Engineering proteins that bind, move, make and break DNA

Cynthia H Collins*, Yohei Yokobayashi[†], Daisuke Umeno[†] and Frances H Arnold[†‡]

Recent protein engineering efforts have generated artificial transcription factors that bind new target DNA sequences and enzymes that modify DNA at new target sites. Zinc-finger-based transcription factors are favored targets for design; important technological advances in their construction and numerous biotechnological applications have been reported. Other notable advances include the generation of endonucleases and recombinases with altered specificities, made by innovative combinatorial and evolutionary protein engineering strategies. An unexpectedly high tolerance to mutation in the active sites of DNA polymerases is being exploited to engineer polymerases to incorporate artificial nucleotides or to display other, nonnatural activities.

Addresses
*Biochemistry and Molecular Biophysics, California Institute of Technology, mail code 210-41, Pasadena, California 91125, USA
[†]Division of Chemistry and Chemical Engineering, California Institute of Technology, mail code 210-41, Pasadena, California 91125, USA
[‡]e-mail: frances@cheme.caltech.edu

Current Opinion in Biotechnology 2003, **14**:371–378

This review comes from a themed issue on
Protein technologies and commercial enzymes
Edited by Gjalt Huisman and Stephen Sligar

0958-1669/$ – see front matter
© 2003 Elsevier Science Ltd. All rights reserved.

DOI 10.1016/S0958-1669(03)00091-0

Abbreviations

bp	base pair
CSR	compartmentalized self-replication
dNTP	deoxyribonucleotide triphosphate
PCR	polymerase chain reaction
pol I	polymerase I
rNTP	ribonucleotide triphosphate
ZFP	zinc-finger protein

Introduction

Proteins involved in DNA recognition, manipulation and synthesis could provide a multitude of tools for studying gene function, genetic engineering, molecular biology and gene therapy. Engineered versions of these proteins would have the potential to modulate the expression of any gene of interest or to rearrange chromosomal DNA at any site within a genome. The utility of proteins that interact with DNA has been limited in many cases by their target sequence specificities. Thus, significant efforts have been made to engineer them to target alternate DNA sequences. DNA polymerases with altered fidelity and the ability to incorporate modified bases would enable new technologies for gene amplification, mutagenesis, and specific labeling. By engineering proteins that bind and modify DNA, we can also gain insights into the molecular mechanisms of maintenance, control and modification of genetic information. This review covers recent efforts to change the binding and catalytic specificities of transcription factors, endonucleases, recombinases, and DNA polymerases.

α-Helical DNA-binding proteins

The first report of engineering protein–DNA interactions involved modification of the helix–turn–helix DNA-binding domain of the 434 repressor so that it recognized the P22 operator [1]. Replacing the amino acids along the face of the recognition helix that make contacts with DNA with those from the corresponding positions in the P22 homolog created an engineered 434 repressor that bound the P22 operator with affinity similar to that of the wild-type P22 repressor for its cognate operator sequence. The engineered repressor no longer bound the 434 operator. Furthermore, heterodimers formed upon coexpression of the wild-type and modified 434 repressors could bind hybrid operators of 434 and P22 operator half-sites [2]. Single-chain proteins that are covalent dimers of the DNA-binding domains from different transcription factors have also been shown to bind to hybrid operators [3–9]. More recently, Liang *et al.* [10] constructed several single-chain heterodimers using engineered 434 repressor DNA-binding domains (identified in previous rational design and selection experiments), which also bound operators made up of half-sites bound by the individual domains. To achieve high DNA-binding affinities ($K_d = 10^{-11}$–10^{-10} M), the single-chain proteins developed by Liang and colleagues require that the six base pair (bp) sequence between the two half-sites contains mostly A and T nucleotides. Apart from this study, there is little recent published work on engineering helix–turn–helix DNA-binding proteins. Thus, we cannot draw any conclusions about the ease of changing their binding specificities.

Zinc-finger proteins

One of the most abundant protein motifs in eukaryotes, the zinc-finger domain, has been the favored scaffold for engineering novel DNA-binding proteins [11,12]. A Cys_2–His_2 zinc finger consists of ~30 amino acid residues in a ββα fold stabilized by the coordination of two cysteine and two histidine residues to a zinc ion. Although each zinc finger typically recognizes only three DNA bases, multiple fingers can be linked in tandem so that the resulting multifinger protein can recognize longer

sequences. This modularity is attractive because it opens the possibility of generating DNA-binding proteins of arbitrary sequence specificities by fusing pre-made fingers that each recognize any one of the 64 possible DNA triplets. Although this simplest vision of combinatorial zinc-finger protein (ZFP) design has only been partially realized, significant advances in design strategies have enabled the construction of artificial polydactyl ZFPs with diverse sequence specificities.

Phage display has been by far the most widely used technique to design (or discover) ZFPs with novel sequence specificities. Typically, a set of amino acids that contact the DNA in one finger of the multifinger protein is randomly mutated, and the mutant fingers are selected for binding to a desired DNA triplet, often with negative selection against binding to other triplets. This approach identified a set of zinc-finger domains that can recognize the 16 possible combinations of 5′-GNN-3′ [13]. Extending this approach to zinc fingers that recognize the remaining 48 triplets, however, has been complicated by cross-subsite interactions in which DNA bases are contacted by amino acid residues from neighboring fingers. Dreier *et al.* [14] extended the repertoire of custom, interchangeable fingers to recognize 5′-ANN-3′ triplets by eliminating a cross-subsite contact from the finger fused to the C terminus. Other recent refinements of this approach [15,16] further facilitate the ZFP design process.

An alternative strategy for selecting artificial ZFPs using phage display was recently described by Isalan and colleagues [17••] who constructed two libraries of three-finger proteins based on transcription factor Zif268. Each library contained half the wild-type Zif268 sequence (one and a half fingers); the remaining half harbored randomized amino acids that contact five of the nine bases in the recognition sequence. The libraries were selected in parallel for binding to double-stranded DNA the sequence of which contained four bases recognized by unmodified Zif268 and five bases of the target sequence. The selected half-libraries were recombined *in vitro* and further selected for binding to the full target DNA sequence. This approach allows the selection of high-affinity three-finger domains optimized for cross-subsite interactions. It also allows high-throughput selection of multiple ZFPs, because the half-libraries used for the initial selection can be used universally. The authors report that the entire selection process takes approximately two weeks and is amenable to automation.

Although overshadowed by the intense activity in phage display based selection of ZFPs, notable progress using a rational design strategy was recently reported by Sera and Uranga [18]. A nondegenerate recognition code table that assigns specific amino acids at positions −1, 2, 3, and 6 (relative to the start of the recognition helix) of a zinc finger to arbitrary 4 bp sequences was devised. The antisense base of the fourth base pair is contacted by the amino acid in position 2 of the first finger, whereas the sense base (which is also the first base of the second, overlapping 4 bp unit) is contacted by position 6 of the second finger. Proposing that one can design artificial ZFPs for arbitrary target sequences using the universal table, Sera and Uranga tested ten three-finger proteins targeted to different 10 bp sequences. Five exhibited nanomolar affinities towards the desired sequences, and the functional constructs were reported to discriminate single base-pair changes. The results suggest a preference for GC-rich sequences. With further refinement, this approach may complement the existing combinatorial design strategies, which still require substantial labor to construct ZFPs with novel sequence specificities.

With these advances in design and the rapid accumulation of knowledge regarding their DNA-binding properties, custom-designed ZFPs are now finding applications as artificial transcription regulators. Genes of interest can be activated or repressed in cells transfected with ZFPs fused to appropriate effector domains. Recent progress includes controlling gene expression in plants [19–22], inhibiting virus replication by targeting critical regulatory processes [23,24], and activating a gene involved in angiogenesis in a mouse model [25]. Blancafort and colleagues [26••] recently described the large-scale screening of cells transfected with ZFP transcription activator libraries for various phenotypic markers, demonstrating a promising new tool for functional genomics. These efforts have elucidated some important criteria for successful *in vivo* applications of ZFP-based transcription factors. It is critically important, for example, that the target DNA sequence be within the chromatin-accessible region. Liu *et al.* [27] clearly demonstrated this point in their work which identified accessible regions of the genomic DNA (which may differ among cell types) using a DNase I hypersensitivity assay. It also appears that ZFPs with six fingers function better than those with three fingers in most cases, most likely due to stronger binding and slower degradation.

Restriction enzymes

Restriction endonucleases are indispensable in today's molecular biology. Years of screening various microbial sources have yielded hundreds of restriction enzymes that are capable of recognizing specific DNA sequences four to eight bases long and which cleave phosphodiester bonds within or adjacent to the recognition site. Engineered restriction enzymes that recognize altered or expanded sequences are needed for applications in biotechnology and medicine. However, modifying the sequence specificities of restriction enzymes has proved challenging. For example, a recent attempt by Lanio and coworkers [28] to rationally expand the recognition sequence of *Eco*RV based on available structural data

yielded variants with altered selectivity, but not the predicted one. This group had previously shown that directed evolution could yield variants that prefer AT-rich flanking sites over GC-flanked sites [29]. The formidable challenges of rational engineering have led several researchers to turn to combinatorial or evolutionary methods, involving generation of random or directed mutant libraries coupled with screening or selection. However, because most restriction and other DNA-modifying enzymes do not exhibit the modular separation of functions that, for example, the ZFP-based transcription factors do, laboratory evolution of these enzymes must face the serious challenge of retaining catalytic activity while changing substrate sequence specificity. Phage display is therefore not useful, unless it can also select for catalysis [30•].

Samuelson and Xu [31] narrowed the substrate specificity of the promiscuous restriction enzyme *Bst*YI that cleaves four DNA sequences 5′-(A/G)GATC(C/T)-3′ with similar efficiency. *Bgl*II *N*4-cytosine methyltransferase was used to protect 5′-AGATCT-3′ sites in the genome of host bacteria transformed with random mutants of *Bst*YI. Cells containing *Bst*YI mutants that retain activity toward the other three substrate sequences do not survive, due to damage to the genomic DNA. Mutant enzymes in the surviving cells, however, may or may not have retained activity toward 5′-AGATCT-3′ sites, which were protected by methylation. The mutants that survived the initial selection were subsequently tested for their ability to cleave the targeted 5′-AGATCT-3′ sequence using *in vivo* and *in vitro* screens. With further recombination and analysis of some functional mutations, a variant with at least 12-fold greater catalytic efficiency towards the targeted sequence was found. This particular variant, however, lost a significant fraction of its specific activity.

Random mutagenesis coupled with well-designed genetic assays allowed Seligman *et al.* [32] to identify several mutations that altered the target sequence specificity of homing endonuclease I-*Cre*I. Homing endonucleases are encoded in introns or expressed as inteins of certain genes and are involved in the lateral gene transfer of their own genetic elements to the alleles that lack the intervening sequences [33]. The researchers incorporated the 22 bp homing site and mutated analogs into F′ plasmids containing kanamycin resistance and *lac*Z genes. I-*Cre*I mutants that cleaved the wild-type homing sequence or its analogs would lose kanamycin resistance or yield white colonies in media containing 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal). Screening of mostly single amino acid mutants of I-*Cre*I at positions known to make direct contacts with DNA bases identified several with altered or relaxed sequence specificities.

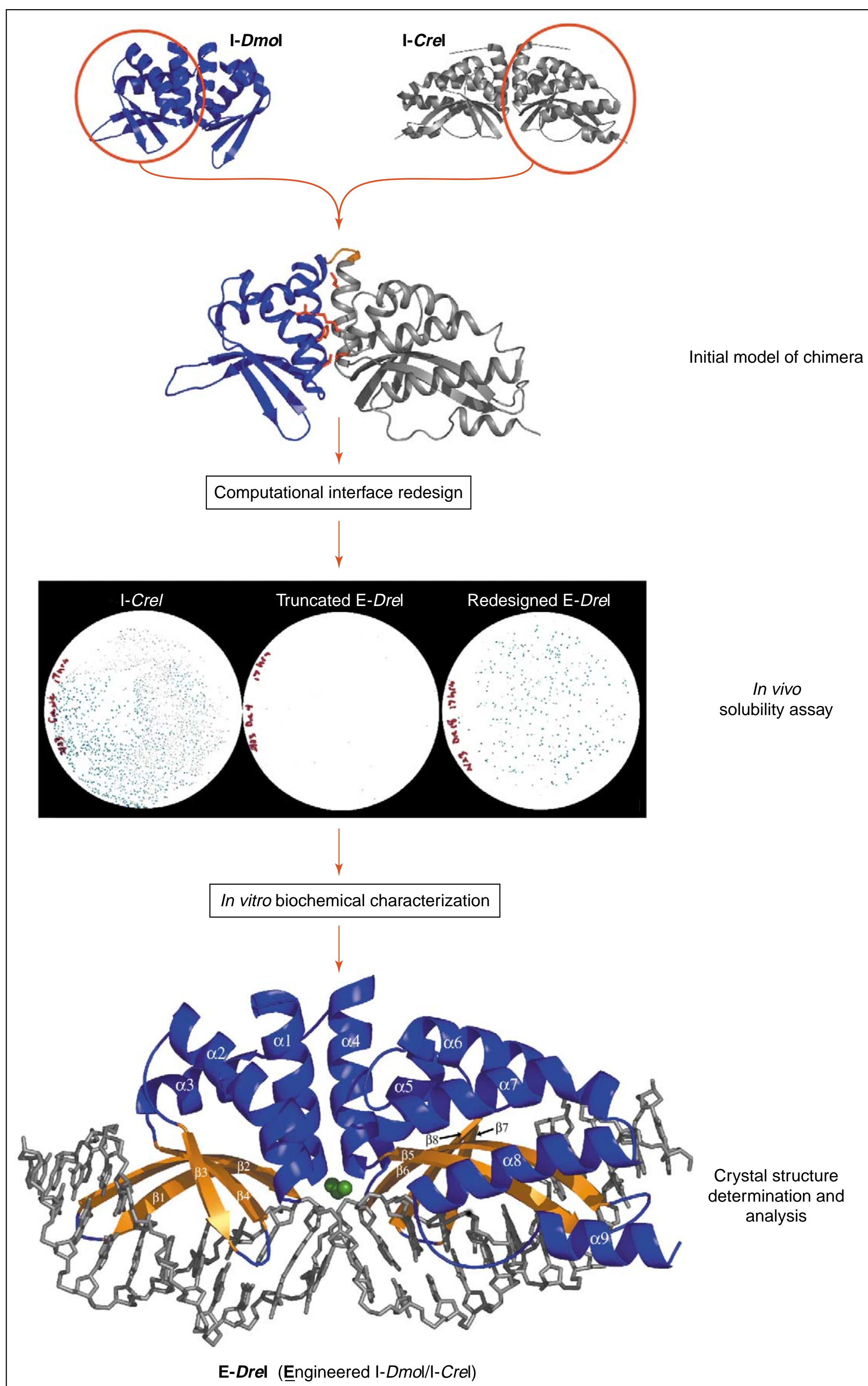
A more drastic modification of restriction sequence specificity was achieved by swapping domains of two homing endonucleases that naturally function as homodimers. Chevalier and colleagues [34••] took advantage of the fact that endonucleases of the LAGLIDADG family share a characteristic dimerization interface mediated by an α-helical pair to construct a hybrid enzyme stabilized by this common interface. An artificial single-chain enzyme that recognizes and cleaves hybrid DNA sequences was made by fusing the N-terminal domain of I-*Dmo*I to a I-*Cre*I monomer via a short peptide linker (Figure 1). The hydrophobic interface of the chimeric protein, E-*Dre*I, was optimized for packing using an automated computational algorithm. Sixteen candidates were constructed and screened for solubility *in vivo*, and three were overexpressed and purified for further biochemical analysis *in vitro*. All three specifically recognized and cleaved 23 bp hybrid DNA sequences derived from the target sequences of the parental enzymes. The single turnover rate (k_{cat}) of one chimera was comparable with that of the wild-type I-*Cre*I, but the dissociation constant (K_d) towards the target DNA was two orders of magnitude higher than that of I-*Cre*I towards its natural substrate sequence.

Recombinases

Recombinases integrate, excise, invert or translocate DNA based on the relative location and orientation of a target DNA site, which is typically palindromic and has a short spacer between half-sites. Used for site-specific recombination of DNA in prokaryotes, yeast, flies and mammals [35–37], their applications are currently limited by their strict target-site specificities. Evolutionary methods have been used with considerable success to engineer recombinases to target alternative DNA sites. Most screening methods require the mutant recombinase to excise a region of DNA that either places a reporter gene downstream of a constitutive promoter or removes a reporter gene that would be expressed in the absence of recombinase activity. Changes in the expression of the reporter gene are used to assess whether or not the mutant recombinase is functional on the desired DNA target site [38••,39•,40•,41].

The usual outcome of such experiments is relaxation of the target-site specificity, unless additional screening identifies those mutants that no longer function at the wild-type recombination site. Santoro and Schultz [40•] devised a method that allowed them to screen for Cre mutants that recombined at an engineered target site or that no longer recombined at the wild-type target site. When they alternated screening for recognition of the engineered site with screening for the inability to recognize the wild-type site in subsequent generations, they identified mutants with shifted DNA sequence specificities. When they did not specifically screen for loss of wild-type function, they always found mutants that were able to recombine at both the engineered and wild-type DNA sites. Towards the same goal of shifting rather than relaxing recombinase specificity, Buchholz and Stewart

Figure 1

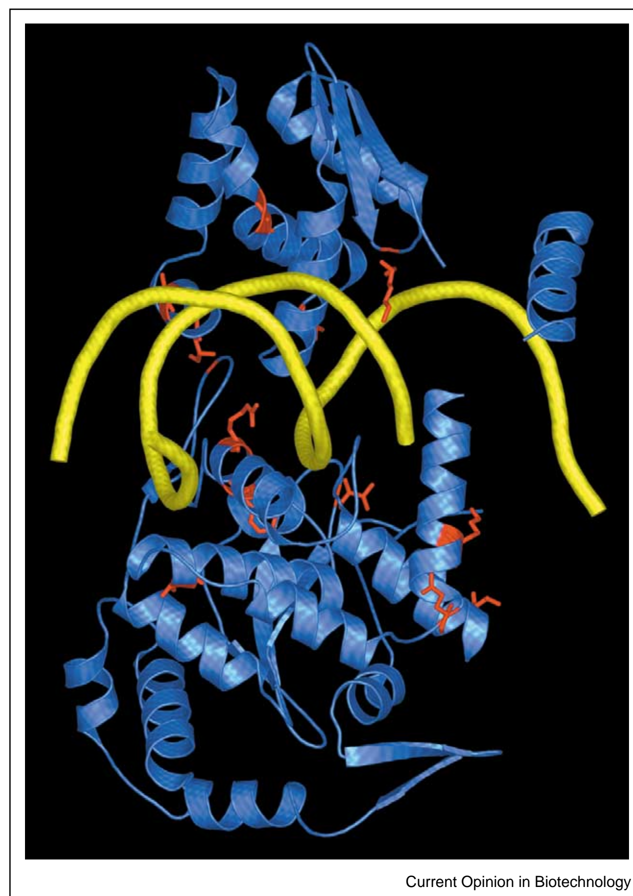


[42] placed the wild-type (loxP) and engineered (loxH) recombination sites on a single plasmid spaced in such a way that only one site could be recombined. Their competition-based approach generated Cre mutants with higher specificity for the engineered sequence. The strategy used in the most recent work on Flp recombinase includes a dual reporter system [38••,39•], which allows both wild-type and mutant-site recombination to be assessed simultaneously. This screen is based on the removal of two reporter genes by the recombinase. One reporter plasmid contains mutant recombination sites flanking the *lacZα* gene, while a second reporter contains the wild-type recombination sites flanking the gene for red fluorescent protein (RFP). The authors identified Flp mutants that were active only at the mutant sites by selecting colonies expressing RFP but not β-galactosidase.

Several groups have used two different stepwise approaches to alter substrate specificity. In one approach, the recognition of the palindromic repeats is changed first followed by that of the directional spacer sequence [42]. In the second approach, mutants are progressively required to act at a target site with more mutations [38••]. Such stepwise strategies are likely to prove necessary as recombinases are engineered to recognize more drastically altered target sequences. Interestingly, Voziyanov *et al.* [38••] found that the amino acid substitutions found in single target site mutants could not be recombined to produce a Flp variant that was active on a target site containing two mutations. Recombination with wild-type Flp and another round of random mutagenesis and screening, however, identified Flp variants able to recombine at the target sequence containing both mutations. One mutant showed a clear preference for this site over the wild-type and single-mutant sites.

Most efforts to change DNA sequence specificity have focused on mutating amino acid residues that make direct contact with the DNA. The identification of several non-contact positions that play key roles in determining the substrate specificity of Cre and Flp, however, indicates that this may not be the best strategy [38••,41]. The mutations shown to alter DNA target specificity of Flp (Figure 2) are clearly not limited to those that make direct contact with the DNA. Screening or selecting random mutagenesis libraries may be a useful addition to future efforts to engineer DNA sequence specificity.

Figure 2



Distribution of mutations that modulate DNA-binding specificity in Flp recombinase [38••]. A monomer of Flp (blue) bound to DNA (yellow) is shown. The amino acid residues identified in Flp variants with shifted DNA-binding specificity are in red. The distribution of these amino acids indicates that residues that do not make direct contact with DNA can also modulate binding specificity. Therefore, when screening for proteins with altered DNA-binding specificity, targeting only those residues that make contact with the DNA may not always be the best search strategy.

DNA polymerases

DNA polymerases are integral to many molecular biology techniques, including sequencing, labeling, modification, amplification, detection, and random mutagenesis of targeted DNA. Potential uses for DNA polymerases also include the synthesis of DNA-based or DNA-like polymeric materials. Needs for higher-performance and

(Figure 1 Legend) Rational design of an artificial, domain-swapped homing endonuclease [34••]. A chimeric homing endonuclease was made by swapping the N-terminal domain of I-*Dmo*I with a subunit of I-*Cre*I. The initial chimera was insoluble. Computational methods were used to identify 16 constructs with redesigned interactions between the two domains. These constructs were screened using an *in vivo* protein folding assay in which the chimeras were covalently linked to the LacZα peptide. Soluble E-*Dre*I/*lacZα* constructs expressed in *E. coli* complemented the lacZω fragment to form blue colonies; expression of insoluble E-*DreI*/*lac*Zα constructs yielded white colonies. Three examples of this assay are shown: soluble I-*Cre*I, an insoluble E-*Dre*I construct with clashing interface residues truncated, and a final E-*Dre*I construct containing a redesigned interface. Biochemical experiments showed that the selected E-*Dre*I construct is both active and highly specific. The structure of E-*Dre*I complexed to its DNA target site was solved to 2.4 Å resolution. (Figure reproduced from [34••] with permission.)

specialized polymerases have driven polymerase engineering efforts focused on altering properties such as processivity, activity, stability, and fidelity. Recent experiments have overwhelmingly used combinatorial or evolutionary approaches. Screens or selection systems for polymerase engineering, like those used for engineering recombinase and endonuclease activities, must be based on enzyme function and not just on substrate binding.

Loeb and colleagues [43••] have conducted systematic mutational analyses of *Escherichia coli* and *Taq* polymerase I (pol I) to elucidate the molecular basis of their replication fidelity and substrate tolerance. Their experiments couple intensive mutagenesis of amino acid residues that contact the incoming deoxynucleotide triphosphates (dNTPs) with functional complementation of a polymerase-mutant *E. coli* [43••] or yeast strain [44]. Sequence analysis of some of the thousands of polymerase variants that are catalytically active has shown that several highly conserved residues are, nonetheless, tolerant to mutation. Hundreds of active mutants have been individually characterized to determine their activity, fidelity, and substrate specificity. Various interesting polymerases, such as error-prone *E. coli* pol I variants [45] and a *Taq* pol I that preferentially incorporates ribonucleotide triphosphates (rNTPs; by approximately 1000-fold) [46] have been identified. This approach, of intensive mutagenesis and genetic selection with subsequent biochemical analysis, was also used to analyze mouse polymerase β and resulted in the discovery of a variant with approximately 25-fold increased catalytic activity [47]. Although the active sites of DNA polymerases are strictly conserved in nature, where the selective pressure is apparently much more stringent than the laboratory genetic complementation, mutations in these sites can alter polymerase properties without destroying catalytic ability.

Polymerases involved in DNA repair are attractive for their broad substrate tolerance. Polymerase η is known for efficiently bypassing bulky lesions, such as cis-syn thymine dimmers, and is one of the most error-prone of the polymerases (~10% error frequency). Glick *et al.* [44] developed a genetic selection, where active mutants rescue a UV-sensitive yeast strain deficient in its DNA repair system, to find functional polymerase η variants. From the functional variants, the authors isolated one with fourfold improved activity [44] and several better able to incorporate fluorescent dNTP analogs [48]. They also isolated mutants with 15-fold higher replication fidelity [49]. At present, these error-prone polymerases are not very practical for biotechnology applications due to their extremely slow polymerase activity (~10 000 times slower than typical pol Is). However, the unique abilities of these enzymes to bypass damaged or irregular sites have proven useful in combination with other polymerases [50].

The fact that DNA polymerases can amplify their own genes establishes a link between genotype and phenotype, which can be used in evolutionary engineering to identify polymerases that are better self-replicators. Holliger and coworkers [51••] used the technique of compartmentalized self-replication (CSR). Polymerase variants are generated by *in vitro* mutagenesis and transformation into *E. coli* cells. The mutant polymerases are then encapsulated individually in droplets in a water/oil emulsion. In the droplet compartments, which also contain the components required for the polymerase chain reaction (PCR), the polymerases amplify only their own genes; here, improved function directly translates into gene amplification. Three cycles of CSR, in which the compartmentalized library pool was treated for a progressively longer time at 99°C (up to 15 min) before PCR, generated a *Taq* polymerase I with a half-life 11-fold greater than that of wild-type at 97.5°C. CSR also identified a *Taq* pol I variant with >130-fold increased resistance to heparin, a general DNA polymerase inhibitor.

A novel application of phage display allowed Romesburg and colleagues [30•] to select for polymerases that bind to rNTPs and act as RNA polymerases. They created a mutant library of *Taq* DNA pol I by fusion to the phage pIII coat protein. The substrate DNA template/primer duplexes were attached to other, adjacent pIII coat proteins. Polymerases that could extend the attached oligonucleotide primer by incorporating rNTP and biotinylated rUTP were selectively recovered using streptavidin-coated magnetic beads. Four rounds of screening isolated mutants that incorporate rNTPs virtually as efficiently as the wild-type enzyme incorporates dNTP substrates. Notably, each rNTP was incorporated with similar efficiency, although the wild-type *Taq* pol I and variants reported elsewhere [46] show very poor incorporation of rUTP. None of these polymerases, however, show good processivity.

DNA replication, the raison d'etre of DNA polymerases, is an extraordinarily accurate process. The rate of somatic mutation in mammalian cells is estimated to be about 10^{-10} per cell per duplication, and even very subtle changes in polymerase properties could lead to genetic instability. Nonetheless, the studies discussed here show that polymerases are quite robust to mutation. While the evolutionary implications of this fact remain elusive, these enzymes have significant engineering potential.

Conclusions

Engineering the target specificities of proteins that bind and modify DNA has proven challenging, but by no means impossible. Innovative structure-based and evolutionary design strategies have generated new transcription factors, restriction enzymes, recombinases and polymerases, and we can begin to envision engineering proteins capable of binding and acting at any (accessible) target DNA sequence. With these proteins will come the

power to control the expression of any gene or to recombine, excise or incorporate new DNA at specified sites within a genome. Engineered DNA polymerases will lead to improvements in such essential processes as PCR and mutant library generation, as well as fundamentally new applications. It is clear that we are just beginning to examine and understand how 'designable' these proteins are and to identify the most effective methods for engineering them. Current work, however, indicates a level of functional plasticity which, coupled with the multitude of potential applications, promises significant advances in the next few years.

Acknowledgements

We acknowledge support from the US National Science Foundation (Biological Information Technology and Storage (BITS) Grant EIA-0130613) and the US Defence Advanced Research Project Agency (DARPA) under award N66001-02-1-8929 (Disclaimer: any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the DARPA). YY also acknowledges support from a Japan Society for the Promotion of Science Postdoctoral Fellowship for Research Abroad.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Wharton RP, Ptashne M: **Changing the binding-specificity of a repressor by redesigning an α-helix**. *Nature* 1985, **316**:601-605.
2. Hollis M, Valenzuela D, Pioli D, Wharton R, Ptashne M: **A repressor heterodimer binds to a chimeric operator**. *Proc Natl Acad Sci USA* 1988, **85**:5834-5838.
3. Jana R, Hazbun TR, Fields JD, Mossing MC: **Single-chain λ Cro repressors confirm high intrinsic dimer-DNA affinity**. *Biochemistry* 1998, **37**:6446-6455.
4. Sieber M, Allemann RK: **Single chain dimers of MASH-1 bind DNA with enhanced affinity**. *Nucleic Acids Res* 1998, **26**:1408-1413.
5. Kuntz MA, Shapiro DJ: **Dimerizing the estrogen receptor DNA binding domain enhances binding to estrogen response elements**. *J Biol Chem* 1997, **272**:27949-27956.
6. Simoncsits A, Chen JQ, Percipalle P, Wang SL, Toro I, Pongor S: **Single-chain repressors containing engineered DNA-binding domains of the phage 434 repressor recognize symmetric or asymmetric DNA operators**. *J Mol Biol* 1997, **267**:118-131.
7. Gates CM, Stemmer WPC, Kaptein R, Schatz PJ: **Affinity selective isolation of ligands from peptide libraries through display on a lac repressor 'headpiece dimer'**. *J Mol Biol* 1996, **255**:373-386.
8. Robinson CR, Sauer RT: **Covalent attachment of Arc repressor subunits by a peptide linker enhances affinity for operator DNA**. *Biochemistry* 1996, **35**:109-116.
9. Percipalle P, Simoncsits A, Zakhariev S, Guarnaccia C, Sanchez R, Pongor S: **Rationally designed helix-turn-helix proteins and their conformational changes upon DNA-binding**. *EMBO J* 1995, **14**:3200-3205.
10. Liang TB, Chen JQ, Tjornhammar ML, Pongor S, Simoncsits A: **Modular construction of extended DNA recognition surfaces: mutant DNA-binding domains of the 434 repressor as building blocks**. *Protein Eng* 2001, **14**:591-599.
11. Pabo CO, Peisach E, Grant RA: **Design and selection of novel Cys_2His_2 zinc finger proteins**. *Annu Rev Biochem* 2001, **70**:313-340.
12. Beerli RR, Barbas CF III: **Engineering polydactyl zinc-finger transcription factors**. *Nat Biotechnol* 2002, **20**:135-141.
13. Segal DJ, Dreier B, Beerli RR, Barbas CF III: **Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences**. *Proc Natl Acad Sci USA* 1999, **96**:2758-2763.
14. Dreier B, Beerli RR, Segal DJ, Flippin JD, Barbas CF III: **Development of zinc finger domains for recognition of the 5′-ANN-3′ family of DNA sequences and their use in the construction of artificial transcription factors**. *J Biol Chem* 2001, **276**:29466-29478.
15. Segal DJ, Beerli RR, Blancafort P, Dreier B, Effertz K, Huber A, Koksch B, Lund CV, Magnenat L, Valente D *et al.*: **Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins**. *Biochemistry* 2003, **42**:2137-2148.
16. Liu Q, Xia Z, Zhong X, Case CC: **Validated zinc finger protein designs for all 16 GNN DNA triplet targets**. *J Biol Chem* 2002, **277**:3850-3856.
17. •• Isalan M, Klug A, Choo Y: **A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter**. *Nat Biotechnol* 2001, **19**:656-660.

This paper describes an efficient strategy for the high-throughput selection of three-finger ZFPs that can recognize arbitrary 9 bp DNA sequences, on the basis of phage display technology.

18. Sera T, Uranga C: **Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table**. *Biochemistry* 2002, **41**:7074-7081.
19. Sanchez JP, Ullman C, Moore M, Choo Y, Chua NH: **Regulation of gene expression in *Arabidopsis thaliana* by artificial zinc finger chimeras**. *Plant Cell Physiol* 2002, **43**:1465-1472.
20. Stege JT, Guan X, Ho T, Beachy RN, Barbas CF III: **Controlling gene expression in plants using synthetic zinc finger transcription factors**. *Plant J* 2002, **32**:1077-1086.
21. Guan X, Stege J, Kim M, Dahmani Z, Fan N, Heifetz P, Barbas CF III, Briggs SP: **Heritable endogenous gene regulation in plants with designed polydactyl zinc finger transcription factors**. *Proc Natl Acad Sci USA* 2002, **99**:13296-13301.
22. Ordiz MI, Barbas CF III, Beachy RN: **Regulation of transgene expression in plants with polydactyl zinc finger transcription factors**. *Proc Natl Acad Sci USA* 2002, **99**:13290-13295.
23. Reynolds L, Ullman C, Moore M, Isalan M, West MJ, Clapham P, Klug A, Choo Y: **Repression of the HIV-1 5′ LTR promoter and inhibition of HIV-1 replication by using engineered zinc-finger transcription factors**. *Proc Natl Acad Sci USA* 2003, **100**:1615-1620.
24. Papworth M, Moore M, Isalan M, Minczuk M, Choo Y, Klug A: **Inhibition of herpes simplex virus 1 gene expression by designer zinc-finger transcription factors**. *Proc Natl Acad Sci USA* 2003, **100**:1621-1626.
25. Rebar EJ, Huang Y, Hickey R, Nath AK, Meoli D, Nath S, Chen B, Xu L, Liang Y, Jamieson AC *et al.*: **Induction of angiogenesis in a mouse model using engineered transcription factors**. *Nat Med* 2002, **8**:1427-1432.
26. •• Blancafort P, Magnenat L, Barbas CF: **Scanning the human genome with combinatorial transcription factor libraries**. *Nat Biotechnol* 2003, **21**:269-274.

This study demonstrates whole genome scanning for specific phenotypes using combinatorial artificial ZFP transcription factor libraries.

27. Liu PQ, Rebar EJ, Zhang L, Liu Q, Jamieson AC, Liang Y, Qi H, Li PX, Chen B, Mendel MC *et al.*: **Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A**. *J Biol Chem* 2001, **276**:11323-11334.
28. Lanio T, Jeltsch A, Pingoud A: **On the possibilities and limitations of rational protein design to expand the specificity of restriction enzymes: a case study employing *Eco*RV as the target**. *Protein Eng* 2000, **13**:275-281.
29. Lanio T, Jeltsch A, Pingoud A: **Towards the design of rare cutting restriction endonucleases: using directed evolution to**

generate variants of *Eco*RV differing in their substrate specificity by two orders of magnitude. *J Mol Biol* 1998, **283**:59-69.

30. Xia G, Chen LJ, Sera T, Fa M, Schultz PG, Romesberg FE: **Directed**
• **evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase**. *Proc Natl Acad Sci USA* 2002, **99**:6597-6602.

This paper describes a clever phage display approach to select for RNA polymerase activity.

31. Samuelson JC, Xu SY: **Directed evolution of restriction endonuclease *Bst*YI to achieve increased substrate specificity**. *J Mol Biol* 2002, **319**:673-683.

32. Seligman LM, Chisholm KM, Chevalier BS, Chadsey MS, Edwards ST, Savage JH, Veillet AL: **Mutations altering the cleavage specificity of a homing endonuclease**. *Nucleic Acids Res* 2002, **30**:3870-3879.

33. Chevalier BS, Stoddard BL: **Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility**. *Nucleic Acids Res* 2001, **29**:3757-3774.

34. Chevalier BS, Kortemme T, Chadsey MS, Baker D, Monnat RJ,
•• Stoddard BL: **Design, activity, and structure of a highly specific artificial endonuclease**. *Mol Cell* 2002, **10**:895-905.

An impressive protein engineering effort to construct chimeric endonucleases with novel, predicted specificities using an automated computational algorithm. The crystal structure of one of the chimeric enzymes supports the design process.

35. Mills AA, Bradley A: **From mouse to man: generating megabase chromosome rearrangements**. *Trends Genet* 2001, **17**:331-339.

36. Rodriguez CI, Buchholz F, Galloway J, Sequerra R, Kasper J, Ayala R, Stewart AF, Dymecki SM: **High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP**. *Nat Genet* 2000, **25**:139-140.

37. Kilby NJ, Snaith MR, Murray JAH: **Site-specific recombinases – tools for genome engineering**. *Trends Genet* 1993, **9**:413-421.

38. Voziyanov Y, Konieczka JH, Stewart AF, Jayaram M: **Stepwise**
•• **manipulation of DNA specificity in Flp: progressively adapting Flp to individual and combinatorial mutations in its target site**. *J Mol Biol* 2003, **326**:65-76.

This work combines rational and evolutionary design strategies to engineer the DNA target-site specificity of Flp recombinase. A powerful double-reporter screen was used to identify mutants with altered substrate specificities.

39. Voziyanov Y, Stewart AF, Jayaram M: **A dual reporter screening**
• **system identifies the amino acid at position 82 in Flp site-specific recombinase as a determinant for target specificity**. *Nucleic Acids Res* 2002, **30**:1656-1663.

This paper describes a dual reporter screen for identifying Flp recombinase variants that recognize altered target DNA sequences. One reporter gene is expressed constitutively until it is excised by a recombinase that recognizes the wild-type target sequence. A second reporter gene is constitutively expressed until it is excised by a recombinase that recognizes a novel target DNA sequence. This system was used to identify mutants with either relaxed or shifted substrate specificity. By comparing these proteins, the authors identified an amino acid that plays an important role in determining substrate specificity.

40. Santoro SW, Schultz PG: **Directed evolution of the site**
• **specificity of Cre recombinase**. *Proc Natl Acad Sci USA* 2002, **99**:4185-4190.

The authors used a directed evolution strategy to identify Cre recombinase variants that recognize altered DNA target sites. In separate screens, they identified Cre mutants that recognized the novel target sites (positive screening) and variants that could not recognize the wild-type site (negative screening). A combination of positive and negative screening was required to switch specificity.

41. Rufer AW, Sauer B: **Non-contact positions impose site selectivity on Cre recombinase**. *Nucleic Acids Res* 2002, **30**:2764-2771.

42. Buchholz F, Stewart AF: **Alteration of Cre recombinase site specificity by substrate-linked protein evolution**. *Nat Biotechnol* 2001, **19**:1047-1052.

43. Patel PH, Suzuki M, Adman E, Shinkai A, Loeb LA: **Prokaryotic**
•• **DNA polymerase I: evolution, structure, and 'base flipping' mechanism for nucleotide selection**. *J Mol Biol* 2001, **308**:823-837.

An excellent review of the structure, catalytic mechanism, and evolution of pol I, providing a detailed discussion of the molecular basis of replication fidelity. The unexpectedly high mutability of highly conserved active-site residues in DNA polymerases is discussed in relationship to their biological functions and evolution.

44. Glick E, Vigna KL, Loeb LA: **Mutations in human DNA polymerase eta motif II alter bypass of DNA lesions**. *EMBO J* 2001, **20**:7303-7312.

45. Shinkai A, Loeb LA: ***In vivo* mutagenesis by *Escherichia coli* DNA polymerase I — Ile709 in motif A functions in base selection**. *J Biol Chem* 2001, **276**:46759-46764.

46. Patel PH, Loeb LA: **Multiple amino acid substitutions allow DNA polymerases to synthesize RNA**. *J Biol Chem* 2000, **275**:40266-40272.

47. Skandalis A, Loeb LA: **Enzymatic properties of rat DNA polymerase β mutants obtained by randomized mutagenesis**. *Nucleic Acids Res* 2001, **29**:2418-2426.

48. Glick E, Anderson JP, Loeb LA: ***In vitro* production and screening of DNA polymerase eta mutants for catalytic diversity**. *Biotechniques* 2002, **33**:1136.

49. Glick E, Chau JS, Vigna KL, McCulloch SD, Adman ET, Kunkel TA, Loeb LA: **Amino acid substitution at conserved tyrosine 52 alters fidelity and bypass efficiency of human polymerase H**. *J Biol Chem* 2003, **21**:19341-19346.

50. Tae EJL, Wu YQ, Xia G, Schultz PG, Romesberg FE: **Efforts toward expansion of the genetic alphabet: replication of DNA with three base pairs**. *J Am Chem Soc* 2001, **123**:7439-7440.

51. Ghadessy FJ, Ong JL, Holliger P: **Directed evolution of**
•• **polymerase function by compartmentalized self-replication**. *Proc Natl Acad Sci USA* 2001, **98**:4552-4557.

This paper demonstrates the use of CSR for evolving polymerase function. With a heat-stable water-in-oil emulsion system to encapsulate mutant polymerases with PCR reagents, each DNA polymerase can be amplified according to its activity. This system was used to evolve polymerases with higher thermostability or resistance to an inhibitor.

Erratum
Engineering proteins that bind, move, make and break DNA☆

Cynthia H Collins, Yohei Yokobayashi, Daisuke Umeno and Frances H Arnold

Current Opinion in Biotechnology 2003, **14**:371–378

0958-1669/$ – see front matter
© 2003 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.copbio.2003.10.002

In this article, published in the August 2003 issue of *Current Opinion in Biotechnology*, the author's discussion of engineering α-helical DNA-binding domains failed to note earlier reports of engineering protein–DNA interactions [1–3]. In particular, two studies reported altering the DNA-binding specificities of the Mnt repressor [1] and catabolite activator protein [2], using genetic selection methods, whereas a third described the substitution of an entire α-helix to alter the DNA-binding specificity of the 434 repressor protein [3]. The authors regret this significant oversight.

References

1. Youderian P, Vershon A, Bouvier S, Sauer RT, Susskind MM: **Changing the DNA-binding specificity of a repressor**. *Cell* 1983, **35**:777-783.
2. Ebright RH, Cossart P, Gicquelsanzey B, Beckwith J: **Mutations that alter the DNA-sequence specificity of the catabolite gene activator protein of *E. coli***. *Nature* 1984, **311**:232-235.
3. Wharton RP, Brown EL, Ptashne M: **Substituting an α-helix switches the sequence-specific DNA interactions of a repressor**. *Cell* 1984, **38**:361-369.

☆ PII of original article: S0958-1669(03)00091-0