Protein Engineering, Design & Selection vol. 26 no. 7 pp. 471–478, 2013
Published online June 10, 2013 doi:10.1093/protein/gzt024

Engineered DNA ligases with improved activities *in vitro*

Robert H.Wilson[1†], Susan K.Morton[1†], Heather Deiderick[2], Monica L.Gerth[1,3], Hayden A.Paul[1], Ilana Gerber[1], Ankita Patel[1], Andrew D.Ellington[4], Scott P.Hunicke-Smith[2] and Wayne M.Patrick[1,3,5]

[1]Institute of Natural Sciences, Massey University, Auckland 0632, New Zealand, [2]Genome Sequencing and Analysis Facility, University of Texas, Austin, TX 78712, USA, [3]Department of Biochemistry, University of Otago, Dunedin 9054, New Zealand and [4]Department of Chemistry and Biochemistry, University of Texas, Austin, TX 78712, USA

[5]To whom correspondence should be addressed.
E-mail: wayne.patrick@otago.ac.nz

Received February 28, 2013; revised April 24, 2013; accepted May 9, 2013

Edited by Matthew Cordes

The DNA ligase from bacteriophage T4 is one of the most widely used enzymes in molecular biology. It has evolved to seal single-stranded nicks in double-stranded DNA, but not to join double-stranded fragments with cohesive or blunt ends. Its poor activity *in vitro*, particularly with blunt-ended substrates, can lead to failed or sub-optimal experimental outcomes. We have fused T4 DNA ligase to seven different DNA-binding proteins, including eukaryotic transcription factors, bacterial DNA repair proteins and archaeal DNA-binding domains. Representatives from each of these classes improved the activity of T4 DNA ligase, by up to 7-fold, in agarose gel-based screens for cohesive- and blunt-ended fragment joining. Overall, the most active variants were p50-ligase (i.e. NF-κB p50 fused to T4 DNA ligase) and ligase-cTF (T4 DNA ligase fused to an artificial, chimeric transcription factor). Ligase-cTF out-performed T4 DNA ligase by ∼160% in blunt end 'vector + insert' cloning assays, and p50-ligase showed an improvement of a similar magnitude when it was used to construct a library for Illumina sequencing. The activity of the *Escherichia coli* DNA ligase was also enhanced by fusion to p50. Together, these results suggest that our protein design strategy is a generalizable one for engineering improved DNA ligases.

Keywords: DNA-binding protein/DNA ligase/enzyme engineering/fusion protein

Introduction

DNA ligases catalyze the formation of new phosphodiester linkages in DNA, through the condensation of adjacent 3′-hydroxyl and 5′-phosphate termini. The first step in the ligation reaction is the formation of a ligase-AMP covalent intermediate. DNA ligases are classified into two families—ATP-dependent and NAD^+-dependent—according to the identity of their AMP donor. DNA ligases are essential for DNA replication, repair and recombination in all organisms. Consequently, their mechanism has been well studied and numerous structures have been solved (Tomkinson *et al.*, 2006; Pascal, 2008; Shuman, 2009).

DNA ligases are critical for many applications in molecular biology and biotechnology (Lohman *et al.*, 2011a). For decades, they have been used in the construction of recombinant DNA molecules (i.e. cloning), and for mutation detection using the ligation chain reaction (Barany, 1991). More recently, a method for the isothermal assembly of very large DNA fragments (Gibson *et al.*, 2009), which utilizes the thermostable *Taq* DNA ligase, has been described and widely adopted. DNA ligases can also be used in gene synthesis (Bang and Church, 2008). They are essential in many next-generation sequencing (NGS) methods, either for adapter ligation during sample preparation (e.g. Illumina, 454 and Ion Torrent sequencing), or for the sequencing reaction itself (e.g. SOLiD sequencing).

The most commonly used DNA ligase in these applications is the ATP-dependent enzyme from bacteriophage T4, which was also one of the first to be discovered (Weiss and Richardson, 1967). The T4 DNA ligase is the product of gene *30* (*lig*) and it is essential for sealing Okazaki fragments during replication of the phage genome (Miller *et al.*, 2003). Consistent with its physiological role, *in vitro* it is highly efficient at sealing single-stranded nicks in duplex DNA. Early experiments suggested a K_M of 1.5 nM for 5′-phosphates in nicked DNA substrates (Weiss *et al.*, 1968), while a more recent study with a homogeneous nicked DNA substrate yielded steady state kinetic parameters of $k_{cat} = 0.4\ s^{-1}$ and $k_{cat}/K_M = 1.5 \times 10^8\ M^{-1}\ s^{-1}$ (Lohman *et al.*, 2011b).

While T4 DNA ligase has evolved to be a nick-sealing enzyme, it can also join double-stranded DNA (dsDNA) fragments that have complementary, overhanging, single-stranded ('cohesive') ends. Further, it is the only commercially available ligase that can join blunt-ended DNA duplexes in the absence of macromolecular enhancers such as polyethylene glycol (Sgaramella *et al.*, 1970). It is the ligation of cohesive- or blunt-ended dsDNA fragments that is most commonly required in molecular biology protocols. However, T4 DNA ligase is comparatively poor at catalyzing these reactions. In particular, its K_M for cohesive ends has been estimated to be 0.6 μM, while the K_M for blunt-ended dsDNA substrates is ∼50 μM (Sugino *et al.*, 1977). The turnover numbers for fragment joining reactions are also lower than for nick sealing (Sugino *et al.*, 1977), meaning that T4 DNA ligase is approximately five orders of magnitude less efficient at joining blunt-ended duplexes than it is at sealing nicks. It is similarly inefficient at ligating fragments with single base overhangs (Lohman *et al.*, 2011a).

For practitioners of molecular biology, the poor kinetic parameters of T4 DNA ligase usually manifest as failed cloning experiments, or sub-optimal libraries for Illumina and 454 sequencing runs. The aim of this study was to engineer DNA ligase variants with improved *in vitro* activities that would be useful for molecular biologists. Previously, the genetic fusion of a sequence non-specific dsDNA-binding protein (Sso7d from *Sulfolobus solfataricus*) to two DNA polymerases (*Taq* and *Pfu*) was shown to result in chimeric enzymes with increased

[†]These authors contributed equally to this work.

Table I. DNA-binding proteins that were fused to T4 DNA ligase

Protein	Source organism	Source of the gene	Reference
Eukaryotic transcription factors			
NF-κB p50	*Homo sapiens*	Plasmid pRES112	Patrick and Blackburn (2005)
NFAT	*Mus musculus*	Plasmid pHAR503[a]	de Lumley *et al.* (2004)
cTF	*M.musculus/H.sapiens*	Plasmid pSAN101[a]	de Lumley *et al.* (2004)
Bacterial DNA repair proteins			
PprA	*Deinococcus radiodurans*	Synthetic	Narumi *et al.* (2004)
Ku	*Mycobacterium tuberculosis*	Strain H37Rv, genomic DNA[b]	Weller *et al.* (2002)
Archaeal DNA-binding domains			
Sso7d	*Sulfolobus solfataricus*	Synthetic	Wang *et al.* (2004)
$[(HhH)_2]_2$	*Methanopyrus kandleri*	Synthetic	de Vega *et al.* (2010)

[a]Plasmids pHAR503 and pSAN101 were kind gifts from Prof. Jonathan Blackburn (University of Cape Town, South Africa).
[b]*Mycobacterium tuberculosis* H37Rv genomic DNA was a generous gift from Dr Stephanie Dawes (University of Auckland, New Zealand).

processivity and improved performance in polymerase chain reaction (PCR) amplifications (Wang *et al.*, 2004). Inspired by this work, we set out to construct and test a variety of chimeras, each comprising a DNA ligase fused to a DNA-binding protein. While processivity is not a relevant consideration for improving ligation activity, we hypothesized that increasing the affinity of the DNA ligase for its dsDNA substrates would be advantageous. We demonstrate that the activities of both the ATP-dependent T4 DNA ligase, and the NAD^+-dependent *Escherichia coli* DNA ligase, can be modulated by fusion to eukaryotic transcription factors, bacterial DNA repair proteins, and archaeal DNA-binding domains. Practical benefits include improvements of ∼160% both in colony counts when cloning blunt-ended fragments, and in reads when constructing a library for Illumina sequencing.

Materials and methods

Materials

Restriction enzymes and the NEBNext DNA Library Prep Master Mix Set for Illumina were from New England Biolabs (Ipswich, MA, USA). Vent or Phusion polymerases (both New England Biolabs) were used for all PCRs, except where noted otherwise. All oligonucleotide primers were from Integrated DNA Technologies (Coralville, IA, USA); their sequences are listed in Supplementary Table SI. Isopropyl-β-D-thiogalactopyranoside (IPTG) was from Gold Biotechnology (St Louis, MO, USA). Protease inhibitor cocktail, chicken egg white lysozyme and proteinase K were from Sigma Chemical Co. (St Louis, MO, USA). Talon metal affinity resin was from Clontech (Mountain View, CA, USA). 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) was from Melford Laboratories (Ipswich, Suffolk, UK). *Escherichia coli* strain DH5α-E (Invitrogen, Carlsbad, CA, USA) was used for all cloning and protein expression.

Construction of ligase fusion proteins

The T4 DNA ligase gene (*lig*) was from plasmid pRBL (Ren *et al.*, 1997), which was a gift from Assoc. Prof. Ichiro Matsumura (Emory University, Atlanta, GA, USA). The *E.coli* DNA ligase gene (*ligA*) was from the ASKA collection of open reading frames (Kitagawa *et al.*, 2005). T4 DNA ligase was fused to DNA-binding proteins from a variety of sources (Table I); *E.coli* DNA ligase was fused to NF-κB p50 only. Standard cloning techniques were used to construct a total of 17 plasmids, for the expression of T4 DNA ligase, *E.coli* DNA ligase and 15 fusion proteins. In all cases, the IPTG-inducible expression vector used was pCA24N (Kitagawa *et al.*, 2005). Full details of the cloning strategies used to construct each plasmid are provided in the Supplementary data.

Construction of ligase-cTF linker variants

The linker between the T4 DNA ligase and DNA-binding domains of ligase-cTF was replaced with four other linkers, which we abbreviate Natural Linker Short (NLS), Natural Linker Long (NLL), Flexible Linker (FL), and Helical Linker (HL). The methods used to construct vectors for expressing ligase-NLS-cTF, ligase-NLL-cTF, ligase-FL-cTF and ligase-HL-cTF are described in the Supplementary data.

Purification of ligase fusion proteins

The DNA ligases were all cloned with His_6 tags at their N-termini, to facilitate purification by immobilized metal affinity chromatography (IMAC). The expression and purification protocol is provided in the Supplementary data.

Gel-based activity assays

The ligation activities of each fusion protein were ranked using semi-quantitative, gel-based assays. For cohesive-ended ligation tests, a 1277-bp PCR product was generated by amplifying the ASKA plasmid pCA24N-*ompC* (Kitagawa *et al.*, 2005) with the primers pCA24N.for and pCA24N.rev (Supplementary Table SI). The PCR product was cleaved with SpeI, yielding two fragments of very similar size (638 and 639 bp). The two products were co-purified from an agarose gel using the Qiaquick Gel Extraction Kit (Qiagen, Valencia, CA, USA). For the T4 DNA ligase variants, each 10 μl activity assay contained 150 ng of this DNA substrate (∼40 nM ligatable DNA ends) and a ligase (2 μM), in 1× New England Biolabs T4 DNA Ligase Reaction Buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP, pH 7.5). Reactions were incubated for 10 min at 16°C, before being stopped by heating at 65°C for a further 15 min. The samples were purified using Qiagen MinElute columns, and then run on an ethidium bromide-stained agarose gel. Activity was defined as the appearance of the 1277-bp ligated product, and the disappearance of the 638-/639-bp substrate. Band intensities were quantified using the ImageJ 1.46r software package (http://rsbweb.nih.gov/ij/). For the variants of *E.coli* ligase, enzyme (2 μM) was incubated with 170 ng of the DNA

substrate (~40 nM ligatable ends) in 10 μl of 1× New England Biolabs *E.coli* DNA Ligase Reaction Buffer (30 mM Tris-HCl, 4 mM $MgCl_2$, 1 mM dithiothreitol, 26 μM NAD^+, 50 μg ml^{-1} BSA, pH 8.0). Reactions were incubated at 16°C for 16 h, before being heat-inactivated (65°C, 20 min) and run on an agarose gel.

To test for ligation activity with blunt-ended substrates, the ASKA plasmid pCA24N-*tig* (Kitagawa *et al.*, 2005) was cleaved with SfiI and SmaI, yielding fragments of 5232, 717 and 589 bp. The 717-bp fragment was purified from an agarose gel using the Qiaquick Gel Extraction Kit. Activity assays (10 μl) comprised 150 ng of this DNA substrate (~30 nM ligatable ends) with ligase (2 μM), in 1× New England Biolabs T4 DNA Ligase Reaction Buffer. Reactions were incubated at 16°C for 20 min, then heat-inactivated at 65°C for 15 min, purified with Qiagen MinElute columns, and run on an ethidium bromide-stained agarose gel. Activity was measured as the appearance of a 1434-bp ligated product, and the disappearance of the 717-bp substrate. Assays for activity of the ligase-cTF linker variants were in volumes of 15 μl, with 190 ng of the blunt-ended substrate (~30 nM ligatable ends), ligase at 0.3 μM and 1× New England Biolabs T4 DNA Ligase Reaction Buffer. Reactions were incubated at 16°C for 20 min, then heat-inactivated (65°C, 20 min) and run on an agarose gel.

Cloning assay

DNA ligase activities were assessed by using blue-white screening in a 'vector + insert' ligation assay. The vector backbone in the assay was pUC18, linearized with restriction enzyme SmaI to generate blunt ends. A suitable blunt-ended insert fragment was generated in two stages. First, the ASKA plasmid pCA24N-*cpdB* (Kitagawa *et al.*, 2005) was amplified in a PCR with the primers pCA24N.for and pCA24N.rev (Supplementary Table SI). Next, the PCR product was digested with MscI to liberate a blunt-ended fragment of the *cpdB* gene (739 bp), which was purified from an agarose gel using the Qiaquick Gel Extraction Kit. Each 20 μl ligation reaction contained 100 ng of the pUC18 vector and 83 ng of the *cpdB* insert (a three-fold molar excess of insert over vector). Each reaction also contained a ligase (1 μM) and was carried out in 1× New England Biolabs T4 DNA Ligase Reaction Buffer. Reactions were incubated at 22°C for 1 h, then heat-inactivated at 65°C for 15 min. Proteinase K was added to 1 mg ml^{-1} and the samples were incubated for a further 20 min at 50°C. Next, the products of each reaction were purified using the EZNA Cycle Pure Kit from Omega Bio-Tek (Norcross, GA, USA). The reaction products were treated with SmaI (20 U, 25°C, 1.5 h) to linearize any pUC18 that had been recircularized without the *cpdB* insert. The SmaI was inactivated by heating at 65°C for 20 min and then aliquots of each reaction were used to transform *E.coli* DH5α-E by electroporation. The transformed cells were allowed to recover in SOC medium for 1 h at 37°C, before aliquots were spread on LB-agar plates containing ampicillin (100 μg ml^{-1}), IPTG (100 μM) and X-gal (40 μg ml^{-1}). Successful ligation of the *cpdB* insert to the pUC18 vector gave rise to white colonies, which were counted after plates had been incubated at 37°C for 16–18 h. Each ligase was tested in triplicate.

NGS library preparation and sequencing

T4 DNA ligase, ligase-cTF and p50-ligase were tested for their ability to ligate bar-coded adaptors onto dA-tailed dsDNA fragments, for subsequent Illumina sequencing. T4 DNA ligase was from the NEBNext DNA Library Prep Master Mix Set for Illumina. The ligase fusion proteins were purified by Enzymatics Inc. (Beverly, MA, USA); yields and specific activities were quantified using their in-house Unit Characterization Assay for cohesive end joining (http://www.enzymatics.com/product_l6030-lc-l.htm). Approximately 3 μg of *E.coli* genomic DNA was sheared using a Covaris S2 sonicator with a 6 × 126 mm MicroTube for four cycles with 10% duty, intensity setting 4, and 200 cycles/burst for 13 s/cycle. End repair and dA tailing of the fragments were carried out using reagents from the NEBNext DNA Library Prep Set 1 kit. After dA tailing, the DNA sample was split into three tubes, for ligation of equimolar ratios of three different adaptor sequences in each tube. Each ligation reaction comprised one-third of the sheared, dA-tailed DNA, 2 μl of unique bar-coded index adaptors (50 μM) and 10 000 U of ligase, in 1× Enzymatics Rapid Ligation Buffer (66 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, 1 mM ATP, 7.5% PEG 6000, pH 7.6). Ligation reactions were carried out at 20°C for 15 min. The ligation reactions were pooled, purified using the Qiagen MinElute PCR Purification Kit, and fragments from 450 to 500 bp were selected using an Invitrogen SizeSelect 2% E-Gel. The size-selected DNA fragments were then amplified for four cycles using the PCR master mix from the NEBNext DNA Library Prep Set 1 kit and primers specific for the adaptor sequences. The products were purified with another Qiagen MinElute column. Final library sizes were verified to be between 450 and 500 bp and free of adaptor contamination using an Agilent Bioanalyzer 2100. The library concentration was determined by qPCR using the Kapa Sybr Fast Universal 2× qPCR Master Mix, before paired-end 100 bp sequencing on an Illumina HiSeq 2000. Fastq files were generated using the Illumina Casava 1.8 suite, which separated the sequence reads based on the index sequences provided.

Results

Design rationale for T4 DNA ligase fusion proteins

We began by designing an Sso7d-ligase fusion protein, analogous to the chimeric DNA polymerases that had been described previously (Wang *et al.*, 2004). Sso7d is a small (7 kDa), monomeric and highly thermostable protein that binds dsDNA in a sequence-independent manner, with a dissociation constant (K_D) of <10 μM (Baumann *et al.*, 1994; Gao *et al.*, 1998). We expected that the Sso7d domain would effectively recruit its ligase fusion partner to dsDNA. However, we also predicted that a high degree of conformational flexibility for the ligase moiety, relative to Sso7d, would be advantageous for effecting ligation. Therefore, we used a glycine-rich sequence (Gly-Thr-Ser-Gly-Gly-Gly-Ser-Gly-Gly-Gly) as the linker between the two fusion partners. This linker was also used in all subsequent fusion proteins, unless noted otherwise.

Sso7d has an unfolding temperature above 90°C and it is highly resistant to chemical denaturation (Gao *et al.*, 1998). We anticipated that this may be problematic, because most ligation protocols involve thermal inactivation and/or chemical denaturation of the ligase, to dissociate it from its ligated product and to prevent it from inhibiting downstream steps

such as bacterial transformation. Therefore, we sought alternative fusion partners for T4 DNA ligase. The p50 fragment of the human transcription factor NF-κB associates strongly with dsDNA *in vitro* and it can tolerate the presence of fusion partners at either terminus (Speight *et al.*, 2001; Patrick and Blackburn, 2005). A detailed biophysical analysis (de Lumley *et al.*, 2004) showed that p50 binds a 10-bp palindromic consensus sequence as a dimer, with extremely high affinity ($K_D = 8$ pM), but that it also binds non-specific dsDNA strongly ($K_D = 5.7$ nM). The same study showed that a monomeric murine homolog of p50, nuclear factor of activated T-cells (NFAT), bound dsDNA less tightly and with less sequence specificity ($K_D = 40$ nM and $K_D = 11$ nM, for non-specific DNA and the 5 bp NF-κB half-site, respectively). The authors also constructed a chimera comprising the DNA-binding domain of NFAT and the dimerization domain of p50 (de Lumley *et al.*, 2004). They termed this chimera NFAT-Ala-p50, but for brevity we abbreviate it 'cTF' (for 'chimeric transcription factor'). cTF bound DNA exclusively as a monomer, and showed almost no preference for the p50 palindome ($K_D = 28$ nM) over non-specific DNA ($K_D = 40$ nM). The various DNA binding modes, affinities and specificities of p50, NFAT and cTF made it likely that they would alter the performance of T4 DNA ligase, albeit in ways that were difficult to predict *a priori*. Hence, we constructed the six corresponding fusion proteins (p50-ligase, ligase-p50, NFAT-ligase, ligase-NFAT, cTF-ligase and ligase-cTF).

We identified additional fusion partner candidates from pathways of double-strand break repair in bacteria. One such candidate was the PprA protein from the radiation-resistant bacterium, *Deinococcus radiodurans* (Narumi *et al.*, 2004). Loss of PprA sensitizes *D.radiodurans* to DNA damage, and the protein appears to localize to broken DNA ends *in vitro* (Murakami *et al.*, 2006). It has also been reported that free PprA can stimulate the activities of the T4 and *E.coli* DNA ligases in *trans* (Narumi *et al.*, 2004). Similarly, the *Mycobacterium tuberculosis* Ku protein binds preferentially to dsDNA ends. Ku recruits the ATP-dependent *M.tuberculosis* DNA ligase and stimulates double-strand break repair (Weller *et al.*, 2002). We constructed the PprA-ligase, ligase-PprA, Ku-ligase and ligase-Ku fusion proteins.

Finally, a large number of sequence non-specific DNA-binding proteins contain helix-hairpin-helix (HhH) motifs, most often in the context of larger $(HhH)_2$ domains (Shao and Grishin, 2000). Recently, it was shown that the activity of the ϕ29 DNA polymerase could be improved by fusing it to two $(HhH)_2$ domains from the *Methanopyrus kandleri* topoisomerase V enzyme (de Vega *et al.*, 2010). Therefore, we constructed the $[(HhH)_2]_2$-ligase and ligase-$[(HhH)_2]_2$ fusion proteins.

In total, we expressed and purified 14 His_6-tagged proteins, including T4 DNA ligase itself, for our initial activity assays. IMAC was used to recover each protein from the soluble cell lysate after IPTG-induced over-expression. Proteins were purified to >95% homogeneity (as judged by SDS-PAGE) and yields were typically 10–20 mg per litre of bacterial culture. The exceptions were ligase-NFAT, Ku-ligase, ligase-Ku and cTF-ligase, the yields of which were 1–5 mg l^{-1} of culture. All four of these proteins were prone to aggregate during storage at 4°C, with precipitation being most rapid (<1 h) in the case of cTF-ligase. Consequently, these proteins were stored at −80°C, and thawed immediately prior to their use.

Activity screens identify improved DNA ligases

Assay design. We designed and implemented agarose gel-based assays for cohesive- and blunt-ended fragment joining, in which two substrate molecules could be ligated, but further concatemerization was not possible. This simplified downstream data analysis, because there were single bands corresponding to the substrate and the ligated product in each assay. For cohesive-ended fragment joining, we used PCR to amplify *ompC* from the ASKA library of open reading frames (Kitagawa *et al.*, 2005). This open reading frame was chosen solely because it contained an SpeI restriction site, positioned such that the PCR product could be cut into two almost identically-sized fragments (638 and 639 bp) that co-migrated on agarose gels. Because the PCR primers were not phosphorylated, ligase-catalyzed re-joining of the PCR product could only occur at the cohesive ends that had been generated by SpeI digestion. The substrate for blunt-ended fragment joining assays was a SmaI/SfiI restriction fragment. SfiI cleaves a non-palindromic sequence to leave a 3-base overhang, which is not a substrate for cohesive end joining. Hence, ligation of this substrate could only occur at the blunt ends generated by SmaI digestion. The conditions in each assay were designed to approximate those used in most molecular biology protocols (Lohman *et al.*, 2011a); that is, the concentration of the dsDNA substrate was comparatively low (15 ng μl^{-1}) and the concentration of the DNA ligase was comparatively high (2 μM).

Cohesive-ended fragment ligation. The results of the screen for cohesive-ended ligation activity are shown in Fig. 1A. Under the conditions of the assay, T4 DNA ligase was effective at ligating most of the substrate (lanes 14 and 19 of Fig. 1A). The ligase fusion proteins displayed a range of activities, from almost undetectably low (Ku-ligase and ligase-Ku), to catalyzing the near-complete conversion of substrate to product (Sso7d-ligase, ligase-cTF and p50-ligase). The poor performance of the Ku fusion proteins is consistent with the observation that Ku inhibits T4 DNA ligase in *trans* (Weller *et al.*, 2002), presumably by hindering the access of the ligase to the DNA ends that are to be joined. Unsurprisingly, the other aggregation-prone ligases (cTF-ligase and ligase-NFAT) also showed weak activity. On the other hand, the reciprocal fusion proteins (ligase-cTF and NFAT-ligase) both possessed greater activity than T4 DNA ligase.

Blunt-ended fragment ligation. As with the cohesive-ended fragments, the ligases showed marked variation in their abilities to join blunt-ended DNA (Fig. 1B). On average, T4 DNA ligase only ligated a minority of the substrate molecules over the course of the 20 min assay. Eight of the 13 engineered ligases, including at least one of the fusions to each DNA-binding protein apart from Ku, out-performed T4 DNA ligase in this screen. The best were Sso7d-ligase and p50-ligase (lanes 2 and 5 in Fig. 1B), each of which converted almost all of the substrate to product. Notably, the optimal orientation for each fusion varied. For example, p50-ligase was more active than ligase-p50 (compare lanes 5 and 6 in Fig. 1B), while ligase-PprA outperformed PprA-ligase (lanes 10 and 11). The implications are that T4 DNA ligase is tolerant of fusions at either of its termini, and that the optimal orientation of the fusion partner, relative to the ligase, needs to be determined empirically.

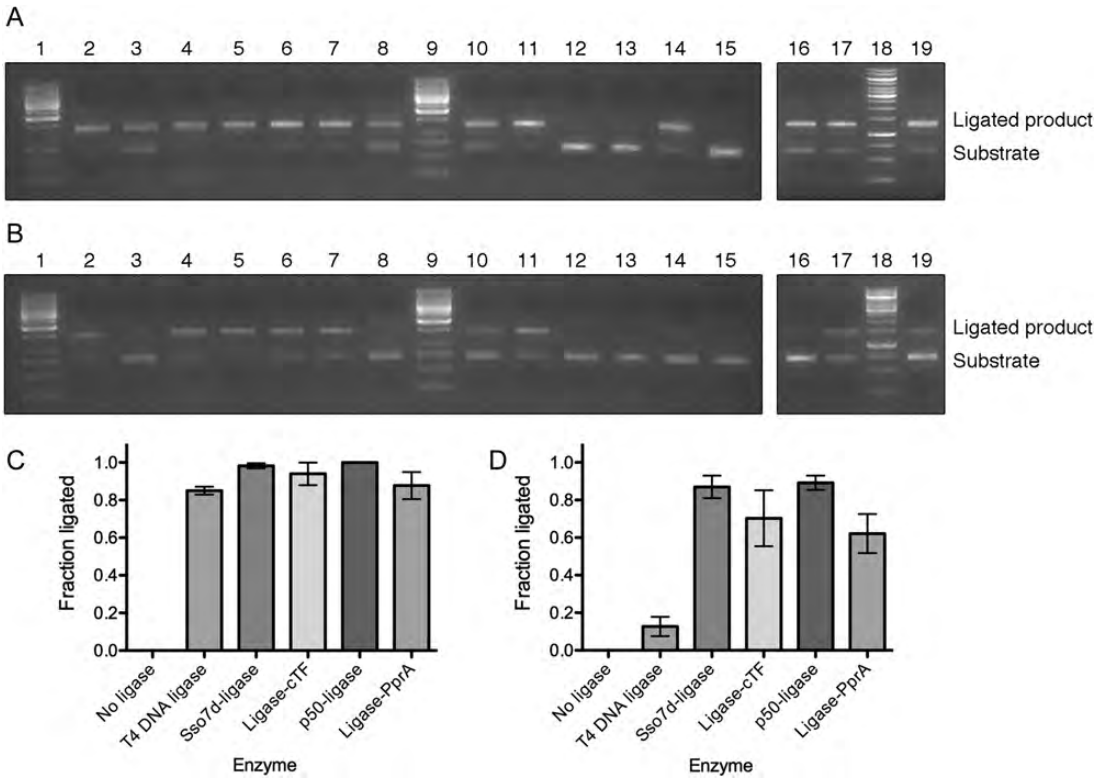

Fig. 1. Gel-based screens for ligase activity. (**A**) Ligation of cohesive-ended PCR products (638/639 bp; lower band) to yield a product of 1277 bp (upper band). Lanes are numbered and loaded as follows: 1, molecular weight marker; 2, ligation catalyzed by Sso7d-ligase; 3, cTF-ligase; 4, ligase-cTF; 5, p50-ligase; 6, ligase-p50; 7, NFAT-ligase; 8, ligase-NFAT; 9, molecular weight marker; 10, PprA-ligase; 11, ligase-PprA; 12, Ku-ligase; 13, ligase-Ku; 14, T4 DNA ligase; 15, no ligase control; 16, $[(HhH)_2]_2$-ligase; 17, ligase-$[(HhH)_2]_2$; 18, molecular weight marker; 19, T4 DNA ligase. (**B**) Ligase-catalyzed joining of a blunt-ended 717-bp restriction fragment (lower band) to give a 1434-bp product (upper band). Lanes are numbered and loaded as described in (A). (**C**) Fraction of cohesive-ended substrate that was ligated to product, for selected ligases from (A) and (B). Data are the means ($\pm$SEM) of two independent experiments, conducted under identical conditions to the assay in (A). (**D**) Fraction of blunt-ended substrate that was ligated to product, for selected ligases from (A) and (B). Data are the means ($\pm$SEM) of two independent experiments, conducted under identical conditions to the assay in (B).

Densitometry. The agarose gel-based screens (Fig. 1A and B) suggested that the most active ligase variants were Sso7d-ligase, ligase-cTF, p50-ligase and ligase-PprA. Fresh batches of these enzymes were retested under identical conditions, and densitometry was used to quantify their improvements over T4 DNA ligase. Improvements in cohesive end joining were small, because T4 DNA ligase was active enough to ligate ~85% of the substrate in the assay (Fig. 1C). In contrast, T4 DNA ligase only converted ~13% of the blunt-ended substrate to product (Fig. 1D). Sso7d-ligase and p50-ligase reproducibly ligated ~90% of this substrate, corresponding to improvements of almost 7-fold in blunt-ended ligation. These improvements over T4 DNA ligase were highly significant in two-tailed unpaired *t*-tests ($P = 0.01$ for Sso7d-ligase; $P < 0.01$ for p50-ligase). On average, ligase-cTF and ligase-PprA each ligated ~5-fold more of the blunt-ended substrate than T4 DNA ligase; however, the results for these fusion proteins were more variable. This variability was reflected in the decreased statistical support for their improvements over T4 DNA ligase ($P = 0.07$ for ligase-cTF; $P = 0.05$ for ligase-PprA).

Blunt end cloning assay

In most laboratories, the primary use for T4 DNA ligase remains in the construction of recombinant plasmids. We used T4 DNA ligase and the four best fusion proteins to ligate an arbitrary, blunt-ended dsDNA fragment into a standard cloning vector (pUC18) that had been linearized by digestion with SmaI. Blue-white screening enabled the enumeration of successful cloning events (Fig. 2). On average in independent triplicates, ligase-cTF gave the greatest number of white colonies (mean = 1234 colonies). This was an improvement of ~160% over T4 DNA ligase, which gave an average of 748 colonies per replicate assay. All three of the ligase-cTF colony counts were higher than any of the three T4 DNA ligase counts, although the assay-to-assay variation in the T4 DNA ligase data meant that the difference between the means was only moderately significant in a two-tailed unpaired *t* test

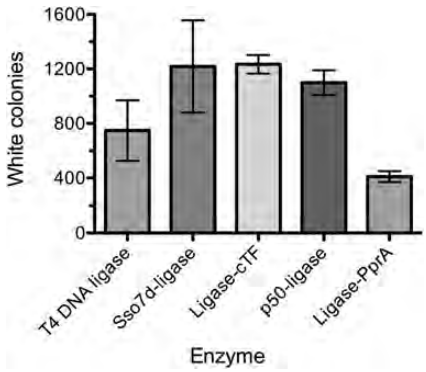

Fig. 2. Cloning assay. A 739-bp, blunt-ended insert was cloned into pUC18 and the ligation products were used to transform *E.coli* DH5α-E. Colony counts are the means ($\pm$SEM) of independent triplicates.

($P = 0.10$). Sso7d-ligase also yielded highly variable colony counts, resulting in no statistically significant improvement over T4 DNA ligase ($P = 0.31$). This variability is likely to have arisen due to the extreme stability of Sso7d, which makes it difficult to remove from the ligated reaction products. During assay development, Sso7d-ligase appeared to perform 3- to 10-fold worse than T4 DNA ligase. Incorporating a Proteinase K step into the reaction clean-up, in an effort to disrupt the protein-DNA interaction, recovered the colony counts to the levels reported here. Even with the Proteinase K treatment, however, cloning with ligase-PprA gave an average of 55% as many colonies as T4 DNA ligase, though this difference also lacked statistical support ($P = 0.21$).

Engineered ligases in NGS library preparation

There is a growing demand for DNA ligases in NGS library construction. For Illumina sequencing, dsDNA samples are prepared by random fragmentation, end repair and then the addition of a single nucleotide dA overhang at the 3′ end of each strand. A DNA ligase is required that can effectively join these substrates with adaptors that carry single nucleotide 3′ dT overhangs. We compared the activities of T4 DNA ligase, ligase-cTF and p50-ligase in this scenario. Each of the ligases was used to ligate three different sets of bar-coded adapters to randomly fragmented *E.coli* genomic DNA. The samples were pooled after the ligation step, and then sequenced, so that the number of reads with each bar code represented the relative efficiency of the DNA ligase that was used with that adaptor. The Illumina sequencing run gave a total of 3.10×10^7 bar-coded reads. The most efficient ligase was p50-ligase, which contributed 49.1% of the total reads to this pool (mean ± SEM for the three p50-ligase bar codes was $(5.1 \pm 0.4) \times 10^6$ reads). This was an improvement of ~160% over T4 DNA ligase (Fig. 3), and this difference was highly significant in a two-tailed unpaired *t*-test ($P = 0.01$). In contrast, ligase-cTF was comparatively inefficient in this test (Fig. 3).

Scope for further engineering

Linker variants. All of the engineered proteins that were described above contained a glycine-rich linker (GTSGGG SGGG) between the ligase and its fusion partner. In a final design iteration, we replaced this sequence in ligase-cTF with four alternate linkers. The first, which we termed Natural Linker Long (NLL), was the inter-domain linker from *Cellulomonas fimi* endoglucanase A. Its amino acid sequence was GPTPTP TPTTPTPTPTPTPTPTPTG and it has been proposed to adopt an extended, kinked and rigid conformation (Shen *et al.*, 1991). We also introduced a shorter variant of NLL, Natural Linker Short (NLS), with the sequence GTPTPTPTPTGEF (Gustavsson *et al.*, 2001). In addition, we tested two artificial linkers, designed previously (Arai *et al.*, 2001): Flexible Linker (FL; sequence LSGGGGSGGGGSGGGGSGGGGSAAA); and Helical Linker (HL; sequence GPLAEAAAKEAAAKEAAAKEAAAKAA APG). Ligase-cTF and the four linker variants were tested for their blunt end joining activity, using the agarose gel-based assay and each ligase at a concentration of 0.3 μM. Ligase-NLS-cTF was the most efficient ligase under these conditions, with ~2-fold greater activity than ligase-cTF (Table II).

Fusion of p50 to the E.coli *DNA ligase.* Finally, we assessed whether a DNA-binding fusion partner could enhance the activity of an NAD^+-dependent DNA ligase. We fused p50 to the *E.coli* DNA ligase, LigA. LigA is known to repair single-stranded nicks and to join cohesive ends (Lohman *et al.*, 2011a). The purified LigA protein, together with p50-LigA and LigA-p50, were tested in the gel-based assay for cohesive end joining. As expected, LigA was reasonably efficient at ligating the cohesive-ended substrate (Table II). However the two fusion proteins, and particularly p50-LigA, were even more active, suggesting that our engineering approach will be generalizable to other ligases.

Discussion

It has been shown that fusing dsDNA-binding proteins to DNA polymerases can improve their performances *in vitro* (Wang *et al.*, 2004; Sun *et al.*, 2006; de Vega *et al.*, 2010). In this work, we demonstrated that the activities of DNA ligases can be enhanced by taking a similar approach. Our primary goal was to improve the activity of T4 DNA ligase, in scenarios that are relevant for molecular biologists. We hypothesized that the unfavourable kinetic parameters of T4 DNA ligase could be overcome by artificially increasing its overall affinity for dsDNA. At the outset, however, it was unclear how DNA binding affinity would trade off with other factors that might affect ligation efficiency, such as the sequence specificity of the protein-DNA interaction, the potential for the fusion partner to block access of the ligase to its substrates (i.e. to the DNA ends), and whether overly stable protein-

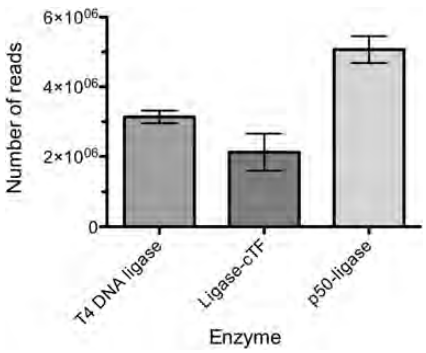

Fig. 3. Ligase efficiency at adaptor ligation for subsequent Illumina sequencing. Each enzyme was used to ligate three bar-coded adaptors onto dA-tailed dsDNA fragments. For each ligase, the mean number of reads per barcode (± SEM) is plotted.

Table II. Ligation efficiencies of second-generation DNA ligases

Ligase	Substrate	Fraction ligated[a]
Ligase-cTF linker variants		
Ligase-cTF	Blunt ended	0.27
Ligase-NLS-cTF	Blunt ended	0.61
Ligase-NLL-cTF	Blunt ended	0.12
Ligase-FL-cTF	Blunt ended	0.23
Ligase-HL-cTF	Blunt ended	0.32
E.coli LigA fusion proteins		
LigA	Cohesive ended	0.81
p50-LigA	Cohesive ended	0.99
LigA-p50	Cohesive ended	0.95

[a]Estimated from agarose gel-based assays and densitometry. The standard errors from two experimental replicates were all less than ± 20% of the reported mean values.

DNA interactions might inhibit downstream steps in common protocols. The protein design problem was also rendered more difficult because the three-dimensional structure of T4 DNA ligase has not yet been determined.

With these considerations in mind, we undertook a broad and empirical search for suitable DNA-binding fusion partners. We identified seven candidates, from all three domains of life (Table I). Our initial activity screen (Fig. 1) showed that the majority of these fusion partners could increase the activity of T4 DNA ligase, especially in joining blunt-ended dsDNA fragments. The most promising fusion proteins from the screen, in rank order according to the data in Fig. 1C and D, were p50-ligase, Sso7d-ligase, ligase-cTF and ligase-PprA. The composition of this set emphasized the value of our somewhat heuristic engineering strategy: it contained chimeras with each possible orientation of ligase and DNA-binding protein; and the DNA-binding fusion partners were a eukaryotic transcription factor (p50), an archaeal protein (Sso7d), a bacterial DNA repair protein (PprA), and an artificial protein that was itself a chimera (cTF).

Further testing showed that different ligases were best for different applications. The two best fusion proteins in the gel-based screen (p50-ligase and Sso7d-ligase) were out-performed by ligase-cTF in the blunt end cloning assay (Fig. 2). Moreover, the 5- to 7-fold improvements over T4 DNA ligase in joining blunt-ended fragments (Fig. 1D) were reduced to $<$2-fold improvements in blunt end cloning (Fig. 2). Fragment joining comprises a single intermolecular ligation event, whereas 'vector + insert' cloning requires an intermolecular ligation step, followed by an intramolecular ligation step (to circularize the insert-containing plasmid). Our data suggest that the ligase fusion proteins may be better, proportionately, at the intermolecular step.

While ligase-cTF was the best ligase for cloning blunt-ended fragments, it was less efficient at adaptor ligation for NGS library preparation (Fig. 3). In contrast, p50-ligase significantly out-performed T4 DNA ligase at adaptor ligation, by $\sim$160%. Overall, these two ligases with eukaryotic DNA-binding partners emerged as the most promising for further development, with additional evidence of their utility coming from the increased activity of a linker variant (ligase-NLS-cTF), and the ability of p50 to enhance the activity of the *E.coli* DNA ligase (Table II). On the other hand, Sso7d-ligase was difficult to dissociate from its ligated product, and this decreased its usefulness. Similarly, none of the chimeras that contained the bacterial DNA repair proteins (PprA and Ku) systematically out-performed T4 DNA ligase, although ligase-PprA showed some promise (Fig. 1). The positioning of PprA and Ku at the ends of DNA duplexes (Weller *et al.*, 2002; Murakami *et al.*, 2006) appears to provide a steric block that interferes with the ability of the fused ligase to catalyze end joining.

We have demonstrated the utility of our ligases in common molecular biology protocols. However, more detailed kinetic characterization of the end-joining reactions that they catalyze will be required to understand the underlying mechanism(s) of their improved performance. To date, an analysis of this kind has only been reported for the nick-sealing activity of T4 DNA ligase (Lohman *et al.*, 2011b), and not the end-joining activities. The authors observed burst phase kinetics for the nick-sealing reaction, consistent with product release or a post-ligation conformational change being the rate limiting step. It is unlikely that the same post-ligation step(s) will be rate limiting for the (much slower) end-joining reactions. Nevertheless, assessing the mechanistic impacts of the fusion partners will be valuable for informing future designs.

There is increasing interest in the use and optimization of application-specific DNA polymerases, particularly for efficient and unbiased amplification of NGS libraries (Aird *et al.*, 2011; Quail *et al.*, 2012). Novel DNA polymerases have been engineered for a number of specific applications (Kranaster and Marx, 2010) and many companies now offer extensive portfolios of highly specialized polymerase products. Here, we have shown that our protein design strategy is a facile one for engineering bespoke, application-specific DNA ligases.

Supplementary data

Supplementary data are available at *PEDS* online.

Acknowledgements

We thank Erik Quandt (University of Texas, Austin) for initiating the collaboration between ADE and WMP. We also thank Stephen Picone and his staff at Enzymatics Inc. for purifying the ligases that were used in the NGS experiment.

Conflict of interest: Patent applications related to this work have been filed, with authors R.H.W., S.K.M. and W.M.P. as inventors.

Funding

This work was supported by grants from the Massey University Research Fund and the New Zealand Ministry of Business, Innovation and Employment's Pre-Seed Accelerator Fund.

References

Aird,D., Ross,M.G., Chen,W.S., Danielsson,M., Fennell,T., Russ,C., Jaffe,D.B., Nusbaum,C. and Gnirke,A. (2011) *Genome Biol.*, **12**, R18.
Arai,R., Ueda,H., Kitayama,A., Kamiya,N. and Nagamune,T. (2001) *Protein Eng.*, **14**, 529–532.
Bang,D. and Church,G.M. (2008) *Nat. Methods*, **5**, 37–39.
Barany,F. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 189–193.
Baumann,H., Knapp,S., Lundback,T., Ladenstein,R. and Hard,T. (1994) *Nat. Struct. Biol.*, **1**, 808–819.
de Lumley,M., Hart,D.J., Cooper,M.A., Symeonides,S. and Blackburn,J.M. (2004) *J. Mol. Biol.*, **339**, 1059–1075.
de Vega,M., Lazaro,J.M., Mencia,M., Blanco,L. and Salas,M. (2010) *Proc. Natl Acad. Sci. USA*, **107**, 16506–16511.
Gao,Y.G., Su,S.Y., Robinson,H., Padmanabhan,S., Lim,L., McCrary,B.S., Edmondson,S.P., Shriver,J.W. and Wang,A.H. (1998) *Nat. Struct. Biol.*, **5**, 782–786.
Gibson,D.G., Young,L., Chuang,R.Y., Venter,J.C., Hutchison,C.A., 3rd. and Smith,H.O. (2009) *Nat. Methods*, **6**, 343–345.
Gustavsson,M., Lehtio,J., Denman,S., Teeri,T.T., Hult,K. and Martinelle,M. (2001) *Protein Eng.*, **14**, 711–715.
Kitagawa,M., Ara,T., Arifuzzaman,M., Ioka-Nakamichi,T., Inamoto,E., Toyonaga,H. and Mori,H. (2005) *DNA Res.*, **12**, 291–299.
Kranaster,R. and Marx,A. (2010) *ChemBioChem*, **11**, 2077–2084.
Lohman,G.J., Tabor,S. and Nichols,N.M. (2011a) *Curr. Protoc. Mol. Biol.*, **94**, 3.14.1–3.14.7.
Lohman,G.J.S., Chen,L.X. and Evans,T.C. (2011b) *J. Biol. Chem.*, **286**, 44187–44196.
Miller,E.S., Kutter,E., Mosig,G., Arisaka,F., Kunisawa,T. and Ruger,W. (2003) *Microbiol. Mol. Biol. Rev.*, **67**, 86–156.
Murakami,M., Narumi,I., Satoh,K., Furukawa,A. and Hayata,I. (2006) *Biochim. Biophys. Acta*, **1764**, 20–23.
Narumi,I., Satoh,K., Cui,S., Funayama,T., Kitayama,S. and Watanabe,H. (2004) *Mol. Microbiol.*, **54**, 278–285.
Pascal,J.M. (2008) *Curr. Opin. Struct. Biol.*, **18**, 96–105.

Patrick,W.M. and Blackburn,J.M. (2005) *FEBS J.*, **272**, 3684–3697.
Quail,M.A., Otto,T.D., Gu,Y., Harris,S.R., Skelly,T.F., McQuillan,J.A., Swerdlow,H.P. and Oyola,S.O. (2012) *Nat. Methods*, **9**, 10–11.
Ren,Z.J., Baumann,R.G. and Black,L.W. (1997) *Gene*, **195**, 303–311.
Sgaramella,V., Van de Sande,J.H. and Khorana,H.G. (1970) *Proc. Natl Acad. Sci. USA*, **67**, 1468–1475.
Shao,X. and Grishin,N.V. (2000) *Nucleic Acids Res.*, **28**, 2643–2650.
Shen,H., Schmuck,M., Pilz,I., Gilkes,N.R., Kilburn,D.G., Miller,R.C., Jr. and Warren,R.A. (1991) *J. Biol. Chem.*, **266**, 11335–11340.
Shuman,S. (2009) *J. Biol. Chem.*, **284**, 17365–17369.
Speight,R.E., Hart,D.J., Sutherland,J.D. and Blackburn,J.M. (2001) *Chem. Biol.*, **8**, 951–965.
Sugino,A., Goodman,H.M., Heyneker,H.L., Shine,J., Boyer,H.W. and Cozzarelli,N.R. (1977) *J. Biol. Chem.*, **252**, 3987–3994.
Sun,S., Geng,L. and Shamoo,Y. (2006) *Proteins*, **65**, 231–238.
Tomkinson,A.E., Vijayakumar,S., Pascal,J.M. and Ellenberger,T. (2006) *Chem. Rev.*, **106**, 687–699.
Wang,Y., Prosen,D.E., Mei,L., Sullivan,J.C., Finney,M. and Vander Horn,P.B. (2004) *Nucleic Acids Res.*, **32**, 1197–1207.
Weiss,B. and Richardson,C.C. (1967) *Proc. Natl Acad. Sci. USA*, **57**, 1021–1028.
Weiss,B., Jacquemin-Sablon,A., Live,T.R., Fareed,G.C. and Richardson,C.C. (1968) *J. Biol. Chem.*, **243**, 4543–4555.
Weller,G.R., Kysela,B., Roy,R., *et al*. (2002) *Science*, **297**, 1686–1689.