 2014.

[49] R. Kasiviswanathan, J.-H. Shin, E. Melamud, and Z. Kelman, "Biochemical characterization of the Methanothermobacter thermautotrophicus minichromosome maintenance (MCM) helicase N-terminal domains," *The Journal of Biological Chemistry*, vol. 279, no. 27, pp. 28358–28366, 2004.

[50] B. W. Graham, G. D. Schauer, S. H. Leuba, and M. A. Trakselis, "Steric exclusion and wrapping of the excluded DNA strand occurs along discrete external binding paths during MCM helicase unwinding," *Nucleic Acids Research*, vol. 39, no. 15, pp. 6585–6595, 2011.

[51] T. Petojevic, J. J. Pesavento, A. Costa et al., "Cdc45 (cell division cycle protein 45) guards the gate of the Eukaryote Replisome helicase stabilizing leading strand engagement," *Proceedings of the National Academy of Sciences*, vol. 112, no. 3, pp. E249–E258, 2015.

[52] E. Rothenberg, M. A. Trakselis, S. D. Bell, and T. Ha, "MCM forked substrate specificity involves dynamic interaction with the 5′-tail," *Journal of Biological Chemistry*, vol. 282, no. 47, pp. 34229–34234, 2007.

[53] C. F. S. Hardy, O. Dryga, S. Seematter, P. M. B. Pahl, and R. A. Sclafani, "mcm5/cdc46-bob1 bypasses the requirement for the S phase activator Cdc7p," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 7, pp. 3151–3155, 1997.

[54] E. R. Jenkinson, A. Costa, A. P. Leech, A. Patwardhan, S. Onesti, and J. P. J. Chong, "Mutations in subdomain B of the minichromosome maintenance (MCM) helicase affect DNA binding and modulate conformational transitions," *Journal of Biological Chemistry*, vol. 284, no. 9, pp. 5654–5661, 2009.

[55] A. Poplawski, B. Grabowski, S. E. Long, and Z. Kelman, "The zinc finger domain of the archaeal minichromosome maintenance protein is required for helicase activity," *The Journal of Biological Chemistry*, vol. 276, no. 52, pp. 49371–49377, 2001.

[56] B. K. Tye, "MCM proteins in DNA replication," *Annual Review of Biochemistry*, vol. 68, pp. 649–686, 1999.

[57] A. G. Murzin, "OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences," *The EMBO Journal*, vol. 12, no. 3, pp. 861–867, 1993.

[58] N. Sakakibara, R. Kasiviswanathan, and Z. Kelman, "Mutational analysis of conserved aspartic acid residues in the *Methanothermobacter thermautotrophicus* MCM helicase," *Extremophiles*, vol. 15, no. 2, pp. 245–252, 2011.

[59] E. R. Barry, J. E. Lovett, A. Costa, S. M. Lea, and S. D. Bell, "Intersubunit allosteric communication mediated by a conserved loop in the MCM helicase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 4, pp. 1051–1056, 2009.

[60] N. Sakakibara, R. Kasiviswanathan, E. Melamud, M. Han, F. P. Schwarz, and Z. Kelman, "Coupling of DNA binding and helicase activity is mediated by a conserved loop in the MCM protein," *Nucleic Acids Research*, vol. 36, no. 4, pp. 1309–1320, 2008.

[61] N. Shima, A. Alcaraz, I. Liachko et al., "A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice," *Nature Genetics*, vol. 39, no. 1, pp. 93–98, 2007.

[62] V. G. Gorgoulis, L.-V. F. Vassiliou, P. Karakaidos et al., "Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions," *Nature*, vol. 434, no. 7035, pp. 907–913, 2005.

[63] C. Lengauer, K. W. Kinzler, and B. Vogelstein, "Genetic instabilities in human cancers," *Nature*, vol. 396, no. 6712, pp. 643–649, 1998.

[64] S. Negrini, V. G. Gorgoulis, and T. D. Halazonetis, "Genomic instability—an evolving hallmark of cancer," *Nature Reviews Molecular Cell Biology*, vol. 11, no. 3, pp. 220–228, 2010.

[65] M. J. Davey, C. Indiani, and M. O'Donnell, "Reconstitution of the Mcm2-7p heterohexamer, subunit arrangement, and ATP site architecture," *The Journal of Biological Chemistry*, vol. 278, no. 7, pp. 4491–4499, 2003.

[66] M. J. Moreau, A. T. McGeoch, A. R. Lowe, L. S. Itzhaki, and S. D. Bell, "ATPase site architecture and helicase mechanism of an archaeal MCM," *Molecular Cell*, vol. 28, no. 2, pp. 304–314, 2007.

[67] N. Sakakibara, F. P. Schwarz, and Z. Kelman, "ATP hydrolysis and DNA binding confer thermostability on the MCM helicase," *Biochemistry*, vol. 48, no. 11, pp. 2330–2339, 2009.

[68] Z. Wei, C. Liu, X. Wu et al., "Characterization and structure determination of the cdt1 binding domain of human minichromosome maintenance (Mcm) 6," *The Journal of Biological Chemistry*, vol. 285, no. 17, pp. 12469–12473, 2010.

[69] P. D. Robertson, B. Chagot, W. J. Chazin, and B. F. Eichman, "Solution NMR structure of the C-terminal DNA binding domain of Mcm10 reveals a conserved MCM motif," *The Journal of Biological Chemistry*, vol. 285, no. 30, pp. 22942–22949, 2010.

[70] P. D. Robertson, E. M. Warren, H. Zhang et al., "Domain architecture and biochemical characterization of vertebrate Mcm10," *The Journal of Biological Chemistry*, vol. 283, no. 6, pp. 3338–3348, 2008.

[71] C. Wiedemann, P. Bellstedt, C. Herbst, M. Görlach, and R. Ramachandran, "An approach to sequential NMR assignments of proteins: application to chemical shift restraint-based structure prediction," *Journal of Biomolecular NMR*, vol. 59, no. 4, pp. 211–217, 2014.

[72] E. R. Jenkinson and J. P. J. Chong, "Minichromosome maintenance helicase activity is controlled by N- and C-terminal motifs and requires the ATPase domain helix-2 insert," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 20, pp. 7613–7618, 2006.

[73] A. T. McGeoch, M. A. Trakselis, R. A. Laskey, and S. D. Bell, "Organization of the archaeal MCM complex on DNA and implications for the helicase mechanism," *Nature Structural and Molecular Biology*, vol. 12, no. 9, pp. 756–762, 2005.

[74] B. Pucci, M. De Felice, M. Rossi, S. Onesti, and F. M. Pisani, "Amino acids of the Sulfolobus solfataricus mini-chromosome maintenance-like DNA helicase involved in DNA binding/remodeling," *Journal of Biological Chemistry*, vol. 279, no. 47, pp. 49222–49228, 2004.

[75] M. Pan, T. J. Santangelo, Z. Li, J. N. Reeve, and Z. Kelman, "Thermococcus kodakarensis encodes three MCM homologs but only one is essential," *Nucleic Acids Research*, vol. 39, no. 22, pp. 9671–9680, 2011.

[76] X. Zhang and D. B. Wigley, "The 'glutamate switch' provides a link between ATPase activity and ligand binding in AAA+ proteins," *Nature Structural and Molecular Biology*, vol. 15, no. 11, pp. 1223–1227, 2008.

[77] E. J. Enemark and L. Joshua-Tor, "Mechanism of DNA translocation in a replicative hexameric helicase," *Nature*, vol. 442, no. 7100, pp. 270–275, 2006.

[78] D. Gai, R. Zhao, D. Li, C. V. Finkielstein, and X. S. Chen, "Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen," *Cell*, vol. 119, no. 1, pp. 47–60, 2004.

[79] J. P. Erzberger and J. M. Berger, "Evolutionary relationships and structural mechanisms of AAA+ proteins," *Annual Review of Biophysics and Biomolecular Structure*, vol. 35, no. 1, pp. 93–114, 2006.

[80] R. J. Fletcher, J. Shen, Y. Gómez-Llorente, C. San Martín, J. M. Carazo, and X. S. Chen, "Double hexamer disruption and biochemical activities of Methanobacterium thermoautotrophicum MCM," *The Journal of Biological Chemistry*, vol. 280, no. 51, pp. 42405–42410, 2005.

[81] R. A. Sclafani, R. J. Fletcher, and X. S. Chen, "Two heads are better than one: regulation of DNA replication by hexameric helicases," *Genes and Development*, vol. 18, no. 17, pp. 2039–2045, 2004.

[82] O. Itsathitphaisarn, R. A. Wing, W. K. Eliason, J. Wang, and T. A. Steitz, "The hexameric helicase DnaB adopts a nonplanar conformation during translocation," *Cell*, vol. 151, no. 2, pp. 267–277, 2012.

[83] N. D. Thomsen and J. M. Berger, "Running in reverse: the structural basis for translocation polarity in hexameric helicases," *Cell*, vol. 139, no. 3, pp. 523–534, 2009.

[84] R. P. Leon, M. Tecklenburg, and R. A. Sclafani, "Functional conservation of β-hairpin DNA binding domains in the Mcm protein of *Methanobacterium thermoautotrophicum* and the Mcm5 protein of *Saccharomyces cerevisiae*," *Genetics*, vol. 179, no. 4, pp. 1757–1768, 2008.

[85] N. D. Thomsen and J. M. Berger, "Structural frameworks for considering microbial protein- and nucleic acid-dependent motor ATPases," *Molecular Microbiology*, vol. 69, no. 5, pp. 1071–1090, 2008.

[86] Z. Yu, D. Feng, and C. Liang, "Pairwise interactions of the six human MCM protein subunits," *Journal of Molecular Biology*, vol. 340, no. 5, pp. 1197–1206, 2004.

[87] E. F. Pettersen, T. D. Goddard, C. C. Huang et al., "UCSF Chimera—a visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.

[88] The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.

[89] J. M. Berger, "SnapShot: nucleic acid helicases and translocases," *Cell*, vol. 134, no. 5, pp. 888–888.e1, 2008.

Hindawi Publishing Corporation
Archaea
Volume 2015, Article ID 978632, 6 pages
http://dx.doi.org/10.1155/2015/978632

Research Article

A Novel Highly Thermostable Multifunctional Beta-Glycosidase from Crenarchaeon *Acidilobus saccharovorans*

Vadim M. Gumerov, Andrey L. Rakitin, Andrey V. Mardanov, and Nikolai V. Ravin

Centre "Bioengineering", Russian Academy of Sciences, Moscow 117312, Russia

Correspondence should be addressed to Nikolai V. Ravin; nravin@biengi.ac.ru

Received 10 March 2015; Revised 25 June 2015; Accepted 5 July 2015

Academic Editor: Frédéric Pecorari

We expressed a putative β-galactosidase Asac_1390 from hyperthermophilic crenarchaeon *Acidilobus saccharovorans* in *Escherichia coli* and purified the recombinant enzyme. Asac_1390 is composed of 490 amino acid residues and showed high sequence similarity to family 1 glycoside hydrolases from various thermophilic Crenarchaeota. The maximum activity was observed at pH 6.0 and 93°C. The half-life of the enzyme at 90°C was about 7 hours. Asac_1390 displayed high tolerance to glucose and exhibits hydrolytic activity towards cellobiose and various aryl glucosides. The hydrolytic activity with *p*-nitrophenyl (pNP) substrates followed the order pNP-β-D-galactopyranoside (328 U mg^{-1}), pNP-β-D-glucopyranoside (246 U mg^{-1}), pNP-β-D-xylopyranoside (72 U mg^{-1}), and pNP-β-D-mannopyranoside (28 U mg^{-1}). Thus the enzyme was actually a multifunctional β-glycosidase. Therefore, the utilization of Asac_1390 may contribute to facilitating the efficient degradation of lignocellulosic biomass and help enhance bioconversion processes.

1. Introduction

Lignocellulosic biomass, predominantly composed of cellulose, hemicellulose, and lignin, is the most abundant renewable resource on earth, and its degradation to soluble sugars is a key issue for the production of biobased chemicals and biofuels. Cellulose and hemicellulose consist of polysaccharides, and because of their recalcitrance, the enzymatic conversion of these substrates into simple sugars requires many steps.

Microorganisms capable of degrading cellulose possess three types of enzymes which act synergistically [1, 2]: (1) endoglucanases (EC 3.2.1.4), (2) exoglucanases also known as cellobiohydrolases (EC 3.2.1.91), and (3) β-glucosidases (EC 3.2.1.21). Endoglucanases randomly cleave the internal bonds of cellulosic materials, releasing oligosaccharides of various lengths and thus generating new ends of the polysaccharide chains. The cellobiohydrolases act progressively on these chain ends generating short cellooligosaccharides or cellobiose. β-Glucosidases hydrolyze cellooligosaccharides or cellobiose to glucose units [3, 4]. Hemicellulose is broken down into soluble xylose or other types of monosaccharides by hemicellulases such as endo-1,4-β-xylanase (EC 3.2.1.8), β-xylosidase (EC 3.2.1.37), α-L-arabinofuranosidase (EC 3.2.1.55), endo-1,4-β-mannanase (EC 3.2.1.78), β-mannosidase (EC 3.2.1.25), and β-galactosidase (EC 3.2.1.23) [5].

The presence of β-glucosidases is very important in cellulose hydrolysis process because they perform a rate-limiting step by preventing the accumulation of cellobiose, which inhibits the activities of most endo- and exoglucanases [6]. Moreover, most β-glucosidases are inhibited by glucose [7]. Therefore, β-glucosidases, with high specific activity and glucose tolerance, could improve the efficiency of cellulolytic enzyme complexes.

Thermostable enzymes have several advantages in lignocellulose degradation processes because they are able to withstand rough reaction conditions at elevated temperature and allow elongated hydrolysis time due to higher stability and reduced risk of contamination. High temperature promotes high activity of these enzymes and increases the solubility of substrates in the aqueous phase. Although a number of thermostable β-glucosidases have been identified [8–16], there are only several examples of glucose-tolerant thermostable β-glucosidases [13, 14].

Acidilobus saccharovorans 345-15^T (DSM 16705) is an anaerobic, organotrophic, thermoacidophilic crenarchaeon with a pH range from 2.5 to 5.8 (optimum at pH 3.5 to 4) and a temperature range from 60 to 90°C (optimum at 80 to 85°C), isolated from an acidic hot spring of Uzon Caldera, Kamchatka, Russia [17]. The complete genome of *A. saccharovorans* was sequenced [18], and the presence of genes for extra- and intracellular glycoside hydrolases correlates with the growth of *A. saccharovorans* on carbohydrates. Particularly, the gene Asac_1390 encoding putative β-galactosidase (EC 3.2.1.23) was identified [18]. In this paper we report the cloning, expression, and biochemical characterization of this enzyme appearing to be a multifunctional β-glycosidase.

2. Materials and Methods

2.1. Cloning of A. saccharovorans Gene Asac_1390. The β-galactosidase gene Asac_1390 was amplified from *A. saccharovorans* genomic DNA by PCR using primers F2_NcoI (TTCCATGGCAGTTACCTTCCCAAA) and R2_BglII (5′-TTAGATCTGGATCTACCAGGCGCT-3′); the PCR products were digested with NcoI and BglII and inserted into pQE60 (Qiagen) at NcoI and BglII sites, yielding the plasmid pQE60_Asac1390.

2.2. Expression and Purification of Recombinant Enzyme Asac_1390. Plasmid pQE60_Asac1390 was transformed into *Escherichia coli* strain DLT1270 carrying plasmid pRARE2 (Novagen). Recombinant strain was grown at 37°C in Luria-Bertani medium (LB) supplemented with ampicillin and induced to express recombinant xylanases by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1.0 mM at OD_{600} approximately 0.6 and incubated further at 37°C for 20 hours. The grown cells were harvested by centrifugation at 5,000 g for 20 min at 4°C, washed with 0.1 M sodium phosphate buffer, pH 7.0, and resuspended in 20 mL of the same buffer with lysozyme (1 mg/mL). The cells were incubated 30 min at 4°C and then disrupted by sonication. The insoluble debris was removed by centrifugation at 5,000 g for 20 min at 4°C. The supernatant was then incubated in a water bath for 30 min at 75°C and then for 30 min at 85°C (double heat treatment) and cooled on ice. Denatured proteins of *E. coli* were removed by centrifugation at 12,000 g for 20 min at 4°C. The protein sample was dialysed against 25 mM phosphate buffer (pH 7.0) at 4°C for 3 h.

The purity of the purified protein was examined by SDS-PAGE (10%) and its concentration was determined by the Bradford method using bovine serum albumin (BSA) as a standard.

2.3. Assay of β-Galactosidase Activity. The β-galactosidase activity was analyzed using *o*-nitrophenyl-β-D-galactopyranoside (oNPGal) as substrate following modified Craven method [19]. 1.076 mL of the substrate solution (0.7 mg/mL oNPGal in 0.1 M sodium phosphate buffer, pH 7.0) was preincubated at the appropriate temperature for 5 min, and the reaction was initiated by the addition of 0.05 mL of enzyme (0.08 μg). The reaction was stopped with 0.375 mL of cold 1 M Na_2CO_3 solution. A blank, containing 0.05 mL of 0.1 M sodium phosphate buffer instead of enzyme solution, was used to correct for the thermal hydrolysis of oNPGal. The amount of released *o*-nitrophenol was measured at 420 nm. One unit (U) of enzyme activity was defined as the amount of enzyme required to liberate 1 μmol of *o*-nitrophenol per min under described conditions.

2.4. Biochemical Characterization of Recombinant Enzyme Asac_1390. To examine the effects of pH and temperature on β-galactosidase activity, pH values were varied from 3.0 to 8.0 at 80°C using 0.1 M acetate buffer (pH 3.0 to 5.0) and 0.1 M sodium phosphate buffer (pH 6.0 to 8.0). Similarly, β-galactosidase assay was done at various temperatures (50–100°C) and pH 7.0 to determine the optimum temperature for Asac_1390 activity.

The thermostability of Asac_1390 was assessed by preincubating the enzyme at 90°C in sodium phosphate buffer, pH 7.0. The samples were collected at the desired intervals and assayed for the residual β-galactosidase activity in 0.1 M sodium phosphate buffer (pH 7.0) at 50°C.

The substrate specificity of Asac_1390 was determined using oNPGal, *p*-nitrophenyl-β-D-galactopyranoside (pNPGal), *p*-nitrophenyl-β-D-glucopyranoside (pNPGlu), *p*-nitrophenyl-β-D-xylopyranoside (pNPXyl), and pNP-β-D-mannopyranoside (pNPMan). The reactions were performed with 2.21 mM of each substrate in 0.1 M sodium phosphate buffer (pH 7.0) at 93°C and the activity was measured by release of pNP (at 405 nm) and oNP (at 420 nm).

When cellobiose was used as a substrate (0.5% w/v), the amount of glucose released was determined with a Sucrose/D-Glucose/D-Fructose Kit (R-Biopharm AG, Germany) according to the manufacturer's protocol. The cellobiose hydrolysis reactions were performed in 0.1 M sodium phosphate buffer (pH 6.0) at 50°C. In this assay, 1 U of activity is defined as the amount of enzyme which is required to release 1 μmol of glucose per minute under test conditions.

Various concentrations of oNPGal, pNPGal, pNPGlu, pNPXyl, and pNPMan (from 0.277 to 2.21 mM for oNPGal, pNPGal, pNPGlu, and pNPMan and from 0.074 to 1.23 mM for pNPXyl) were used to determine kinetic parameters of Asac_1390. The reactions were performed in 0.1 M sodium phosphate buffer (pH 7.0) at 93°C. The enzyme kinetic parameters, K_m (mM), V_{max} (U/mg), k_{cat} (c^{-1}), and k_{cat}/K_m, were calculated from Hanes-Woolf plot of Michaelis-Menten equation.

The effects of glucose on β-galactosidase activity were investigated at the concentrations of glucose from 10 to 100 mM using 2.21 mM oNPGal at 80°C in 0.1 M phosphate buffer (pH 7.0). The relative activity was defined as the relative value to the maximum activity without glucose. The type of inhibition and inhibition constants for glucose were determined by fitting to Cornish-Bowden and Dixon plots [20] using various concentrations of glucose (from 0 to 400 mM) with various concentrations of oNPGal (from 0.05 to 0.5 mM) as a substrate.

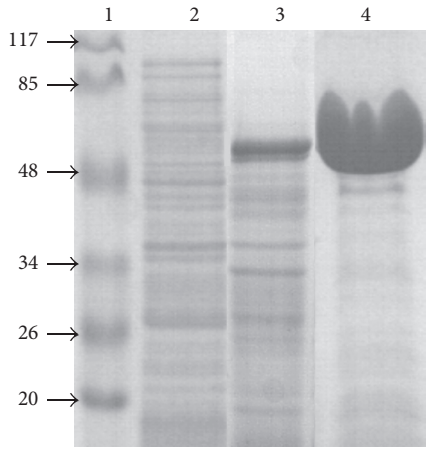


Figure 1: Expression and purification of recombinant glycosidase Asac_1390. SDS-PAGE was done using a 10.0% polyacrylamide gel; proteins were stained with Coomassie Brilliant Blue R-250. Lanes: 1—molecular weight markers (sizes are shown in kDa); 2—total protein from uninduced cells; 3—total protein from induced cells; 4—purified recombinant Asac_1390.

3. Results and Discussion

3.1. Gene Cloning and Enzyme Expression. The amino acid sequence of a putative β-galactosidase Asac_1390 from *A. saccharovorans* exhibited 54–71% identities with the glycoside hydrolases from the thermophilic archaea of the genera *Caldivirga*, *Sulfolobus*, *Vulcanisaeta*, *Thermoproteus*, *Ignisphaera*, *Thermoplasma*, *Thermosphaera*, *Picrophilus*, *Thermococcus*, and *Pyrococcus*, annotated as β-galactosidases or β-glucosidases. A Blastp search of the amino acid sequence of Asac_1390 suggested that residues 1–46 contain a signature typical of glycosyl hydrolase (GH) family 1 (Pfam00232). Particularly, Asac_1390 is 63% identical to β-glycosidase from *Sulfolobus acidocaldarius* that was found to exhibit activities toward β-glucosides, β-galactosides, and β-fucosides [15]. The results suggest that the enzyme could have broad substrate specificity. Asac_1390 does not have a predicted secretion signal suggesting an intracellular action of this enzyme.

Asac_1390 gene with the same sequence as that reported in GenBank (CP001742) was cloned and expressed in *E. coli*. The recombinant enzyme was expressed with a yield of about 30% of the total soluble protein and was purified from crude *E. coli* extracts via two-step heat treatment to a purity of above 95%. The purified protein appeared in SDS-PAGE analysis as a single band with a molecular mass of approximately 55 kDa (Figure 1), consistent with the calculated value of 55,521 kDa based on the 490 amino acid residues of Asac_1390.

3.2. Effects of pH and Temperature on the Enzyme Activity. The β-galactosidase activity was examined over a pH range of 3.0 to 8.0 at 80°C. Maximum activity was observed at pH 6.0 in sodium phosphate buffer. At pH 5.0 and 7.0, the activity was approximately 70% of the maximum (Figure 2(a)). The effect of temperature on enzyme activity was investigated in 0.1 M sodium phosphate buffer, pH 7.0 (Figure 2(b)). The maximum activity was recorded at 93°C; at temperature of 100°C the activity was about 80% of the maximum.

Table 1: Hydrolytic activity of Asac_1390 with different substrates.

Substrate	Specific activity (U mg^{-1})
oNPGal	325 ± 10
pNPGal	328 ± 10
pNPGlu	246 ± 6
pNPXyl	72 ± 1
pNPMan	28 ± 2

The reaction was carried out in 100 mM sodium phosphate buffer (pH 7.0) at 93°C. Data represent the means of three separate experiments.

The thermal stability of Asac_1390 was examined by measuring the activity over time at 90°C. The samples were withdrawn at various time intervals and assayed in 0.1 M sodium phosphate buffer (pH 7.0) at 50°C for 20 min (Figure 3). The Asac_1390 enzyme appeared to be extremely thermostable; the half-life of the enzyme at 90°C was about 7 hours.

These values of the temperature and pH optima of Asac_1390 are typical for β-galactosidases and β-glucosidases from hyperthermophilic archaea, including enzymes from *Sulfolobus solfataricus* (95°C and pH 6.5, [21]), *Pyrococcus furiosus* (100°C and pH 5.0, [22]), *S. acidocaldarius* (90°C and pH 5.5, [15]), and *Thermococcus kodakarensis* (100°C and pH 6.5, [23]). In terms of thermal inactivation, Asac_1390 is one of the most thermostable β-glycosidases with the half-life of 7 h at 90°C. It showed a higher stability than β-glycosidases from *S. solfataricus* and *S. acidocaldarius*, but a lower stability than those from *P. furiosus* (85 h at 100°C) and *T. kodakaraensis* (18 h at 90°C).

3.3. Effect of Glucose on the Activity of Asac_1390. The effects of glucose on β-galactosidase activity during oNPGal hydrolysis were investigated at various concentrations of D-glucose. The addition of 10, 50, and 100 mM of sugar reduced the activity of Asac_1390 to 80%, 76%, and 65%, respectively. Thus, glucose had little effect on hydrolytic activity.

The Cornish-Bowden and Dixon plots demonstrated a mixed type of inhibition of Asac_1390 by glucose. The dissociation constant of the EIS complex ($K_i' = 158$ mM) was significantly lower than that for EI complex ($K_i = 500$ mM). Glucose was reported to be a competitive inhibitor of β-glycosidase from *S. solfataricus* [21] with the inhibition constant K_i of 96 mM, while it has little effect on the β-glucosidase from *Pyrococcus furiosus* with an apparent K_i of 300 mM [22].

3.4. Substrate Specificity and Kinetics of Asac_1390. The hydrolytic activity of Asac_1390 was investigated with various aryl glycosides (Table 1). For the pNP substrates the highest activity was observed for pNPGal, followed by pNPGlu and pNPXyl. The hydrolysis of pNPMan was the least effective. The activity of the enzyme for oNPGal was about the same as for pNPGal, indicating that the enzyme equally and efficiently hydrolyzed β-1-2 and β-1-4 linkages. The Michaelis-Menten

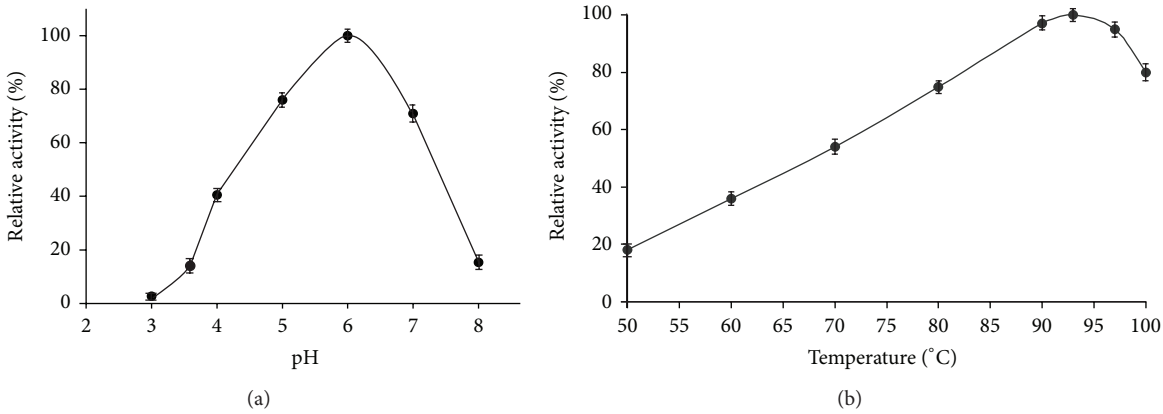


Figure 2: Effects of pH (a) and temperature (b) on the β-galactosidase activity of recombinant Asac_1390. Data represent the means of three experiments and error bars represent standard deviation.

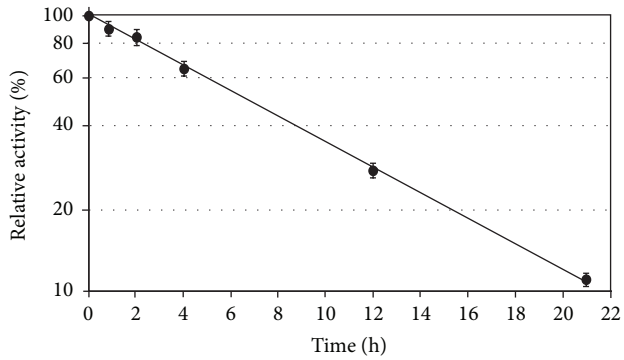


Figure 3: Thermal inactivation of recombinant Asac_1390. Purified Asac_1390 was preincubated at 90°C followed by β-galactosidase assay. Data represent the means of three experiments and error bars represent standard deviation.

Table 2: Kinetic parameters of Asac_1390.

Substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} mM^{-1})	V_{max} (U mg^{-1})
oNPGal	2.0	434.0	217.0	468.9
pNPGal	2.9	461.8	159.8	499.3
pNPGlu	0.24	318.5	1327.1	343.9
pNPXyl	4.65	114.6	24.6	123.5
pNPMan	2.4	58.2	24.2	62.8

The reaction was carried out in 100 mM sodium phosphate buffer (pH 7.0) at 93°C. Data represent the means of three separate experiments.

constants (K_m), turnover numbers (k_{cat}), and catalytic efficiencies (k_{cat}/K_m) for oNPGal, pNPGal, pNPGlu, pNPXyl, and pNPMan are presented in Table 2. The value of k_{cat}/K_m for pNPGlu was much higher than that obtained with pNPGal, indicating that Asac_1390 was not a β-galactosidase as it was annotated, but rather a β-glucosidase.

The Asac_1390 enzyme has high specific β-galactosidase and β-glucosidase activity (246–328 U mg^{-1} on different substrates), a high affinity with substrate (K_m = 0.24 mM for pNPGlu), and high catalytic activity (k_{cat}/K_m = 1327 s^{-1} mM^{-1} for pNPGlu). These values are among the highest among archaeal enzymes of this class. Some more active β-galactosidases from thermophilic bacteria are known, but usually they are less thermostable (e.g., 30,400 U mg^{-1} and half-life of 1.5 h at 90°C in case of β-glucosidase *Thermotoga petrophila* [24]).

Taking into account that some microbial GH1 family β-glycosidases had low specific activity for cellobiose, we investigated the hydrolytic activity of Asac_1390 with this natural β-glucosidase substrate. The observed activities of Asac_1390 with cellobiose and pNPGlu at 50°C and pH 6.0 were 68 U mg^{-1} and 53 U mg^{-1}, respectively. Typically, glucose-tolerant microbial β-glucosidases have considerably lower specific activity for cellobiose than for pNPGlu, while the known exceptions (e.g., β-glucosidase from *Thermoanaerobacterium thermosaccharolyticum*) are less thermostable than Asac_1390 [14].

The broad substrate specificity of Asac_1390 is an interesting and important feature of this enzyme. Although initially annotated as β-galactosidase, this enzyme exhibits hydrolytic activity towards cellobiose and various aryl glucosides: high activity with pNPGal and pNPGlu, followed with pNPXyl and pNPMan. Such broad substrate specificity β-glycosidases are known among Archaea [25], including *S. acidocaldarius* and *S. solfataricus* [15, 26]. Analysis of recently determined three-dimensional structure of Asac_1390 [27] could help to reveal molecular features defining substrate specificity of the enzyme.

Multifunctionality of Asac_1390 makes it very promising for application in enzymatic hydrolysis of lignocellulose biomass. β-Xylosidases and β-glucosidases are responsible for the last steps of the hydrolysis of xylan and cellulose: cleavage of xylobiose to xylose [28] and cellobiose to glucose [1]. The activity of β-mannosidase is also useful since this enzyme participates in the production of fermentable sugar from another component of hemicellulosic materials, mannan [29]. For instance, *Trichoderma reesei* is a well-known cellulase-overproducing filamentous fungus which secretes

several cellulolytic enzymes. However, β-glucosidase activity in *T. reesei* is partly mycelium-bound and obviously limits the enzyme performance in commercial *T. reesei* preparations [30]. Supplementation of enzyme complex produced by *T. reesei* with highly active β-glucosidase, exhibiting also β-xylosidase, β-mannosidase, and β-galactosidase hydrolytic activities, could improve efficiency of processing of lignocellulose biomass. Asac_1390 is particularly suitable for these purposes due to its resistance to inhibition by the main reaction product, glucose.

4. Conclusions

In this study, we expressed and characterized a thermostable recombinant glycosyl hydrolase Asac_1390 from *A. saccharovorans*. This enzyme is optimally active at high temperature (93°C) and pH 6.0 and is highly thermostable. Asac_1390 is a multifunctional β-glycosidase exhibiting activities of β-glucosidase, β-galactosidase, β-xylosidase, and β-mannosidase. The broad substrate specificity and resistance to inhibition by glucose make the new enzyme promising for application in enzymatic degradation of lignocellulosic materials.

Disclosure

The present address of Dr. Vadim M. Gumerov is Central European Institute of Technology, Masaryk University, Brno, Czech Republic.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the program "Molecular and cellular biology" of the Russian Academy of Sciences and by the Ministry of Education and Science of the Russian Federation (Projects 16.512.11.2234 and RFMEFI57514X0001).

References

[1] P. Béguin, "Molecular biology of cellulose degradation," *Annual Review of Microbiology*, vol. 44, pp. 219–248, 1990.

[2] T. M. Wood, S. I. McCrae, and K. M. Bhat, "The mechanism of fungal cellulase action. Synergism between enzyme components of *Penicillium pinophilum* cellulase in solubilizing hydrogen bound-ordered cellulose," *The Biochemical Journal*, vol. 260, no. 1, pp. 37–43, 1989.

[3] D. D. Y. Ryu and M. Mandels, "Cellulases: biosynthesis and applications," *Enzyme and Microbial Technology*, vol. 2, no. 2, pp. 91–102, 1980.

[4] T. M. Wood, "Properties of cellulolytic enzyme systems," *Biochemical Society Transactions*, vol. 13, no. 2, pp. 407–410, 1985.

[5] Y.-H. P. Zhang, M. E. Himmel, and J. R. Mielenz, "Outlook for cellulase improvement: screening and selection strategies," *Biotechnology Advances*, vol. 24, no. 5, pp. 452–481, 2006.

[6] Z. Xin, Q. Yinbo, and G. Peiji, "Acceleration of ethanol production from paper mill waste fiber by supplementation with β-glucosidase," *Enzyme and Microbial Technology*, vol. 15, no. 1, pp. 62–65, 1993.

[7] J. Hong, M. R. Ladisch, C. Gong, P. C. Wankat, and G. T. Tsao, "Combined product and substrate inhibition equation for cellobiase," *Biotechnology and Bioengineering*, vol. 23, no. 12, pp. 2779–2788, 1981.

[8] N. Aït, N. Creuzet, and J. Cattanéo, "Characterization and purification of thermostable β-glucosidase from *Clostridium thermocellum*," *Biochemical and Biophysical Research Communications*, vol. 90, no. 2, pp. 537–546, 1979.

[9] B. Klippel and G. Antranikian, "Lignocellulose converting enzymes from thermophiles," in *Extremophiles Handbook*, K. Horikoshi, Ed., pp. 444–474, Springer, Tokyo, Japan, 2011.

[10] X.-H. Li, R. Bhaskar, H.-J. Yang, D. Wang, and Y.-G. Miao, "Screening and identification of new isolate: thermostable *Escherichia coli* with novel thermoalkalotolerant cellulases," *Current Microbiology*, vol. 59, no. 4, pp. 393–399, 2009.

[11] M. L. Patchett, R. M. Daniel, and H. W. Morgan, "Purification and properties of a stable β-glucosidase from an extremely thermophilic anaerobic bacterium," *Biochemical Journal*, vol. 243, no. 3, pp. 779–787, 1987.

[12] W. G. B. Voorhorst, R. I. L. Eggen, E. J. Luesink, and W. M. de Vos, "Characterization of the celB gene coding for β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its expression and site-directed mutation in *Escherichia coli*," *Journal of Bacteriology*, vol. 177, no. 24, pp. 7105–7111, 1995.

[13] D. Jabbour, B. Klippel, and G. Antranikian, "A novel thermostable and glucose-tolerant β-glucosidase from *Fervidobacterium islandicum*," *Applied Microbiology and Biotechnology*, vol. 93, no. 5, pp. 1947–1956, 2012.

[14] J. Pei, Q. Pang, L. Zhao, S. Fan, and H. Shi, "Thermoanaerobacterium thermosaccharolyticum β-glucosidase: a glucose-tolerant enzyme with high specific activity for cellobiose," *Biotechnology for Biofuels*, vol. 5, no. 1, article 31, 2012.

[15] A.-R. Park, H.-J. Kim, J.-K. Lee, and D.-K. Oh, "Hydrolysis and transglycosylation activity of a thermostable recombinant β-glycosidase from *Sulfolobus acidocaldarius*," *Applied Biochemistry and Biotechnology*, vol. 160, no. 8, pp. 2236–2247, 2010.

[16] H.-J. Kim, A.-R. Park, J.-K. Lee, and D.-K. Oh, "Characterization of an acid-labile, thermostable β-glycosidase from *Thermoplasma acidophilum*," *Biotechnology Letters*, vol. 31, no. 9, pp. 1457–1462, 2009.

[17] M. I. Prokofeva, N. A. Kostrikina, T. V. Kolganova et al., "Isolation of the anaerobic thermoacidophilic crenarchaeote *Acidilobus saccharovorans* sp. nov. and proposal of *Acidilobales* ord. nov., including *Acidilobaceae* fam. nov. and *Caldisphaeraceae* fam. nov," *International Journal of Systematic and Evolutionary Microbiology*, vol. 59, no. 12, pp. 3116–3122, 2009.

[18] A. V. Mardanov, V. A. Svetlitchnyi, A. V. Beletsky et al., "The genome sequence of the crenarchaeon *Acidilobus saccharovorans* supports a new order, *Acidilobales*, and suggests an important ecological role in terrestrial acidic hot springs," *Applied and Environmental Microbiology*, vol. 76, no. 16, pp. 5652–5657, 2010.

[19] G. R. Craven, E. Steers, and C. B. Enfinsen, "Purification, composition and molecular weight of the β-galactosidase of *Escherichia coli* K12," *The Journal of Biological Chemistry*, vol. 240, pp. 2468–2477, 1965.

[20] A. C. Bowden, "A simple graphical method for determining the inhibition constants of mixed, uncompetitive and non competitive inhibitors," *Biochemical Journal*, vol. 137, no. 1, pp. 143–144, 1974.

[21] F. M. Pisani, R. Rella, C. A. Raia et al., "Thermostable β-galactosidase from the archaebacterium *Sulfolobus solfataricus*. Purification and properties," *European Journal of Biochemistry*, vol. 187, no. 2, pp. 321–328, 1990.

[22] S. W. M. Kengen, E. J. Luesink, A. J. M. Stams, and A. J. B. Zehnder, "Purification and characterization of an extremely thermostable β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*," *European Journal of Biochemistry*, vol. 213, no. 1, pp. 305–312, 1993.

[23] S. Ezaki, K. Miyaoku, K.-I. Nishi et al., "Gene analysis and enzymatic properties of thermostable β-glycosidase from *Pyrococcus kodakaraensis* KOD1," *Journal of Bioscience and Bioengineering*, vol. 88, no. 2, pp. 130–135, 1999.

[24] I. U. Haq, M. A. Khan, B. Muneer et al., "Cloning, characterization and molecular docking of a highly thermostable β-1,4-glucosidase from *Thermotoga petrophila*," *Biotechnology Letters*, vol. 34, no. 9, pp. 1703–1709, 2012.

[25] Y. Bhatia, S. Mishra, and V. S. Bisaria, "Microbial β-glucosidases: cloning, properties, and applications," *Critical Reviews in Biotechnology*, vol. 22, no. 4, pp. 375–407, 2002.

[26] D. W. Grogan, "Evidence that β-galactosidase of *Sulfolobus solfataricus* is only one of several activities of a thermostable β-D-glycosidase," *Applied and Environmental Microbiology*, vol. 57, no. 6, pp. 1644–1649, 1991.

[27] A. A. Trofimov, K. M. Polyakov, A. V. Tikhonov et al., "Structures of β-glycosidase from *Acidilobus saccharovorans* in complexes with tris and glycerol," *Doklady Biochemistry and Biophysics*, vol. 449, no. 1, pp. 99–101, 2013.

[28] P. Biely, "Microbial xylanolytic systems," *Trends in Biotechnology*, vol. 3, no. 11, pp. 286–290, 1985.

[29] L. R. S. Moreira and E. X. F. Filho, "An overview of mannan structure and mannan-degrading enzyme systems," *Applied Microbiology and Biotechnology*, vol. 79, no. 2, pp. 165–178, 2008.

[30] L. Viikari, M. Alapuranen, T. Puranen, J. Vehmaanperä, and M. Siika-Aho, "Thermostable enzymes in lignocellulose hydrolysis," *Advances in Biochemical Engineering/Biotechnology*, vol. 108, pp. 121–145, 2007.

Hindawi Publishing Corporation
Archaea
Volume 2015, Article ID 267570, 20 pages
http://dx.doi.org/10.1155/2015/267570

Review Article

From Structure-Function Analyses to Protein Engineering for Practical Applications of DNA Ligase

Maiko Tanabe,[1] Yoshizumi Ishino,[2] and Hirokazu Nishida[1]

[1]*Central Research Laboratory, Hitachi Ltd., 1-280 Higashi-koigakubo, Kokubunji, Tokyo 185-8601, Japan*
[2]*Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Sciences, Faculty of Agriculture, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka-shi, Fukuoka 812-8581, Japan*

Correspondence should be addressed to Yoshizumi Ishino; ishino@agr.kyushu-u.ac.jp
and Hirokazu Nishida; hirokazu.nishida.ab@hitachi.com

Received 20 February 2015; Accepted 18 May 2015

Academic Editor: Frédéric Pecorari

DNA ligases are indispensable in all living cells and ubiquitous in all organs. DNA ligases are broadly utilized in molecular biology research fields, such as genetic engineering and DNA sequencing technologies. Here we review the utilization of DNA ligases in a variety of *in vitro* gene manipulations, developed over the past several decades. During this period, fewer protein engineering attempts for DNA ligases have been made, as compared to those for DNA polymerases. We summarize the recent progress in the elucidation of the DNA ligation mechanisms obtained from the tertiary structures solved thus far, in each step of the ligation reaction scheme. We also present some examples of engineered DNA ligases, developed from the viewpoint of their three-dimensional structures.

1. Introduction

DNA ligases are critical DNA replication and repair enzymes; they have been widely used in molecular biology and biotechnology applications, such as cloning and next-generation DNA sequencing [1, 2]. DNA ligases catalyze the joining of adjacent 3′-hydroxyl and 5′-phosphorylated DNA termini in duplex DNA. All DNA ligases accept nicked dsDNA and homologous, cohesive ends as substrates, although the minimum length of the overlap required for efficient ligation varies widely. Some ligases, most notably T4 DNA ligase, also accept fully base-paired (blunt end) substrates for *in vitro* ligation reactions.

DNA ligases share a common mechanism and a high degree of structural similarity with other members of the nucleotidyltransferase superfamily, including RNA ligases and RNA-capping enzymes [3]. Like other nucleotidyltransferases, DNA ligases utilize an ATP molecule to activate the enzyme, and thus DNA ligases are prepared to react with DNA substrates [4].

The details of each step in the DNA ligation reaction sequence have been clarified by many efforts performed by numerous researchers. Above all, Shuman's group has made extensive contributions toward the elucidation of the DNA ligation mechanisms [5]. We describe the exemplified DNA ligation schemes in particular, in the second section of this review.

Since the discovery of DNA ligase, this enzyme has been widely utilized in several types of molecular biology and biotechnology applications. DNA ligase is frequently utilized in combination with restriction enzymes in recombinant DNA experiments. The enzyme has been applied in DNA sequencing methods in diverse ways, by taking advantage of the fact that DNA ligase does not require dNTPs as substrates for its function, since large amounts of dNTPs in solution, a general situation in DNA polymerase reactions, sometimes inhibit effective measurements and cause misinterpretations. DNA ligase is also utilized in analytical methods for protein-protein interactions, as developed by Landegren et al. [6]. These methods are widely applied for quick and effective *in situ* investigations of protein-protein interactions [7]. We will describe the details of the utilization of DNA ligase in several aspects of genetic engineering technology and molecular and cell biology in the third section.

Since the first report of the X-ray crystal structure of the ATP-dependent DNA ligase from bacteriophage T7 [8], the crystallographic analyses of the archaeal and eukaryotic ATP-dependent DNA ligases were subsequently reported during the first decade of the new century [9–11]. The ATP-dependent DNA ligases from Archaea and Eukarya comprise three domains, but, surprisingly, the relative arrangements of the three domains were completely different from each other in these reports [9–11], reflecting the dynamic domain motion in the 3-step ligation sequence. Based on these structural transitions in the ligation sequence, several attempts toward improving the enzymatic reaction have been made by mutating the enzymes. In the fourth and the fifth sections, we summarize the structural information about DNA ligase and the mutation-based improvements of its enzymatic efficacy.

2. Basis of the DNA Ligase Reaction

The Gellert, Lehman, Richardson, and Hurwitz laboratories discovered DNA ligases in 1967 and 1968 [12–15]. By joining the 3′-OH and 5′-phosphate termini (nicks) to form a phosphodiester bond, DNA ligases play an important role in maintaining genome integrity. They are essential for DNA replication and repair in all organisms [16, 17]. Furthermore, DNA ligases have long been a critical reagent in the development of molecular cloning and many subsequent aspects of DNA biotechnology.

All DNA ligation reactions entail sequential nucleotidyltransfer steps [18]. The reaction mechanism can be divided into three distinct catalytic events: the first (step 1), the activation of the enzyme through the covalent addition of AMP to the conserved catalytic lysine of the ligase, accompanied by the release of PPi or nicotinamide mononucleotide from the cofactor (ATP or NAD^+); the second (step 2), the binding of the ligase-adenylylate to the substrate DNA and the transfer of AMP from the ligase to the 5′-phosphoryl group of the nick on the DNA; and the third (step 3), the formation of the phosphodiester-bond with the concomitant release of free AMP from the DNA-adenylylate intermediate (Figure 1). All three chemical steps depend on a divalent cation. The DNA ligases are grouped into two families, according to the cofactor dependence for the reaction. ATP-dependent DNA ligases are found in viruses, bacteria, Eukarya, and Archaea [18, 19], whereas NAD^+-dependent DNA ligases are primarily present in bacteria and entomopox viruses [19, 20].

Sequence analyses of DNA ligases revealed that they share six motifs (I, III, IIIa, IV, V, and VI), which are also conserved in the nucleotidyltransferase superfamily members, including RNA ligase [21], tRNA ligase [22], and mRNA-capping enzymes [23], except for motif VI. In the ATP-dependent DNA ligases, these six motifs align well among the enzymes from viruses, bacteriophages, Archaea, and Eukarya, without long gaps or insertions (Figure 2). The sequences of the ATP-dependent DNA ligases showed that the amino acids in motifs I, III, IIIa, IV, V, and VI contact the substrates, AMP or DNA. The important roles of the individual amino acids in these motifs were verified by alanine scanning mutational studies and confirmed as essential for one or more steps of the ligation pathway [24]. For example, motif I (K**x**DG**x**R)

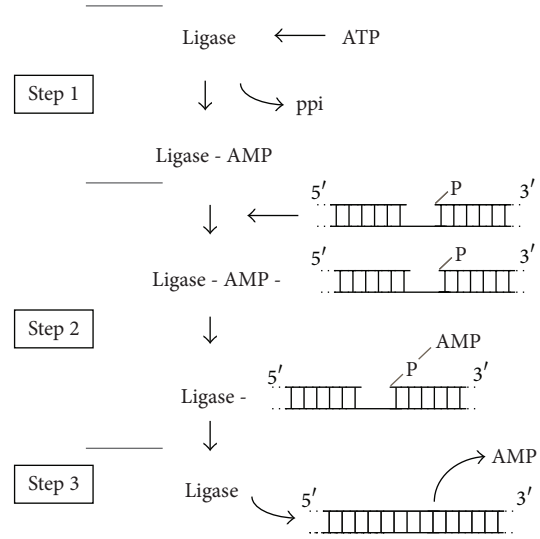


Figure 1: Schematic diagram of the three-step reaction catalyzed by ATP-dependent DNA ligases. The three-step reaction catalyzed by DNA ligase (Ligase) results in the serial transfer of AMP (adenosine 5′-monophosphate) to an active site lysine (step 1) and then to the 5′-PO_4 end of DNA (step 2). During step 3, the 3′-OH end of a second DNA strand attacks the 5′-PO_4, to release AMP and generate the ligated DNA product.

contains the lysine that becomes covalently linked to AMP in the first step of the ligase reaction [25, 26]. Furthermore, the arginine and lysine side chains in motif VI (RxDK) orient the PPi leaving group apically relative to the attacking motif I lysine, during step 1 enzyme nucleotidylation reaction [27, 28].

3. Utilization of DNA Ligases in Genetic Research Technologies

The *in vitro* manipulation of DNA has been broadly applied to studies in molecular genetics, microbiology, immunology, and oncology and also to practical uses in clinical assays [29]. This technology was developed with DNA-related enzymes, which were discovered by research on the molecular mechanisms of the replication and repair of DNA. DNA replication is initiated by DNA primase, which synthesizes small RNA primers that are subsequently elongated by DNA polymerases [30]. Based on its primer extension activity, a DNA polymerase can be utilized in the polymerase chain reaction (PCR), which amplifies a single copy of DNA by several orders of magnitude in a brief *in vitro* reaction [31]. In addition, DNA ligase plays a role in joining Okazaki fragments during lagging strand maturation [32]. Based on its activity to join two DNA strands *in vitro*, DNA ligase has been contributing to recombinant DNA technology, for example, DNA cloning [33]. In addition, DNA ligase is broadly utilized for ligase-mediated mutation detection methods and DNA sequencing.

3.1. Oligonucleotide Ligation Assay (OLA). DNA ligase can only join two DNA strands when they are perfectly

Figure 2: Conserved sequence elements define a superfamily of covalent nucleotidyltransferases. Six collinear sequence elements, designated as motifs I, III, IIIa, IV, V, and VI, are conserved in capping enzymes and ligases. The amino acid sequences are aligned for the DNA ligases from *Chlorella* virus (ChV), bacteriophage T7 (T7), *Pyrococcus furiosus* (Pfu), *Sulfolobus solfataricus* (Sso), and human (Hu). The number of intervening amino acid residues is indicated. Conserved residues are highlighted in red (nucleotidyltransferases) and blue (DNA and RNA ligases).

hybridized to a complementary DNA sequence. Even a single base pair mismatch between two strands significantly decreases the efficiency of the strand joining reaction [34]. In 1988, the Oligonucleotide Ligation Assay (OLA) was developed, as a useful method to detect the genotype of the target DNA by utilizing this characteristic [6, 35]. This was the first example of the development of a method for gene analysis by using DNA ligase.

OLA consists of two steps. First, the two probes to be hybridized with the target DNA at the sequence of interest must lie directly adjacent to one another. In this case, one probe must contain a reporter group, which is either ^{32}P- or fluorescently labeled, and the other must include a recognition group for immobilization, such as a biotin group, which could be captured by streptavidin immobilized on a solid support. Second, the neighboring probes that are perfectly hybridized to the target DNA at the sequence of interest are joined by DNA ligase. Then, the ligation signal is subsequently captured by streptavidin and detected by autoradiography or fluorography (Figure 3). This assay takes advantage of the enzymatic accuracy in two events, the hybridization of the probes and the joining by DNA ligase, resulting in the reliable analysis of the clinical genotype.

A breakthrough in OLA for practical use occurred in 1990, when PCR was coupled to the assay prior to the ligation reaction (PCR-OLA) [36]. PCR-OLA amplification only for the target DNA enhanced the sensitivity of the assay, enabling the nonradioactive detection of the OLA results. Recently,

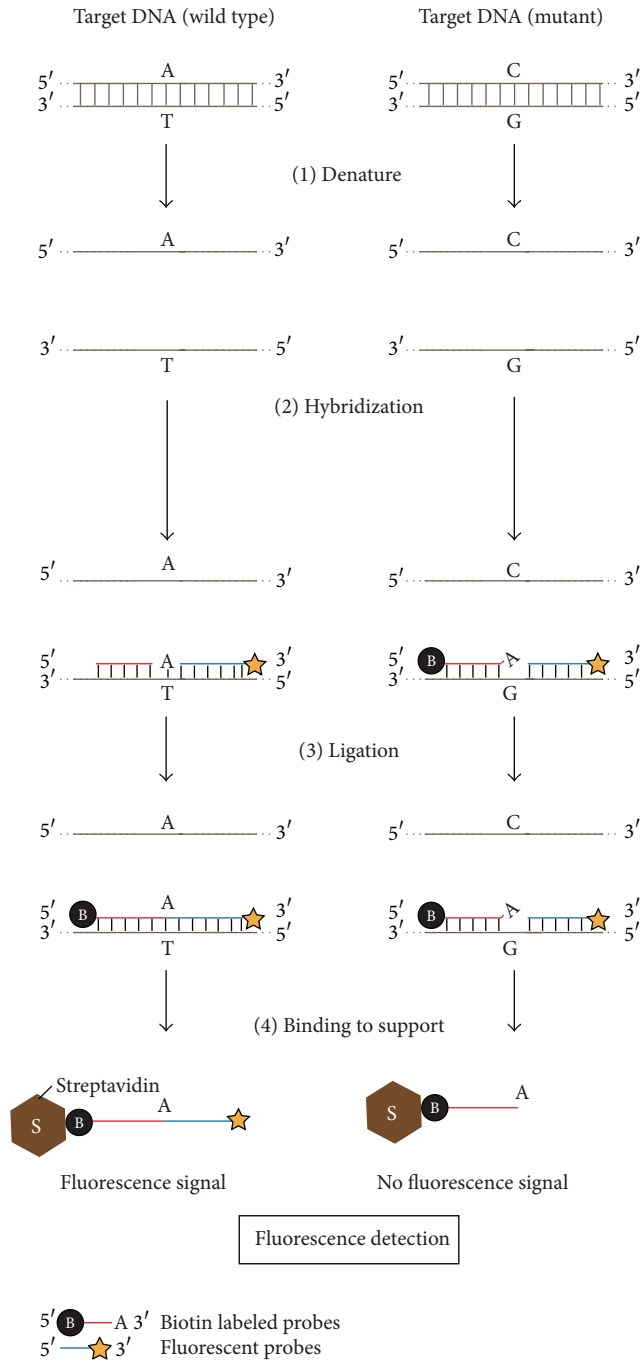


Figure 3: Schematic diagrams for the Oligonucleotide Ligation Assay (OLA). DNA fragments from the normal type and the mutant type (at the single nucleotide polymorphism (SNP) site) are shown as reaction templates. (1) Denature: samples are heated to 85–95°C for one to several minutes to denature (separate into single strands) the target DNA. (2) Hybridization: two oligonucleotide probes, 5′-biotin labeled probes and 3′-fluorescent (or radiolabeled) probes, which are complementary to the normal-type target, hybridize to the target DNA fragment. (3) Ligation: adjacent probes that are perfectly complementary to the target (left) are connected by DNA ligase. (4) Binding to support: the fluorescently labeled ligation samples are immobilized to streptavidin and detected by fluorography only if ligated to biotinylated oligonucleotides that can be bound to streptavidin on a solid support (left).

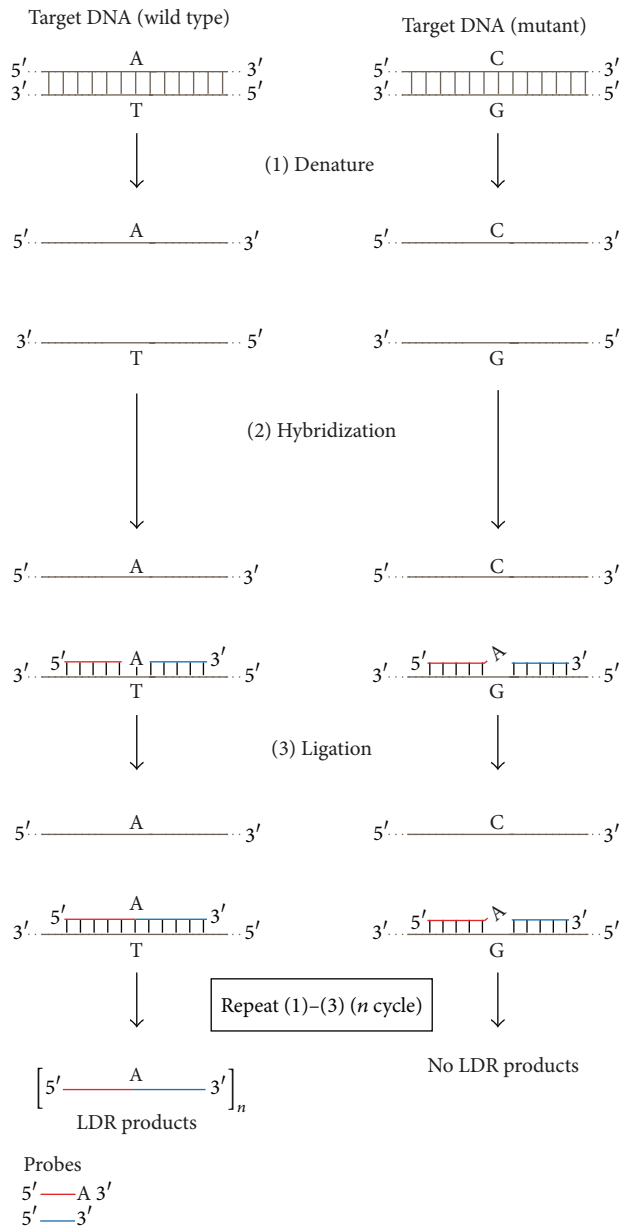


Figure 4: Schematic diagrams for the Ligation Detection Reaction (LDR) processes. DNA fragments from the normal type and the mutant type (at the single nucleotide polymorphism (SNP) site) are shown as reaction templates. (1) The denaturation process is the same as in Figure 3 (1). (2) Hybridization: two oligonucleotide probes, complementary to the normal-type target, hybridize to the target DNA fragment. (3) Ligation: adjacent probes that are perfectly complementary to the target (left) are connected by DNA ligase, and thermocycling linearly amplifies the amount of product formed. In the case of the mutant type template, the presence of a single-base mismatch at the junction inhibits the ligation, and therefore no ligated LDR probes are formed (right).

a sensitive, specific, and high-throughput OLA was developed for the detection of genotypic human immunodeficiency virus type 1 (HIV-1) resistance to a Food and Drug Administration-approved protease [37]. This report revealed that OLA can be used for the detection of the HIV-1 genotype (wild type or mutant) as a genetic diagnosis method.

3.2. DNA Ligase-Mediated Cycling of the Ligation Reaction. Further spreading of DNA ligase-mediated technologies was achieved by cycling the ligation reaction, like PCR, using a thermostable DNA ligase, which became available commercially in 1990. The thermostable DNA ligase enables the cycling OLA reaction. OLA reaction method with thermal cycling procedure developed by using thermostable DNA ligase is often specifically termed Ligation Detection Reaction (LDR) (Figure 4). Through repeated denaturation, annealing, and ligation, the target signal is amplified linearly [38–40]. Ligation Chain Reaction (LCR) [38, 39, 41] and Ligation Amplification Reaction (LAR) [42] are also performed by repeated cycles of heat denaturation of a DNA template

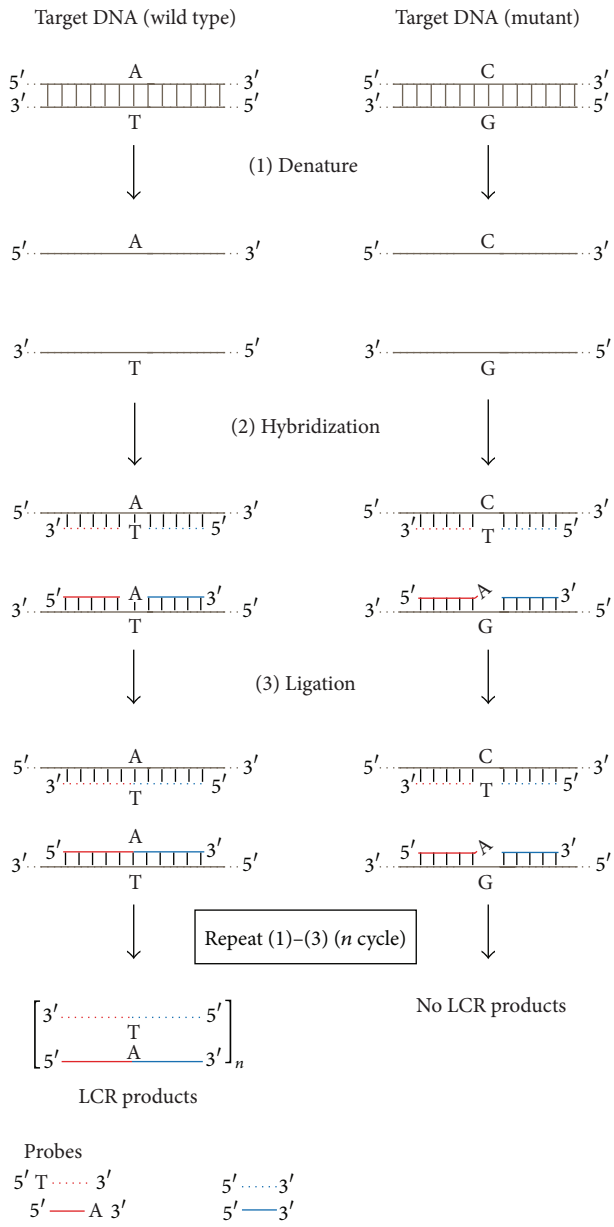


Figure 5: Schematic diagrams for the Ligation Chain Reaction (LCR) and Ligation Amplification Reaction (LAR) processes. (Here, we describe LCR.) DNA fragments from the normal type and the mutant type (at the single nucleotide polymorphism (SNP) site) are shown as reaction templates. (1) The denaturation process is the same as in Figure 3 (1). (2) Hybridization: four oligonucleotide probes, complementary to the normal-type target, hybridize to the target DNA fragment, and (3) Ligation: adjacent probes that are perfectly complementary to the target (left) are connected by DNA ligase. Ligated LCR probes from the first round of the ligation become the targets for the subsequent round using another probe set (dotted lines) complementary to the first set. Thus, the amount of ligated LCR probes increases exponentially. In the case of the mutant type template, the presence of a single-base mismatch at the junction inhibits the ligation, and therefore no ligated LCR probes are formed (right).

containing the target sequence, annealing of probes, and ligation processes. After annealing the first set of two adjacent oligonucleotide probes to the target DNA sequence in a unique manner, a second set of complementary oligonucleotide probes that hybridize to the sequence opposite to the target DNA sequence is applied (Figure 5). Thereafter, a thermostable DNA ligase will covalently link each pair of adjacent probes that are completely hybridized at the junction of the adjacent probes. Through the repeated sequential processes, including denaturation, annealing, and ligation, the target signal is amplified exponentially. These methods require a thermostable DNA ligase to allow the ligation to occur under temperature conditions that prevent mismatches from hybridizing to form acceptable substrates. The thermostable

DNA ligase made this assay useful for DNA-based diagnostic tests for inherited diseases in clinical laboratories. The ligation-mediated detection of point mutations by thermostable DNA ligase is now used to detect hepatitis B virus (HBV) mutants [43–45], colon tumor microsatellite sequence alterations [46], the mutation responsible for bovine leukocyte adhesion deficiency [47], AZT-resistance HIV mutants [48], mutations in the *ras* family of oncogenes [49], extended spectrum β-lactamase resistance in bacteria [50], and ganciclovir-resistant cytomegalovirus mutants [51].

3.3. Padlock Probe and Rolling-Circle Amplification (RCA). Padlock probes and RCA are alternative methods for detecting known sequence variants in DNA and particularly for detecting single nucleotide polymorphisms, by using a thermostable DNA ligase. The padlock probe has target recognition sequences situated at both the 5′- and 3′-ends, connected by an intervening sequence that can include the sequence for detection. When the probe is hybridized to a target sequence, the two ends are brought adjacent to each other, and then a thermostable DNA ligase covalently joins the two ends and circularizes the probe. This circularized probe is wound around the target strand in a manner similar to a padlock, driven by the helical nature of the double-stranded DNA [52] (Figure 6). This probe is used for the localization of signals in *in situ* analyses. For example, a padlock probe was able to detect repeated alphoid sequences in metaphase chromosomes in a living cell, demonstrating its utility for *in situ* studies [53–55]. RCA is also known as a simple and efficient isothermal enzymatic process that utilizes strand-displacement DNA and RNA polymerases, like Phi29, Bst, and Vent exo-DNA polymerases for DNA and T7 RNA polymerase for RNA, to generate long single stranded DNAs and RNAs containing tens to hundreds of tandem repeats that are complementary to the circular template [56–58]. The padlock principle has been combined with RCA for mutation detection, in which a primer annealing to the linker region initiates rolling circle replication (Figure 6), because the padlock approach alone is not sufficient to detect single nucleotide differences in a single copy gene *in situ* [59–61]. Point mutations in the cystic fibrosis transmembrane conductance regulator (CFTR), p53, BRCA1, and Gorlin syndrome genes were visualized in interphase nuclei and DNA fibers by RCA [62]. These results demonstrated that ligase-mediated mutation detection methods coupled with RCA are able to reveal single nucleotide differences in a single cell. The ability to detect mutations in a cellular milieu has important implications for cancer research and diagnosis.

In another aspect of the RCA with ligation methods, Proximity Ligation Assay (PLA) is known for one of the potent techniques for detecting individual proteins or protein complexes *in situ*, by using antibodies with attached DNA strands that participate in ligation and subsequent RCA reactions [63]. PLA can be used for detecting reactions in which identification of target molecules depends on two recognition events. Formation of a proper target complex results in the formation of a circular DNA strand by exogenously added DNA ligase. This circular DNA is used to template a locally restricted RCA reaction, which generates an elongating ssDNA rolling-up in a ball that can be detected when hybridized with fluorescence-labeled probes [7]. The binding to a target interacting complex by two different antibodies with attached oligonucleotides individually (referred to as proximity probes) is followed by the addition of two more oligonucleotides that are then ligated into a circular

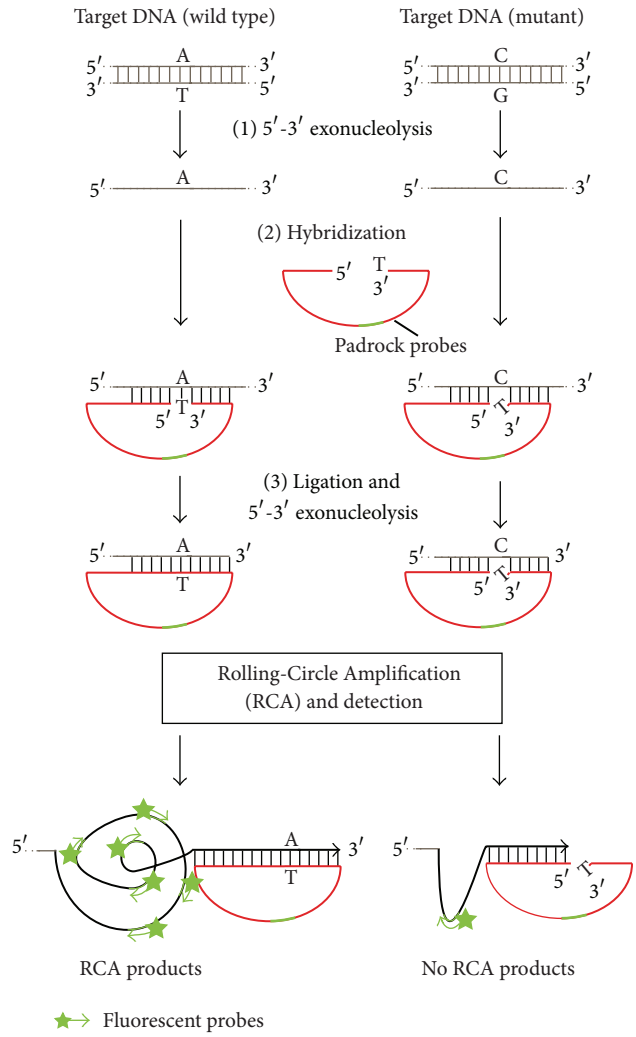


Figure 6: Schematic representation of the target-primed Rolling-Circle Amplification (RCA) of circularized padlock probes. (1) 5′–3′ exonucleolysis: the target DNA is restriction digested 3′ of the target sequence and irreversibly made single stranded by strand-specific 5′–3′ exonucleolysis. (2) Hybridization: padlock probes (red) with the tag sequence (green) are hybridized to their target sequences. (3) Ligation & 5′–3′ exonucleolysis: adjacent padlock probes that are perfectly complementary to the target (left) are connected by DNA ligase, thus locking the probe onto the target molecule. After ligation, the RCA is initiated by the Phi29 DNA polymerase, by turning the target molecule into a primer through the 3′–5′ exonucleolysis of any 3′ end protruding beyond the padlock probe hybridization site. The padlock probe then serves as the template for DNA synthesis. The RCA product (black) is detected through the hybridization of fluorescently labeled probes (green) to tag sequences specific for the padlock probe. Arrowheads indicate 3′ ends of the RCA reaction.

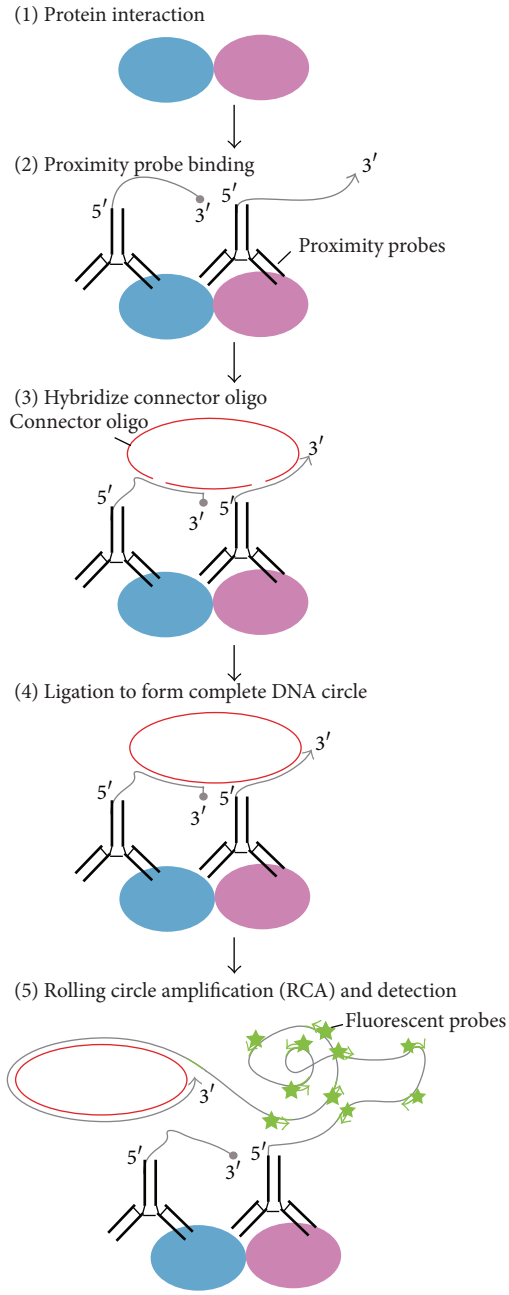


Figure 7: Schematic representation of *in situ* Proximity Ligation Assay (PLA). (1) Protein interaction: the complex is formed between target proteins (light blue and pink). (2) Proximity probe binding: antibodies (black) with individual proximity probes (glay) bind to each of the target protein complexes. (3) Hybridization of connector oligo-DNAs: target protein complex serves to template the hybridization of connector oligo-DNAs (red). (4) Ligation to form complete DNA circle: adjacent connector oligos that are perfectly complementary to the target template are connected by DNA ligase. (5) Rolling Circle Amplification (RCA) and detection: after ligation, the RCA is initiated by the Phi29 DNA polymerase, by turning the target molecule primed by one of the proximity probes. The RCA product is detected through the hybridization of fluorescently labeled probes (green). Arrowheads indicate 3′ ends, and roundheads means 3′ ends modified with 2′ *O*-methyl RNA.

DNA strand by the function of DNA ligase. The circular DNA strand is templated by the oligonucleotides attached to antibodies (Figure 7). Next, one of the antibody-bound oligonucleotides is utilized as a primer of the succeeding RCA reaction, resulting in the generation of an ssDNA rolling circle product. The rolling circle is covalently attached to one of the proximity probes. A 60 min RCA by Phi29 DNA polymerase results in a 1000-fold amplification of the 100 nt DNA circle, producing an around 100 kb rolling circle product [63]. The rolling circle product is then visualized by hybridization of fluorescence-labeled complementary oligonucleotide detection probes *in situ*.

3.4. Next-Generation DNA Sequencing by Using DNA Ligase. Recently, Next Generation Sequencing (NGS) technologies, such as the 454 FLX pyrosequencer (Roche) [64], Illumina genome analyzer (Illumina) [65], and Sequence by Oligonucleotide Ligation and Detection (SOLiD) sequencer (Life Technologies) [66], have revolutionized genomic and genetic research. Although these platforms are quite diverse, in terms of their sequencing biochemistry and sequence detection, their workflows are conceptually similar to each other. The array is prepared by random fragmentation of DNA, followed by *in vitro* ligation of common adaptor sequences by DNA ligase. DNA fragments with adapters are amplified by DNA polymerase and are detected by each platform. The detection platforms rely on sequencing by DNA synthetic methods, that is, the serial extension of primed templates. The enzyme driving the DNA synthesis can be either a DNA polymerase or a DNA ligase. The 454 FLX Pyrosequencer and Illumina analyzer use DNA polymerase methods with pyrosequencing [67] and reversible dye terminator technology, respectively [68]. In contrast, SOLiD uses DNA ligase and a unique approach to sequence the amplified fragments [69]. A universal primer complementary to the adaptor sequence is hybridized to the array of amplicons. Each cycle of sequencing involves the ligation of a degenerate, fluorescently labeled 8-mer probe set (Figure 8). The octamer mixture is structured, and the type of nucleotide (A, T, G, or C) at the specific position within the 8-mer probe set correlates with the type of fluorescent label. After the ligation of the 8-mer probes, images are acquired in four channels, resulting in the effective collection of sequencing data for the $3'$ end positions of the probes across all template-bearing beads. Then, the octamer is chemically cleaved between positions 3 and 4, to remove the fluorescent label. Progressive rounds of octamer ligation enable the sequencing of every 5th base (e.g., bases 1, 6, 11, and 16 of the template strand). After several cycles of probe ligation reactions, the extended primer is denatured to reset the system. Subsequent iterations of this process can be applied to a different set of positions (e.g., bases 0, 5, 10, and 15 of the template strand), by using different mixtures of octamers in which the nucleotides at the different position are correlated with the label colors. An additional feature of this platform involves the use of two-base encoding, an error-correction scheme in which two adjacent bases, rather than a single base, are correlated with the label. Each base position is then queried twice (in a set of 2 bp interrogated in a given cycle) and thus miscalls can be more readily identified [70].

4. Structural Transition of the DNA Ligase in the Reaction Sequence

In Section 2, we showed that the sequence alignments clearly revealed several homologous regions among ATP-dependent DNA ligases and RNA capping enzymes, indicating the presence of five conserved motifs (I, III, IIIa, IV, and V) in these proteins. This fact suggests that the nucleotidyltransferase superfamily members, including ATP-dependent DNA ligases, may share a similar protein fold and a common reaction mechanism. Next, we will present a series of crystal structures of ATP-dependent DNA ligases and describe the structural and functional insights into the mechanisms of the enzymes.

4.1. Structural Studies of ATP-Dependent DNA Ligase. ATP-dependent DNA ligases show considerable variation in their molecular sizes, which range from 41 kDa (bacteriophage T7) [71] to 102 kDa (human DNA ligase I (hLigI)) [72]. Despite this wide size variation, it is clear from the sequence alignments that the smaller enzymes constitute a common core structure that is conserved across all ATP-dependent ligases [73]. The crystal structures of the ATP-dependent DNA ligases, such as bacteriophage T7 DNA ligase (T7Lig) complexed with ATP [8, 74] and *Chlorella* virus DNA ligase (ChvLig) with covalently bound AMP [75], have been solved (Figure 9(a), top). These DNA ligases adopt a common architecture of two distinct domains: the adenylylation domain (AdD) and the oligonucleotide/oligosaccharide-binding-fold domain (OBD), which are jointly called the catalytic core domains. The catalytic core domains are the minimal entity for the nick-joining activity, as seen in the ChvLig and T7Lig enzymes [8, 75]. The AdD contains the catalytic lysine residue that forms a covalent bond with AMP, which is derived from the substrate ATP. The Lys238 and Lys240 residues of T7Lig (Lys188 and Lys186 of ChvLig) within the AdD are essential for the adenylylation and nick sealing activities [76, 77]. The Lys240 residue forms a photo-crosslinking adduct with the $5'$-terminal nucleotide of the nick, implying its direct involvement in binding the phosphate of the nick [78]. The OBD is connected to the AdD via the conserved motif V [5, 79] and is similar to other DNA and RNA binding proteins [80, 81]. The OBD is observed in the related nucleotidyltransferase, the mRNA capping enzyme from *Chlorella* virus, which undergoes opened-closed conformational changes during catalysis. The OBD domain of the enzyme was found to move towards AdD and close the nucleotide-binding pocket [80, 81].

In contrast, three crystal structures of large ATP-dependent DNA ligases, which mainly ligate the nicks between Okazaki fragment in DNA replication in Eukarya and Archaea, with an additional N-terminal DNA-binding domain (DBD), have been reported (Figure 9(a), bottom) [9–11, 82, 83]. This extra domain is not essential for hLigI activity *in vitro* [84, 85] but is considered to be crucial for detecting a singly nicked dsDNA [9]. The structures of the two archaeal DNA ligases from *Pyrococcus furiosus* (PfuLig) and *Sulfolobus solfataricus* (SsoLig) were determined in the closed [11] and extended forms [10], respectively. The structure of hLigI in complex with DNA [9] revealed that the enzyme entirely encircles the nicked-DNA. All of these DNA ligases are commonly composed of three domains (DBD, AdD, and OBD) in the sequences from the N to C termini. Although the protein folding of each domain is strikingly similar among the three DNA ligases, the relative domain orientations within each enzyme are quite different.

4.2. Structural Transition of DNA Ligase Molecules in Each Reaction Step. Based on the crystal structures described above, a ligation mechanism was proposed, as schematically represented in Figure 9(b). In solution, a DNA ligase partly

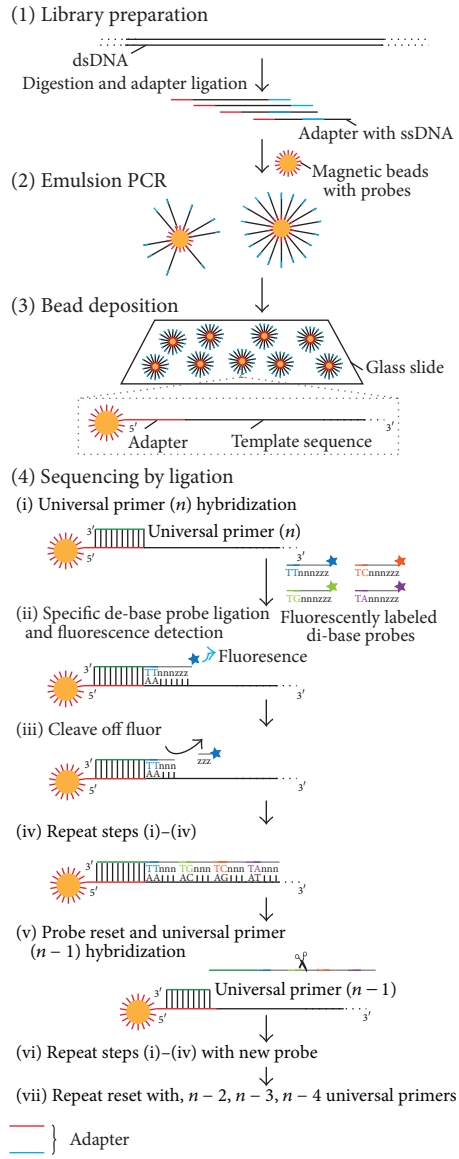


Figure 8: The ligase-mediated sequencing approach of the Sequence by Oligonucleotide Ligation and Detection (SOLiD) sequencer (Life Technologies). (1) Library preparation: two different adapters are ligated to sheared genomic DNA. (2) Emulsion PCR: emulsion PCR is conducted using magnetic beads to generate "bead clones," in which each contains a single nucleic acid species. (3) Bead deposition: the beads are then attached to the surface of a glass slide. (4) Sequencing by ligation: ligase-mediated sequencing begins by annealing a universal primer to the shared adapter sequences on each amplified fragment (i), and then DNA ligase is provided along with specific fluorescently labeled 8-mers, in which the two bases at the 3′ end of the probe are encoded by the attached fluorescent group. Each ligation step is followed by fluorescence detection (ii), after which a regeneration step removes the bases from the ligated 8-mer (including the fluorescent group) (iii), and concomitantly prepares the extended probe for another round of ligation (iv–vii). Since each fluorescent group on a ligated 8-mer identifies a two-base combination, the resulting sequence reads can be screened for base-calling errors versus either true polymorphisms or single base deletions, by aligning the individual reads to a known high-quality reference sequence.

adopts the extended form, as expected from the SsoLig crystal structure and the small angle X-ray scattering (SAXS) analysis [10]. In the absence of DNA and AMP, the OBD of SsoLig is turned away from the AdD in an open conformation, resembling that seen in the crystal structures of the compact and two-domain (AdD-OBD) DNA ligases from *Chlorella* virus and T7 (Figure 9(a)). The closed conformation of the catalytic core domains, which represents the active conformation for step 1, is observed in the crystal structures of PfuLig, as proposed previously [10, 11]. In step 2, tight binding between the enzyme and the substrate DNA occurs, as revealed by the hLigI–DNA complex crystal structure [9]. These structures depict three different phosphoryl transfer reactions, and the flexible multidomain structure of DNA ligases facilitates the adoption of different enzyme conformations during the course of the reaction (Figure 9(b)) [10]. A superimposition

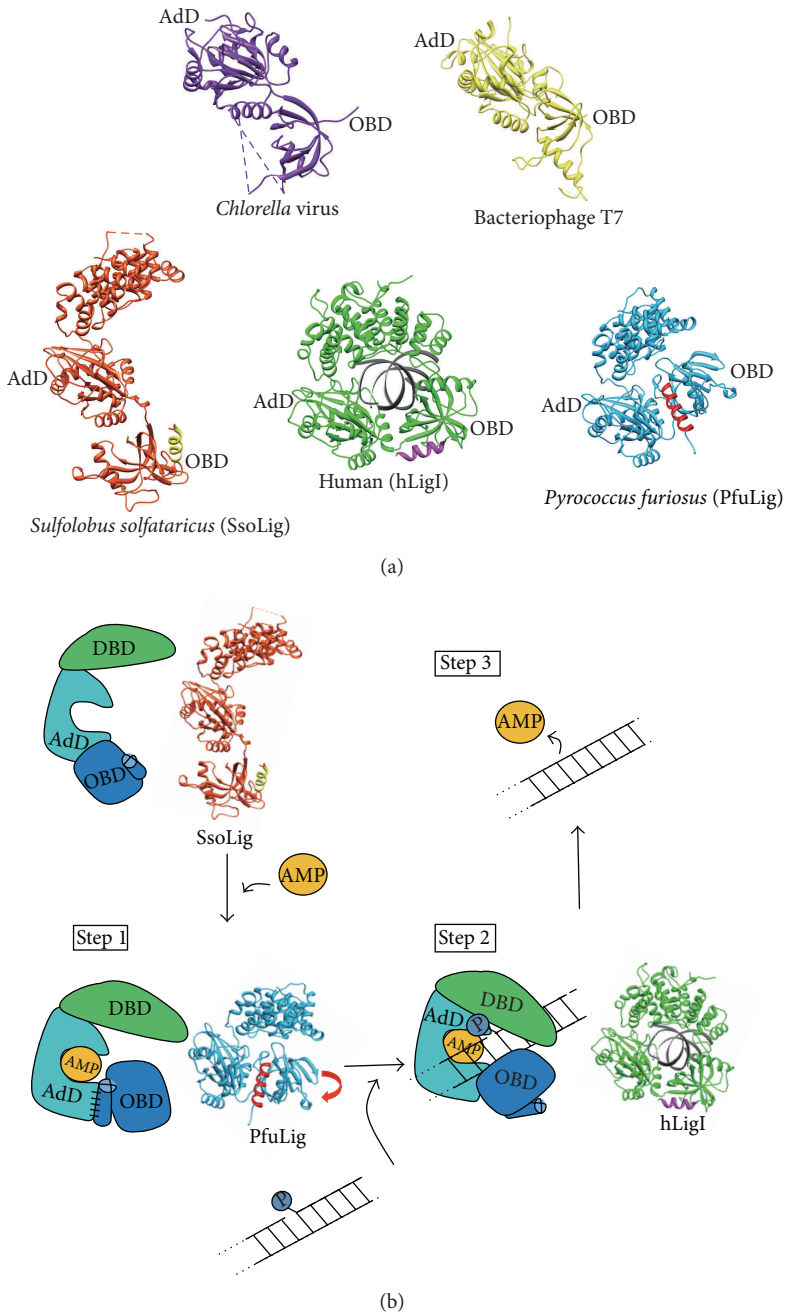


Figure 9: Crystal structures of ATP-dependent DNA ligases. (a) The viral and bacterial ATP-dependent DNA ligases from *Chlorella* virus (left, blue) and bacteriophage T7 (right, yellow) are two-domain ligases. These ligases revealed the common structures of two distinct domains, designated as the adenylylation domain (AdD) and the oligonucleotide/oligosaccharide-binding-fold domain (OBD), which are jointly called the catalytic core domains. In these two structures, the catalytic core domains adopted an open form. The three-domain structure of DNA ligase is characteristic of the archaeal (*Sulfolobus solfataricus*: SsoLig (orange, left) and *Pyrococcus furiosus*: PfuLig (blue, right)) and eukaryotic (human: hLigI (green, center)) DNA ligases. The AdD contains most of the catalytic residues and is assisted by two flanking domains that also bind to DNA, the N-terminal DNA-binding domain (DBD) and the C-terminal OBD. Each C-terminal helix is highlighted in yellow (SsoLig), magenta (hLigI), and red (PfuLig). The figure was prepared using Chimera [105]. (b) Schematic diagram of the structure-based ligation mechanism. Before and after the ligation reaction, DNA ligase adopts the extended conformation, as expected from the SsoLig crystal structure. In step 1, the closed conformation of the catalytic core domains at the carboxyl terminus in PfuLig creates a small compartment, which holds a covalently bound AMP molecule. In step 2, the interactions with the substrate DNA must be stabilized for the enzyme to adopt a compact conformation, as seen in the crystal structure of hLigI bound to the nicked-DNA. Finally, the ligation of the DNA strands and the subsequent release of AMP and DNA strands from DNA ligase occur during step 3.

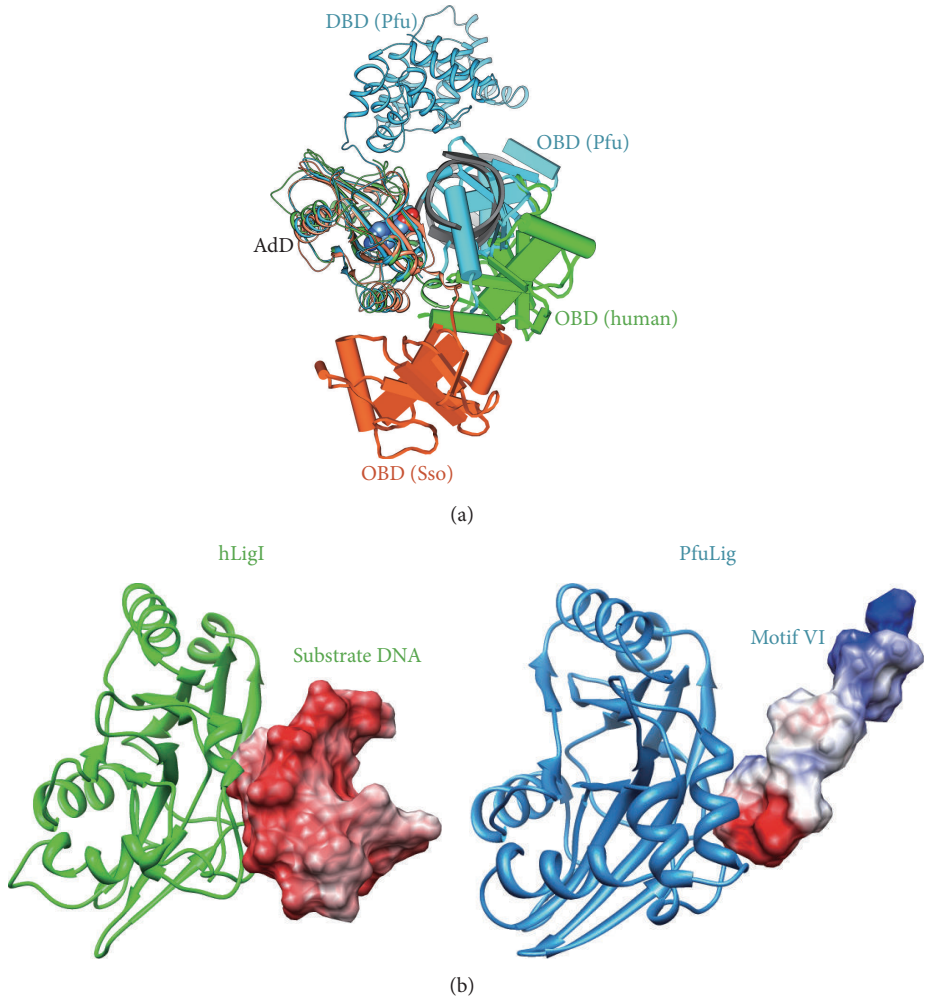


Figure 10: Superimposed ribbon diagrams. (a) Ribbon diagrams of PfuLig (2CFM, blue), hLigI (1X9N, green), and SsoLig (2HIV, orange), in which the AdD from each ligase were maximally overlapped with each other. The DBD from hLigI and SsoLig were omitted, for clarity. The arrangements of the OBD relative to the AdD in each ligase are apparently different from each other. Notably, the OBD from PfuLig is closely bound to AdD and is replaced by the bound-DNA substrate (grey) in hLigI. (b) Superimposed diagrams of adenylation domains from PfuLig and hLigI, flanked by surface representations of motif VI (PfuLig) and the upstream region of the substrate DNA (hLigI), in which the electrostatic distributions (positive charge, blue; negative charge, red) are mapped onto the molecular surfaces. Since the electrostatic distributions on the surfaces nearby the AdD of motif VI and the DNA are both negatively charged, the replacement of motif VI with DNA may easily occur.

of the AdDs in PfuLig, SsoLig, and hLigI revealed that the arrangements of the OBD relative to the AdD in each ligase are apparently different from each other. Notably, the OBD from PfuLig is closely bound to the AdD and is replaced by the bound-DNA substrate in hLigI (Figure 10(a)). Ribbon diagrams of the AdDs from hLigI and PfuLig, together with the adjacent surface representations of the substrate DNA (hLigI) and motif VI (PfuLig), are shown in Figure 10(b). The electrostatic distribution on the motif VI surface facing AdD is negative, suggesting that motif VI is a molecular mimic of the incoming DNA substrate [11]. This finding indicates that the ligation could be completed simply by the smooth replacement between the substrate DNA and motif VI. Indeed, in the closed structure of PfuLig, motif VI approaches the active site in the absence of DNA, whereas it occupies a position in the region upstream from the nicked DNA in the structure of hLigI.

4.3. Interface of the Mandatory Domains (AdD and OBD) for the Enzymatic Reaction. In the ATP-dependent DNA ligases, the enzyme captures ATP to adenylylate a Lys residue in the first step, in which motif VI is also indispensable [27]. Motif VI in PfuLig provides several basic residues located around the AMP binding pocket. This electron density contributes to forming the compartment that traps the AMP molecule (Figure 11). In particular, R531 and K534 in motif VI, specific to the DNA ligases in Archaea and Eukarya, contribute to the basic surface within the pocket (Figure 11) [11].

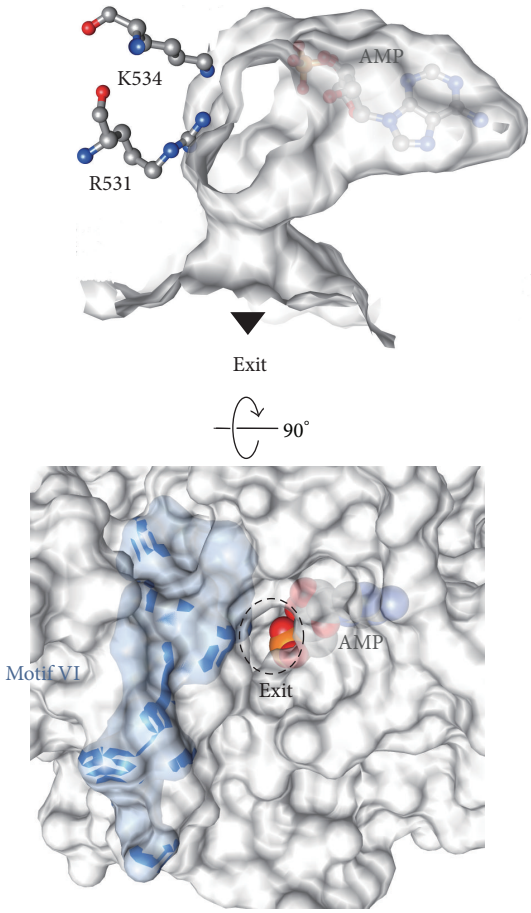


Figure 11: Top: the AMP-binding pocket viewed from the inside of the molecule. A solvent-accessible surface was created without the bound AMP, and then the final refined AMP model was superimposed. The noncovalently bound AMP is trapped within a closed compartment, which conforms well to the shape of the AMP molecule. Bottom: surface representation around the hole in the active site pocket. The surface of motif VI is colored cyan. The bound AMP is depicted as a sphere model. The phosphate (phosphorus, orange; oxygen, red) group is visible through a small "exit."

The electrostatic potential distributions of the AdD-OBD interacting surfaces (Figure 12) revealed that the predominant charge distribution on each surface was oppositely charged, and this may be involved in stabilizing the closed conformation of the two catalytic core domains [11]. There is a small exit of the pocket through the closed conformation of the two domains (Figure 11), and the diameter of the exit is about 5.5 Å and could accommodate the PPi molecule but not the AMP moiety. The exit may serve for the spontaneous release of the PPi product. This notion is consistent with the fact that the DNA ligase from *Pyrococcus horikoshii*, which is more than 90% identical to PfuLig, cannot use NAD^+ as a nucleotide cofactor [86], because the pocket is too small to accommodate the nicotinamide ring to be released from the active site. Presumably, this compact "reaction room" of PfuLig would have been specifically designed to complete the conversion process from ATP to AMP within the closed pocket [11].

4.4. Role of the C-Terminal Helix Commonly Observed in the Archaeal and Eukaryotic DNA Ligases. The overall architectures of the three domains in hLigI and PfuLig are similar to each other (Figure 9(a)), although the sequence identities between the corresponding fragments of hLigI and PfuLig are moderate (32%). The crystal structures and sequence analyses revealed that the eukaryotic and archaeal DNA ligases possess a C-terminal helix shortly after motif VI (Figures 9(a) and 13(a)). The sequence extension harboring the C-terminal helix is conserved among the eukaryotic and archaeal DNA ligases, whereas the ligases from viruses (*Chlorella* virus (ChV)) and bacteriophages (bacteriophage T7 (T7)) lack this long extension (Figure 13(a)). The C-terminal helix is located at the boundary of the AdD and the OBD in the closed structure of PfuLig [11]. In the case of hLigI, this helix is distant from the domain interface, because of the bound DNA substrate (Figure 9(a)). In fact, a structural comparison between the DNA-hLigI complex and PfuLig suggested that the C-terminal helix might play a crucial role in switching from the tightly closed form to the DNA-substrate-bound form. The C-terminal helix of PfuLig connects the AdD and the OBD through five polar or ionic interactions (Figure 13(b)). This feature suggests that the helix might play a critical role in stabilizing the closed conformation of the catalytic core domains during step 1.

5. Engineered DNA Ligases with Improved Abilities

As described in Section 3, DNA ligases are critical for many applications in molecular biology. Improvements in the enzymatic ability, such as the nick sealing efficiency or fidelity, will provide better performance in the current methods for ligase-mediated mutation detection and NGS, as described above. Numerous improvements of DNA polymerases have been reported [87–89], whereas fewer are known for DNA ligases. Here we describe some reports on improvements of the enzymatic performances of DNA ligases.

5.1. Improvement of the Nick-Sealing Activity by Fusion with the Archaeal DNA Binding Domain. The ATP-dependent DNA ligase from bacteriophage T4 (T4Lig) has evolved to be a nick-sealing enzyme. It can also join double-stranded DNA (dsDNA) fragments with cohesive ends [90]. Furthermore, it is the only commercially available DNA ligase that can join blunt-ended DNA duplexes *in vitro*, in the absence of macromolecular enhancers such as polyethylene glycol [91, 92]. The ligation reaction of cohesive or blunt-ended dsDNA fragments is prerequisite for NGS, which requires the *in vitro* ligation of common adaptor sequences. However, the turnover numbers for the fragment joining reactions of T4Lig are lower than those for nick sealing, and T4Lig is approximately five orders of magnitude less efficient in joining blunt-ended duplexes than nick-sealing [92, 93]. T4Lig is also inefficient for ligating fragments with single base overhangs [94]. The poor fragment joining activity of T4Lig often leads to NGS failures. In order to improve the efficiency of fragment joining by T4Lig, Wilson et al. constructed a variety of chimeras of T4Lig fused with the DNA-binding

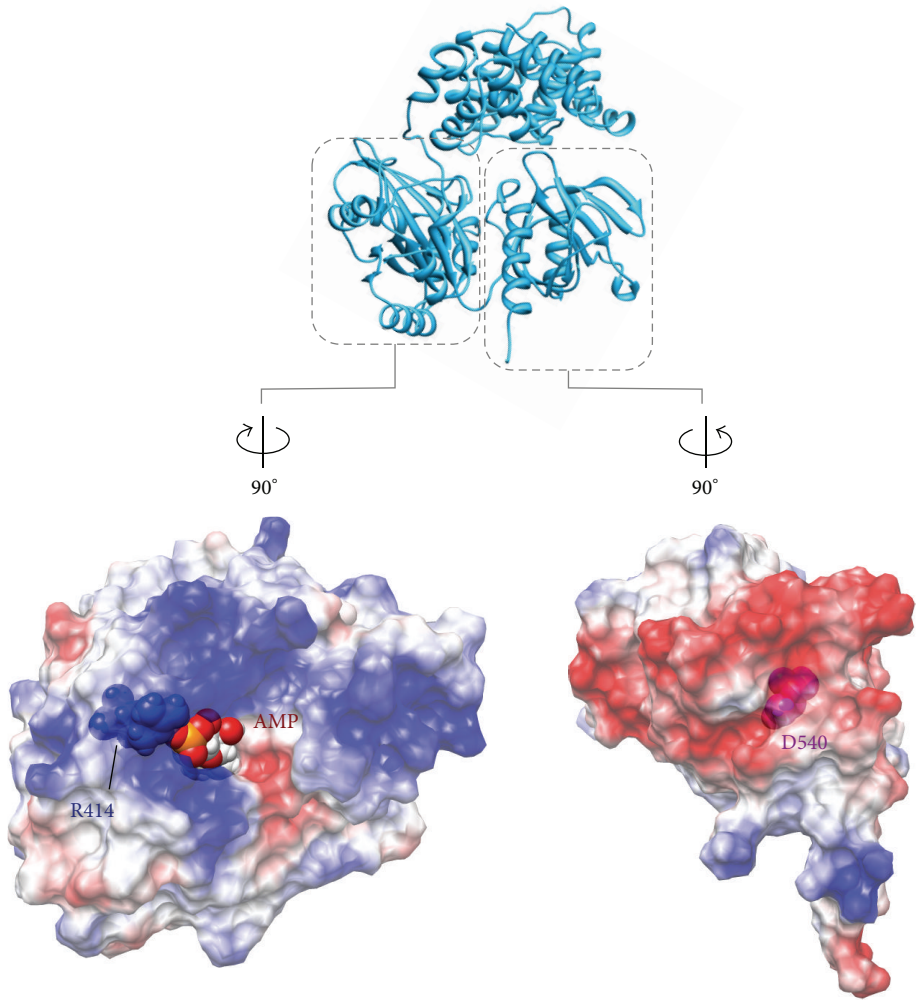


Figure 12: The electrostatic distributions on the interacting surfaces of AdD (bottom, left) and OBD (bottom, right), in which one of the ionic interaction pairs (Arg414 and Asp540) is highlighted as sphere models.

domains from other enzymes [92]. This mutational strategy was inspired by a previous report, in which the genetic fusion of a sequence of a nonspecific archaeal DNA binding protein (Sso7d from *Sulfolobus solfataricus*) to Pfu DNA polymerases resulted in considerably increased processivity and improved performance in polymerase chain reaction (PCR) amplifications [88]. The fusion of T4Lig with Sso7d showed a 60% increase in performance over T4Lig in the cloning of a blunt-ended fragment [92].

5.2. Improvement of the Fidelity of Thermostable DNA Ligase by Site-Directed Mutagenesis. The specificity of DNA ligase is exploited in LCR and LDR analyses to distinguish single base mutations associated with genetic diseases. The ligation fidelity of T4Lig is improved by the presence of spermidine and by high salt and low ligase concentrations [6, 34]. However, the increased fidelity of T4Lig by adjusting the reaction conditions is not sufficient for the ideal detection of SNPs [6, 34, 95, 96]. The thermostable DNA ligases from *Thermus aquaticus* (Taq) and *Thermus thermophilus* (Tth) exhibited far greater fidelity than that reported for T4Lig [39, 97]. The ligation reactions at the higher temperature prevented mismatched hybridizations in the substrates. Luo et al. reported the further improvement of the fidelity of TthLig by site-directed mutagenesis. The fidelity of DNA polymerases was decreased by site-directed mutagenesis at the motif associated with primer-template binding or the exoIII motif [98–100]. However, the exoIII motif mutants, also known as "antimutator" strains, showed increased fidelity in which the balanced activities of the polymerizing and $3' \rightarrow 5'$ exonuclease reactions might improve the overall fidelity [99–102]. Two mutant ligases, K294R and K294P, were identified with fidelities increased by ~4-fold and 11-fold, respectively, besides retaining their nick sealing activities [97].

5.3. Structure-Based Mutational Study of an Archaeal DNA Ligase towards Improvement of the Ligation Activity. As mentioned in the previous section, thermostable DNA ligases are

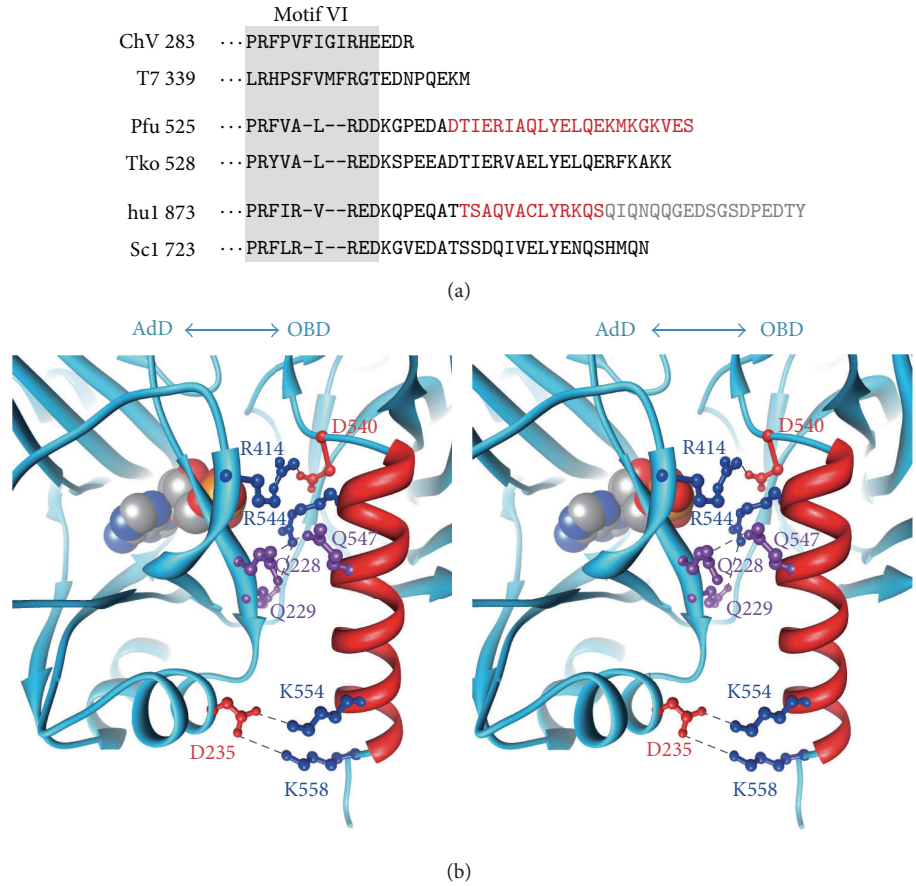


Figure 13: Roles of the C-terminal extension helix. (a) Alignment of the sequences of motif VI and the C-terminal extension. The sequences of the DNA ligases from a virus (ChV, *Chlorella* virus), bacteriophage (T7, bacteriophage T7), Archaea (Pfu, *Pyrococcus furiosus*; Tko, *Thermococcus kodakaraensis*), human (hu1, human DNA ligase I), and yeast (Sc1, *Saccharomyces cerevisiae* DNA ligase I) are aligned. The region for motif VI is shaded, and the extension helix region, determined by the crystal structures of PfuLig and hLigI, is colored red (residues after Q902 of hLigI were not included in the previously reported crystal structure and are colored gray). (b) Stereo diagram of the interface of the AdD and the OBD in PfuLig. Ball-and-stick models represent the amino acid residues involved in the interaction between the two domains: basic, acidic, and polar residues are colored blue, red, and purple, respectively. The C-terminal helix is colored red. The dotted lines indicate the polar or ionic interactions between AdD and OBD.

utilized in LCR and LDR, which require a heat denaturation step. However, thermostable DNA ligases possess weaker ligation activity at lower temperatures (20–40°C), resulting in the decreased efficiency of the LCR and LDR procedures using short probes with low-melting temperatures. Here we present a structure-guided mutational analysis of the hyperthermostable DNA ligase from *P. furiosus*, in order to improve the ligation efficiency with the thermostability [103]. In Section 4.4, we showed that the C-terminal helix in the closed structure observed in PfuLig connects the AdD and the OBD via five polar or ionic interactions (Figure 13(b)). In contrast, the crystal structures of hLigI and SsoLig revealed that the relative arrangements of the OBD and the AdD are quite different from that of PfuLig (Figure 9(a)). Taken together, we hypothesized that the binding efficiency of the DNA ligase to the DNA substrate would increase by reducing the ionic interactions between the AdD and the OBD, resulting in improved ligation efficiency. The ionic residues involved in the interactions between the AdD and the OBD were selected and mutated to create a series of mutants in which certain ionic residues are replaced or deleted. First, the series of the alanine mutants (1ala, 2ala, and 3ala), in which the ionic residues at the C-terminal helix are replaced with alanine residues, were prepared, and the series of deletion mutants (d4, d8, and d15) were also created. Then the initial reaction rates of these mutants were measured at the maximum temperature (60°C). The initial reaction rates were increased according to the number of replaced or deleted ionic residues (Figure 14(a)) [104].

Next, we found that the initial reaction rate was largely improved when the fourth ionic residue (Asp540) from the C-terminus was mutated, in addition to the 3ala mutations (D540A/3ala). Therefore the single alanine mutant of Asp540 (D540A) was created and assessed the effect of the mutation on the ligation efficiency. The reaction rate of D540A was almost the same as that of D540A/3ala [103], suggesting that the effect of the mutation of Asp540 dominates the improvement of the alanine mutations of the ionic residues

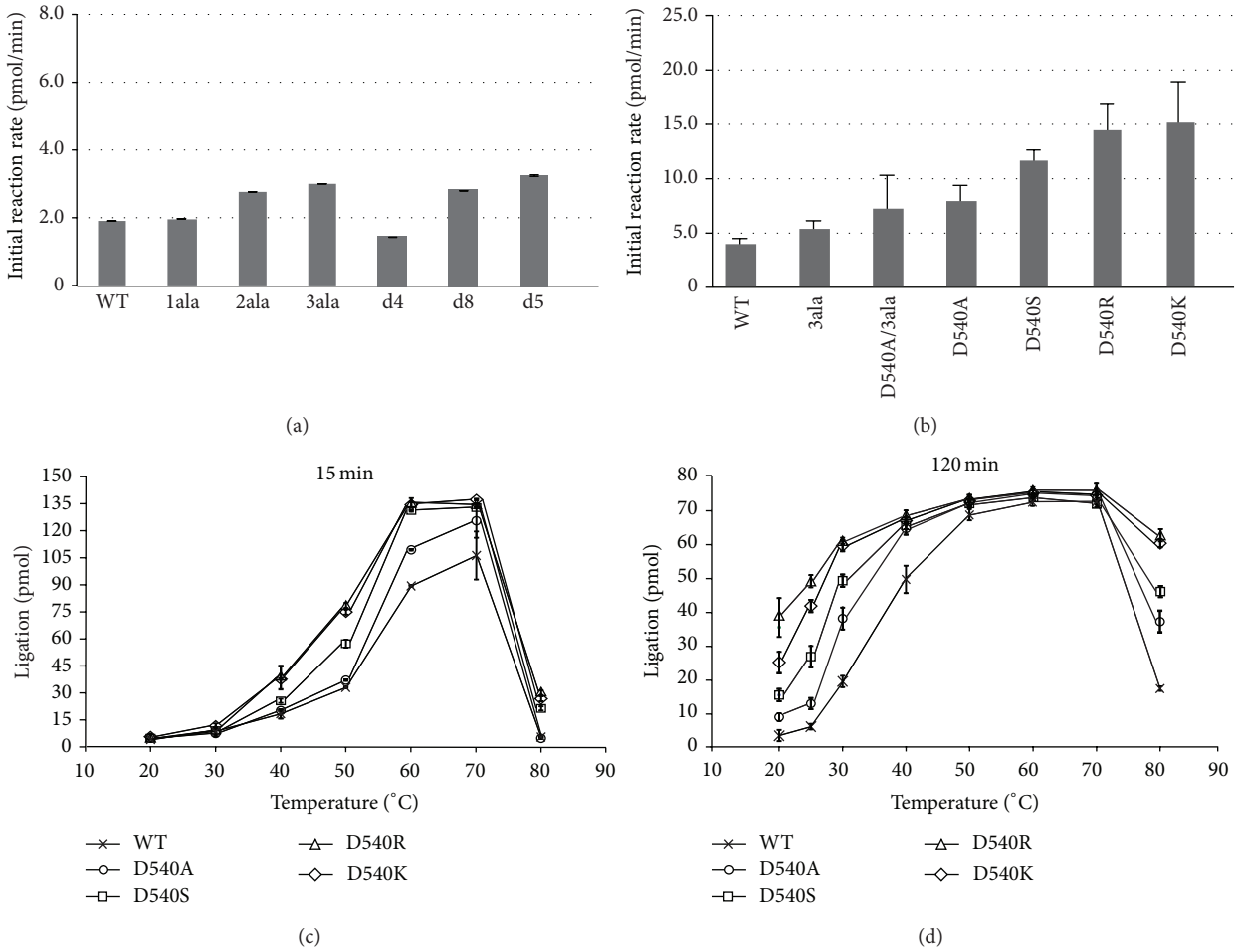


Figure 14: Nick-joining activities of the wild type and mutant proteins. (a) Initial reaction rates for the wild type (WT), K558A (1ala), K554A/K558A (2ala), Q547A/K554A/K558A (3ala), C-terminal 4 residue-deleted mutant d4 (open circles), d8 (open squares), and d15 (open triangles) enzymes, represented by time course data. (b) Initial reaction rates for WT, 3ala, D540A/3ala, D540A, D540S, D540R, and D540K. (c) Nick-joining activities of Asp540 mutants at each temperature for 15 min. The extents of nick ligation by D540A (open circles), D540S (open squares), D540R (open triangles), and D540K (open diamonds) were plotted as a function of time. (d) Nick-joining activities of Asp540 mutants at each temperature for 120 min. Error bars represent the standard deviation ($n = 3$) [103].

on the C-terminal helix. Asp540 is located at the AdD-interacting surface of the OBD and forms an ionic pair with Arg414, in the vicinity of the AMP binding site in the AdD (Figure 12).

A series of Asp540 mutants, in which Asp540 was replaced with serine (D540S), lysine (D540K), and arginine (D540R), were also constructed, and their ligation efficiencies were evaluated. As a result, the initial reaction rates of D540K and D540R were similarly the fastest among all of the mutants, followed by D540S, then D540A, and finally the wild type (Figure 14(b)). The evaluation of the ligation efficiencies of these Asp540 mutants over the broad temperature range from 20 to 80°C revealed that D540K and D540R were also the most productive. Thus, we successfully generated the PfuLig mutant with highly improved ligation efficiency over a broad temperature range, while maintaining its useful thermostability, by replacing Asp540 with a basic residue such as lysine or arginine (Figures 14(c) and 14(d)) [103].

Further investigation into the effects of the replacement of Asp540 with a basic residue on the ligation process revealed that the replacement contributed to both the adenylylation and DNA binding steps. These results suggested that the increase in the number of basic residues in the proximity of the active site was advantageous for the adenylylation step, in which the negatively charged pyrophosphate is pulled away from the ATP molecule in the active site pocket and that the alteration to the basic residue was also efficient for the AdD/OBD domain opening, caused by the repulsion between either Arg540 or Lys 540 and Arg414, which formed the ion pair with Asp540 (Figure 12) [103].

6. Conclusions

Archaea live in extreme conditions, even in the temperature range of 80°C to 100°C, and have yielded many useful enzymes for gene technology, such as hyperthermostable

DNA polymerases and DNA ligases. In this review, we have described the utilization of thermostable DNA ligases in common molecular biology protocols and the structural mechanism of DNA ligase and shown some examples of protein-engineered DNA ligases. Improvements of these enzymes are potentially effective for advancing the existing methods mentioned above and beneficial for constructing novel biotechnological methods. Several protein engineering strategies, such as fusion with other proteins or site-direct mutagenesis based on structural information, may facilitate the construction of application-specific DNA ligases in the future. The enzymes can be improved artificially, based on their three-dimensional structures, even if they have naturally evolved for a long period of time.

As many Archaea thrive in extreme environments, it will be very interesting to learn how the fidelity mechanisms of these extremophilic organisms have adapted to overcome these harsh conditions. Therefore, archaeal enzymes will continue to play an important role in future research and will be employed in numerous aspects of gene technology.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors thank Professor K. Morikawa (Kyoto University) and Dr. S. Ishino (Kyushu University) for their support. Yoshizumi Ishino was supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan (Grant nos. 23310152 and 26242075). The authors are also grateful to John Wiley & Sons Ltd., for the permission for the reuse of the figure (Figure 14) from the previous paper [103].

References

[1] B. Weiss and C. C. Richardson, "Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from Escherichia coli infected with T4 bacteriophage," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 57, no. 4, pp. 1021–1028, 1967.

[2] H.-M. Eun, "DNA ligases," in *Enzymology Primer for Recombinant DNA Technology*, pp. 109–133, Academic Press, San Diego, Calif, USA, 1996.

[3] J. M. Pascal, "DNA and RNA ligases: structural variations and shared mechanisms," *Current Opinion in Structural Biology*, vol. 18, no. 1, pp. 96–105, 2008.

[4] S. Shuman and B. Schwer, "RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyl transferases," *Molecular Microbiology*, vol. 17, no. 3, pp. 405–410, 1995.

[5] S. Shuman, "Closing the gap on DNA ligase," *Structure*, vol. 4, no. 6, pp. 653–656, 1996.

[6] U. Landegren, R. Kaiser, J. Sanders, and L. Hood, "A ligase-mediated gene detection technique," *Science*, vol. 241, no. 4869, pp. 1077–1080, 1988.

[7] O. Söderberg, M. Gullberg, M. Jarvius et al., "Direct observation of individual endogenous protein complexes *in situ* by proximity ligation," *Nature Methods*, vol. 3, no. 12, pp. 995–1000, 2006.

[8] H. S. Subramanya, A. J. Doherty, S. R. Ashford, and D. B. Wigley, "Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7," *Cell*, vol. 85, no. 4, pp. 607–615, 1996.

[9] J. M. Pascal, P. J. O'Brien, A. E. Tomkinson, and T. Ellenberger, "Human DNA ligase I completely encircles and partially unwinds nicked DNA," *Nature*, vol. 432, no. 7016, pp. 473–478, 2004.

[10] J. M. Pascal, O. V. Tsodikov, G. L. Hura et al., "A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA," *Molecular Cell*, vol. 24, no. 2, pp. 279–291, 2006.

[11] H. Nishida, S. Kiyonari, Y. Ishino, and K. Morikawa, "The closed structure of an archaeal DNA ligase from *Pyrococcus furiosus*," *Journal of Molecular Biology*, vol. 360, no. 5, pp. 956–967, 2006.

[12] S. B. Zimmerman, J. W. Little, C. K. Oshinsky, and M. Gellert, "Enzymatic joining of DNA strands: a novel reaction of diphosphopyridine nucleotide," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 57, no. 6, pp. 1841–1848, 1967.

[13] I. R. Lehman, "DNA ligase: structure, mechanism, and function," *Science*, vol. 186, no. 4166, pp. 790–797, 1974.

[14] M. J. Engler and C. C. Richardson, "DNA ligases," in *The Enzymes*, P. D. Boyer, Ed., vol. 15, pp. 3–29, Academic Press, New York, NY, USA, 1982.

[15] P. Sadowski, B. Ginsberg, A. Yudelevich, L. Feiner, and J. Hurwitz, "Enzymatic mechanisms of the repair and breakage of DNA," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 33, pp. 165–177, 1968.

[16] T. Lindahl and R. D. Wood, "Quality control by DNA repair," *Science*, vol. 286, no. 5446, pp. 1897–1905, 1999.

[17] S. Waga and B. Stillman, "The DNA replication fork in eukaryotic cells," *Annual Review of Biochemistry*, vol. 67, pp. 721–751, 1998.

[18] T. Ellenberger and A. E. Tomkinson, "Eukaryotic DNA ligases: structural and functional insights," *Annual Review of Biochemistry*, vol. 77, pp. 313–338, 2008.

[19] A. Wilkinson, J. Day, and R. Bowater, "Bacterial DNA ligases," *Molecular Microbiology*, vol. 40, no. 6, pp. 1241–1248, 2001.

[20] V. Sriskanda, R. W. Moyer, and S. Shuman, "NAD^+-dependent DNA ligase encoded by a eukaryotic virus," *The Journal of Biological Chemistry*, vol. 276, no. 39, pp. 36100–36109, 2001.

[21] Z. A. Wood, R. S. Sabatini, and S. L. Hajduk, "RNA ligase: picking up the pieces," *Molecular Cell*, vol. 13, no. 4, pp. 455–456, 2004.

[22] L. K. Wang and S. Shuman, "Structure-function analysis of yeast tRNA ligase," *RNA*, vol. 11, no. 6, pp. 966–975, 2005.

[23] R. Sawaya and S. Shuman, "Mutational analysis of the guanylyltransferase component of mammalian mRNA capping enzyme," *Biochemistry*, vol. 42, no. 27, pp. 8240–8249, 2003.

[24] S. Shuman, "DNA ligases: progress and prospects," *The Journal of Biological Chemistry*, vol. 284, no. 26, pp. 17365–17369, 2009.

[25] V. Sriskanda and S. Shuman, "Chlorella virus DNA ligase: nick recognition and mutational analysis," *Nucleic Acids Research*, vol. 26, no. 2, pp. 525–531, 1998.

[26] V. Sriskanda and S. Shuman, "Role of nucleotidyltransferase motifs I, III and IV in the catalysis of phosphodiester bond formation by Chlorella virus DNA ligase," *Nucleic Acids Research*, vol. 30, no. 4, pp. 903–911, 2002.

[27] V. Sriskanda and S. Shuman, "Mutational analysis of Chlorella virus DNA ligase: Catalytic roles of domain I and motif VI," *Nucleic Acids Research*, vol. 26, no. 20, pp. 4618–4625, 1998.

[28] P. Samai and S. Shuman, "Kinetic analysis of DNA strand joining by Chlorella virus DNA ligase and the role of nucleotidyltransferase motif VI in ligase adenylylation," *The Journal of Biological Chemistry*, vol. 287, no. 34, pp. 28609–28618, 2012.

[29] C. A. Foy and H. C. Parkes, "Emerging homogeneous DNA-based technologies in the clinical laboratory," *Clinical Chemistry*, vol. 47, no. 6, pp. 990–1000, 2001.

[30] R. C. Conaway and I. R. Lehman, "A DNA primase activity associated with DNA polymerase alpha from *Drosophila melanogaster* embryos," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, no. 8, pp. 2523–2527, 1982.

[31] H. A. Erlich, "Polymerase chain reaction," *Journal of Clinical Immunology*, vol. 9, no. 6, pp. 437–447, 1989.

[32] D. S. Levin, W. Bai, N. Yao, M. O'Donnell, and A. E. Tomkinson, "An interaction between DNA ligase I and proliferating cell nuclear antigen: implications for Okazaki fragment synthesis and joining," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 24, pp. 12863–12868, 1997.

[33] N. G. Copeland, N. A. Jenkins, and D. L. Court, "Recombineering: a powerful new tool for mouse functional genomics," *Nature Reviews Genetics*, vol. 2, no. 10, pp. 769–779, 2001.

[34] D. Y. Wu and R. B. Wallace, "Specificity of the nick-closing activity of bacteriophage T4 DNA ligase," *Gene*, vol. 76, no. 2, pp. 245–254, 1989.

[35] A. M. Alves and F. J. Carr, "Dot blot detection of point mutations with adjacently hybridising synthetic oligonucleotide probes," *Nucleic Acids Research*, vol. 16, no. 17, article 8723, 1988.

[36] D. A. Nickerson, R. Kaiser, S. Lappin, J. Stewart, L. Hood, and U. Landegren, "Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 22, pp. 8923–8927, 1990.

[37] I. A. Beck, M. Mahalanabis, G. Pepper et al., "Rapid and sensitive oligonucleotide ligation assay for detection of mutations in human immunodeficiency virus type 1 associated with high-level resistance to protease inhibitors," *Journal of Clinical Microbiology*, vol. 40, no. 4, pp. 1413–1419, 2002.

[38] F. Barany, "The ligase chain reaction in a PCR world," *Genome Research*, vol. 1, no. 1, pp. 5–16, 1991.

[39] F. Barany, "Genetic disease detection and DNA amplification using cloned thermostable ligase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 1, pp. 189–193, 1991.

[40] M. Khanna, W. Cao, M. Zirvi, P. Paty, and F. Barany, "Ligase detection reaction for identification of low abundance mutations," *Clinical Biochemistry*, vol. 32, no. 4, pp. 287–290, 1999.

[41] H. H. Lee, "Ligase chain reaction," *Biologicals*, vol. 24, no. 3, pp. 197–199, 1996.

[42] D. Y. Wu and R. B. Wallace, "The ligation amplification reaction (LAR)-amplification of specific DNA sequences using sequential rounds of template-dependent ligation," *Genomics*, vol. 4, no. 4, pp. 560–569, 1989.

[43] S. Minamitani, S. Nishiguchi, T. Kuroki, S. Otani, and T. Monna, "Detection by ligase chain reaction of precore mutant of hepatitis B virus," *Hepatology*, vol. 25, no. 1, pp. 216–222, 1997.

[44] V. D. Karthigesu, M. Mendy, M. Fortuin, H. C. Whittle, C. R. Howard, and L. M. C. Allison, "The ligase chain reaction distinguishes hepatitis B virus S-gene variants," *FEMS Microbiology Letters*, vol. 131, no. 2, pp. 127–132, 1995.

[45] C. Osiowy, "Sensitive detection of HBsAG mutants by a gap ligase chain reaction assay," *Journal of Clinical Microbiology*, vol. 40, no. 7, pp. 2566–2571, 2002.

[46] M. Zirvi, T. Nakayama, G. Newman, T. McCaffrey, P. Paty, and F. Barany, "Ligase-based detection of mononucleotide repeat sequences," *Nucleic Acids Research*, vol. 27, no. 24, pp. e40i–e40viii, 1999.

[47] C. A. Batt, P. Wagner, M. Wiedmann, and R. Gilbert, "Detection of bovine leukocyte adhesion deficiency by nonisotopic ligase chain reaction," *Animal Genetics*, vol. 25, no. 2, pp. 95–98, 1994.

[48] K. Abravaya, J. J. Carrino, S. Muldoon, and H. H. Lee, "Detection of point mutations with a modified ligase chain reaction (Gap-LCR)," *Nucleic Acids Research*, vol. 23, no. 4, pp. 675–682, 1995.

[49] V. L. Wilson, Q. Wei, K. R. Wade et al., "Needle-in-a-haystack detection and identification of base substitution mutations in human tissues," *Mutation Research*, vol. 406, no. 2–4, pp. 79–100, 1999.

[50] C. Niederhauser, L. Kaempf, and I. Heinzer, "Use of the ligase detection reaction-polymerase chain reaction to identify point mutations in extended-spectrum beta-lactamases," *European Journal of Clinical Microbiology and Infectious Diseases*, vol. 19, no. 6, pp. 477–480, 2000.

[51] C. Bourgeois, N. Sixt, J. B. Bour, and P. Pothier, "Value of a ligase chain reaction assay for detection of ganciclovir resistance-related mutation 594 in UL97 gene of human cytomegalovirus," *Journal of Virological Methods*, vol. 67, no. 2, pp. 167–175, 1997.

[52] M. Szemes, P. Bonants, M. de Weerdt, J. Baner, U. Landegren, and C. D. Schoen, "Diagnostic application of padlock probes-multiplex detection of plant pathogens using universal microarrays," *Nucleic Acids Research*, vol. 33, no. 8, article e70, 2005.

[53] M. Nilsson, H. Malmgren, M. Samiotaki, M. Kwiatkowski, B. P. Chowdhary, and U. Landegren, "Padlock probes: circularizing oligonucleotides for localized DNA detection," *Science*, vol. 265, no. 5181, pp. 2085–2088, 1994.

[54] M. Nilsson, K. Krejci, J. Koch, M. Kwiatkowski, P. Gustavsson, and U. Landegren, "Padlock probes reveal single-nucleotide differences, parent of origin and in situ distribution of centromeric sequences in human chromosomes 13 and 21," *Nature Genetics*, vol. 16, no. 3, pp. 252–255, 1997.

[55] D.-O. Antson, M. Mendel-Hartvig, U. Landegren, and M. Nilsson, "PCR-generated padlock probes distinguish homologous chromosomes through quantitative fluorescence analysis," *European Journal of Human Genetics*, vol. 11, no. 5, pp. 357–363, 2003.

[56] A. Fire and S.-Q. Xu, "Rolling replication of short DNA circles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 10, pp. 4641–4645, 1995.

[57] D. Liu, S. L. Daubendiek, M. A. Zillman, K. Ryan, and E. T. Kool, "Rolling circle DNA synthesis: small circular oligonucleotides as efficient templates for DNA polymerases," *Journal of the American Chemical Society*, vol. 118, no. 7, pp. 1587–1594, 1996.

[58] S. L. Daubendiek and E. T. Kool, "Generation of catalytic RNAs by rolling transcription of synthetic DNA nanocircles," *Nature Biotechnology*, vol. 15, no. 3, pp. 273–277, 1997.

[59] P. M. Lizardi, X. Huang, Z. Zhu, P. Bray-Ward, D. C. Thomas, and D. C. Ward, "Mutation detection and single-molecule

counting using isothermal rolling-circle amplification," *Nature Genetics*, vol. 19, no. 3, pp. 225–232, 1998.

[60] J. Banér, M. Nilsson, M. Mendel-Hartvig, and U. Landegren, "Signal amplification of padlock probes by rolling circle replication," *Nucleic Acids Research*, vol. 26, no. 22, pp. 5073–5078, 1998.

[61] C. Larsson, J. Koch, A. Nygren et al., "In situ genotyping individual DNA molecules by target-primed rolling-circle amplification of padlock probes," *Nature Methods*, vol. 1, no. 3, pp. 227–232, 2004.

[62] X.-B. Zhong, P. M. Lizardi, X.-H. Huang, P. L. Bray-Ward, and D. C. Ward, "Visualization of oligonucleotide probes and point mutations in interphase nuclei and DNA fibers using rolling circle DNA amplification," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 7, pp. 3940–3945, 2001.

[63] O. Söderberg, K.-J. Leuchowius, M. Gullberg et al., "Characterizing proteins and their interactions in cells and tissues using the *in situ* proximity ligation assay," *Methods*, vol. 45, no. 3, pp. 227–232, 2008.

[64] M. Margulies, M. Egholm, W. E. Altman et al., "Genome sequencing in microfabricated high-density picolitre reactors," *Nature*, vol. 437, no. 7057, pp. 376–380, 2005.

[65] D. R. Bentley, "Whole-genome re-sequencing," *Current Opinion in Genetics & Development*, vol. 16, no. 6, pp. 545–552, 2006.

[66] E. R. Mardis, "Next-generation DNA sequencing methods," *Annual Review of Genomics and Human Genetics*, vol. 9, pp. 387–402, 2008.

[67] M. Ronaghi, M. Uhlén, and P. Nyrén, "A sequencing method based on real-time pyrophosphate," *Science*, vol. 281, no. 5375, pp. 363–365, 1998.

[68] O. Morozova and M. A. Marra, "Applications of next-generation sequencing technologies in functional genomics," *Genomics*, vol. 92, no. 5, pp. 255–264, 2008.

[69] J. Shendure, G. J. Porreca, N. B. Reppas et al., "Accurate multiplex colony sequencing of an evolved bacterial genome," *Science*, vol. 309, no. 5741, pp. 1728–1732, 2005.

[70] J. Shendure and H. Ji, "Next-generation DNA sequencing," *Nature Biotechnology*, vol. 26, no. 10, pp. 1135–1145, 2008.

[71] J. J. Dunn and F. W. Studier, "Nucleotide sequence from the genetic left end of bacteriophage T7 DNA to the beginning of gene 4," *Journal of Molecular Biology*, vol. 148, no. 4, pp. 303–330, 1981.

[72] D. E. Barnes, L. H. Johnston, K. I. Kodama, A. E. Tomkinson, D. D. Lasko, and T. Lindahl, "Human DNA ligase I cDNA: cloning and functional expression in Saccharomyces cerevisiae," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 17, pp. 6679–6683, 1990.

[73] A. Kletzin, "Molecular characterisation of a DNA ligase gene of the extremely thermophilic archaeon *Desulfurolobus ambivalens* shows close phylogenetic relationship to eukaryotic ligases," *Nucleic Acids Research*, vol. 20, no. 20, pp. 5389–5396, 1992.

[74] A. J. Doherty, S. R. Ashford, H. S. Subramanya, and D. B. Wigley, "Bacteriophage T7 DNA ligase: overexpression, purification, crystallization, and characterization," *The Journal of Biological Chemistry*, vol. 271, no. 19, pp. 11083–11089, 1996.

[75] M. Odell, V. Sriskanda, S. Shuman, and D. B. Nikolov, "Crystal structure of eukaryotic DNA ligase-adenylate illuminates the mechanism of nick sensing and strand joining," *Molecular Cell*, vol. 6, no. 5, pp. 1183–1193, 2000.

[76] A. J. Doherty and T. R. Dafforn, "Nick recognition by DNA ligases," *Journal of Molecular Biology*, vol. 296, no. 1, pp. 43–56, 2000.

[77] V. Sriskanda and S. Shuman, "Role of nucleotidyl transferase motif V in strand joining by *Chlorella virus* DNA ligase," *The Journal of Biological Chemistry*, vol. 277, no. 12, pp. 9661–9667, 2002.

[78] D. Suck, "Common fold, common function, common origin?" *Nature Structural & Molecular Biology*, vol. 4, no. 3, pp. 161–165, 1997.

[79] A. G. Murzin, "OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences," *The EMBO Journal*, vol. 12, no. 3, pp. 861–867, 1993.

[80] K. Håkansson, A. J. Doherty, S. Shuman, and D. B. Wigley, "X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes," *Cell*, vol. 89, no. 4, pp. 545–553, 1997.

[81] A. V. Cherepanov and S. de Vries, "Dynamic mechanism of nick recognition by DNA ligase," *European Journal of Biochemistry*, vol. 269, no. 24, pp. 5993–5999, 2002.

[82] D. J. Kim, O. Kim, H.-W. Kim, H. S. Kim, S. J. Lee, and S. W. Suh, "ATP-dependent DNA ligase from *Archaeoglobus fulgidus* displays a tightly closed conformation," *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, vol. 65, no. 6, pp. 544–550, 2009.

[83] T. Petrova, E. Y. Bezsudnova, K. M. Boyko et al., "ATP-dependent DNA ligase from *Thermococcus* sp. 1519 displays a new arrangement of the OB-fold domain," *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, vol. 68, no. 12, pp. 1440–1447, 2012.

[84] M. C. Cardoso, C. Joseph, H. P. Rahn, R. Reusch, B. Nadal-Ginard, and H. Leonhardt, "Mapping and use of a sequence that targets DNA ligase I to sites of DNA replication in vivo," *The Journal of Cell Biology*, vol. 139, no. 3, pp. 579–587, 1997.

[85] C. Prigent, D. D. Lasko, K. Kodama, J. R. Woodgett, and T. Lindahl, "Activation of mammalian DNA ligase I through phosphorylation by casein kinase II," *The EMBO Journal*, vol. 11, no. 8, pp. 2925–2933, 1992.

[86] N. Keppetipola and S. Shuman, "Characterization of a thermophilic ATP-dependent DNA ligase from the euryarchaeon *Pyrococcus horikoshii*," *Journal of Bacteriology*, vol. 187, no. 20, pp. 6902–6908, 2005.

[87] H. Echols and M. F. Goodman, "Fidelity mechanisms in DNA replication," *Annual Review of Biochemistry*, vol. 60, pp. 477–511, 1991.

[88] Y. Wang, D. E. Prosen, L. Mei, J. C. Sullivan, M. Finney, and P. B. Vander Horn, "A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro," *Nucleic Acids Research*, vol. 32, no. 3, pp. 1197–1207, 2004.

[89] T. Kuroita, H. Matsumura, N. Yokota et al., "Structural mechanism for coordination of proofreading and polymerase activities in archaeal DNA polymerases," *Journal of Molecular Biology*, vol. 351, no. 2, pp. 291–298, 2005.

[90] B. H. Pheiffer and S. B. Zimmerman, "Polymer-stimulated ligation: enhanced blunt- or cohesive-end ligation of DNA or deoxyribooligonudcleotides by T4 DNA ugase in polymer solutions," *Nucleic Acids Research*, vol. 11, no. 22, pp. 7853–7871, 1983.

[91] V. Sgaramella, J. H. van de Sande, and H. G. Khorana, "Studies on polynucleotides, C. A novel joining reaction catalyzed by the

T4-polynucleotide ligase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 67, no. 3, pp. 1468–1475, 1970.

[92] R. H. Wilson, S. K. Morton, H. Deiderick et al., "Engineered DNA ligases with improved activities in vitro," *Protein Engineering, Design & Selection*, vol. 26, no. 7, pp. 471–478, 2013.

[93] A. Sugino, H. M. Goodman, H. L. Heyneker, J. Shine, H. W. Boyer, and N. R. Cozzarelli, "Interaction of bacteriophage T4 RNA and DNA ligases in joining of duplex DNA at base paired ends," *The Journal of Biological Chemistry*, vol. 252, no. 11, pp. 3987–3994, 1977.

[94] G. J. S. Lohman, S. Tabor, and N. M. Nichols, "DNA ligases," in *Current Protocols in Molecular Biology*, vol. 94, chapter 3, unit 3.14, pp. 1–7, 2011.

[95] K. Harada and L. E. Orgel, "Unexpected substrate specificity of T4 DNA ligase revealed by *in vitro* selection," *Nucleic Acids Research*, vol. 21, no. 10, pp. 2287–2291, 1993.

[96] C. Goffin, V. Bailly, and W. G. Verly, "Nicks 3′ or 5′ to AP sites or to mispaired bases, and one-nucleotide gaps can be sealed by T4 DNA ligase," *Nucleic Acids Research*, vol. 15, no. 21, pp. 8755–8771, 1987.

[97] J. Luo, D. E. Bergstrom, and F. Barany, "Improving the fidelity of *Thermus thermophilus* DNA ligase," *Nucleic Acids Research*, vol. 24, no. 15, pp. 3071–3078, 1996.

[98] W. A. Beard, S. J. Stahl, H.-R. Kim et al., "Structure/function studies of human immunodeficiency virus type 1 reverse transcriptase. Alanine scanning mutagenesis of an α-helix in the thumb subdomain," *The Journal of Biological Chemistry*, vol. 269, no. 45, pp. 28091–28097, 1994.

[99] L. J. Reha-Krantz, R. L. Nonay, and S. Stocki, "Bacteriophage T4 DNA polymerase mutations that confer sensitivity to the PPi analog phosphonoacetic acid," *Journal of Virology*, vol. 67, no. 1, pp. 60–66, 1993.

[100] L. J. Reha-Krantz and R. L. Nonay, "Motif A of bacteriophage T4 DNA polymerase: role in primer extension and DNA replication fidelity," *The Journal of Biological Chemistry*, vol. 269, no. 8, pp. 5635–5643, 1994.

[101] Q. Dong, W. C. Copeland, and T. S.-F. Wang, "Mutational studies of human DNA polymerase," *The Journal of Biological Chemistry*, vol. 268, no. 32, pp. 24163–24174, 1993.

[102] W. C. Copeland, N. K. Lam, and T. S.-F. Wang, "Fidelity studies of the human DNA polymerase α. The most conserved region among α-like DNA polymerases is responsible for metal-induced infidelity in DNA synthesis," *The Journal of Biological Chemistry*, vol. 268, no. 15, pp. 11041–11049, 1993.

[103] M. Tanabe, S. Ishino, M. Yohda, K. Morikawa, Y. Ishino, and H. Nishida, "Structure-based mutational study of an archaeal DNA ligase towards improvement of ligation activity," *ChemBioChem*, vol. 13, no. 17, pp. 2575–2582, 2012.

[104] M. Tanabe, S. Ishino, Y. Ishino, and H. Nishida, "Mutations of Asp540 and the domain-connecting residues synergistically enhance *Pyrococcus furiosus* DNA ligase activity," *FEBS Letters*, vol. 588, no. 2, pp. 230–235, 2014.

[105] E. F. Pettersen, T. D. Goddard, C. C. Huang et al., "UCSF Chimera—a visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.

Hindawi Publishing Corporation
Archaea
Volume 2015, Article ID 282035, 7 pages
http://dx.doi.org/10.1155/2015/282035

Review Article

Untapped Resources: Biotechnological Potential of Peptides and Secondary Metabolites in Archaea

James C. Charlesworth[1,2] and Brendan P. Burns[1,2]

[1] *School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW 2052, Australia*
[2] *Australian Centre for Astrobiology, University of New South Wales, Sydney, NSW 2052, Australia*

Correspondence should be addressed to Brendan P. Burns; brendan.burns@unsw.edu.au

Received 5 February 2015; Revised 7 July 2015; Accepted 8 July 2015

Academic Editor: Juergen Wiegel

Archaea are an understudied domain of life often found in "extreme" environments in terms of temperature, salinity, and a range of other factors. Archaeal proteins, such as a wide range of enzymes, have adapted to function under these extreme conditions, providing biotechnology with interesting activities to exploit. In addition to producing structural and enzymatic proteins, archaea also produce a range of small peptide molecules (such as archaeocins) and other novel secondary metabolites such as those putatively involved in cell communication (acyl homoserine lactones), which can be exploited for biotechnological purposes. Due to the wide array of metabolites produced there is a great deal of biotechnological potential from antimicrobials such as diketopiperazines and archaeocins, as well as roles in the cosmetics and food industry. In this review we will discuss the diversity of small molecules, both peptide and nonpeptide, produced by archaea and their potential biotechnological applications.

1. Introduction

Archaea are organisms that often thrive in "extreme" environments tolerating conditions that other organisms could not such as a wide range of temperatures from the lakes of Antarctica to the hot springs of Yellowstone, as well as conditions such as high salt of the dead sea or extreme pH conditions of soda lakes in Africa [1–4]. As a result of tolerance to these environmental stressors, proteins from these organisms have been highly valued in biotechnology for their stability and ability to function where other proteins would degrade. However, further research is needed to elucidate, for example, whether metabolites produced by these extremophiles have similar stability and function through chemical modifications such as carboxylation of acyl homoserine lactones discussed below. Enzymes are of particular interest, with many classes being thoroughly investigated such as proteases [5], hydrolases [6], and lipases and esterases [7]. These enzymes have been widely covered in the literature [8, 9], and as such will not be covered in this review.

However, in addition to enzymes already being utilized for biotechnological processes, recent research indicates archaea can produce a wide range of small peptides and secondary metabolites, which could be of considerable interest to biotechnology. Comparatively less is known about biosynthesis in archaea than bacteria; while vitamins and cofactors are readily produced in a number of archaeal strains, there are incomplete biosynthesis pathways possibly due to novel genes being present [10]. The same is true of a number of the metabolites examined in this study despite several having bacterial analogues, and the same genes required for biosynthesis are not present in archaea. Despite a number of recent studies and rise of new technologies such as metabolomics, little is understood regarding the full potential of archaeal metabolites. There are multiple cases of entire metabolite groups being represented at this time by a single archaeal species, most likely due to a lack of study.

In this review we discuss the variety of molecules produced by archaea and their potential role and biosynthesis

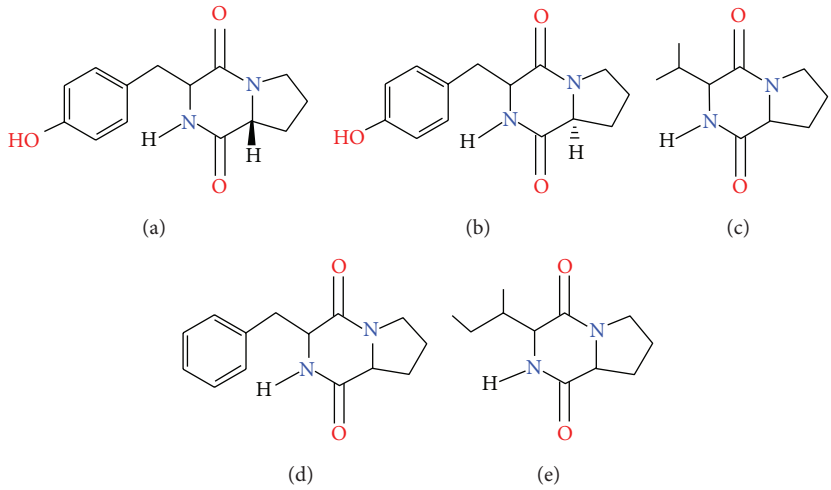


Figure 1: Structure of DKPs known to be produced by the archaeon *H. hispanica*. (a) Cyclo(D-prolyl-L-tyrosine), (b) cyclo(L-prolyl-L-tyrosine), (c) cyclo(L-prolyl-L-valine), (d) cyclo(L-prolyl-L-phenylalanine), and (e) cyclo(L-prolyl-L-isoleucine).

based on bacterial analogues and how these molecules may be exploited for biotechnological gain.

2. Archaeocins

Bacteria have widely produced antimicrobial peptides and proteins termed bacteriocins, which have been utilized in a range of industries. Archaeocins are antibiotic peptides sourced from archaea being found widely amongst haloarchaea (termed halocins) and more recently from the *Sulfolobus* genus (sulfolobicins). Halocin production is thought to be near universal amongst haloarchaea [11], and as a result of this, there is a great deal of diversity within these molecules. Halocins can be divided into two classes based on size: the smaller microhalocins that can be as small as 3.6 kDa to the larger halocins of 35 kDa [12]. The antimicrobial activity of these halocins can also range, with some halocins having narrow range of activity affecting only close relatives, as opposed to a more broadly active A4 halocin capable of inhibiting the growth of *Sulfolobus solfataricus,* a representative of another phylum of archaea [13]. While some bacteriocins are capable of inhibiting archaea [12], there is no confirmed inhibition of bacteria by a halocin, although there have been reports that halophilic archaea are capable of inhibiting halophilic bacteria [14].

One particularly relevant use of halocin producing strains is in the textile industry, which uses considerable amounts of salt in the tanning process. These conditions allow halophiles including some haloarchaea to grow which in turn can damage the product, and halocins have been used to prevent this unwanted growth [15].

Some halocins have also been considered for potential medical use. Halocin H6 has been indicated to have a potential therapeutic benefit in dogs following organ transplant surgery [16]. Halocin production is near universal and relatively few have been fully characterized, in particular their molecular diversity and differences in activity, and it is entirely possible that other therapeutic treatments could result from further study [14]. Another class of archaeocins, Sulfolobicins, were first identified in *Sulfolobus islandicus*, a member of the thermoacidophile group Sulfolobus which grow at low pH range of 2–4 and high temperatures between 65°C and 85°C. *Sulfolobus* species are members of the crenarchaeota phylum as opposed to haloarchaea which are members of the euryarchaeota phylum. Sulfolobicins produced by *S. islandicus* are able to inhibit growth of other closely related strains of *S. islandicus* and other Sulfolobus species. No activity of sulfolobicins against organisms outside of the Sulfolobus phylum has been recorded. Interestingly sulfolobicins appear to be remarkably stable, able to tolerate pH conditions between 3 and 10.7 and as high temperature as 78°C [17]. Sulfolobicins produced by *S. islandicus* appear to be associated with the cytoplasmic membrane and are not generally released into the surrounding medium [18].

3. Diketopiperazines

Diketopiperazines (DKPs) also occasionally referred to as cyclic dipeptides have been observed in many bacteria from a wide range of environments and have only recently been identified in a haloarchaeon, *Haloterrigena hispanica* [19]. A schematic illustrating some of the known DKP structures is shown in Figure 1. It is currently unknown how many archaea produce DKPs, with suggestions that *Natronococcus occultus* is also a potential producer as a result of a biosensor assay which can be activated by DKPs or AHLs [19, 20]. DKPs are synthesized in bacteria using one of two pathways, the nonribosomal peptide synthesis (NRPS) or via a novel class of enzymes called cyclic dipeptide synthases (CDPS) [21]. At present there are no functional NRPS gene clusters or CDPS genes identified in archaea. DKPs have a plethora of useful bioactivities that are of potential significance for industrial and medical purposes, including antibacterial, antifungal, and antiviral as well as antitumor activities. Other activities relating to human physiology have also been reviewed [22].

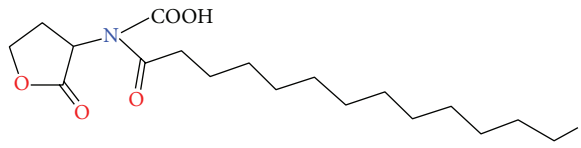


Figure 2: Chemical structure of N-carboxyl-C14-HSL: a putative AHL produced by *Methanosaeta harundinacea*. The carboxylation modification has only been observed in archaea.

Another interesting activity of DKPs is the ability to activate (and sometimes inhibit) quorum sensing systems in bacteria. Quorum sensing refers to genetic organization in which an organism can control when a phenotype is expressed based on the population levels present, that is, activating a phenotype under a high cell density. Inhibition of these quorum sensing systems in bacteria is thought to be a potential therapy for a range of pathogens such as *Pseudomonas aeruginosa* infections of cystic fibrosis patients [23]. There is also potential for quorum sensing blockers to be used more broadly in industry where biofilms can cause biofouling, a problem for a variety of industries, particularly shipping [24]. Biofilms have also been linked to difficulties in implants and catheters [25], as well as contaminating water pipe systems [26]. DKPs sourced from archaea could potentially be used to block these QS systems in order to prevent these biofilms from occurring.

4. Acyl Homoserine Lactones

Acyl homoserine lactones (AHLs) are well described metabolites controlling a range of quorum sensing phenotypes in bacteria and have recently been identified in archaea. AHLs, while predominantly produced by Gram-negative bacteria, have been shown to interact with eukaryotes as well representing a potential cross-kingdom signaling system [27]. It is important to note the AHLs currently identified in *Methanosaeta harundinacea* have a carboxylation modification previously unseen in bacteria (Figure 2); however, these carboxyl AHLs are still able to activate bacterial biosensors [28]. Short chain AHLs typically degrade quite rapidly under a range of conditions archaea may face such as high temperature and alkaline conditions [29], and thus it is likely that (any) AHL-like molecules in archaea would be either longer chain or possess other characteristics that facilitate stability. However, there is little known what properties the carboxyl modification provides for the molecule in the environment when facing such extreme conditions. Currently the only phenotype known to be AHL regulated in archaea is filament production in *M. harundinacea* [28]. An extracellular protease from extreme haloalkaliphilic *Natronococcus occultus* has also been potentially linked to AHL based quorum sensing but this has not been confirmed [20]. Other phenotypes commonly regulated by quorum sensing in bacteria, however, are present in archaea such as antimicrobial production (see above), cellular competence, and biofilm formation.

AHLs and other quorum sensing molecules are often in control of exopolysaccharide production [30], which could potentially be controlled by quorum sensing. The biofilms of haloarchaea and the extremophiles are of particular interest to biotechnology due to the exopolysaccharides and polyhydroxyalkanoates produced during this particular phenotype. These molecules have many potential uses discussed below.

Many phenotypes governed by quorum sensing in bacteria are of particular interest to bioremediation such as the expression of biofilms, flocculation, and enzyme production. It is possible with a greater understanding of archaeal quorum sensing there could be benefits to bioremediation [28–31]. AHLs are synthesized by AHL synthases, using acyl groups taken from the fatty acid synthesis pathway in bacteria via an acyl-carrier protein ACP [32]. Not only do archaea lack any traditional AHL synthase, they also lack any ACPs [33] as they have very different lipid biochemistry to bacteria, having a fatty acid pathway that may function independently of an ACP type protein. While all known AHL synthases in bacteria are members of the GCN5 acetyltransferase family, it has been suggested that FilI, a histidine kinase, is an AHL synthase in *M. harundinacea* [28].

5. Exopolysaccharides and Polyhydroxyalkanoates

Exopolysaccharides (EPS) are high molecular weight carbohydrates produced and released by many different microorganisms including archaea. EPS is thought to provide an organism protection against several environmental insults such as desiccation, predation, or ultraviolet radiation [30]. EPS production in archaea has been primarily researched within the halophile and thermophile groups [34]. EPS molecules have a number of industrial applications such as uses in the food industry as gelling or emulsifying agents. A number of bacterial EPS molecules such as xanthan, curdlan, and dextran are being produced industrially despite a number of these being produced by pathogens such as *Alcaligenes faecalis* [35]. As there are no known archaeal pathogens, it is possible a safer alternative could be found in archaea.

It has been suggested EPS producing *Haloferax* species could be used to remove heavy metal contamination from environments high in salt [36]. The biosorption process of heavy metals by EPS has been documented in other organisms [37]; however, halophilic archaea may be uniquely equipped to survive and produce EPS to reduce contamination in strictly halophilic environments [36].

Another EPS component, polyhydroxyalkanoates, is a water insoluble polymer used as a means of carbon and energy storage in bacteria and archaea. PHAs have received considerable attention in biotechnology as a potential alternative to petrochemical based plastics due to their structural properties and biodegradability [34]. PHAs at present are being produced by recombinant *E. coli* strains; however, there is the suggestion that some halophilic archaea may be ideal producers of PHAs. There are a number of producers of PHAs amongst the haloarchaea [38], the best of which may be *H. medeterrani* with 65% of the dry cell weight being PHA when fed glucose or starch [39]. As *H. medeterrani* needs 22% salt to grow optimally, there is less of a need for sterile

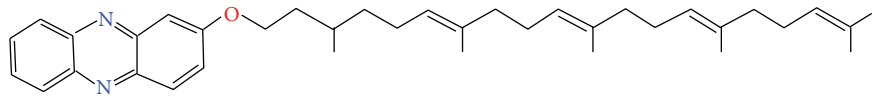


Figure 3: Chemical structure of methanophenazine, a phenazine observed only in archaea.

conditions as there is a reduced chance of contamination [40]. The need for high salt also provides an ease of cell lysis by simply changing to fresh water [40]. Current research revolves around improving the production of PHAs through the knockout of nonessential genes, such as those involved in EPS production [41].

6. Carotenoids

Carotenoids are naturally occurring pigments that are commonly found in haloarchaea and are responsible for the recognizable reddish pigmentation of the organisms. Carotenoids are widely used as food supplements and colouring agents resulting in an estimated market value of $1.02 billion in 2009 [42]. The role of carotenoids in human health is still being researched; however, some potential benefits have been identified, such as the prevention of chronic diseases and disorders such as cancer [43], chronic heart disease [44], and osteoporosis. There are many carotenoids in nature produced by bacteria, eukarya, and archaea, though few are able to be either chemically synthesized or extracted at a high enough level for industry [42]. Canthaxanthin is a carotenoid used by industry as a feed additive for chickens, fish, and crustacean farms and used in cosmetics. Due to the properties of carnathaxin and its many usages there has been continuing research into producing it in larger quantities. The halophilic archaeon *Haloferax alexandrinus* produces canthaxanthin at a high enough level it could be considered for commercial production of canthaxanthin [45]. It has been suggested by using *H. alexandrinus* that there is less of a concern for aseptic culturing as the higher salt conditions reduce the chance for contamination. As for PHAs described earlier, another benefit to using *H. alexandrinus* to produce carnathaxin is the ease at which cells lyse in fresh water making the carnathaxin easier to extract [45].

7. Biosurfactants

Biosurfactants are surfactants produced by a wide range of organisms including archaea and are a mixture of glycolipids, fatty acids, proteins, and sugars. Biosurfactants show many advantages to traditional chemically derived surfactants by being biodegradable, nontoxic, renewable, and active under a range of extreme conditions [46]. Biosurfactants may be able to assist bioremediation of oil spills in soil and water samples, as well as a range of other uses in the food [47], cosmetic, and pharmaceutical industries. Several halophilic archaea have been identified to produce biosurfactants at the highest salt concentration recorded [48]. A species of *Natrialba* was isolated from a solar saltern in Algeria that was found to produce a biosurfactant [49] which was constituted of sugar protein and lipids including rhamnolipids, a class of glycolipid already suggested for use in bioremediation [50] and cosmetics [51]. It was suggested that this *Natrialba* species or possibly other halophilic archaea may be an ideal choice for assisting in the bioremediation of oil spills in saline and hypersaline environments, which can be often contaminated through industrial processes [49].

8. Phenazines

Phenazines are naturally occurring compounds commonly produced by bacterial genera such as *Pseudomonas* and *Streptomyces* and are known for a variety of biological activities such as antibacterial, antiparasitic, and antitumour effects [52]. While there are a variety of bacterially produced phenazines, there is also an example of a phenazine (methanophenazine) produced by an array of *Methanosarcina* including *Methanosarcina mazei* [53]. An example of such a phenazine structure is shown in Figure 3. This methanophenazine appears to play a key role in the metabolism of the organism as an electron carrier, unlike most other phenazines that are typically considered secondary metabolites [53]. It is unknown whether phenazine production in archaea is limited *to M. mazei* or if it is more widespread. The biosynthesis of phenazines in bacteria typically stems from nonribosomal peptide synthesis, a process responsible for many secondary metabolites and natural products, though as mentioned above there are no known NRPS systems detected in archaea.

9. Organic Solutes from Archaea

Organic solutes, also referred to as osmolytes, compatible solutes, and osmoprotectants, are small molecules that assist microorganisms in surviving saline conditions. Organic solutes have several suggested usages in biotechnology as preservatives or cryoprotectants of enzymes and other organic molecules as well as other potential roles such as cosmetics [54]. Organic solutes are widely used by all three kingdoms; dunaliella, an algal species, uses glycine betaine and among bacteria ectoine and glycine betaine are most common [55]. In archaea there are several groups that produce organic solutes such as the halophilic methanogens that produce a range of organic solutes [56]. A novel class of organic solutes, 2-sulfotrehalose, was discovered in the haloalkaphiles such as *N. occultus* [57] and was later found in other related species such as *Natrialba magadii and Halalkalicoccus jeotgali* [58]. Hyperthermophiles also produce a range of compatible solutes considered potentially useful for industry such as trehalose produced by members of the *Sulfolobus* genus which could be used as a cryoprotectant

of mammalian cells, as well as having roles in cosmetics and food industries [59]. Other examples such as diglycerol phosphate produced by *Archaeoglobus fulgidus* [60] or di-myo-inositol phosphate from *Pyrococcus furiosus* have been suggested as protein stabilizers under high temperatures.

10. Conclusions

While archaea apparently lack traditional gene clusters for natural product biosynthesis such as NRPS, archaea still display a wide diversity of metabolites. Although some metabolites produced by archaea are already in use, others would require significantly more research and development before the applications become apparent and/or economically feasible. Despite this, archaea represent an untapped resource in the field of natural product discovery and could contribute to many areas of industry if fully utilized.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] P. Franzmann, E. Stackebrandt, K. Sanderson et al., "*Halobacterium lacusprofundi* sp. nov., a Halophilic bacterium isolated from deep lake, Antarctica," *Systematic and Applied Microbiology*, vol. 11, no. 1, pp. 20–27, 1988.

[2] C. Lapaglia and P. L. Hartzell, "Stress-induced production of biofilm in the hyperthermophile *Archaeoglobus fulgidus*," *Applied and Environmental Microbiology*, vol. 63, no. 8, pp. 3158–3163, 1997.

[3] A. Oren, P. Gurevich, R. T. Gemmell, and A. Teske, "*Halobaculum gomorrense* gen. nov., sp. nov., a novel extremely halophilic archaeon from the dead sea," *International Journal of Systematic Bacteriology*, vol. 45, no. 4, pp. 747–754, 1995.

[4] B. J. Tindall, H. N. M. Ross, and W. D. Grant, "*Natronobacterium* gen. nov. and *Natronococcus* gen. nov., two new genera of haloalkaliphilic archaebacteria," *Systematic and Applied Microbiology*, vol. 5, no. 1, pp. 41–57, 1984.

[5] M. I. Giménez, C. A. Studdert, J. J. Sánchez, and R. E. De Castro, "Extracellular protease of *Natrialba magadii*: purification and biochemical characterization," *Extremophiles*, vol. 4, no. 3, pp. 181–188, 2000.

[6] R. M. Camacho, J. C. Mateos-Díaz, D. M. Diaz-Montaño, O. González-Reynoso, and J. Córdova, "Carboxyl ester hydrolases production and growth of a halophilic archaeon, *Halobacterium* sp. NRC-1," *Extremophiles*, vol. 14, no. 1, pp. 99–106, 2009.

[7] B. Ozcan, G. Ozyilmaz, C. Cokmus, and M. Caliskan, "Characterization of extracellular esterase and lipase activities from five halophilic archaeal strains," *Journal of Industrial Microbiology & Biotechnology*, vol. 36, no. 1, pp. 105–110, 2009.

[8] R. Margesin and F. Schinner, "Potential of halotolerant and halophilic microorganisms for biotechnology," *Extremophiles*, vol. 5, no. 2, pp. 73–83, 2001.

[9] E. Lévêque, Š. Janeček, B. Haye, and A. Belarbi, "Thermophilic archaeal amylolytic enzymes," *Enzyme and Microbial Technology*, vol. 26, no. 1, pp. 3–14, 2000.

[10] M. Falb, K. Müller, L. Königsmaier et al., "Metabolism of halophilic archaea," *Extremophiles*, vol. 12, no. 2, pp. 177–196, 2008.

[11] M. Torreblanca, I. Meseguer, and A. Ventosa, "Production of halocin is a practically universal feature of archaeal halophilic rods," *Letters in Applied Microbiology*, vol. 19, no. 4, pp. 201–205, 1994.

[12] E. O'Connor and R. Shand, "Halocins and sulfolobicins: the emerging story of archaeal protein and peptide antibiotics," *Journal of Industrial Microbiology and Biotechnology*, vol. 28, no. 1, pp. 23–31, 2002.

[13] C. Haseltine, T. Hill, R. Montalvo-Rodriguez, S. K. Kemper, R. F. Shand, and P. Blum, "Secreted euryarchaeal microhalocins kill hyperthermophilic crenarchaea," *Journal of Bacteriology*, vol. 183, no. 1, pp. 287–291, 2001.

[14] R. F. Shand, K. J. Leyva, and R. F. Shand, "Peptide and protein antibiotics from the domain *Archaea*: halocins and sulfolobicins," in *Bacteriocins: Ecology and Evolution*, M. A. Riley and M. A. Chavan, Eds., pp. 93–109, Springer, Berlin, Germany, 2007.

[15] M. Birbir, S. Eryilmaz, and A. Ogan, "Prevention of halophilic microbial damage on brine cured hides by extremely halophilic halocin producer strains," *Journal of the Society of Leather Technologies and Chemists*, vol. 88, no. 3, pp. 99–104, 2004.

[16] J. L. Lequerica, J. E. O'Connor, L. Such et al., "A halocin acting on Na^+/H^+ exchanger of Haloarchaea as a new type of inhibitor in NHE of mammals," *Journal of Physiology and Biochemistry*, vol. 62, no. 4, pp. 253–262, 2006.

[17] A. F. Ellen, O. V. Rohulya, F. Fusetti, M. Wagner, S.-V. Albers, and A. J. M. Driessen, "The sulfolobicin genes of *Sulfolobus acidocaldarius* encode novel antimicrobial proteins," *Journal of Bacteriology*, vol. 193, no. 17, pp. 4380–4387, 2011.

[18] D. Prangishvili, I. Holz, E. Stieger, S. Nickell, J. K. Kristjansson, and W. Zillig, "Sulfolobicins, specific proteinaceous toxins produced by strains of the extremely thermophilic archaeal genus *Sulfolobus*," *Journal of Bacteriology*, vol. 182, no. 10, pp. 2985–2988, 2000.

[19] G. Tommonaro, G. R. Abbamondi, C. Iodice, K. Tait, and S. De Rosa, "Diketopiperazines produced by the halophilic archaeon, *Haloterrigena hispanica*, activate AHL bioreporters," *Microbial Ecology*, vol. 63, no. 3, pp. 490–495, 2012.

[20] R. A. Paggi, C. B. Martone, C. Fuqua, and R. E. de Castro, "Detection of quorum sensing signals in the haloalkaliphilic archaeon *Natronococcus occultus*," *FEMS Microbiology Letters*, vol. 221, no. 1, pp. 49–52, 2003.

[21] P. Belin, M. Moutiez, S. Lautru, J. Seguin, J.-L. Pernodet, and M. Gondry, "The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways," *Natural Product Reports*, vol. 29, no. 9, pp. 961–979, 2012.

[22] M. B. Martins and I. Carvalho, "Diketopiperazines: biological activity and synthesis," *Tetrahedron*, vol. 63, no. 40, pp. 9923–9932, 2007.

[23] M. Hentzer, H. Wu, J. B. Andersen et al., "Attenuation of *Pseudomonas aeruginosa* virulence by quorum sensing inhibitors," *The EMBO Journal*, vol. 22, no. 15, pp. 3803–3815, 2003.

[24] M. P. Schultz, J. A. Bendick, E. R. Holm, and W. M. Hertel, "Economic impact of biofouling on a naval surface ship," *Biofouling*, vol. 27, no. 1, pp. 87–98, 2011.

[25] I. Raad, H. Hanna, Y. Jiang et al., "Comparative activities of daptomycin, linezolid, and tigecycline against catheter-related methicillin-resistant *Staphylococcus* bacteremic isolates embedded in biofilm," *Antimicrobial Agents and Chemotherapy*, vol. 51, no. 5, pp. 1656–1660, 2007.

[26] M. J. Lehtola, I. T. Miettinen, M. M. Keinänen et al., "Microbiology, chemistry and biofilm development in a pilot drinking water distribution system with copper and plastic pipes," *Water Research*, vol. 38, no. 17, pp. 3769–3779, 2004.

[27] I. Joint, K. Tait, and G. Wheeler, "Cross-kingdom signalling: exploitation of bacterial quorum sensing molecules by the green seaweed Ulva," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1483, pp. 1223–1233, 2007.

[28] G. Zhang, F. Zhang, G. Ding et al., "Acyl homoserine lactone-based quorum sensing in a methanogenic archaeon," *The ISME Journal*, vol. 6, pp. 1336–1344, 2012.

[29] E. A. Yates, B. Philipp, C. Buckley et al., "*N*-acylhomoserine lactones undergo lactonolysis in a pH-, temperature-, and acyl chain length-dependent manner during growth of *Yersinia pseudotuberculosis* and *Pseudomonas aeruginosa*," *Infection and Immunity*, vol. 70, no. 10, pp. 5635–5646, 2002.

[30] S. Fröls, "Archaeal biofilms: widespread and complex," *Biochemical Society Transactions*, vol. 41, no. 1, pp. 393–398, 2013.

[31] A. Valle, M. J. Bailey, A. S. Whiteley, and M. Manefield, "N-acyl-l-homoserine lactones (AHLs) affect microbial community composition and function in activated sludge," *Environmental Microbiology*, vol. 6, no. 4, pp. 424–433, 2004.

[32] A. L. Schaefer, D. L. Val, B. L. Hanzelka, J. E. Cronan Jr., and E. P. Greenberg, "Generation of cell-to-cell signals in quorum sensing: acyl homoserine lactone synthase activity of a purified Vibrio fischeri LuxI protein," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 18, pp. 9505–9509, 1996.

[33] J. Lombard, P. López-García, and D. Moreira, "An ACP-independent fatty acid synthesis pathway in archaea: implications for the origin of phospholipids," *Molecular Biology and Evolution*, vol. 29, no. 11, pp. 3261–3265, 2012.

[34] A. Poli, P. Di Donato, G. R. Abbamondi, and B. Nicolaus, "Synthesis, production, and biotechnological applications of exopolysaccharides and polyhydroxyalkanoates by Archaea," *Archaea*, vol. 2011, Article ID 693253, 13 pages, 2011.

[35] B. Nicolaus, M. Kambourova, and E. T. Oner, "Exopolysaccharides from extremophiles: from fundamentals to biotechnology," *Environmental Technology*, vol. 31, no. 10, pp. 1145–1158, 2010.

[36] G. Popescu and L. Dumitru, "Biosorption of some heavy metals from media with high salt concentrations by halophilic *Archaea*," *Biotechnology & Biotechnological Equipment*, vol. 23, supplement 1, pp. 791–795, 2009.

[37] M. P. Acosta, E. Valdman, S. G. F. Leite, F. Battaglini, and S. M. Ruzal, "Biosorption of copper by *Paenibacillus polymyxa* cells and their exopolysaccharide," *World Journal of Microbiology and Biotechnology*, vol. 21, no. 6-7, pp. 1157–1163, 2005.

[38] A. Legat, C. Gruber, K. Zangger, G. Wanner, and H. Stan-Lotter, "Identification of polyhydroxyalkanoates in *Halococcus* and other haloarchaeal species," *Applied Microbiology and Biotechnology*, vol. 87, no. 3, pp. 1119–1127, 2010.

[39] J. Quillaguamán, H. Guzmán, D. Van-Thuoc, and R. Hatti-Kaul, "Synthesis and production of polyhydroxyalkanoates by halophiles: current potential and future prospects," *Applied Microbiology and Biotechnology*, vol. 85, no. 6, pp. 1687–1696, 2010.

[40] M. Koller, P. Hesse, R. Bona, C. Kutschera, A. Atlić, and G. Braunegg, "Potential of various archae- and eubacterial strains as industrial polyhydroxyalkanoate producers from whey," *Macromolecular Bioscience*, vol. 7, no. 2, pp. 218–226, 2007.

[41] D. Zhao, L. Cai, J. Wu et al., "Improving polyhydroxyalkanoate production by knocking out the genes involved in exopolysaccharide biosynthesis in *Haloferax mediterranei*," *Applied Microbiology and Biotechnology*, vol. 97, no. 7, pp. 3027–3036, 2013.

[42] G. K. Chandi and B. S. Gill, "Production and characterization of microbial carotenoids as an alternative to synthetic colors: a review," *International Journal of Food Properties*, vol. 14, no. 3, pp. 503–513, 2011.

[43] T. Tanaka, M. Shnimizu, and H. Moriwaki, "Cancer chemoprevention by carotenoids," *Molecules*, vol. 17, no. 3, pp. 3202–3242, 2012.

[44] H. D. Sesso, J. E. Buring, E. P. Norkus, and J. M. Gaziano, "Plasma lycopene, other carotenoids, and retinol and the risk of cardiovascular disease in women," *American Journal of Clinical Nutrition*, vol. 79, no. 1, pp. 47–53, 2004.

[45] D. Asker and Y. Ohta, "*Haloferax alexandrinus* sp. nov., an extremely halophilic canthaxanthin-producing archaeon from a solar saltern in Alexandria (Egypt)," *International Journal of Systematic and Evolutionary Microbiology*, vol. 52, no. 3, pp. 729–738, 2002.

[46] I. M. Banat, R. S. Makkar, and S. S. Cameotra, "Potential commercial applications of microbial surfactants," *Applied Microbiology and Biotechnology*, vol. 53, no. 5, pp. 495–508, 2000.

[47] D. P. Sachdev and S. S. Cameotra, "Biosurfactants in agriculture," *Applied Microbiology and Biotechnology*, vol. 97, no. 3, pp. 1005–1016, 2013.

[48] S. Kebbouche-Gana, M. L. Gana, S. Khemili et al., "Isolation and characterization of halophilic Archaea able to produce biosurfactants," *Journal of Industrial Microbiology and Biotechnology*, vol. 36, no. 5, pp. 727–738, 2009.

[49] S. Kebbouche-Gana, M. L. Gana, I. Ferrioune et al., "Production of biosurfactant on crude date syrup under saline conditions by entrapped cells of *Natrialba* sp. strain E21, an extremely halophilic bacterium isolated from a solar saltern (Ain Salah, Algeria)," *Extremophiles*, vol. 17, no. 6, pp. 981–993, 2013.

[50] Y. Zhang and R. M. Miller, "Enhanced octadecane dispersion and biodegradation by a *Pseudomonas rhamnolipid* surfactant (biosurfactant)," *Applied and Environmental Microbiology*, vol. 58, no. 10, pp. 3276–3282, 1992.

[51] M. J. Brown, "Biosurfactants for cosmetic applications," *International Journal of Cosmetic Science*, vol. 13, no. 2, pp. 61–64, 1991.

[52] J. B. Laursen and J. Nielsen, "Phenazine natural products: biosynthesis, synthetic analogues, and biological activity," *Chemical Reviews*, vol. 104, no. 3, pp. 1663–1685, 2004.

[53] H.-J. Abken, M. Tietze, J. Brodersen, S. Bäumer, U. Beifuss, and U. Deppenmeier, "Isolation and characterization of methanophenazine and function of phenazines in membrane-bound electron transport of Methanosarcina mazei Go1," *Journal of Bacteriology*, vol. 180, no. 8, pp. 2027–2032, 1998.

[54] G. Lentzen and T. Schwarz, "Extremolytes: natural compounds from extremophiles for versatile applications," *Applied Microbiology and Biotechnology*, vol. 72, no. 4, pp. 623–634, 2006.

[55] A. Oren, "Diversity of halophilic microorganisms: environments, phylogeny, physiology, and applications," *Journal of Industrial Microbiology and Biotechnology*, vol. 28, no. 1, pp. 56–63, 2002.

[56] M.-C. Lai, K. R. Sowers, D. E. Robertson, M. F. Roberts, and R. P. Gunsalus, "Distribution of compatible solutes in the halophilic methanogenic archaebacteria," *Journal of Bacteriology*, vol. 173, no. 17, pp. 5352–5358, 1991.

[57] D. Desmarais, P. E. Jablonski, N. S. Fedarko, and M. F. Roberts, "2-sulfotrehalose, a novel osmolyte in haloalkaliphilic archaea," *Journal of Bacteriology*, vol. 179, no. 10, pp. 3146–3153, 1997.

[58] N. H. Youssef, K. N. Savage-Ashlock, A. L. McCully et al., "Trehalose/2-sulfotrehalose biosynthesis and glycine-betaine uptake are widely spread mechanisms for osmoadaptation in the Halobacteriales," *ISME Journal*, vol. 8, no. 3, pp. 636–649, 2014.

[59] C. Schiraldi, I. Di Lernia, and M. De Rosa, "Trehalose production: exploiting novel approaches," *Trends in Biotechnology*, vol. 20, no. 10, pp. 420–425, 2002.

[60] L. O. Martins, R. Huber, H. Huber, K. O. Stetter, M. S. Da Costa, and H. Santos, "Organic solutes in hyperthermophilic Archaea," *Applied and Environmental Microbiology*, vol. 63, no. 3, pp. 896–902, 1997.

Hindawi Publishing Corporation
Archaea
Volume 2015, Article ID 147671, 10 pages
http://dx.doi.org/10.1155/2015/147671

Review Article

Archaeal Enzymes and Applications in Industrial Biocatalysts

Jennifer A. Littlechild

Henry Wellcome Building for Biocatalysis, Biosciences, College of Life and Environmental Sciences, Stocker Road, Exeter EX4 4QD, UK

Correspondence should be addressed to Jennifer A. Littlechild; j.a.littlechild@exeter.ac.uk

Received 23 July 2015; Accepted 19 August 2015

Academic Editor: Juergen Wiegel

Archaeal enzymes are playing an important role in industrial biotechnology. Many representatives of organisms living in "extreme" conditions, the so-called Extremophiles, belong to the archaeal kingdom of life. This paper will review studies carried by the Exeter group and others regarding archaeal enzymes that have important applications in commercial biocatalysis. Some of these biocatalysts are already being used in large scale industrial processes for the production of optically pure drug intermediates and amino acids and their analogues. Other enzymes have been characterised at laboratory scale regarding their substrate specificity and properties for potential industrial application. The increasing availability of DNA sequences from new archaeal species and metagenomes will provide a continuing resource to identify new enzymes of commercial interest using both bioinformatics and screening approaches.

1. Introduction

The application of enzymes in "White Biotechnology" for the synthesis of industrially important chiral compounds is becoming increasingly important for the pharmaceutical industry. Many companies who were traditionally not incorporating biocatalysis in their drug production programmes are now very keen to develop the technology. Enzyme chemistry can make reactions feasible which are currently unavailable using conventional methods. Use of enzymes for chemical processes is a route to lower energy consumption and reduced waste generation. In addition, the selectivity of enzymatic processes reduces the raw material costs and the safety issues surrounding the production of wasteful biproducts. It is anticipated that optimised enzyme production through further bioprocess intensification will lead to more economically viable and cost effective, sustainable compound production.

The wealth of genome and metagenome data now available makes searching for enzymes using both advanced bioinformatic and substrate screening approaches an area for development. Also more representatives of the different classified enzyme groups are being investigated for their application in industrial biocatalytic processes. The enzyme process is often used in a "cascade reaction" with traditional chemistry synthetic steps. When "nonnatural" industrial substrates are presented to enzymes found in "nature" it has been found that different classes of enzyme can use the same nonnatural compound as a substrate to carry out a specific biotransformation. This makes it difficult to predict which class of enzyme should be best for the biocatalytic process. Also using the enzyme in the presence of solvents or at nonphysiological pH can result in "side reactions" which are different from the normal activity of the enzyme catalyst.

The development of novel, efficient, and cost effective biocatalytic processes in a variety of industries is currently limited by the number of robust, highly selective, and useful biocatalysts. This paper will concentrate on specific novel enzymes from the archaeal kingdom that have been isolated from thermophilic marine and terrestrial environments. Thermophilic enzymes from archaea offer additional novelty in relation to those from thermophilic bacteria since they have been shown to be more primitive enzymes. An example of this is the *Sulfolobus solfataricus* Glyceraldehyde phosphate dehydrogenase (GAPDH) [1] which has the catalytic cysteine on the same secondary structure as other bacterial and eukaryotic ADHs but other residues involved in catalysis are presented into the active site from different secondary structural elements. This enzyme has only 18% sequence identity to other well-characterised GAPDHs and

was thought to have a different overall structure until the crystal structure of the archaeal enzyme was determined. Some archaeal enzymes have evolved by a different route to their bacterial or eukaryotic equivalents in so much that they are a combination of different enzymes such as the L-aminoacylase from *Thermococcus litoralis* [2] that has only L-aminoacylase activity but is related regarding its sequence similarity to a carboxypeptidase enzyme from a *Sulfolobus* species.

Many of the archaeal species have novel metabolic pathways that are not found in other kingdoms of life. For example, some utilise modified versions of the canonical Embden Meyerhof and Entner-Doudoroff pathway involving a large number of novel enzymes [3] and have unusual pentose degradation pathways.

Some archaeal enzymes are more promiscuous in their activity than comparable enzymes from bacteria or eukaryotes: for example, the *S. solfataricus, Picrophilus torridus* glucose/galactose dehydrogenase dehydratase, and gluconate/galactonate dehydratase [4–6]. The archaeon *S. solfataricus* has a promiscuity in 2-keto-3-deoxygluconate aldolase from the Entner-Doudoroff pathway which is able to cleave KDG and D-2-keto-3-deoxygalactonate (KDGal) to produce pyruvate and D-glyceraldehyde. The aldolase also exhibits a lack of stereoselectivity in the reversible condensation reaction of pyruvate and D-glyceraldehyde. An understanding of the structural basis of the promiscuity has been studied [7].

Enzymes isolated from thermophilic archaea usually are more stable to high temperature, presence of solvents, and resistance to proteolysis which together are ideal features for industrial applications. Stability of an enzyme is dependent on maintenance of a functional structure, and the stability of any protein is marginal and equivalent to a small number of molecular interactions [8]. This remains the case with a thermostable protein, the only difference being that the free energy of stabilisation is slightly higher than that of its mesophilic counterpart [9]. The active form of a protein is usually held together by a combination of noncovalent forces including hydrogen bonds, ion pairs, hydrophobic bonds, and Van der Waals interactions. When these interactions are disrupted, for example, by elevated temperatures, both mesophilic and thermophilic proteins unfold into inactive but kinetically stable structures. Once unfolded in this manner the protein is prone to aggregation and chemical modification. Aggregation occurs when the hydrophobic residues of a protein that have been exposed by the polypeptide chain unfolding interact with hydrophobic residues from other protein molecules, which usually follows immediately after unfolding. Chemical modifications of the protein can include cysteine oxidation, deamination of asparagine and glutamine residues, and peptide bond hydrolysis. The unfolding of the protein may be reversible for smaller proteins but is usually irreversible with larger proteins.

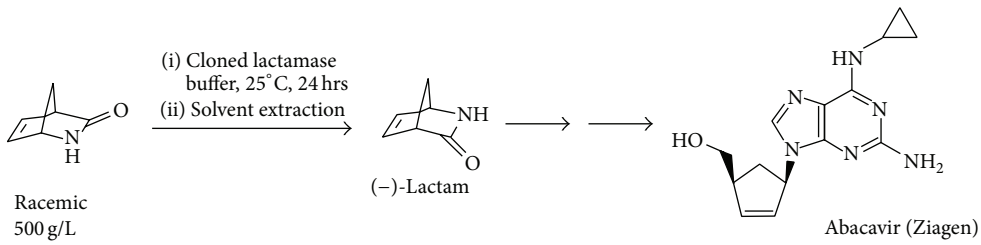
Although the optimal temperature for activity of the thermostable archaeal enzymes is above what would normally be used in an industrial process they can be used at lower temperatures and usually maintain at least 20% of their maximum activity at ambient temperature. There is often an advantage to running a process at temperatures between 50 and 60°C since at this temperature many nonnatural substrates which are insoluble at room temperature become soluble at the higher temperature. The temperature for operation of the industrial process needs to be balanced against the overall economics of the biocatalytic conversion.

Many thermophilic enzymes can be cloned and overexpressed in a soluble form using a mesophilic host (*Escherichia coli*) and can be easily purified from the cell extract by a straight forward heat treatment which precipitates most of the mesophilic proteins. The features responsible for the increased thermophilicity can be identified by studying the biochemical and structural features of a range of purified thermophilic proteins [10]. These include an increase in ionic interactions and often large ionic networks are observed within the protein and at the subunit interfaces. This is observed especially in hyperthermophilic proteins. The α-helices in the protein can be "capped" so that an acidic amino acid is placed to neutralise the charge of the protein at the amino end of the helix and a basic amino acid to neutralise the charge at the acidic end of the helix. Many thermophilic proteins have increased hydrophobicity within their interior and at subunit interfaces. This is especially true for thermophilic proteins from *Sulfolobus* species. They often have increased packing such as additional secondary structures and C-terminal extensions which can pack into the protein to fill unnecessary voids. Thermophilic proteins usually have shorter surface loops and often the internal loops can be stabilised by metal ions. An increased content of proline residues is seen in some thermophilic bacteria such as *Thermus* species which have a high G-C content in their DNA. Generally there is a reduction in amino acids that are unstable at high temperatures such as asparagines and cysteines except where they play an important catalytic role. Finally some species especially some of the aerobic archaea such as *Aeropyrum* species use the introduction of a covalent disulfide bond into the protein to offer the required stability to high temperature.

This paper will address some important industrially relevant biocatalytic reactions which can be carried out using archaeal enzymes.

2. Biocatalytic Industrial Applications

2.1. Production of Carbocyclic Nucleotides. Carbocyclic nucleosides are valuable chemotherapeutic agents such as cardiac vasodilators and are used in the treatment of viral infection. New antiviral compounds are especially important for the treatment of HIV since they act as nucleotide inhibitors of the viral reverse transcriptase enzyme. The viral proteins can readily mutate in order to overcome the inhibition. New inhibitors therefore have to be continually developed. An enzyme from the archaeon *S. solfataricus* MT4 can use the bicyclic synthon (rac)-γ-lactam(2-azabicyclo[2.2.1]hept-5-en-3-one) as a substrate to obtain a single enantiomer of the γ-bicyclic lactam product which is an important building block for the anti-HIV compound, Abacavir (Scheme 1) [11]. This (+)-γ-lactamase was identified in the *Sulfolobus* strain by screening colonies from an expression library for their ability to produce the amino acid product when supplied



Scheme 1: The lactamase catalytic process.

with the racemic γ-lactam. Screening was carried out using genomic libraries using a filter paper overlay. The colonies on the plate that were active showed a brown colouration of the filter paper when the amino acid was produced which had been soaked in ninhydrin stain. Purified enantiomers of the (+) or (−)-γ-lactam were used as substrates to determine the stereospecificity of the enzyme. Another nonthermophilic bacterial (+)-γ-lactamase that can also carry out this reaction has been identified within the bacterium *Delftia acidovorans*. This enzyme is of a different class, structure, and mechanism from the archaeal enzyme but both can use the nonnatural γ-lactam as a substrate. This archaeal γ-lactamase has been cloned and overexpressed in *E. coli* and purified to homogeneity. The molecular mass of the monomer was estimated to be 55 kDa by SDS-PAGE which is consistent with the calculated molecular mass of 55.7 kDa. The native molecular mass was determined to be 110 kDa by gel filtration indicating that the enzyme exists as a dimer in solution. The purified enzyme has been crystallised with a view to determining its three-dimensional structure.

The thermostable archaeal γ-lactamase has a high sequence homology to the signature amidase family of enzymes. It shows similar inhibition patterns as the amidase enzymes towards benzonitrile, phenylmethylsulfonyl fluoride, and heavy metals such as mercury and it is activated by thiol reagents. The enzyme selectively cleaves the (+)-enantiomer from a γ-racemic mixture. It also exhibits general amidase activity by cleaving linear and branched aliphatic and aromatic amides [12, 13].

Alignment of the amino acid sequences of the γ-lactamase from *S. solfataricus* MT4 with 4 amidases from *Pseudomonas chlororaphis* B23, *Rhodococcus* sp. N-771, *R. erythropolis* N-774, and *Rhodococcus rhodochrous* J shows that it has a 41–44% sequence identity to these enzymes. The amidases belong to the signature amidase family as they all contain the consensus sequence GGSS(S/G)GS. The amino acid sequence of the γ-lactamase contains the highly conserved putative catalytic residues of aspartic acid and serine but not the highly conserved cysteine residue [14].

The purified (+)-γ-lactamase enzyme has been immobilised as a cross-linked, polymerised enzyme preparation and packed into microreactors [15]. The thermophilic (+)-γ-lactamase retained 100% of its initial activity at the assay temperature, 80°C, for 6 h and retained 52% activity after 10 h, indicating the advantage of the immobilisation. The high stability of the immobilised enzyme provided the advantage that it could be used to screen many compounds in a microreactor system without denaturation.

2.2. Transfer of Amine Groups. Many drugs are formed from chiral amines and there has been increasing interest from pharmaceutical companies to study the enzymes that can transfer amine groups from one compound to another in a stereospecific way. The transaminases catalyse the transfer of an amino group from an amino acid to a ketoacid [16]. They use the cofactor pyridoxal phosphate (PLP), the biologically active form of vitamin B6. The PLP normally covalently binds to an active site lysine amino acid by a Schiff base (internal aldimine). The mechanism of transaminases is made up of two half-reactions. In the first half-reaction the donor substrate gives its amino group to the cofactor, resulting in a keto acid and enzyme-bound pyridoxamine-phosphate (PMP). In the second half-reaction an amino group is transferred from PMP to an acceptor keto acid, producing an amino acid and restoring the PLP internal aldimine.

The archaeon *S. solfataricus* has been found to be an interesting source of a thermostable transaminase enzyme of group IV Pfam [17]. This pyridoxal phosphate containing enzyme is involved in the nonphosphorylated pathway for serine synthesis which is not found in bacteria and is found in animals and plants. The transaminase reaction that the enzyme carries out is the conversion of L-serine and pyruvate to 3-hydroxypyruvate and alanine. Activity is also shown towards methionine, asparagine, glutamine, phenylalanine, histidine, and tryptophan. The enzyme can be used in combination with transketolase for synthesis of chiral drug intermediates [18].

The dimeric thermophilic archaeal transaminase enzyme structure has been solved in the holo form of the enzyme and in complex with an inhibitor gabaculine and in a substrate complex with phenylpyruvate, the keto product of phenylalanine [17]. Figure 1 shows a cartoon diagram of the dimeric *S. solfataricus* transaminase with the cofactors PLP in the two active sites.

The structural studies with this enzyme have given some insight into the conformational changes around the active site of the transaminase that occur during catalysis and have helped to understand the enzyme's substrate specificity. The most related enzyme to the *S. solfataricus* transaminase is the mesophilic yeast alanine:glyoxylate transaminase (AGAT) which shares 37% amino acid identity [19]. The yeast AGAT enzyme has been reported to have high substrate specificity

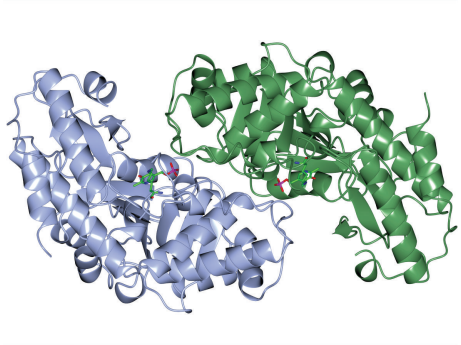


Figure 1: A ribbon diagram of the structure of the *Sulfolobus* transaminase dimer showing each monomer in a different colour as viewed along the molecular twofold axis. The cofactor pyridoxal phosphate (PLP) is shown bound to the active site lysine as a stick model in the two active sites (PDB code 3ZRP). The figure was constructed using CCP4mg [53].

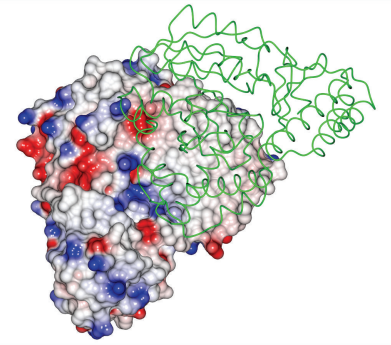


Figure 2: The structure of the *Sulfolobus* transaminase dimer showing one subunit in surface space filling mode where the white regions indicate hydrophobic regions and blue indicates basic regions and red acidic regions. The other subunit is represented as a α-C chain showing that the subunit interface is mainly hydrophobic (PDB code 3ZRP). The figure was constructed using CCP4mg [53].

for small amino acids such as alanine and glyoxylate, unlike the *Sulfolobus* transaminase which exhibits a broad amino acid substrate specificity. The substrate-binding pockets of AGATs are remarkably similar to that of *Sulfolobus* enzyme. The aromatic substrate phenylalanine was modelled into the *Sulfolobus* enzyme active site and was compared with the superimposed yeast AGAT enzyme structure. Most residues in the vicinity of the modelled substrate are conserved between the two proteins. The *Sulfolobus* enzyme has a significantly larger substrate-binding pocket as its loop region between strands 9 and 10 is two amino acids shorter. This gives more space for the bulkier phenylalanine substrate to bind compared with the AGAT enzyme, which is only active towards alanine and glycine. An isoleucine residue in AGAT occupies the space where larger substrates would sit in the active site of the *Sulfolobus* enzyme. The isoleucine is positioned 2.9 Å from the modelled substrate and hinders the binding of any amino acid larger than alanine. In the *S. solfataricus* transaminase enzyme there is a corresponding valine residue in the position of the isoleucine of AGAT which is positioned further away allowing the binding of larger amino acids. These subtle differences between the two enzymes are enough to change the enzymes substrate specificity which is vitally important to understand for its use in commercial applications.

The archaeal transaminase is relatively thermostable for 10 minutes at 70°C and at pH 6.5. Features of the archaeal enzyme that relate to its increased stability when compared with the related AGAT enzyme show that the yeast enzyme has 10 salt bridges compared to 21 salt bridges in the *Sulfolobus* transaminase which includes several 3-4 amino acid networks which offer increased stability. There is a C-terminal extension in the *Sulfolobus* enzyme and shorter surface loops which are all general features that are found in thermophilic enzymes. The *Sulfolobus* transaminase dimer interface is unusual being hydrophobic in nature with few ionic interactions which are generally associated with more thermophilic archaeal enzymes (Figure 2). This *Sulfolobus* serine transaminase is the first example of a thermophilic archaeal serine transaminase to be studied structurally and to show that its properties meet the requirements for the commercial application of the enzyme in biocatalysis.

Figure 3: A ribbon diagram of the structure of the *Sulfolobus* dehalogenase dimer with each monomer shown in a different colour. The monomer has a core domain with a Rossmann-like fold of six-stranded parallel β-strands surrounded by five α-helices and three 3_{10} helices and a subdomain composed of α-helices. A substrate mimic is shown in the two active sites as a stick model, located between the two domains of each monomer (PDB code 2W11). The figure was constructed using CCP4mg [53].

2.3. Removal of Halogen Groups. Another archaeal enzyme of industrial interest is found in *Sulfolobus tokodaii*. This L-haloacid dehalogenase enzyme has been cloned and overexpressed in *E. coli*. It has been characterised biochemically and structurally [20, 21]. The enzyme monomer has two domains. The core domain has a Rossmann fold with a six-stranded parallel β-strand bundle surrounded by five α-helices and three 3_{10} helices. The subdomain is composed of α-helices. The active site is located between the two domains and the native enzyme forms a dimer as shown in Figure 3.

This enzyme has applications for chiral halo-carboxylic acid production and bioremediation. Chiral halo-carboxylic acids are important intermediates in the fine

chemical/pharmaceutical industries. Removal of the halogen group can be carried out by a dehalogenase. The *Sulfolobus* enzyme has the potential to resolve racemic mixtures of bromocarboxylic acids. This L-bromoacid dehalogenase is able to catalyse the conversion of 2-halo-carboxylic acids to the corresponding hydroxyalkanoic acids. It has been shown to display activity towards longer chain substrates than the bacterial *Xanthomonas autotrophicus* dehalogenase [22] with activity seen towards 2-chlorobutyric acid. This is due to a more accessible active site. The enzyme has a maximum activity at 60°C and a half-life of over an hour at 70°C. It is stabilised by a salt bridge and hydrophobic interactions on the subunit interface, helix capping, a more compact subdomain than related enzymes, and shortening of surface loops. Another thermophilic enzyme of this family from hyperthermophilic archaea has addressed the problem of thermostability in a different way. The related *Pyrococcus* dehalogenase (29% sequence identity) whose structure is available from a structural genomics project is a monomeric enzyme stabilised by a disulfide bond [23].

2.4. Hydrolysis and Esterification. Esterases are a class of commonly used enzymes in industrial applications. This is partially due to their inherent stability in organic solvents and the ability to freely reverse the enzyme reaction from hydrolysis to synthesis by the elimination of water that is used in the hydrolysis mechanism. The carboxylesterases catalyse the hydrolysis of the ester bond of relatively small water soluble substrates. The carboxyl esterase NP originally identified from the fungus *Ophiostoma novo-ulmi* is used for the production of the nonsteroidal pain killing drug, naproxen [24]. The racemic naproxen methylester is hydrolysed to the *(S)*-acid which is separated from the *(R)*-methylester to yield *(S)*-naproxen with a 99% ee and a yield of 95%. It is important that new drugs coming to market are of one optical form preventing the problems associated with side effects of the inactive enantiomer.

A thermostable carboxylesterase from the archaeon *Sulfolobus shibatae* has been cloned, sequenced, and overexpressed in *E. coli* [25] (Toogood and Littlechild, unpublished data). The enzyme has a 71 to 77% sequence identity to an esterase from *S. tokodaii* and a carboxylesterase from *S. solfataricus* strain P1, respectively [26, 27]. This enzyme was identified as a serine esterase belonging to the mammalian hormone-sensitive lipase (HSL) family. It contains the conserved putative catalytic triad residues Ser, Asp, and His and is inhibited by the serine hydrolase inhibitors phenylmethylsulfonyl fluoride and benzamidine and partially inhibited by thiol reagents. The enzyme is thermostable, with no loss of activity detected after 24 h at 60°C. The enzyme was able to cleave a variety of *p*-nitrophenyl ester substrates, with the highest activity detected with *p*-nitrophenyl caproate. The carboxylesterase was also tested for its ability to cleave a variety of industrially relevant esters and diesters. It has a preference for substrates containing aromatic groups such as diethyl-2-benzyl malonate, benzyl acetoacetate, and Z-phenylalanine methyl ester. However, it was also able to enantioselectively cleave compounds such as 2-methyl-1,3-propanediol diacetate.

2.5. Resolution of Amino Acids and Amino Acid Analogues. Amino acids can be either of the "so-called" L configuration as found in all proteins or of the D configuration found in bacterial cell walls. The production of the specific L-amino acids and amino acid analogues is important for a variety of purposes. The industrial process to carry out this biocatalytic reaction makes use of an L-aminoacylase enzyme. A thermophilic archaeal version of this enzyme has been cloned and overexpressed from the archaeon *Thermococcus litoralis* [2]. The enzyme is a homotetramer of 43 kDa monomers and has an 82% sequence identity to an aminoacylase from *Pyrococcus furiosus* and 45% sequence identity to a carboxypeptidase from *S. solfataricus*. The enzyme is thermostable, with a half-life of 25 hours at 70°C. Cell-free extracts of the aminoacylase were found to have optimal activity at 85°C in Tris-HCl pH 8.0. Conventional aminoacylase inhibitors, such as mono-tert-butyl malonate, have only a slight effect on its activity. The *T. litoralis* L-aminoacylase has a broad substrate specificity preferring the amino acids: Phe $\gg$ Met > Cys > Ala $\simeq$ Val > Tyr > Propargylglycine > Trp > Pro > Arg. A column bioreactor containing the recombinant *Thermococcus* L-aminoacylase immobilised onto Sepharose beads was constructed with the substrate, N-acetyl-DL-Trp, continuously flowing at 60°C for 10 days. No loss of activity was detected over five days, with 32% activity remaining after 40 days at 60°C [28]. The enzyme has also been immobilized into microreactors by covalent attachment to expoxy resins in channels of the reactor which allow the biocatalytic reaction to be carried out within this high throughput "flow" system [29]. This can be used for rapid screening of substrate specificity and eliminates problems with product inhibition often seen in industrial reactions which are carried out at high substrate concentrations. The *Thermococcus* L-aminoacylase enzyme is now being used in multiton commercial production of L-amino acids and their analogues by Chirotech/Dow Pharma and more recently by Chirotech/Dr. Reddy's for large scale biotransformations [30]. A racemase enzyme has been developed in order to convert the isomer not used by the enzyme to the form that is used which can enable a more efficient process with potentially 100% conversion of a racemic substrate [31].

The hyperthermophilic L-aminoacylase from *P. horikoshii* has also been cloned and overexpressed in *E. coli* [32]. There are differences in substrate specificity between the *Thermococcus* and the *Pyrococcus* enzymes. The substrate N-acetyl-L-phenylalanine is the most favourable substrate for the *Thermococcus* enzyme; however, this substrate is not used by the *Pyrococcus* L-aminoacylase.

2.6. Chiral Alcohol Production. A thermophilic aerobic archaeon *Aeropyrum pernix* is the source of a very stable alcohol dehydrogenase (ADH) enzyme that can be used for chiral alcohol production. This enzyme has been cloned and overexpressed in *E. coli* [33]. The *A. pernix* ADH enzyme is a tetrameric, zinc-containing, type I ADH with a monomer size of 39.5 kDa. It has sequence identity to a related horse liver ADH of 24% and the highest sequence identity to a known structure is 39% to a medium chain ADH from the hyperthermophilic archaeon *S. solfataricus* [34].

Figure 4: A ribbon diagram of the structure of the *Aeropyrum* alcohol dehydrogenase enzyme monomer showing the two different domains: the Rossmann domain which binds the NADH cofactor and the substrate-binding domain. The active site is located between the two domains where the cofactor and a substrate mimic are bound and shown as stick models. The two bound zinc ions are shown as blue spheres: the catalytic zinc in the active site and the structural zinc located in a distal loop. The active enzyme has a quaternary structure of a tetramer of four identical monomers (PDB code1H2B). The Figure was constructed using CCP4mg [53].

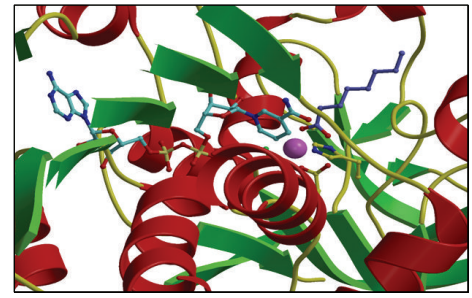


Figure 5: A ribbon diagram highlighting the details of the active site of the *Aeropyrum* ADH enzyme showing the catalytic zinc ion as a sphere together with the bound cofactor NADH and an enzyme inhibitor in stick mode, defining the substrate-binding pocket (PDB code 1H2B). The figure was constructed using CCP4mg [53].

The *A. pernix* enzyme is highly specific for the cofactor NAD(H) and displays activity towards a broad range of alcohols, aldehydes, and ketones, while appearing to show a preference for cyclic substrates. The enzyme is very thermostable with a half-life of 2 hr at 90°C. The maximal activity is beyond 75°C; however, there is still 10% activity at 20°C. The enzyme is solvent stable with over 50% activity retained after incubation with 60% acetonitrile or dioxane. The crystal structure of the enzyme has been determined with an inhibitor bound into the active site [35]. The ADH monomer is formed from a catalytic and a cofactor binding domain, with the overall fold similar to previously solved ADH structures (Figure 4). The 1.62 Å resolution *A. pernix* ADH structure is that of the holo form, with the cofactor NADH bound into the cleft between the two domains. An inhibitor is bound in the active site which has been interpreted to be octanoic acid. This inhibitor is positioned with its carbonyl oxygen forming the fourth ligand of the catalytic zinc ion (Figure 5). The enzyme is stabilised by an ion-pair cluster at the subunit interfaces of the tetramer. There are two zincs bound to the enzyme, one at the active site and another at remote site which appears to stabilise the enzyme. When the zinc does not occupy this second site a disulfide bond is formed to hold the same two protein chains together (Figure 6). It has now been predicted that disulfide bonds do exist to stabilise many cytoplasmic proteins in this aerobic archaeon, *A. pernix* [36].

The enzyme is active against primary and secondary alcohols with optimum chain length of C4-C5. It is most active to large cyclic alcohols such as cycloheptanol and cyclooctanol. The enzyme reaction can be reversed to produce chiral alcohols by changing the pH. Initial experiments demonstrate that the *A. pernix* ADH shows some stereoselectivity in the reverse reaction producing the *(S)* phenylethanol [37] (Guy, 2002). The related *Sulfolobus* ADH has been reported to also be *(S)* selective showing enantiomeric excesses as high as 98% [38] (Raia et al., 2001).

Other thermophilic archaeal alcohol dehydrogenases have been characterised and are of the short chain or aldol keto reductase family such as the short-chain *Pyrococcus furiosus* ADH [39] and the aldol keto reductase *Thermococcus kodakarensis* ADH [40].

2.7. Cleavage of Lactone Rings. The specific cleavage of a lactone ring is an important activity of interest to the pharmaceutical companies. The lactonase enzymes identified to date fall into three structurally diverse groups: the enol lactonases, gluconolactonases, and the quorum sensing lactonases. Phosphotriesterase-like lactonases (PLLs) were identified in the archaeal species *S. solfataricus* and *S. acidocaldarius* [41–44]. These enzymes catalyse the hydrolytic cleavage of the intramolecular ester bond in lactones and acyl-homoserine lactones (AHLs) to give the corresponding hydroxyacylic acids. They also have a promiscuous but significantly lower phosphotriesterase activity towards organophosphate compounds. Recently an enzyme of this class has been identified, cloned, overexpressed, and characterised [45] from *Vulcanisaeta moutnovskia* a hyperthermoacidophilic crenarchaeon that was recently isolated from a solfataric field close to Moutnovsky volcano in Kamchatka (Russia) [46]. The VmutPLL converted lactones and acyl-homoserine lactones (AHLs) with comparable activities. A promiscuous, significantly lower activity was observed to organophosphates and only minor activity was observed with carboxylesters. The catalytic activity strictly depended on bivalent cations (Cd^{2+} > Ni^{2+} > Co^{2+} > Mn^{2+} > Zn^{2+}). The VmutPLL showed a pH optimum around 8.0, a temperature optimum of 80°C, and a half-life of 26 min at 90°C. The enzyme has demonstrated high activity towards linear γ-lactones with hydrophobic side chains of variable lengths. These ranged from γ-butyrolactone (no side chain) and γ-valerolactone which has a methyl side chain and γ-dodecalactone which has a seven carbon side chain. It was shown that the enzyme has activity to whiskey lactone and δ-dodecalactone. No measurable activity was seen for mevalonolactone or δ-decalactone. For two of the substrates

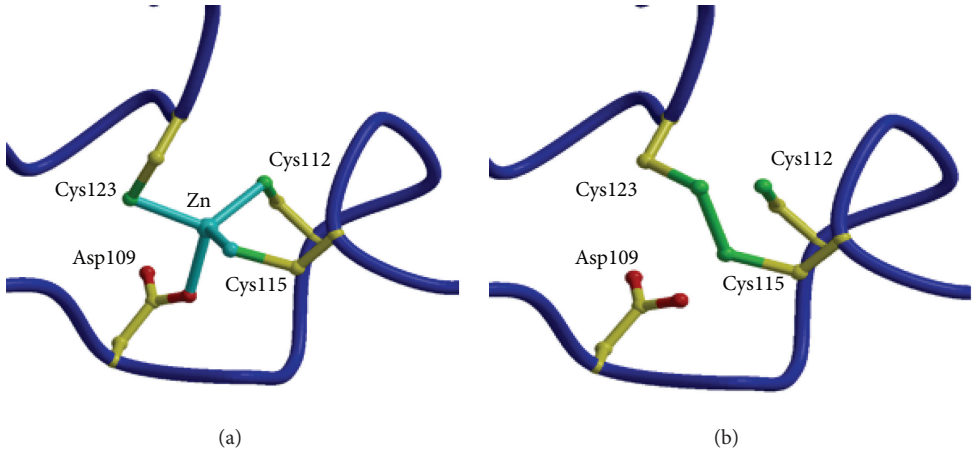


Figure 6: (a) The binding site of the structural zinc ion in the *Aeropyrum* ADH enzyme. (b) When zinc is not present a disulfide bond forms to stabilise the enzyme at this position in its structure (PDB code 1H2B). The figure was constructed with Bobscript [54].

γ-valerolactone and γ-caprolactone, the optical isomeric forms of these were tested to determine stereoselectivity of the enzyme (Table 1). Results indicated that while activity is seen with both isomers the enzyme seems to favour the D form of these substrates.

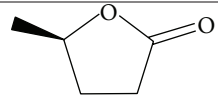
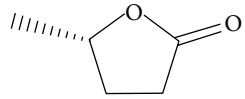
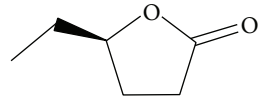
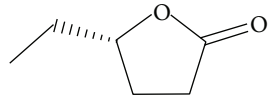
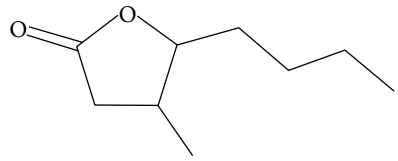
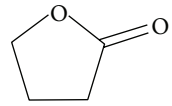
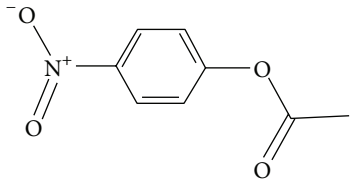
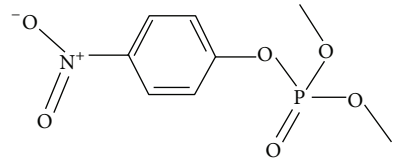
The recent structure of the *V. moutnovskia* lactonase has been carried out in complex with a long chain fatty acid [47] which maps the substrate-binding pocket. This is lined by hydrophobic side chains, which would provide affinity for γ-lactones of any side chain length. The homologous lactonase from the *S. islandicus* is reported to favour γ-lactone substrates with long hydrophobic acyl chains of over four carbons in length [48]. The substrate-binding site of the *S. islandicus* lactonase revealed a number of polar residues at the entrance of the substrate pocket with hydrophobic residues lining the bottom of this pocket. This would disfavour the binding of the smaller γ-lactones in the correct catalytic position. The different distribution of hydrophobic and other residues in the active site pocket between the two related enzymes appears to be responsible for their different catalytic activities. The *V. moutnovskia* lactonase belongs to the amidohydrolase superfamily with a $(\beta/\alpha)_8$-barrel structural fold. In the *V. moutnovskia* enzyme these metals are two cobalt ions [48] that are essential for activity and are located in at the C-terminus of the β barrel. The coordinating residues of the metals are four histidines, one aspartate, and one lysine which are highly conserved in this enzyme family. In the proposed catalytic cycle the metal ions activate a bridging water molecule through proton abstraction. The resulting hydroxide ion then performs a nucleophilic attack on the C_1 of the lactone ring resulting in hydrolysis [49].

The high thermal stability of this enzyme as well as its broad substrate specificity for different lactones AHLs and OPs makes it an interesting new enzyme for commercial biocatalysis. The enzyme is thought to have a natural role in quorum sensing which plays a role in some of the stages involved in biofilm formation [50, 51] and also in the expression of virulence factors that are of interest in medicinal and biotechnological applications [52]. The enzymatic degradation of AHLs catalysed by the PLLs provides potential for the interruption of quorum sensing signalling pathways and the control of microbial communities.

3. Conclusions

This review has summarised some of the important industrial applications using enzymes that have been identified from a variety of different thermophilic archaea. It is expected that the number of these enzymes used industrially will increase due to their inherent stability and novel specificities. The development of new rapid screening technologies and improved bioinformatic approaches in combination with new generation sequencing methodologies of archaeal genomes and metagenomes will provide new enzymes for industrial biotechnology. The enzymes can be cloned and overexpressed in easily grown hosts such as *E. coli* allowing access to sufficient quantities of the purified enzymes for detailed biochemical and structural characterisation. The scale-up of the enzyme production required for commercial applications can be carried out by using a fungal host system which allows export of the proteins into the growth medium for easy downstream processing. The cost of the enzyme biocatalyst is often the most expensive component of the industrial biotransformation and must be matched to the value of the end product. Higher value optically pure compounds which are important as drug intermediates for the pharmaceutical industries will allow a high enzyme price. Other enzymes that are required for the production of bulk chemicals, used as additives for domestic cleaning products, used in food production, or used to supplement biomass degradation processes, generally need to be marketed at a cheaper price and supplied in larger quantities. The stability of the biocatalyst is also an important issue since the enzyme ideally needs to be reused in several repeated biocatalytic cycles. Immobilisation of the enzyme can often increase its stability and allow it to be easily recovered for reuse. Thermostable enzymes are usually more robust under industrial conditions and can be used for repetitive biocatalytic conversions. The increased

Table 1: The kinetic parameters of the activity of the VmutPLL enzyme with different lactone substrates. Taken from Kallnik et al., 2014 [45].

Substrate	Structure	Kcat [s^{-1}]	Km [mM]	Kcat/Km [$s^{-1} * M^{-1}$]
γ-(*R*)-valerolactone		6.23 ± 0.42	4.56 ± 0.51	1327.1 ± 61.3
γ-(*S*)-valerolactone		2.68 ± 0.00	1.95 ± 0.18	1379.95 ± 110.67
γ-(*R*)-caprolactone		3.04 ± 0.01	0.55 ± 0.02	5563.72 ± 140.96
γ-(*S*)-caprolactone		1.89 ± 0.11	0.75 ± 0.04	2531.22 ± 298.99
Whiskey lactone		4.20 ± 0.06	0.93 ± 0.07	4538.63 ± 402.78
γ-Butyrolactone		2.79 ± 0.07	11.57 ± 0.58	241.7 ± 11.18
pNP-acetate		1.66 ± 0.34	8.19 ± 1.10	201.74 ± 20.48
Methyl-paraoxon		1.25 ± 0.40	2.79 ± 0.7	442.58 ± 50.14

use of enzymes from the extremophilic archaea offers the opportunity to access biocatalysts that are naturally stable to a variety of different conditions of temperature, pH, salinity, and pressure making them better suited to different industrial processes.

The use of enzymes in "White Biotechnology" is expected to grow with biobased materials and chemicals produced from emerging technologies predicted to rise globally to over 7.4 million metric tons in 2018 (Lux Research Analysts). The initial process of enzyme discovery and optimisation is still a limiting factor in the adoption of new biobased industrial processes. There is therefore an increasing opportunity to commercialise newly discovered archaeal enzymes for sustainable manufacturing to contribute to the new circular economy.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The author would like to thank the University of Exeter, the Wellcome Trust, the BBSRC, EPSRC, and Technology Strategy Board, UK, for sponsoring research in the J. A. Littlechild group at the Exeter Biocatalysis Centre. The ERA-net BBSRC grant, Thermogene, BB/LOO2035/1, is thanked for studies on the thermophilic transaminase enzymes. The EU Framework 7 grant "HotZyme" entitled Systematic Screening of Organisms from Hot Environments, Grant no. 265933, is thanked for supporting studies on the discovery, characterisation, and sequencing of the novel hyperthermoacidophilic crenarchaeon *Vulcanisaeta moutnovskia* and the study of the lactonase enzyme from this organism as discussed in this paper.

References

[1] M. N. Isupov, T. M. Fleming, A. R. Dalby, G. S. Crowhurst, P. C. Bourne, and J. A. Littlechild, "Crystal structure of the

glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus*," *Journal of Molecular Biology*, vol. 291, no. 3, pp. 651–660, 1999.

[2] H. S. Toogood, E. J. Hollingsworth, R. C. Brown et al., "A thermostable L-aminoacylase from *Thermococcus litoralis*: cloning, overexpression, characterization, and applications in biotransformations," *Extremophiles*, vol. 6, no. 2, pp. 111–122, 2002.

[3] B. Siebers and P. Schönheit, "Unusual pathways and enzymes of central carbohydrate metabolism in Archaea," *Current Opinion in Microbiology*, vol. 8, no. 6, pp. 695–705, 2005.

[4] H. J. Lamble, N. I. Heyer, S. D. Bull, D. W. Hough, and M. J. Danson, "Metabolic pathway promiscuity in the Archaeon *Sulfolobus solfataricus* revealed by studies on glucose dehydrogenase and 2-Keto-3-deoxygluconate aldolase," *Journal of Biological Chemistry*, vol. 278, no. 36, pp. 34066–34072, 2003.

[5] M. J. Danson and D. W. Hough, "Promiscuity in the Archaea. The enzymology of metabolic pathways," *The Biochemist*, vol. 27, pp. 17–21, 2005.

[6] M. Reher, T. Fuhrer, M. Bott, and P. Schönheit, "The nonphosphorylative entner-doudoroff pathway in the thermoacidophilic euryarchaeon *Picrophilus torridus* involves a novel 2-Keto-3-deoxygluconate- specific aldolase," *Journal of Bacteriology*, vol. 192, no. 4, pp. 964–974, 2010.

[7] A. Theodossis, H. Walden, E. J. Westwick et al., "The structural basis for substrate promiscuity in 2-keto-3-deoxygluconate aldolase from the Entner-Doudoroff pathway in *Sulfolobus solfataricus*," *The Journal of Biological Chemistry*, vol. 279, no. 42, pp. 43886–43892, 2004.

[8] R. M. Daniel, M. J. Danson, D. W. Hough, C. K. Lee, M. E. Peterson, and D. A. Cowan, "Enzyme stability and activity at high temperatures," in *Protein Adaptation in Extremophiles*, K. S. Siddiqui and T. Thomas, Eds., pp. 1–34, Nova Science Publishers, New York, NY, USA, 2008.

[9] R. Jaenicke and G. Böhm, "The stability of proteins in extreme environments," *Current Opinion in Structural Biology*, vol. 8, no. 6, pp. 738–748, 1998.

[10] J. Littlechild, P. James, H. Novak, and C. Sayer, *Thermophilic Microbes in Environmental and Industrial Biotechnology, Biotechnology of Thermophiles*, Springer, 2nd edition, 2013.

[11] S. J. Taylor, R. McCague, R. Wisdom et al., "Development of the biocatalytic resolution of 2-azabicyclo[2.2.1]hept-5-en-3-one as an entry to single-enantiomer carbocyclic nucleosides," *Tetrahedron: Asymmetry*, vol. 4, no. 6, pp. 1117–1128, 1993.

[12] H. S. Toogood, R. C. Brown, K. Line et al., "The use of a thermostable signature amidase in the resolution of the bicyclic synthon (rac)-γ-lactam," *Tetrahedron*, vol. 60, no. 3, pp. 711–716, 2004.

[13] I. S. Gonsalvez, M. N. Isupov, and J. A. Littlechild, "Crystallization and preliminary X-ray analysis of a gamma-lactamase," *Acta Crystallographica Section D: Biological Crystallography*, vol. 57, no. 2, pp. 284–286, 2001.

[14] M. Kobayashi, Y. Fujiwara, M. Goda, H. Komeda, and S. Shimizu, "Identification of active sites in amidase: evolutionary relationship between amide bond- and peptide bond-cleaving enzymes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 22, pp. 11986–11991, 1997.

[15] A. M. Hickey, B. Ngamsom, C. Wiles, G. M. Greenway, P. Watts, and J. A. Littlechild, "A microreactor for the study of biotransformations by a cross-linked gamma-lactamase enzyme," *Biotechnology Journal*, vol. 4, no. 4, pp. 510–516, 2009.

[16] P. K. Mehta, T. I. Hale, and P. Christen, "Aminotransferases: demonstration of homology and division into evolutionary subgroups," *European Journal of Biochemistry*, vol. 214, no. 2, pp. 549–561, 1993.

[17] C. Sayer, M. Bommer, J. M. Ward, M. Isupov, and J. A. Littlechild, "Crystal structure and substrate specificity of the thermophilic serine:pyruvate aminotransferase from *Sulfolobus solfataricus*," *Acta Crystallographica, Section D: Biological Crystallography*, vol. 68, no. 7, pp. 763–772, 2012.

[18] B. H. Chen, A. Sayar, U. Kaulmann, P. A. Dalby, J. M. Ward, and J. M. Woodley, "Reaction modelling and simulation to assess the integrated use of transketolase and ω-transaminase for the synthesis of an aminotriol," *Biocatalysis and Biotransformation*, vol. 24, no. 6, pp. 449–457, 2006.

[19] P. Meyer, D. Liger, N. Leulliot et al., "Crystal structure and confirmation of the alanine:glyoxylate aminotransferase activity of the YFL030w yeast protein," *Biochimie*, vol. 87, no. 12, pp. 1041–1047, 2005.

[20] C. A. Rye, M. N. Isupov, A. A. Lebedev, and J. A. Littlechild, "An order-disorder twin crystal of L-2-haloacid dehalogenase from *Sulfolobus tokodaii*," *Acta Crystallographica Section D: Biological Crystallography*, vol. 63, no. 8, pp. 926–930, 2007.

[21] C. A. Rye, M. N. Isupov, A. A. Lebedev, and J. A. Littlechild, "Biochemical and structural studies of a $_L$-haloacid dehalogenase from the thermophilic archaeon *Sulfolobus tokodaii*," *Extremophiles*, vol. 13, no. 1, pp. 179–190, 2009.

[22] J. Van der Ploeg, G. Van Hall, and D. B. Janssen, "Characterization of the haloacid dehalogenase from *Xanthobacter autotrophicus* GJ10 and sequencing of the dhlB gene," *Journal of Bacteriology*, vol. 173, no. 24, pp. 7925–7933, 1991.

[23] R. Arai, M. Kukimoto-Niino, C. Kuroishi, Y. Bessho, M. Shirouzu, and S. Yokoyama, "Crystal structure of the probable haloacid dehalogenase PH0459 from *Pyrococcus horikoshii* OT3," *Protein Science*, vol. 15, no. 2, pp. 373–377, 2006.

[24] M. N. Isupov, A. A. Brindley, E. J. Hollingsworth, G. N. Murshudov, A. A. Vagin, and J. A. Littlechild, "Crystallization and preliminary X-ray diffraction studies of a fungal hydrolase from *Ophiostoma novo-ulmi*," *Acta Crystallographica Section D: Biological Crystallography*, vol. 60, no. 10, pp. 1879–1882, 2004.

[25] S. Huddleston, C. A. Yallop, and B. M. Charalambous, "The identification and partial characterisation of a novel inducible extracellular thermostable esterase from the archaeon *Sulfolobus shibatae*," *Biochemical and Biophysical Research Communications*, vol. 216, no. 2, pp. 495–500, 1995.

[26] A. Morana, N. Di Prizito, V. Aurilia, M. Rossi, and R. Cannio, "A carboxylesterase from the hyperthermophilic archaeon *Sulfolobus solfataricus*: cloning of the gene, characterization of the protein," *Gene*, vol. 283, no. 1-2, pp. 107–115, 2002.

[27] Y.-J. Park, S. Y. Choi, and H.-B. Lee, "A carboxylesterase from the thermoacidophilic archaeon *Sulfolobus solfataricus* P1; purification, characterization, and expression," *Biochimica et Biophysica Acta (BBA)—General Subjects*, vol. 1760, no. 5, pp. 820–828, 2006.

[28] H. S. Toogood, I. N. Taylor, R. C. Brown, S. J. C. Taylor, R. McCague, and J. A. Littlechild, "Immobilisation of the thermostable L-aminoacylase from *Thermococcus litoralis* to generate a reusable industrial biocatalyst," *Biocatalysis and Biotransformation*, vol. 20, no. 4, pp. 241–249, 2002.

[29] B. Ngamsom, A. M. Hickey, G. M. Greenway, J. A. Littlechild, P. Watts, and C. Wiles, "Development of a high throughput screening tool for biotransformations utilising a thermophilic

L-aminoacylase enzyme," *Journal of Molecular Catalysis B: Enzymatic*, vol. 63, no. 1-2, pp. 81–86, 2010.

[30] K. E. Holt, "Biocatalysis and chemocatalysis: a powerful combination for the preparation of enantiomerically pure α-amino acids," *ChemInform*, vol. 36, no. 42, 2005.

[31] S. Baxter, S. Royer, G. Grogan et al., "An improved racemase/acylase biotransformation for the preparation of enantiomerically pure amino acids," *Journal of the American Chemical Society*, vol. 134, no. 47, pp. 19310–19313, 2012.

[32] K. Tanimoto, N. Higashi, M. Nishioka, K. Ishikawa, and M. Taya, "Characterization of thermostable aminoacylase from hyperthermophilic archaeon *Pyrococcus horikoshii*," *The FEBS Journal*, vol. 275, no. 6, pp. 1140–1149, 2008.

[33] J. E. Guy, M. N. Isupov, and J. A. Littlechild, "Crystallization and preliminary X-ray diffraction studies of a novel alcohol dehydrogenase from the hyperthermophilic archaeon *Aeropyrum pernix*," *Acta Crystallographica, Section D: Biological Crystallography*, vol. 59, no. 1, pp. 174–176, 2003.

[34] L. Esposito, F. Sica, C. A. Raia et al., "Crystal structure of the alcohol dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus* at 1.85 Å resolution," *Journal of Molecular Biology*, vol. 318, no. 2, pp. 463–477, 2002.

[35] J. E. Guy, M. N. Isupov, and J. A. Littlechild, "The structure of an alcohol dehydrogenase from the hyperthermophilic archaeon *Aeropyrum pernix*," *Journal of Molecular Biology*, vol. 331, no. 5, pp. 1041–1051, 2003.

[36] P. Mallick, D. R. Boutz, D. Eisenberg, and T. O. Yeates, "Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 15, pp. 9679–9684, 2002.

[37] J. E. Guy, *Studies on a hyperthermophilic alcohol dehydrogenase [Ph.D. thesis]*, University of Exeter, 2004.

[38] C. A. Raia, A. Giordano, and M. Rossi, "Alcohol dehydrogenase from *Sulfolobus solfataricus*," in *Methods in Enzymology*, vol. 331, pp. 176–195, Elsevier, 2001.

[39] J. van der Oost, W. G. B. Voorhorst, S. W. M. Kengen et al., "Genetic and biochemical characterization of a short-chain alcohol dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*," *European Journal of Biochemistry*, vol. 268, no. 10, pp. 3062–3068, 2001.

[40] X. Wu, C. Zhang, I. Orita, T. Imanaka, T. Fukui, and X.-H. Xing, "Thermostable alcohol dehydrogenase from *Thermococcus kodakarensis* KOD1 for enantioselective bioconversion of aromatic secondary alcohols," *Applied and Environmental Microbiology*, vol. 79, no. 7, pp. 2209–2217, 2013.

[41] L. Merone, L. Mandrich, M. Rossi, and G. Manco, "A thermostable phosphotriesterase from the archaeon *Sulfolobus solfataricus*: cloning, overexpression and properties," *Extremophiles*, vol. 9, no. 4, pp. 297–305, 2005.

[42] L. Afriat, C. Roodveldt, G. Manco, and D. S. Tawfik, "The latent promiscuity of newly identified microbial lactonases is linked to a recently diverged phosphotriesterase," *Biochemistry*, vol. 45, no. 46, pp. 13677–13686, 2006.

[43] E. Porzio, L. Merone, L. Mandrich, M. Rossi, and G. Manco, "A new phosphotriesterase from *Sulfolobus acidocaldarius* and its comparison with the homologue from *Sulfolobus solfataricus*," *Biochimie*, vol. 89, no. 5, pp. 625–636, 2007.

[44] L. Merone, L. Mandrich, M. Rossi, and G. Manco, "Enzymes with phosphotriesterase and lactonase activities in Archaea," *Current Chemical Biology*, vol. 2, no. 3, pp. 237–248, 2008.

[45] V. Kallnik, A. Bunescu, C. Sayer et al., "Characterization of a phosphotriesterase-like lactonase from the hyperthermoacidophilic crenarchaeon *Vulcanisaeta moutnovskia*," *Journal of Biotechnology*, vol. 190, pp. 11–17, 2014.

[46] V. M. Gumerov, A. V. Mardanov, A. V. Beletsky et al., "Complete genome sequence of *Vulcanisaeta moutnovskia* strain 768-28, a novel member of the hyperthermophilic crenarchaeal genus *Vulcanisaeta*," *Journal of Bacteriology*, vol. 193, no. 9, pp. 2355–2356, 2011.

[47] J. Hiblot, J. Bzdrenga, C. Champion, E. Chabriere, and M. Elias, "Crystal structure of *Vmo*Lac, a tentative quorum quenching lactonase from the extremophilic crenarchaeon *Vulcanisaeta moutnovskia*," *Scientific Reports*, vol. 5, p. 8372, 2015.

[48] J. Hiblot, G. Gotthard, E. Chabriere, and M. Elias, "Structural and enzymatic characterization of the lactonase *Sis*Lac from *Sulfolobus islandicus*," *PLoS ONE*, vol. 7, no. 10, Article ID e47028, 2012.

[49] A. N. Bigley and F. M. Raushel, "Catalytic mechanisms for phosphotriesterases," *Biochimica et Biophysica Acta (BBA)—Proteins and Proteomics*, vol. 1834, no. 1, pp. 443–453, 2013.

[50] M. J. Kirisits and M. R. Parsek, "Does *Pseudomonas aeruginosa* use intercellular signalling to build biofilm communities?" *Cellular Microbiology*, vol. 8, no. 12, pp. 1841–1849, 2006.

[51] T. R. De Kievit, "Quorum sensing in *Pseudomonas aeruginosa* biofilms," *Environmental Microbiology*, vol. 11, no. 2, pp. 279–288, 2009.

[52] J. Hiblot, G. Gotthard, E. Chabriere, and M. Elias, "Structural and Enzymatic characterization of the lactonase SisLac from *Sulfolobus islandicus*," *PLoS ONE*, vol. 7, no. 10, Article ID e47028, 2012.

[53] S. McNicholas, E. Potterton, K. S. Wilson, and M. E. M. Noble, "Presenting your structures: the CCP4mg molecular-graphics software," *Acta Crystallographica Section D: Biological Crystallography*, vol. 67, no. 4, pp. 386–394, 2011.

[54] R. M. Esnouf, "An extensively modified version of MolScript that includes greatly enhanced coloring capabilities," *Journal of Molecular Graphics and Modelling*, vol. 15, no. 2, pp. 132–134, 1997.

Hindawi Publishing Corporation
Archaea
Volume 2015, Article ID 170571, 10 pages
http://dx.doi.org/10.1155/2015/170571

Review Article

Archaeal Nucleic Acid Ligases and Their Potential in Biotechnology

Cecilia R. Chambers and Wayne M. Patrick

Department of Biochemistry, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand

Correspondence should be addressed to Wayne M. Patrick; wayne.patrick@otago.ac.nz

Received 13 March 2015; Accepted 18 May 2015

Academic Editor: Frédéric Pecorari

With their ability to catalyse the formation of phosphodiester linkages, DNA ligases and RNA ligases are essential tools for many protocols in molecular biology and biotechnology. Currently, the nucleic acid ligases from bacteriophage T4 are used extensively in these protocols. In this review, we argue that the nucleic acid ligases from Archaea represent a largely untapped pool of enzymes with diverse and potentially favourable properties for new and emerging biotechnological applications. We summarise the current state of knowledge on archaeal DNA and RNA ligases, which makes apparent the relative scarcity of information on *in vitro* activities that are of most relevance to biotechnologists (such as the ability to join blunt- or cohesive-ended, double-stranded DNA fragments). We highlight the existing biotechnological applications of archaeal DNA ligases and RNA ligases. Finally, we draw attention to recent experiments in which protein engineering was used to modify the activities of the DNA ligase from *Pyrococcus furiosus* and the RNA ligase from *Methanothermobacter thermautotrophicus*, thus demonstrating the potential for further work in this area.

1. Introduction

DNA and RNA ligases are ubiquitous enzymes that catalyse the formation of phosphodiester bonds between opposing 5′-phosphate and 3′-hydroxyl termini in nucleic acids [1–3]. Their activities are essential for central biological processes, including DNA replication and recombination, rearrangement of immunoglobulin genes, and RNA editing and repair. Their activities *in vitro* have also been exploited in numerous molecular biology protocols, making them critical tools for modern biotechnology.

Together with the RNA capping enzymes and tRNA ligases, DNA and RNA ligases constitute the nucleotidyl transferase superfamily [4]. All of the enzymes in this superfamily catalyse phosphodiester bond formation in a conserved, three-step mechanism that utilises either ATP, GTP, or NAD^+ as a high-energy cofactor [4–6]. In the first step, nucleophilic attack on the α-phosphate of the cofactor by an active site lysine yields a ligase-AMP intermediate. Secondly, the AMP is transferred to the 5′-phosphate of one polynucleotide strand, resulting in an adenylated nucleic acid intermediate (with the active site lysine as the leaving group). Finally, the 3′-hydroxyl group of the second polynucleotide strand attacks the 5′-phosphate of the opposing strand, joining the two strands with a new phosphodiester bond and liberating AMP.

Species from the domain Archaea not only survive, but thrive, under conditions of extreme temperature, salinity, pH, and pressure. Evolution in these extreme environments has resulted in archaeal proteins that have properties of value to biotechnologists, including stability and activity under a range of comparatively harsh *in vitro* conditions. A familiar example is the widespread use of the *Pyrococcus furiosus* DNA polymerase in the Polymerase Chain Reaction (PCR) [7], in which its thermostability and processivity also make it valuable for related protocols such as QuikChange mutagenesis [8, 9]. In this review, we turn the attention to archaeal nucleic acid ligases. We summarise the current state of knowledge about these enzymes, including their existing applications in biotechnology, and we argue that they offer a largely untapped pool of activities for use in "next generation" molecular biology protocols.

2. DNA Ligases

In vivo, DNA ligases catalyse the formation of phosphodiester bonds at single-stranded nicks in double-stranded DNA. This activity is critical for maintaining genomic integrity during DNA replication, DNA recombination, and DNA excision repair [2, 6]. They are essential in all organisms and they are conventionally classified into two families according to their cofactor specificity. ATP-dependent ligases (EC 6.5.1.1) are typically found in Eukarya, Archaea, and viruses (including bacteriophages), while the NAD^+-dependent DNA ligases (EC 6.5.1.2) are typically found in bacteria and some eukaryotic viruses. There are, however, exceptions to this rule. Most notably, the archaeal species *Haloferax volcanii* possesses two active DNA ligases: one ATP-dependent (LigA) and the other NAD^+-dependent (LigN) [10].

DNA ligases are essential for numerous applications in molecular biology and biotechnology. For decades, DNA ligases have been used to construct recombinant DNA molecules (i.e., cloning) and for genetic disease detection using the ligation chain reaction [11]. More recently, the DNA ligase from the bacterium, *Thermus aquaticus* (*Taq*), has become important for Gibson assembly. This is an isothermal, one-pot method for assembling overlapping DNA molecules without the use of restriction enzymes [12]. A number of next generation sequencing methods also depend on DNA ligases [13, 14], either for adapter ligation during sample preparation (e.g., Illumina and 454 sequencing) or for the sequencing reaction itself (SOLiD sequencing).

With its ability to ligate both cohesive- and blunt-ended, double-stranded DNA molecules [15], the most commonly used DNA ligase in biotechnology is the ATP-dependent enzyme from bacteriophage T4. However, it is only weakly active for the ligation of blunt-ended fragments [16] and it is irreversibly inactivated at 65°C. It is also inactive at NaCl concentrations above ~150 mM. We posit that thermostable archaeal DNA ligases could be well suited for use in some or all of the above applications. For example, the ligation chain reaction requires a ligase that is stable at temperatures above 90°C, and Archaea may provide biotechnologists with superior alternatives to *Taq* ligase for Gibson assembly.

3. Archaeal DNA Ligases

To date, fewer than 25 archaeal DNA ligases have been characterised to any extent at all, and in general data on them is limited compared with DNA ligases from other domains of life [17, 18]. Table 1 summarises the current state of knowledge and highlights the diverse range of properties possessed by archaeal DNA ligases. Data for T4 DNA ligase are also included, for comparison.

As noted above, DNA ligases are usually classified based on their strict cofactor specificity for either ATP or NAD^+. Interestingly, a number of archaeal DNA ligases can utilise multiple cofactors. Sequence homology suggested that the DNA ligases from *Thermococcus kodakaraensis, T. fumicolans,* and *T. onnurineus* belong to the ATP-dependent family (EC 6.5.1.1); however, *in vitro* characterisation of each has shown that they are able to use either ATP or NAD^+ as their cofactor [19–21]. The ATP-dependent DNA ligase from *Sulfophobococcus zilligii* also shows relative activity of 63% when the cofactor is switched from ATP to GTP [22]. Other than the *S. zilligii* DNA ligase, activity with GTP has only been described for RNA capping enzymes [4]. A number of archaeal DNA ligases, including the *S. zilligii* enzyme, are also able to use ADP (Table 1).

One hypothesis for the undifferentiated nucleotide specificities of archaeal DNA ligases is that they have retained a trait from the ancient common ancestor of the ATP- and NAD^+-dependent enzymes. This ancestor may have used ADP as a cofactor [23], as the ADP moiety is common to both ATP and NAD^+. However, it has also been noted that direct evidence of ADP utilisation by DNA ligases is minimal [24]. Another proposal is that ATP is comparatively unstable at high temperatures, and this provided the selection pressure for evolution of thermophilic ligases with specificity for alternative cofactors such as ADP and GTP [22].

DNA ligases employ multidomain architectures in order to catalyse phosphodiester bond formation; however, there is variation in the number and identity of the domains they possess [14]. To date, the structures of six archaeal DNA ligases have been solved, from *Archaeoglobus fulgidus* [25], *Pyrococcus furiosus* [26, 27], *Sulfolobus solfataricus* [28], *S. zilligii* [29], *Thermococcus sibiricus* [30], and *Thermococcus* sp. 1519 [31]. Each enzyme comprises three domains: the adenylation domain (AdD), the oligonucleotide-binding domain (OBD), and the N-terminal DNA-binding domain (DBD). The AdD contains the six motifs (I, III, IIIb, IV, V, and VI) that are characteristic of the nucleotidyl transferase superfamily [32]. The AdD and OBD are minimally required for activity and together they are referred to as the catalytic core. The N-terminal DBD is unique to the eukaryotic and archaeal DNA ligases and is thought to play roles in maintaining an active conformation of the catalytic core, as well as distorting the DNA substrate [33].

Elucidation of the unbound [34] and DNA-bound [35] structures of the ATP-dependent ligase from *Chlorella* virus has highlighted the importance of large conformational changes during the catalytic cycle of DNA ligases. During DNA binding the OBD translocates by >60 Å and rotates nearly 180° around a swivel point, in order to fit into the minor groove of the DNA substrate. No archaeal DNA ligases have had their structures solved in complex with DNA; however, OBDs have been captured adopting three different conformations (Figure 1). The *S. solfataricus* enzyme exhibited an open and extended conformation in which the OBD was turned away from the AdD (Figure 1(a)); the overall structure resembled that of the DNA ligase from bacteriophage T7 [36]. In contrast, the *Thermococcus* sp. 1519 ligase structure (Figure 1(b)) adopted an intermediate conformation in which the OBD was rotated anticlockwise around the AdD by ~90° compared to the open extended conformation, although this rotation was insufficient to introduce any hydrogen bonds or salt bridges between the OBD and the other domains. A further 120° rotation of the OBD yields a closed conformation, as observed in the structures of the DNA ligases from *P. furiosus* (Figure 1(c)), *A. fulgidus,* and *T. sibiricus*. In these structures a C-terminal

Table 1: Properties of archaeal DNA ligases.

Organism	Growth conditions	UniProt ID	PDB ID	Cofactor	T_{opt}	Reference
Bacteriophage						
Bacteriophage T4	Mesophile	P00970	—	ATP	37°C	[16, 37]
Crenarchaeota						
Aeropyrum pernix	Hyperthermophile	Q9YD18	—	ATP, ADP	70°C	[38]
Desulfurolobus ambivalens	Acidophile/thermophile	Q02093	—		NR	[39]
Staphylothermus marinus	Hyperthermophile	A3DP49	—	ATP, ADP	75°C	[40]
Sulfolobus acidocaldarius	Acidophile/thermophile	Q4JAM1	—	ATP	85°C	[41]
Sulfolobus shibatae	Acidophile/thermophile	Q9P9K9	—	ATP, dATP	60–80°C	[42]
Sulfolobus solfataricus	Acidophile/thermophile	Q980T8	2HIX, 2HIV	ATP	NR	[28]
Sulfophobococcus zilligii	Hyperthermophile	D2CJS7	—	ATP, ADP, GTP	75°C	[22, 29]
Euryarchaeota						
Archaeoglobus fulgidus	Hyperthermophile	O29632	3GDE	ATP	NR	[25]
Ferroplasma acidarmanus	Acidophile	S0AR65	—	ATP, dATP	30°C	[43]
Ferroplasma acidiphilum	Acidophile	Q2PCE4	—	ATP, NAD^+	40°C	[41]
Haloferax volcanii (LigA)	Halophile	D4GYZ4	—	ATP	NR	[10]
Haloferax volcanii (LigN)	Halophile	D4GY98	—	NAD^+	45°C	[10, 44, 45]
Methanothermobacter thermautotrophicus	Thermophile	Q50566	—	ATP, dATP	60°C	[46]
Picrophilus torridus	Acidophile	Q6L195	—	ATP, NAD^+	65°C	[41]
Pyrococcus furiosus	Hyperthermophile	P56709	2CFM	ATP	55°C	[26, 27, 47]
Pyrococcus horikoshii	Hyperthermophile	O59288	—	ATP	70–90°C	[23]
Thermococcus sp. 1519	Thermophile	C0LJI8	3RR5	ATP	70°C	[31, 48, 49]
Thermococcus fumicolans	Hyperthermophile	Q9HH07	—	ATP, NAD^+	65°C	[19]
Thermococcus kodakaraensis	Thermophile	Q9HHC4	—	ATP, NAD^+	65°C	[20, 50]
Thermococcus onnurineus	Hyperthermophile	B6YTR4	—	ATP, NAD^+	80°C	[21]
Thermococcus sibiricus	Hyperthermophile	C6A2U9	4EQ5	ATP	NR	[30]
Thermoplasma acidophilum	Acidophile	Q9HJ26	—	ATP, NAD^+	65°C	[41]

T_{opt}: temperature optimum for the DNA ligase *in vitro*; NR: not reported.

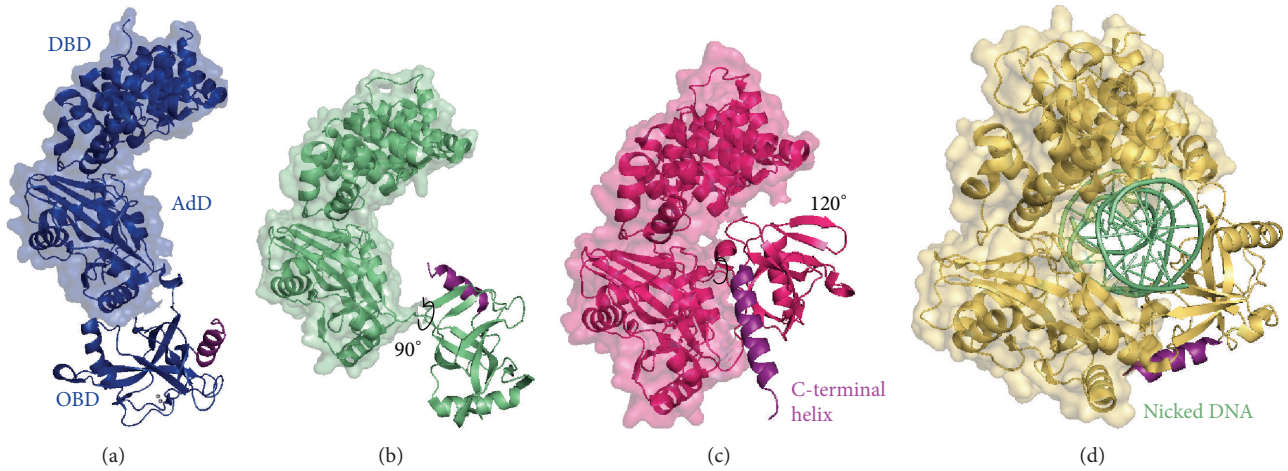


Figure 1: Structures of the DNA ligases from: (a) *Sulfolobus solfataricus* (PDB ID 2HIX); (b) *Thermococcus* sp. 1519 (PDB ID 3RR5); (c) *Pyrococcus furiosus* (PDB ID 2CFM); and (d) *Homo sapiens* (PDB ID 1X9N). The positions of the DNA binding domain (DBD), adenylation domain (AdD), and oligonucleotide binding domain (OBD) are indicated in the *S. solfataricus* structure. While the DBD and AdD occupy equivalent positions in all of the structures, the *S. solfataricus* enzyme shows an open extended conformation of the OBD, the *Thermococcus* sp. 1519 enzyme shows a 90° anticlockwise rotation of the OBD relative to the *S. solfataricus* enzyme (indicated by a black arrow) resulting in the intermediate conformation, and the *P. furiosus* ligase has a 120° rotation exhibiting the closed conformation. Human DNA ligase 1 is shown bound to a nicked DNA substrate (light green). The C-terminal helix of each enzyme is highlighted in purple.

helix (Figure 1), found after conserved motif VI, stabilises the closed conformation by mediating several ionic interactions between the OBD and the AdD [26]. This additional helix occupies the cleft between the AdD and OBD in the archaeal unbound structures, but it is displaced in the DNA-bound structure of human DNA ligase I [33].

The domain arrangements of the archaeal ligase structures all differ substantially from those of the DNA-bound structures obtained for human DNA ligases I and III [33, 51], where the three domains encircle the DNA substrate (Figure 1(d)). The emerging picture is one in which conformational flexibility is critical for the correct functioning of archaeal DNA ligases. However, to date, no structures of archaeal DNA ligases in complex with DNA have been solved; therefore it is unknown whether the differences in unbound structures (Figures 1(a)–1(c)) correlate with differences in domain orientations when the DNA substrate is bound. The protein dynamics associated with catalysis at the growth temperatures of the host cells (70–100°C) also remain to be elucidated.

4. Biotechnological Applications of Archaeal DNA Ligases

Given their primary physiological role in DNA repair, it is unsurprising that most archaeal DNA ligases have only been assayed for their ability to seal single-stranded nicks in double-stranded DNA. Of more interest for biotechnological applications is the ability to ligate double-stranded, cohesive-, or blunt-ended fragments. These activities have been reported for four archaeal DNA ligases. The enzymes from *Aeropyrum pernix*, *Staphylothermus marinus*, *Thermococcus* sp. 1519, and *T. fumicolans* have all been shown to perform ligation of cohesive-ended fragments. In addition, the *S. marinus* and *T. fumicolans* DNA ligases could also join blunt-ended fragments [19, 40]. Thus, it seems likely that further characterisation of archaeal DNA ligases should yield a pool of enzymes with potential utility in molecular biology and biotechnology.

The most immediate applications for archaeal DNA ligases are likely to be those that exploit their high temperature optima (typically 50–100°C; Table 1). For example, the DNA ligase from *Thermococcus* sp. 1519 is most active at 60–70°C and it is able to ligate DNA fragments with long cohesive ends (12-nucleotide overhangs), but not fragments with shorter (4-nucleotide) cohesive ends or with blunt ends [48]. While it remains to be tested, this combination of properties would appear to make it a promising tool for Gibson assembly [12]. This protocol has rapidly emerged as the dominant method for restriction enzyme-independent assembly of DNA fragments in synthetic biology and it is currently performed at 50°C. We speculate that archaeal DNA ligases, such as the one from *Thermococcus* sp. 1519, may drive the development of new methods, with the promise that ligation at higher temperatures (60–70°C) would decrease the number of incorrect ligation events that arise from misannealing of fragments with short overhangs.

In a similar vein, the DNA ligase from *Staphylothermus marinus* has a half-life of almost 3 h at 100°C and catalyses a variety of ligation reactions with cohesive- and blunt-ended fragments [40]. This extremely thermostable enzyme could find utility in the ligase chain reaction (LCR) for detection of single nucleotide polymorphisms, as it is able to survive the high temperature denaturation steps (~95°C) in the thermal cycling protocol. More generally, it has been shown that thermostable proteins are ideal starting points for protein

engineering, as they are more tolerant of mutations and thus yield more functional variants on mutation than their mesophilic homologues [52].

5. Engineering an Improved Archaeal DNA Ligase

Despite the ubiquity of DNA ligases in molecular biology, very few attempts have been made to enhance their properties by protein engineering. To date only the DNA ligase from bacteriophage T4 [53] and one archaeal DNA ligase, from *Pyrococcus furiosus* [27, 47], have been targeted.

Nishida and colleagues have successfully used their structural insights [26] to enhance the activity of the *P. furiosus* DNA ligase through structure-guided mutagenesis. In particular, they have targeted the C-terminal helix that interacts with the OBD and the AdD to stabilise the closed conformation of the enzyme (Figure 1(c)). To begin, five polar residues from the OBD (Asp540, Arg544, Gln547, Lys554, and Lys558), each of which contributed to interactions with the AdD, were mutated to alanine [47]. The hypothesis was that destabilising the interdomain interaction would facilitate increased motion of the OBD and thus increase activity by "unlocking" the enzyme. Of the five selected residues, mutation of Asp540, located at the N-terminus of the helical extension, exerted the greatest effect. Further mutagenesis at this position showed that the Asp540 → Arg (D540R) substitution gave optimal activity, over a broadened temperature range (20–80°C).

In proof-of-concept ligation-amplification experiments, the authors showed that the engineered ligase (with the D540R mutation) outperformed the wild type at two temperatures. At 60°C, maximum amplification of ligated DNA product was achieved after only 3 cycles with the mutant but took 10 cycles with the wild type enzyme. At 30°C, the engineered enzyme gave maximum product yield after 5 cycles of ligation-amplification, whereas the product yield with the wild type ligase was only ~30% as great, even after 10 cycles [47]. Further, a series of insightful biophysics experiments showed that the introduction of a positively charged arginine residue, in place of the negatively charged Asp540, had accelerated both the formation of the covalent ligase-AMP intermediate and binding of the nicked DNA substrate [47]. This work is also the basis of a granted US patent, number 8,137,943.

In a follow-up study, the same group has recently used further mutagenesis to eliminate the ionic interactions between the AdD and the OBD entirely [27]. *P. furiosus* DNA ligase carrying either three point mutations (D540R/K554A/K558A), or D540R plus deletion of the final four amino acids of the C-terminal helix, had further enhanced nick-joining activities. While it is yet to be tested with other archaeal DNA ligases, this design approach, releasing the interactions of the C-terminal helix with the AdD and OBD domains, appears as though it could be a generally applicable one for enhancing activity.

6. RNA Ligases

RNA ligases (EC 6.5.1.3) are RNA end-joining enzymes involved in RNA repair, splicing, and editing pathways. In contrast to the ubiquitous DNA ligases, RNA ligases have a narrower phylogenetic distribution. Sequence similarity searches have found RNA ligases in all three domains of life but only in a subset of species [54].

RNA ligases are typically classified into two broad families. The Rnl1 family includes the eponymous RNA ligase 1 (Rnl1) from bacteriophage T4 [3] and the tRNA ligases from fungi, yeasts, and plants [5, 55, 56]. These enzymes repair breaks that have been introduced into single-stranded RNA by site-specific nucleases. The Rnl2 family contains the bacteriophage T4 RNA ligase 2 (Rnl2) and the RNA-editing ligases from the protozoans *Trypanosoma* and *Leishmania*. These enzymes are primarily associated with sealing nicks in RNA made duplex by the presence of a bridging complementary strand [54, 57, 58]. While the RNA ligases share the same six conserved nucleotidyl transferase motifs as DNA ligases, their overall levels of sequence conservation are low. In general, this makes family classification both more difficult and less meaningful.

Like DNA ligases, RNA ligases are also important in molecular biology. T4 RNA ligases 1 and 2 have become essential for a subset of rapid amplification of cDNA ends (RACE) protocols, 3′ RNA labelling, and most importantly, at present, the preparation of microRNA (miRNA) sequencing libraries. ATP-dependent RNA ligases capable of forming phosphodiester bonds between 5′-phosphate and 3′-hydroxyl termini are of most use in these protocols; therefore, they will be the focus of the following sections. For completeness, we also note that two noncanonical RNA ligases from the archaeal species *Pyrococcus horikoshii* have also been reported. The first is a putative 2′-5′ RNA ligase, the structure of which has been solved [59]. The second, RtcB, is an unusual ligase that joins either 2′,3′-cyclic phosphate or 3′-phosphate termini to 5′-hydroxyl termini. Its structure, its mechanism, and its interaction with a novel protein cofactor (Archease) have recently been characterised in detail [60, 61].

7. Archaeal RNA Ligases

The first detailed biochemical characterisation of an archaeal RNA ligase was reported in 2008, when an open reading frame from *Pyrococcus abyssi*, previously annotated as encoding a DNA ligase, was found to encode an RNA ligase instead [62]. Previously, it had been assumed that archaeal RNA ligases were likely to be Rnl2-like enzymes, as they showed similarly variant nucleotidyl transferase motifs as T4 Rnl2 [54]. However, the structure of the *P. abyssi* RNA ligase was a marginally closer structural homologue of T4 Rnl1 (secondary structure matching Z-score of 6.4, and RMSD of 2.78 Å over 200 aligned residues) than T4 Rnl2 (Z-score 6.2, RMSD 2.39 Å over 164 aligned residues) [62]. Further, the *P. abyssi* RNA ligase was active with single-stranded RNA substrates, but not double-stranded RNA, similar to T4 Rnl1.

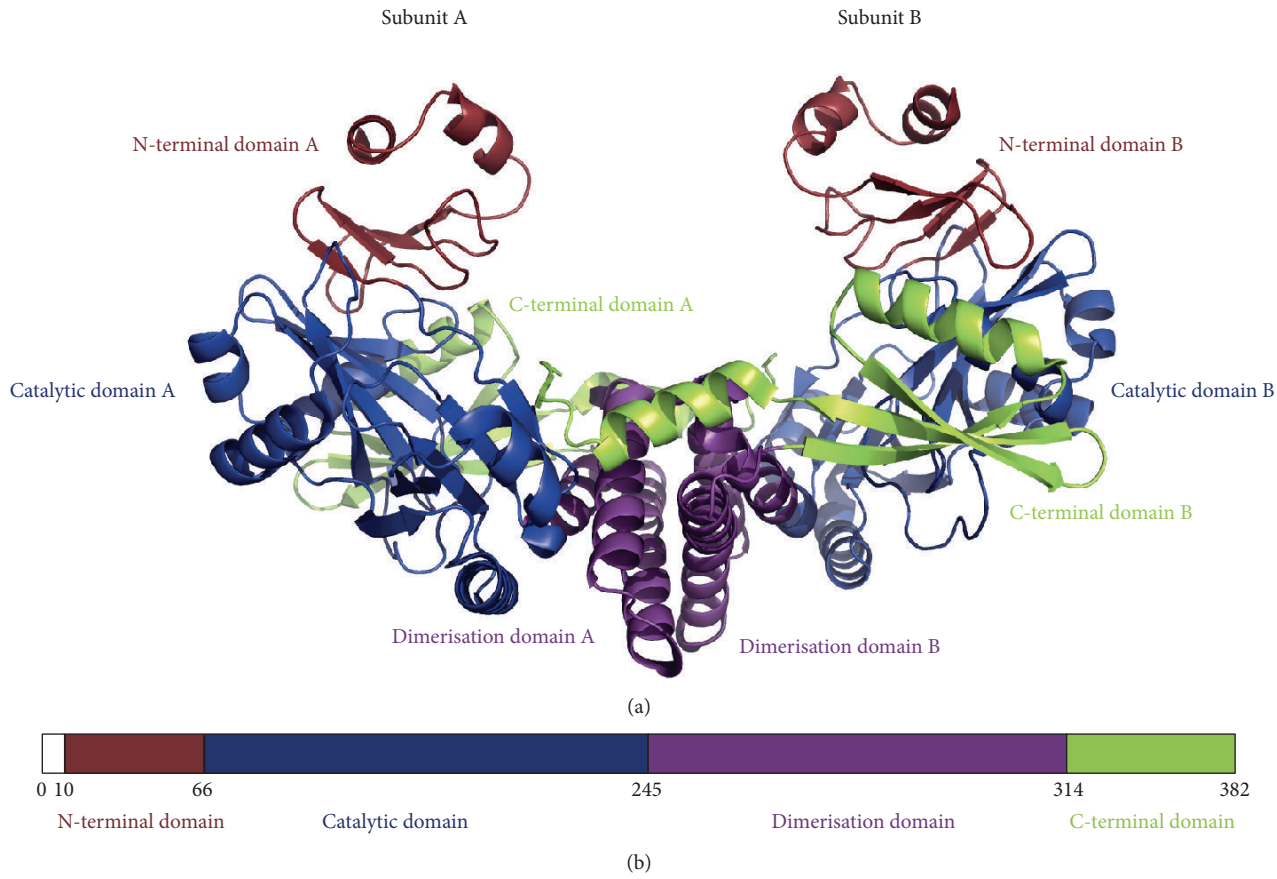


Figure 2: The *Pyrococcus abyssi* RNA ligase. (a) The homodimeric structure of the enzyme (PDB ID 2VUG). The four domains of each monomer are coloured individually and labelled. (b) Schematic diagram of the domain boundaries, with amino acid numbering shown.

Unlike the monomeric mesophilic ligases, X-ray crystallography revealed a homodimeric structure for the *P. abyssi* RNA ligase [62]. Each monomer comprised four domains: an N-terminal domain, a catalytic domain, a dimerisation domain, and a C-terminal domain (Figure 2). The catalytic domain showed structural similarities with other members of the nucleotidyl transferase superfamily. The N-terminal domain resembled that of T4 Rnl1 and has only been observed in these two enzymes to date. The C-terminal domain was all α-helical, had no structural homologues, and was not ascribed a function. The dimerisation domain had structural similarity to the copper-binding domain of the amyloid precursor protein, although the metal binding residues are absent in the ligase [62].

The role of dimerisation in the *P. abyssi* RNA ligase is not known. It has been proposed that it may be functionally important for facilitating two symmetric and simultaneous ligation events, such as splicing and intron circularisation [62]. More generally, oligomerisation is a common adaptation associated with thermophily in archaeal proteins [64]. This strategy is thought to increase the rigidity of the individual subunits and promote tighter packing of the hydrophobic core.

To date, the only other archaeal RNA ligase to be characterised biochemically is the one from the species *Methanobacterium thermoautotrophicum* [63], which is now generally known as *Methanothermobacter thermautotrophicus* [65]. The properties of the *P. abyssi* and *M. thermautotrophicus* RNA ligases are summarised in Table 2.

8. RNA Ligases in Biotechnology

In addition to their roles *in vivo*, RNA ligases have become important tools in molecular biology [66]. Shortly after the discovery of T4 Rnl1, new protocols were established, including 3′-end biotin and fluorophore labelling [67], RNA ligase mediated rapid amplification of cDNA ends (RLM-RACE) [68], oligonucleotide synthesis [69], and 5′ nucleotide modifications of both RNA and DNA [70].

More recently RNA ligases have become essential for constructing sequencing libraries of microRNAs (miRNAs) and other small RNAs. During library preparation, T4 RNA ligases are used to join 5′- and 3′-adaptors to the RNA substrates, so that the adaptor sequences can be used for priming during reverse transcription and PCR [71–73]. The emerging realisation that miRNAs, small regulatory RNAs

Table 2: Properties of archaeal RNA ligases.

Organism	UniProt ID	PDB ID	Co-Factor	T_{opt}	Properties	Reference
Pyrococcus abyssi	Q9UYG2	2VUG	ATP	85–90°C	Efficient at forming adenylated RNA but not so efficient at strand joining. Forms circular products.	[62]
Methanothermobacter thermautotrophicus	O27289	—	ATP	65°C	Circularises single-stranded RNA and single-stranded DNA.	[63]

T_{opt}: temperature optimum for the RNA ligase *in vitro*.

involved in posttranscriptional regulation [74], have numerous biological functions and whose misregulation have been implicated in a number of diseases [75], has meant that high-throughput screening has become an invaluable tool for both the discovery and profiling of miRNA expression. Therefore RNA ligases capable of producing high quality sequencing libraries, representative of the original miRNA population in a sample, are of great interest.

Unfortunately, it has become increasingly apparent that miRNA sequencing datasets are prone to severe biases [76] and that the adaptor ligation step is a key contributor. One limitation is the unwanted production of circular by-products [66, 76]. The evolutionarily conserved function of RNA ligases *in vivo* is to seal nicks in RNA hairpin loops (such as those in cleaved tRNA molecules). *In vitro*, this results in a propensity to circularise the RNA substrates, preventing adaptor ligation. Another limitation with the T4 RNA ligases is that they are biased towards ligating particular RNA sequences [77, 78] which can lead to the miscalculation of miRNA abundance by up to 4 orders of magnitude [73]. This ligation bias is not a result of primary sequence preference but instead a bias against RNA secondary structure [79]. Therefore there is growing interest in characterising thermostable RNA ligases that are active at temperatures sufficient to denature RNA secondary structures [80].

9. Biotechnological Roles for Archaeal RNA Ligases

Archaeal RNA ligases have found some use in molecular biology protocols. The *M. thermautotrophicus* RNA ligase has the ability to adenylate both single-stranded RNA and single-stranded DNA (Table 2) and it has been used to 5′-adenylate single-stranded DNA adapters for use in the construction of miRNA sequencing libraries [81]. Previously either a chemical synthesis protocol [82] or a methodology involving T4 DNA ligase was used for this adenylation step [83]; however, T4 DNA ligase does not accumulate sufficient adenylated products and the synthesis method was expensive. On the other hand, the *M. thermautotrophicus* RNA ligase accumulates high quantities of the adenylated intermediates (AppRNA and AppDNA) when an excess of ATP is used in the reaction, making it an ideal substitute. This enzyme is currently available commercially as a component of a 5′ DNA adenylation kit (from New England Biolabs).

While the *M. thermautotrophicus* RNA ligase is highly active as a 5′ adenylation enzyme, a single point mutation (Lys97 → Ala) resulted in an enzyme that was unable to perform adenylation at all, but which retained the ability to form phosphodiester bonds [81]. This has enabled the development of a two-step protocol in which the wild type enzyme is used to adenylate DNA adapters in an initial reaction. The adenylated adaptors can then be incubated with the pool of target miRNA molecules and the mutated ligase. The result is ligation of preadenylated adaptors to the RNA substrates, with no possibility of circularising the (nonadenylated) RNA [81]. The ability of the *M. thermautotrophicus* RNA ligase (and the K97A mutant) to function at 65°C also helps to remove ligation bias associated with RNA secondary structures. In order to implement this protocol, the mutated enzyme is commercially available (as the Thermostable 5′ AppDNA/RNA Ligase from New England Biolabs).

10. Concluding Remarks

In this review, we have summarised the current literature on archaeal nucleic acid ligases. We have highlighted the relative dearth of knowledge on these enzymes, while discussing characteristics that are likely to make them valuable additions to the biotechnologist's toolbox in future. In particular, archaeal enzymes are archetypically thermostable. DNA ligases that are stable and active at elevated temperatures are becoming critical for emerging technologies such as Gibson assembly (which underpins synthetic biology) [12], while thermostable RNA ligases offer the promise of constructing unbiased miRNA sequencing libraries [76]. Moreover, the thermostability of archaeal enzymes makes them ideal starting points for protein engineering [52]. Recent experiments to engineer the *Pyrococcus furiosus* DNA ligase [27, 47] and the *Methanothermobacter thermautotrophicus* RNA ligase [81] demonstrate the great potential for further work in this area.

Overall, the pool of archaeal nucleic acid ligases is diverse but currently undersampled. We anticipate that its further exploration will lead to the discovery of new enzymes with

favourable properties for molecular biology and biotechnology, which in turn will drive the development of new methodologies.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by a Smart Ideas Grant from the New Zealand Ministry of Business, Innovation and Employment. Wayne M. Patrick is the grateful recipient of a Rutherford Discovery Fellowship.

References

[1] B. Weiss and C. C. Richardson, "Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from *Escherichia coli* infected with T4 bacteriophage," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 57, no. 4, pp. 1021–1028, 1967.

[2] I. R. Lehman, "DNA ligase: structure, mechanism, and function," *Science*, vol. 186, no. 4166, pp. 790–797, 1974.

[3] R. Silber, V. G. Malathi, and J. Hurwitz, "Purification and properties of bacteriophage T4-induced RNA ligase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 69, no. 10, pp. 3009–3013, 1972.

[4] S. Shuman and B. Schwer, "RNA capping enzyme and DNA ligase: A superfamily of covalent nucleotidyl transferases," *Molecular Microbiology*, vol. 17, no. 3, pp. 405–410, 1995.

[5] L. K. Wang and S. Shuman, "Structure-function analysis of yeast tRNA ligase," *RNA*, vol. 11, no. 6, pp. 966–975, 2005.

[6] S. Shuman, "DNA ligases: progress and prospects," *The Journal of Biological Chemistry*, vol. 284, no. 26, pp. 17365–17369, 2009.

[7] K. S. Lundberg, D. D. Shoemaker, M. W. W. Adams, J. M. Short, J. A. Sorge, and E. J. Mathur, "High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*," *Gene*, vol. 108, no. 1, pp. 1–6, 1991.

[8] C. Papworth, J. C. Bauer, J. Braman, and D. A. Wright, "Site-directed mutagenesis in one day with >80% efficiency," *Strategies*, vol. 9, pp. 3–4, 1996.

[9] H. H. Hogrefe, J. Cline, G. L. Youngblood, and R. M. Allen, "Creating randomized amino acid libraries with the Quikchange multi site-directed mutagenesis kit," *BioTechniques*, vol. 33, no. 5, pp. 1158–1165, 2002.

[10] A. Zhao, F. C. Gray, and S. A. MacNeill, "ATP- and NAD+-dependent DNA ligases share an essential function in the halophilic archaeon Haloferax volcanii," *Molecular Microbiology*, vol. 59, no. 3, pp. 743–752, 2006.

[11] F. Barany, "Genetic disease detection and DNA amplification using cloned thermostable ligase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 1, pp. 189–193, 1991.

[12] D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, "Enzymatic assembly of DNA molecules up to several hundred kilobases," *Nature Methods*, vol. 6, no. 5, pp. 343–345, 2009.

[13] M. A. Quail, I. Kozarewa, F. Smith et al., "A large genome center's improvements to the Illumina sequencing system," *Nature Methods*, vol. 5, no. 12, pp. 1005–1010, 2008.

[14] G. J. Lohman, S. Tabor, and N. M. Nichols, "DNA ligases," in *Current Protocols in Molecular Biology*, vol. 94, unit 3.14, pp. 11–17, 2011.

[15] V. Sgaramella, J. H. van de Sande, and H. G. Khorana, "Studies on polynucleotides, C. A novel joining reaction catalyzed by the T4-polynucleotide ligase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 67, no. 3, pp. 1468–1475, 1970.

[16] A. Sugino, H. M. Goodman, H. L. Heyneker, J. Shine, H. W. Boyer, and N. R. Cozzarelli, "Interaction of bacteriophage T4 RNA and DNA ligases in joining of duplex DNA at base-paired ends," *The Journal of Biological Chemistry*, vol. 252, no. 11, pp. 3987–3994, 1977.

[17] A. Wilkinson, J. Day, and R. Bowater, "Bacterial DNA ligases," *Molecular Microbiology*, vol. 40, no. 6, pp. 1241–1248, 2001.

[18] A. E. Tomkinson, S. Vijayakumar, J. M. Pascal, and T. Ellenberger, "DNA ligases: structure, reaction mechanism, and function," *Chemical Reviews*, vol. 106, no. 2, pp. 687–699, 2006.

[19] J.-L. Rolland, Y. Gueguen, C. Persillon, J.-M. Masson, and J. Dietrich, "Characterization of a thermophilic DNA ligase from the archaeon *Thermococcus fumicolans*," *FEMS Microbiology Letters*, vol. 236, no. 2, pp. 267–273, 2004.

[20] M. Nakatani, S. Ezaki, H. Atomi, and T. Imanaka, "A DNA ligase from a hyperthermophilic archaeon with unique cofactor specificity," *Journal of Bacteriology*, vol. 182, no. 22, pp. 6424–6433, 2000.

[21] Y. J. Kim, H. S. Lee, S. S. Bae et al., "Cloning, expression, and characterization of a DNA ligase from a hyperthermophilic archaeon *Thermococcus* sp.," *Biotechnology Letters*, vol. 28, no. 6, pp. 401–407, 2006.

[22] Y. Sun, M. S. Seo, J. H. Kim et al., "Novel DNA ligase with broad nucleotide cofactor specificity from the hyperthermophilic crenarchaeon *Sulfophobococcus zilligii*: influence of ancestral DNA ligase on cofactor utilization," *Environmental Microbiology*, vol. 10, no. 12, pp. 3212–3224, 2008.

[23] N. Keppetipola and S. Shuman, "Characterization of a thermophilic ATP-dependent DNA ligase from the euryarchaeon *Pyrococcus horikoshii*," *Journal of Bacteriology*, vol. 187, no. 20, pp. 6902–6908, 2005.

[24] K. Takano, A. Aoi, Y. Koga, and S. Kanaya, "Evolvability of thermophilic proteins from archaea and bacteria," *Biochemistry*, vol. 52, no. 28, pp. 4774–4780, 2013.

[25] D. J. Kim, O. Kim, H.-W. Kim, H. S. Kim, S. J. Lee, and S. W. Suh, "ATP-dependent DNA ligase from *Archaeoglobus fulgidus* displays a tightly closed conformation," *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, vol. 65, no. 6, pp. 544–550, 2009.

[26] H. Nishida, S. Kiyonari, Y. Ishino, and K. Morikawa, "The closed structure of an archaeal DNA ligase from *Pyrococcus furiosus*," *Journal of Molecular Biology*, vol. 360, no. 5, pp. 956–967, 2006.

[27] M. Tanabe, S. Ishino, Y. Ishino, and H. Nishida, "Mutations of Asp540 and the domain-connecting residues synergistically enhance *Pyrococcus furiosus* DNA ligase activity," *FEBS Letters*, vol. 588, no. 2, pp. 230–235, 2014.

[28] J. M. Pascal, O. V. Tsodikov, G. L. Hura et al., "A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA," *Molecular Cell*, vol. 24, no. 2, pp. 279–291, 2006.

[29] S. Supangat, Y. J. An, Y. Sun, S.-T. Kwon, and S.-S. Cha, "Purification, crystallization and preliminary crystallographic analysis of a multiple cofactor-dependent DNA ligase from *Sulfophobococcus zilligii*," *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, vol. 66, no. 12, pp. 1583–1585, 2010.

[30] T. E. Petrova, E. Y. Bezsudnova, B. D. Dorokhov et al., "Expression, purification, crystallization and preliminary crystallographic analysis of a thermostable DNA ligase from the archaeon *Thermococcus sibiricus*," *Acta Crystallographica, Section F: Structural Biology and Crystallization Communications*, vol. 68, no. 2, pp. 163–165, 2012.

[31] T. Petrova, E. Y. Bezsudnova, K. M. Boyko et al., "ATP-dependent DNA ligase from *Thermococcus* sp. 1519 displays a new arrangement of the OB-fold domain," *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, vol. 68, no. 12, pp. 1440–1447, 2012.

[32] S. Shuman, Y. Liu, and B. Schwer, "Covalent catalysis in nucleotidyl transfer reactions: essential motifs in *Saccharomyces cerevisiae* RNA capping enzyme are conserved in *Schizosaccharomyces pombe* and viral capping enzymes and among polynucleotide ligases," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 25, pp. 12046–12050, 1994.

[33] J. M. Pascal, P. J. O'Brien, A. E. Tomkinson, and T. Ellenberger, "Human DNA ligase I completely encircles and partially unwinds nicked DNA," *Nature*, vol. 432, no. 7016, pp. 473–478, 2004.

[34] M. Odell, V. Sriskanda, S. Shuman, and D. B. Nikolov, "Crystal structure of eukaryotic DNA ligase-adenylate illuminates the mechanism of nick sensing and strand joining," *Molecular Cell*, vol. 6, no. 5, pp. 1183–1193, 2000.

[35] P. A. Nair, J. Nandakumar, P. Smith, M. Odell, C. D. Lima, and S. Shuman, "Structural basis for nick recognition by a minimal pluripotent DNA ligase," *Nature Structural and Molecular Biology*, vol. 14, no. 8, pp. 770–778, 2007.

[36] H. S. Subramanya, A. J. Doherty, S. R. Ashford, and D. B. Wigley, "Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7," *Cell*, vol. 85, no. 4, pp. 607–615, 1996.

[37] G. J. S. Lohman, L. Chen, and T. C. Evans Jr., "Kinetic characterization of single strand break ligation in duplex DNA by T4 DNA ligase," *The Journal of Biological Chemistry*, vol. 286, no. 51, pp. 44187–44196, 2011.

[38] S.-J. Jeon and K. Ishikawa, "A novel ADP-dependent DNA ligase from *Aeropyrum pernix K1*," *FEBS Letters*, vol. 550, no. 1–3, pp. 69–73, 2003.

[39] A. Kletzin, "Molecular characterisation of a DNA ligase gene of the extremely thermophilic archaeon *Desulfurolobus ambivalens* shows close phylogenetic relationship to eukaryotic ligases," *Nucleic Acids Research*, vol. 20, no. 20, pp. 5389–5396, 1992.

[40] M. S. Seo, Y. J. Kim, J. J. Choi et al., "Cloning and expression of a DNA ligase from the hyperthermophilic archaeon *Staphylothermus marinus* and properties of the enzyme," *Journal of Biotechnology*, vol. 128, no. 3, pp. 519–530, 2007.

[41] M. Ferrer, O. V. Golyshina, A. Beloqui et al., "A purple acidophilic di-ferric DNA ligase from *Ferroplasma*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 26, pp. 8878–8883, 2008.

[42] X. Lai, H. Shao, F. Hao, and L. Huang, "Biochemical characterization of an ATP-dependent DNA ligase from the hyperthermophilic crenarchaeon *Sulfolobus shibatae*," *Extremophiles*, vol. 6, no. 6, pp. 469–477, 2002.

[43] B. R. Jackson, C. Noble, M. Lavesa-Curto, P. L. Bond, and R. P. Bowater, "Characterization of an ATP-dependent DNA ligase from the acidophilic archaeon '*Ferroplasma acidarmanus*' Fer1," *Extremophiles*, vol. 11, no. 2, pp. 315–327, 2007.

[44] L. Poidevin and S. A. MacNeill, "Biochemical characterisation of LigN, an NAD^+-dependent DNA ligase from the halophilic euryarchaeon *Haloferax volcanii* that displays maximal *in vitro* activity at high salt concentrations," *BMC Molecular Biology*, vol. 7, article 44, 2006.

[45] G. Ortega, A. Laín, X. Tadeo, B. López-Méndez, D. Castaño, and O. Millet, "Halophilic enzyme activation induced by salts," *Scientific Reports*, vol. 1, article 6, 2011.

[46] V. Sriskanda, Z. Kelman, J. Hurwitz, and S. Shuman, "Characterization of an ATP-dependent DNA ligase from the the thermophilic archaeon *Methanobacterium thermoautotrophicum*," *Nucleic Acids Research*, vol. 28, no. 11, pp. 2221–2228, 2000.

[47] M. Tanabe, S. Ishino, M. Yohda, K. Morikawa, Y. Ishino, and H. Nishida, "Structure-based mutational study of an archaeal DNA ligase towards improvement of ligation activity," *ChemBioChem*, vol. 13, no. 17, pp. 2575–2582, 2012.

[48] V. A. Smagin, A. V. Mardanov, E. A. Bonch-Osmolovskaia, and N. V. Ravin, "Isolation and characteristics of new thermostable DNA ligase from archaea of the genus *Thermococcus*," *Prikladnaia Biokhimiia i Mikrobiologiia*, vol. 44, no. 5, pp. 523–528, 2008.

[49] E. Y. Bezsudnova, M. V. Kovalchuk, A. V. Mardanov et al., "Overexpression, purification and crystallization of a thermostable DNA ligase from the archaeon *Thermococcus* sp. 1519," *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, vol. 65, no. 4, pp. 368–371, 2009.

[50] M. Nakatani, S. Ezaki, H. Atomi, and T. Imanaka, "Substrate recognition and fidelity of strand joining by an archaeal DNA ligase," *European Journal of Biochemistry*, vol. 269, no. 2, pp. 650–656, 2002.

[51] E. Cotner-Gohara, I.-K. Kim, M. Hammel, J. A. Tainer, A. E. Tomkinson, and T. Ellenberger, "Human DNA ligase III recognizes DNA ends by dynamic switching between two DNA-bound states," *Biochemistry*, vol. 49, no. 29, pp. 6165–6176, 2010.

[52] W. Besenmatter, P. Kast, and D. Hilvert, "Relative tolerance of mesostable and thermostable protein homologs to extensive mutation," *Proteins: Structure, Function and Genetics*, vol. 66, no. 2, pp. 500–506, 2007.

[53] R. H. Wilson, S. K. Morton, H. Deiderick et al., "Engineered DNA ligases with improved activities *in vitro*," *Protein Engineering, Design and Selection*, vol. 26, no. 7, pp. 471–478, 2013.

[54] C. K. Ho and S. Shuman, "Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 20, pp. 12709–12714, 2002.

[55] J. Abelson, C. R. Trotta, and H. Li, "tRNA splicing," *The Journal of Biological Chemistry*, vol. 273, no. 21, pp. 12685–12688, 1998.

[56] M. Englert and H. Beier, "Plant tRNA ligases are multifunctional enzymes that have diverged in sequence and substrate specificity from RNA ligases of other phylogenetic origins," *Nucleic Acids Research*, vol. 33, no. 1, pp. 388–399, 2005.

[57] V. Blanc, J. D. Alfonzo, R. Aphasizhev, and L. Simpson, "The mitochondrial RNA ligase from *Leishmania tarentolae* can join RNA molecules bridged by a complementary RNA," *Journal of Biological Chemistry*, vol. 274, no. 34, pp. 24289–24296, 1999.

[58] S. S. Palazzo, A. K. Panigrahi, R. P. Igo Jr., R. Salavati, and K. Stuart, "Kinetoplastid RNA editing ligases: complex association, characterization, and substrate requirements," *Molecular and Biochemical Parasitology*, vol. 127, no. 2, pp. 161–167, 2003.

[59] P. H. Rehse and T. H. Tahirov, "Structure of a putative 2′-5′ RNA ligase from *Pyrococcus horikoshii*," *Acta Crystallographica Section D: Biological Crystallography*, vol. 61, no. 9, pp. 1207–1212, 2005.

[60] K. K. Desai, C. A. Bingman, G. N. Phillips Jr., and R. T. Raines, "Structures of the noncanonical RNA ligase rtcb reveal the mechanism of histidine guanylylation," *Biochemistry*, vol. 52, no. 15, pp. 2518–2525, 2013.

[61] K. K. Desai, C. L. Cheng, C. A. Bingman, G. N. Phillips Jr., and R. T. Raines, "A tRNA splicing operon: archease endows RtcB with dual GTP/ATP cofactor specificity and accelerates RNA ligation," *Nucleic Acids Research*, vol. 42, no. 6, pp. 3931–3942, 2014.

[62] M. A. Brooks, L. Meslet-Cladiére, M. Graille et al., "The structure of an archaeal homodimeric ligase which has RNA circularization activity," *Protein Science*, vol. 17, no. 8, pp. 1336–1345, 2008.

[63] C. Torchia, Y. Takagi, and C. K. Ho, "Archaeal RNA ligase is a homodimeric protein that catalyzes intramolecular ligation of single-stranded RNA and DNA," *Nucleic Acids Research*, vol. 36, no. 19, pp. 6218–6227, 2008.

[64] C. J. Reed, H. Lewis, E. Trejo, V. Winston, and C. Evilia, "Protein adaptations in archaeal extremophiles," *Archaea*, vol. 2013, Article ID 373275, 14 pages, 2013.

[65] A. Wasserfallen, J. Nölling, P. Pfister, J. Reeve, and E. C. De Macario, "Phylogenetic analysis of 18 thermophilic *Methanobacterium* isolates supports the proposals to create a new genus, *Methanothermobacter* gen. nov., and to reclassify several isolates in three species, *Methanothermobacter thermautotrophicus* comb. nov., *Methanothermobacter wolfeii* comb. nov., and *Methanothermobacter marburgensis* sp. nov.," *International Journal of Systematic and Evolutionary Microbiology*, vol. 50, no. 1, pp. 43–53, 2000.

[66] E. Paredes, M. Evans, and S. R. Das, "RNA labeling, conjugation and ligation," *Methods*, vol. 54, no. 2, pp. 251–259, 2011.

[67] R. W. Richardson and R. I. Gumport, "Biotin and fluorescent labeling of RNA using T4 RNA ligase," *Nucleic Acids Research*, vol. 11, no. 18, pp. 6167–6184, 1983.

[68] X. Liu and M. A. Gorovsky, "Mapping the 5′ and 3′ ends of *Tetrahymena thermophila* mRNAs using RNA ligase mediated amplification of cDNA ends (RLM-RACE)," *Nucleic Acids Research*, vol. 21, no. 21, pp. 4954–4960, 1993.

[69] S. Kaluz, M. Kaluzova, and A. P. F. Flint, "Enzymatically produced composite primers: An application of T4 RNA ligase—coupled primers to PCR," *BioTechniques*, vol. 19, no. 2, pp. 182–186, 1995.

[70] Y. Kinoshita, K. Nishigaki, and Y. Husimi, "Fluorescence-, isotope- or biotin-labeling of the 5′-end of single-stranded DNA/RNA using T4 RNA ligase," *Nucleic Acids Research*, vol. 25, no. 18, pp. 3747–3748, 1997.

[71] C. Lu, B. C. Meyers, and P. J. Green, "Construction of small RNA cDNA libraries for deep sequencing," *Methods*, vol. 43, no. 2, pp. 110–117, 2007.

[72] M. Hafner, P. Landgraf, J. Ludwig et al., "Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing," *Methods*, vol. 44, no. 1, pp. 3–12, 2008.

[73] S. E. V. Linsen, E. de Wit, G. Janssens et al., "Limitations and possibilities of small RNA digital gene expression profiling," *Nature Methods*, vol. 6, no. 7, pp. 474–476, 2009.

[74] N. C. Lau, L. P. Lim, E. G. Weinstein, and D. P. Bartel, "An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*," *Science*, vol. 294, no. 5543, pp. 858–862, 2001.

[75] T.-C. Chang and J. T. Mendell, "MicroRNAs in vertebrate physiology and human disease," *Annual Review of Genomics and Human Genetics*, vol. 8, pp. 215–239, 2007.

[76] C. A. Raabe, T.-H. Tang, J. Brosius, and T. S. Rozhdestvensky, "Biases in small RNA deep sequencing data," *Nucleic Acids Research*, vol. 42, no. 3, pp. 1414–1426, 2014.

[77] A. D. Jayaprakash, O. Jabado, B. D. Brown, and R. Sachidanandam, "Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing," *Nucleic Acids Research*, vol. 39, no. 21, article e141, 2011.

[78] M. Hafner, N. Renwick, M. Brown et al., "RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries," *RNA*, vol. 17, no. 9, pp. 1697–1712, 2011.

[79] F. Zhuang, R. T. Fuchs, Z. Sun, Y. Zheng, and G. B. Robb, "Structural bias in T4 RNA ligase-mediated 3′-adapter ligation," *Nucleic Acids Research*, vol. 40, no. 7, p. e54, 2012.

[80] T. Blondal, A. Thorisdottir, U. Unnsteinsdottir et al., "Isolation and characterization of a thermostable RNA ligase 1 from a *Thermus scotoductus* bacteriophage TS2126 with good single-stranded DNA ligation properties," *Nucleic Acids Research*, vol. 33, no. 1, pp. 135–142, 2005.

[81] A. M. Zhelkovsky and L. A. McReynolds, "Structure-function analysis of *Methanobacterium thermoautotrophicum* RNA ligase - engineering a thermostable ATP independent enzyme," *BMC Molecular Biology*, vol. 13, article 24, 2012.

[82] Q. Dai, M. Saikia, N.-S. Li, T. Pan, and J. A. Piccirilli, "Efficient chemical synthesis of AppDNA by adenylation of immobilized DNA-5′-monophosphate," *Organic Letters*, vol. 11, no. 5, pp. 1067–1070, 2009.

[83] W. Chiuman and Y. Li, "Making AppDNA using T4 DNA ligase," *Bioorganic Chemistry*, vol. 30, no. 5, pp. 332–349, 2002.