Molecular Cell, Vol. 10, 895–905, October, 2002, Copyright ©2002 by Cell Press

Design, Activity, and Structure of a Highly Specific Artificial Endonuclease

Brett S. Chevalier,[1,3] Tanja Kortemme,[2] Meggen S. Chadsey,[3] David Baker,[2] Raymond J. Monnat, Jr.,[3] and Barry L. Stoddard[1,4]
[1]Fred Hutchinson Cancer Research Center and Graduate Program in Molecular and Cell Biology
University of Washington
1100 Fairview Avenue N. A3-023
Seattle, Washington 98109
[2]Howard Hughes Medical Institute and Department of Biochemistry
University of Washington
Box 357350
[3]Department of Pathology and Department of Genome Sciences
University of Washington
Box 357705
Seattle, Washington 98195

Summary

We have generated an artificial highly specific endonuclease by fusing domains of homing endonucleases I-DmoI and I-CreI and creating a new 1400 Å^2 protein interface between these domains. Protein engineering was accomplished by combining computational redesign and an in vivo protein-folding screen. The resulting enzyme, E-DreI (Engineered I-DmoI/I-CreI), binds a long chimeric DNA target site with nanomolar affinity, cleaving it precisely at a rate equivalent to its natural parents. The structure of an E-DreI/DNA complex demonstrates the accuracy of the protein interface redesign algorithm and reveals how catalytic function is maintained during the creation of the new endonuclease. These results indicate that it may be possible to generate novel highly specific DNA binding proteins from homing endonucleases.

Introduction

Evolution is a tinkerer, continually adapting and reusing proteins to accomplish new tasks (Jacob, 1977). The recycling of protein domains has been revealed by genetic, developmental, and, most recently, whole genome sequence analyses (Carroll et al., 2001). Evolution generates new functional proteins through the recombination and fusion of existing proteins. This has been accomplished both through the linkage of independently folded protein domains with flexible peptide linkers and through the more intimate fusion of protein domains via highly specific protein interfaces. Reproducing the process of domain linkage is relatively straightforward in the laboratory. In contrast, the design of novel protein interfaces is extremely difficult.

Recent developments in computational protein design algorithms (Pokala and Handel, 2001) and in directed protein evolution strategies (Farinas et al., 2001; Schmidt-Dannert, 2001) offer great promise for protein engineering and molecular interface design. Computational protein design successes have included the redesign of protein cores, the introduction of metal binding sites into proteins, and increases in protein stability (Dahiyat and Mayo, 1997; Pokala and Handel, 2001). Particularly notable have been the complete redesign of a $\beta\beta\alpha$ protein motif (Dahiyat and Mayo, 1997), design of novel helical bundle topologies (Harbury et al., 1998), and rational construction of a protein with enzyme-like properties (Bolon and Mayo, 2001). There has also been substantial progress in the selection of enzyme variants that display altered substrate specificities or physical properties (Farinas et al., 2001; Schmidt-Dannert, 2001).

DNA binding proteins are attractive targets for protein redesign. Nature is particularly adept at creating novel DNA binding proteins by tethering independently folded protein domains; successful attempts to mimic this strategy have been reported. For example, nonspecific nuclease domains have been tethered to sequence-specific DNA binding modules such as zinc-fingers (Smith et al., 1999, 2000) and used in vivo to stimulate homologous recombination (Bibikova et al., 2001). Such constructs are potentially useful for the creation of gene-specific reagents (a single protein that recognizes a unique site in a genome) but generally lack the ability to specifically act at a single unique phosphodiester bond or base pair within the DNA target site (Smith et al., 1999).

Catalytic specificity might be improved in engineered DNA binding proteins by embedding DNA binding and catalytic activities within a single structural unit, such as a type II restriction endonuclease. However, efforts to increase the length of sequence readout or alter the specificity of these enzymes have resulted in the loss of catalytic activity or overall diminution of specificity (Lanio et al., 2000) due to the tight interdependence of enzyme structure, substrate recognition, and catalysis.

Homing endonucleases may be intrinsically more amenable to redesign than type II restriction endonucleases. These enzymes have been found in *Eubacteria*, *Archea,* and single-cell eukaryotes, where they are encoded by and promote the lateral transfer of mobile introns (Belfort and Perlman, 1995; Belfort et al., 1995; Belfort and Roberts, 1997; Chevalier and Stoddard, 2001). Homing endonucleases bind long (15–40 bp) DNA target sites, ensuring extremely high specificity while tolerating small numbers of single base pair polymorphisms in those sites (Chevalier and Stoddard, 2001). This combination makes these proteins highly sequence-specific (recognizing as few as 1 in 10^9 random sequences) and excellent candidates from which to engineer new, sequence-specific DNA binding proteins.

Four homing endonuclease families have been defined on the basis of conserved protein motifs (Chevalier and Stoddard, 2001). Members of the LAGLIDADG enzyme family appear particularly amenable to reengineering. This endonuclease family consists of several hundred members that bind and cleave diverse DNA homing sites (Dalgaard et al., 1997). LAGLIDADG endonucleases

[4]Correspondence: bstoddar@fred.fhcrc.org

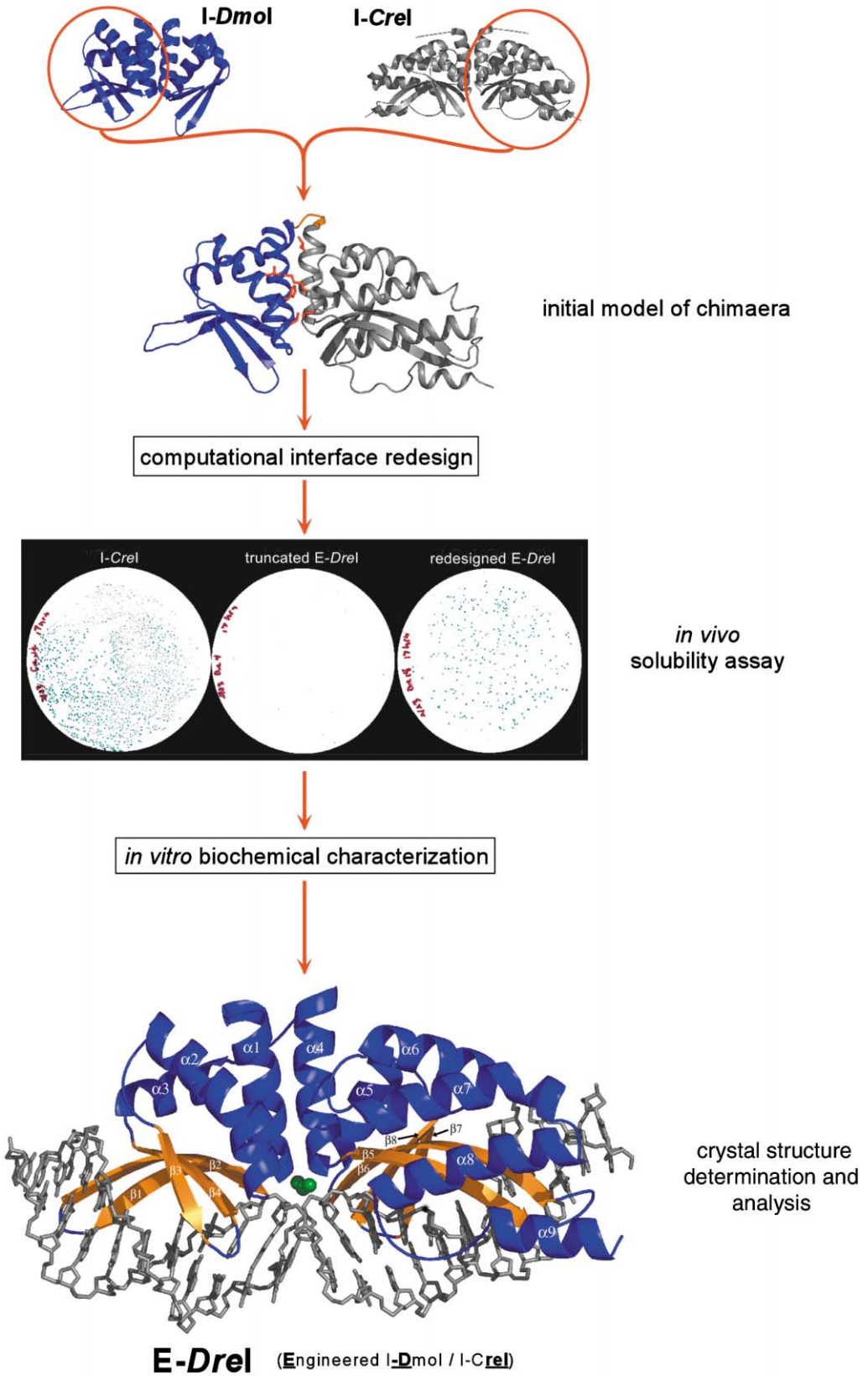

Figure 1. Overall Design Strategy for E-DreI

The N-terminal domain of I-DmoI (RCSB1B24) was swapped for a subunit in I-CreI (1G9Z). Alanine substitutions were made that eliminated obvious steric clashes across the domain interface; these were insoluble. The interface was optimized through computational design, followed by an in vivo protein-folding assay in which resulting constructs were linked to the lacZα peptide. Constructs expressing soluble E-DreI/lacZα

Table 1. Predicted Contributions to the Interface Free Energy of Each Stage of Design

	Predicted Contributions to the Interface Free Energy (kcal/mol)					Accessible Surface Area Buried in the Interface (Å^2)
	LJ_{attr}	LJ_{rep}	Solv	H Bond	ΔG_{tot}	
I-CreI X-ray structure	−34.4	3.3	14.9	−6.4	−22.6	1867
I-DmoI X-ray structure	−22.6	2.4	3.8	−1.7	−18.1	1431
Initial model of E-DreI chimera (with all wild-type side chains)	−21.6	36.2	7.2	−4.4	17.4	1294
Model of chimera with truncated side chains at interface	−14.0	4.6	3.9	−4.7	−10.2	1013
E-DreI (computational redesign)	−26.1	3.5	5.5	−8.2	−25.3	1441
E-DreI X-ray structure	−25.5	2.1	6.7	−3.1	−19.8	1461

Energies are computed as described in Kortemme and Baker (2002). LJ_{attr}, attractive component of the Lennard-Jones potential; LJ_{rep}, repulsive component of the Lennard-Jones potential; Sovl, solvation energy; H Bond, hydrogen bonding energy; ΔG_{tot}, total interface free energy. Accessible surface area buried in the interface is computed with a probe radius of 1.4 Å using WHATIF (Vriend, 1990).

are relatively small protein homodimers or monomers composed of separate domains (Chevalier and Stoddard, 2001; Dalgaard et al., 1997) in which LAGLIDADG motifs form structurally conserved, tightly associated α-helical pairs at the center of hydrophobic domain interface (Duan et al., 1997; Heath et al., 1997; Ichiyanagi et al., 2000; Jurica et al., 1998). Binding of DNA target sites is dictated by independent sets of interactions made between individual domains or subunits to individual DNA half-sites (Jurica et al., 1998). The active sites are directly juxtaposed at the enzyme domain interface and share a catalytic divalent cation so that they must maintain their physical association in order to cleave DNA substrates (Chevalier et al., 2001).

A variety of approaches has been used either to identify DNA target site variants that are still recognized by wild-type homing endonucleases (Argast et al., 1998; Bryk et al., 1995; Gruen et al., 2002; Guo et al., 2000; Silva et al., 1999) or endonuclease variants that can cleave mutant DNA homing sites that are not recognized by wild-type enzyme (Seligman et al., 2002). Such strategies allow a homing endonuclease to be altered to recognize minor variants of its natural DNA target site. A more radical strategy for the reengineering of homing endonucleases, allowing investigators to make larger changes in site specificity, is suggested by phylogenetic analyses: the LAGLIDADG family has undergone expansion and diversification by the repeated fusion of unrelated enzyme domains (Chevalier and Stoddard, 2001). We reasoned that it might be possible to mimic this evolutionary history by fusing the DNA binding domains of two highly divergent LAGLIDADG homing endonucleases to generate a catalytically active chimeric endonuclease. The correspondence between the designed and experimentally determined structural features of the new endonuclease, called E-DreI, demonstrate the feasibility of this approach for the generation of new, highly sequence-specific DNA binding proteins.

Results and Discussion

The Creation of E-DreI

The experimental strategy is outlined in Figure 1. Structural modeling indicated that it should be possible to create a novel chimaeric endonuclease by fusing the N-terminal domain of I-DmoI to an I-CreI monomer, repacking the new protein interface to facilitate efficient folding and intimate domain association, and then inserting a short peptide linker to create an enzyme monomer. Analysis of the initial protein model identified 14 domain interface residues exhibiting poor interatomic contacts. An automated computational design protocol was used to search through possible new interface sequence combinations for these residues. The best predicted E-DreI interface variants as determined by computational redesign (16 total constructs, each containing between 8 and 12 altered residues in the interface) were generated and screened in vivo to insure proper folding and solubility. Biochemical characterization of several soluble E-DreI variant proteins revealed that each was able to bind and cleave a specific, 23 bp chimeric DNA target site with high specificity and wild-type kinetic activity. We determined the cocrystal structure of one of these active variants bound to target site DNA in order to determine the accuracy of prediction of the computational redesign method and to characterize the artificial endonuclease.

The availability of X-ray crystal structures of I-DmoI (Silva et al., 1999) and of I-CreI (Chevalier et al., 2001; Jurica et al., 1998) allowed us to generate a detailed starting model of E-DreI. The N-terminal large domain of I-DmoI was substituted for a single subunit of the I-CreI homodimer to create the initial scaffold for an enzyme chimera. Superposition of the backbone atoms in the two conserved LAGLIDADG helices at the Dmo/Cre interface was used to orient and align the docked I-DmoI domain. Further analysis of this modeled protein revealed a large though unoptimized protein interface of

complemented the *E. coli* lacZω fragment to form blue colonies; constructs expressing insoluble E-DreI/lacZα yielded white colonies. Three examples of this assay are shown: fully soluble I-CreI, an insoluble E-DreI construct with clashing interface residues truncated, and a final E-DreI construct containing a redesigned interface. Biochemical characterization was performed (Figure 2), and E-DreI was shown to be an enzymatically active, highly specific endonuclease. The structure of E-DreI complexed to its DNA target site was solved to 2.4 Å resolution. This and all structural figures were created and rendered using PYMOL (www.pymol.org).

approximately 1300 Å^2 with numerous steric side chain clashes and a very unfavorable overall repulsive interface free energy (Table 1). One residue (L108) was located within a central LAGLIDADG helix in the domain interface, while the remaining five residues (L47, H51, L55, K193, and L194) were located within the interface near the helices. The residues at these six positions were substituted with alanine residues (or, in one case, an aspartate residue) to minimize steric clashes in an attempt to promote the formation of a more stable domain interface. However, this alanine-substituted or side chain-truncated E-DreI variant was insoluble (data not shown). The most likely explanation for this was failure to form a stable domain interface due to the presence of structural cavities; we therefore performed a complete automated redesign of the domain interface.

The computational interface redesign focused on the six residues exhibiting steric clashes in the original model and was extended to include eight additional residues predicted to contribute substantially to the interface free energy (A12, Y13, L17, I19, I52, E105, Y109, and F113). At each of these fourteen sites, a library of possible side chain conformations spanning 19 potential amino acids (all but cysteine) in different backbone-dependent rotameric states (about 500 rotamers per sequence position) was created. The interaction of all rotamers with the surrounding, fixed portion of the molecule (including the polypeptide backbone and all side chains not subjected to sequence design), and all pairwise rotamer-rotamer energies were computed using a free energy function which includes van der Waals interactions, solvation effects, explicit hydrogen-bonding interactions, and statistical terms representing the backbone-dependent internal free energies of amino acid rotamers (Kortemme and Baker, 2002; Kuhlman and Baker, 2000).

A Monte Carlo-simulated annealing procedure, in which a move consists of the random replacement of a single rotamer with an alternative rotamer from the library, was then used to search through the 8×10^{17} sequence combinations (with 6×10^{37} total rotamer combinations) to identify particularly low free energy amino acid sequences. Since the Monte Carlo protocol does not guarantee finding a global free energy minimum, we performed 1000 separate sequence design runs using two polypeptide backbone models with slightly different relative orientations of the LAGLIDADG helices in the domain interface. This procedure yielded a family of sequences with different amino acid choices at each of the 14 design positions. Native residues were consistently best at 3 of 14 positions, and a single new residue was consistently best at a fourth position. Two to five different residues were identified at the remaining ten positions, to yield a total of 51,840 possible combinations. This set of solutions was reduced by eliminating sequence changes likely to affect active site residues and by reducing redundancy (for example, if F and Y were computationally selected at a position, we continued with only one residue based on whether a neighboring atom could form an H bond). These steps reduced the solutions from 35 possible amino acid substitutions over 10 positions to 25 substitutions over 9 positions, or 1152 position/residue combinations. In a final computational step, the interface free energy of all 1152 combinations was exhaustively enumerated using optimized rotamer conformations for each sequence. Consistently top scoring interface free energies were found for sixteen different sequences with the following substitutions as compared with wild-type: A12A or Y; Y13Y; L17L; I19W; L47L or W; H51H or F; I52I; L55R; E105R; A108A; Y109Y; F113I; K193N or Y; L194F.

In addition to the redesign of the enzyme interface, a short peptide linker was inserted between the I-DmoI and I-CreI domains to generate a monomeric protein. The linker chosen for this purpose, -NGN-, resembled the -NMR- linker found in native I-DmoI but contained glycine and asparagine residues in order to exploit the high β-turn propensity of NG- and GN-containing peptides (Hutchinson and Thornton, 1994). The 16 enzyme variants described above were generated by site-directed mutagenesis and screened for in vivo folding and solubility. The solubility screen utilizes a blue/white colony color difference that reflects protein solubility-dependent LacZα complementation in *E. coli* (Wigley et al., 2001). In order to perform this screen, we fused a lacZα peptide to E-DreI variants and then expressed each in *E. coli* cells expressing a lacZω protein partner. In this screen insoluble E-DreI/lacZα constructs form inclusion bodies, fail to complement lacZω, and give rise to white colonies. Conversely, soluble E-DreI/lacZα constructs complement lacω to give rise to blue colonies on X-gal indicator plates (Figure 1).

The alanine-truncated version of E-DreI gave rise exclusively to white colonies in this protein-folding and solubility screen. In contrast, strongly lacZ-positive blue colonies indistinguishable from those generated by an I-CreI/lacZα positive control were generated when we expressed different E-DreI/lacZα variants whose interface had been computationally redesigned. The interface residue substitutions indicated by computational redesign were incorporated over several rounds of site-directed mutagenesis. We observed an increase in the fraction and intensity of blue colonies with the incorporation of each successive substitution (data not shown). The most soluble E-DreI constructs displayed predicted interface free energies similar to parent structures I-CreI and I-DmoI, in contrast to the initial and alanine-truncated forms of E-DreI that had poor predicted interface free energies and were largely insoluble (Table 1). These results indicate that the computational interface redesign protocol generated several protein variants that are able to fold.

E-DreI is a Novel Endonuclease with Altered Specificity

In order to determine the binding and catalytic activities of E-DreI on different DNA target sites, we overexpressed and purified three of the most highly soluble E-DreI variants as identified by the in vivo protein solubility assay described above. All three proteins were soluble and purified by heparin affinity and size exclusion chromatography. All three proteins were stable at 4°C at a concentration of ~5 mg/ml in buffer containing 5% glycerol, 150 mM NaCl, 1 mM $CaCl_2$, and 50 mM Tris (pH 8.0). Since E-DreI is a two-domain chimeric monomer composed of I-DmoI and I-CreI domains, the most likely E-DreI target site would be a chimera of the I-DmoI and

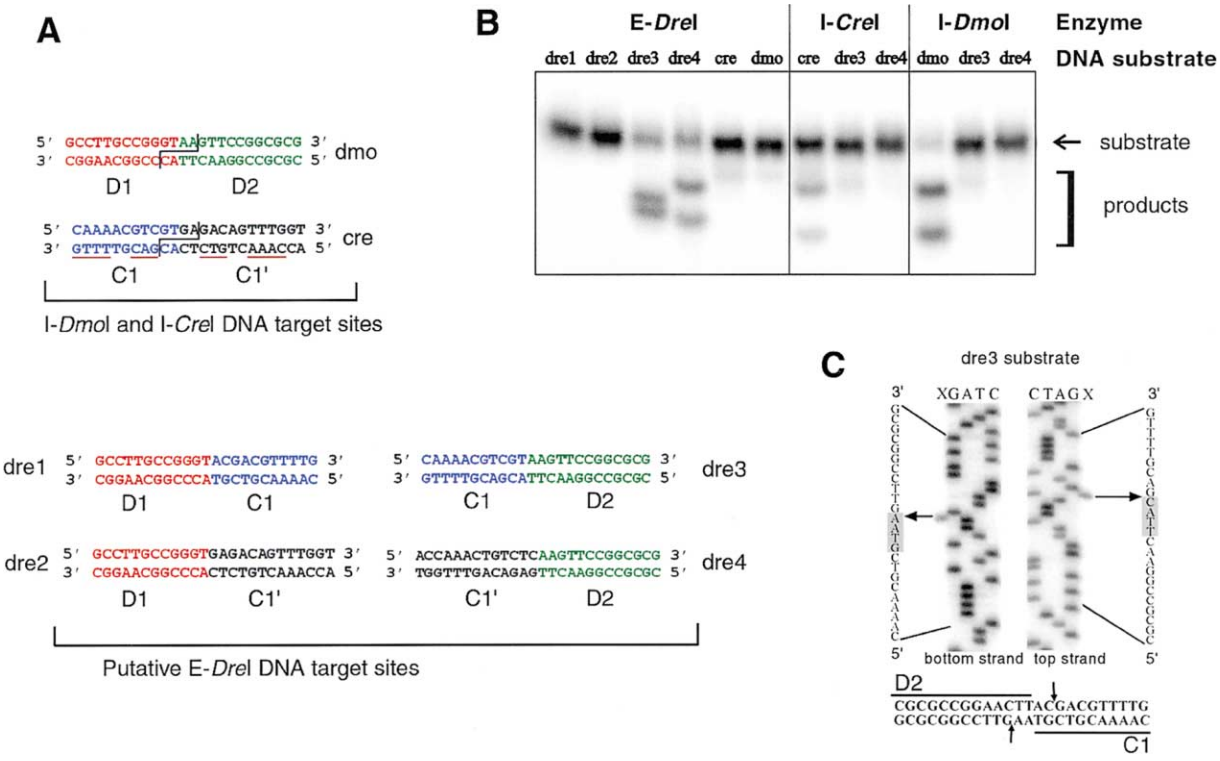

Figure 2. Biochemical Characterization of E-DreI

(A) The dmo and cre sites (targets of I-DmoI and I-CreI, respectively) differ considerably; dmo is asymmetric, while cre is nearly palindromic. All four potential chimeric sites were used for initial cleavage assays.

(B) E-DreI, I-CreI, and I-DmoI activity against different target site DNAs. E-DreI cleaves dre3 and dre4 substrates (which both contain the same dmo half-site) but does not cleave dre1, dre2, cre, or dmo target sites. I-CreI and I-DmoI cleave their respective target sites but do not cleave any of the dre sites (dre1 and dre2 not shown).

(C) Mapping of E-DreI scissile phosphate positions in the dre3 target site. E-DreI cleaves target site strands to generate four base, 3′ cohesive ends that are characteristic of all known LAGLIDADG enzymes (identical results for dre4 are not shown).

I-CreI target sites (Figure 2A). The two native homing sites (termed dmo and cre, respectively) can be considered as four distinct half-sites, with the center of each target site defined by the middle of the four base overhang generated by cleavage. The native dmo site is asymmetric, and the two half-sites are referred to here as D1 and D2. The cre site is pseudo-palindromic, and we refer to the two half-sites as C1 and C1′. Four chimeric sites can be generated from these four half-sites (Figure 2A); we term these sites dre1 (D1:C1), dre2 (D1:C1′), dre3 (D2:C1), and dre4 (D2:C1′).

Each of the three E-DreI variants cleaved target sites dre3 and dre4 but was unable to cleave the dre1 or dre2 target sites or the native dmo or cre target sites (Figure 2B). Conversely, purified I-DmoI or I-CreI did not cleave any of the four dre target sites. The dre3 and dre4 sites each contain the same dmo half-site and one of the two cre half-sites: thus the N-terminal domain of I-DmoI recognizes only the D2 dmo half site (which we did not know upon beginning this project) while the C-terminal domain from I-CreI recognizes either cre half-site, as expected for a domain from an endonuclease homodimer (Figure 2). The behavior of E-DreI indicates that it is a novel, highly sequence-specific endonuclease that displays altered DNA target site specificity.

We chose a single variant for more thorough characterization. This construct contains eight computationally designed point mutations at the domain interface (I19W, H51F, L55R, E105R, L108A, F113I, K193N, and L194F). Dre16 cleaves its target site precisely at one phosphodiester bond on each DNA strand, separated by four base pairs in the target site DNA to generate four base, 3′-extended cohesive ends (Figure 2C). This end geometry is identical to all other characterized LAGLIDADG homing endonucleases. Dre16 displays a dissociation constant (K_D) of 100 ± 5 nM as determined by gel shift assays (data not shown), two orders magnitude lower than the 1 nM dissociation constant of native I-CreI (Wang et al., 1997). The estimated single turnover catalytic rate (k_{cat}*) (Halford et al., 1980) of E-DreI is nearly identical to native I-CreI: $k_{cat}^* = 0.04\ min^{-1}$ for E-DreI versus $0.03\ min^{-1}$ for I-CreI (R.J.M., additional data not shown; M. Turmel and C. Lemieux, personal communication).

E-DreI Structural Analysis

The structure of E-DreI was determined by X-ray crystallography using data collected at the Advanced Light Source Synchrotron beamline 5.0.2 to 2.4 Å resolution (R_{work}/R_{free} = 0.231/0.256; Figure 1). Within each asymmetric unit of the $P3_1$ unit cell, four copies of the E-DreI/DNA complex are visible: two are well ordered and have

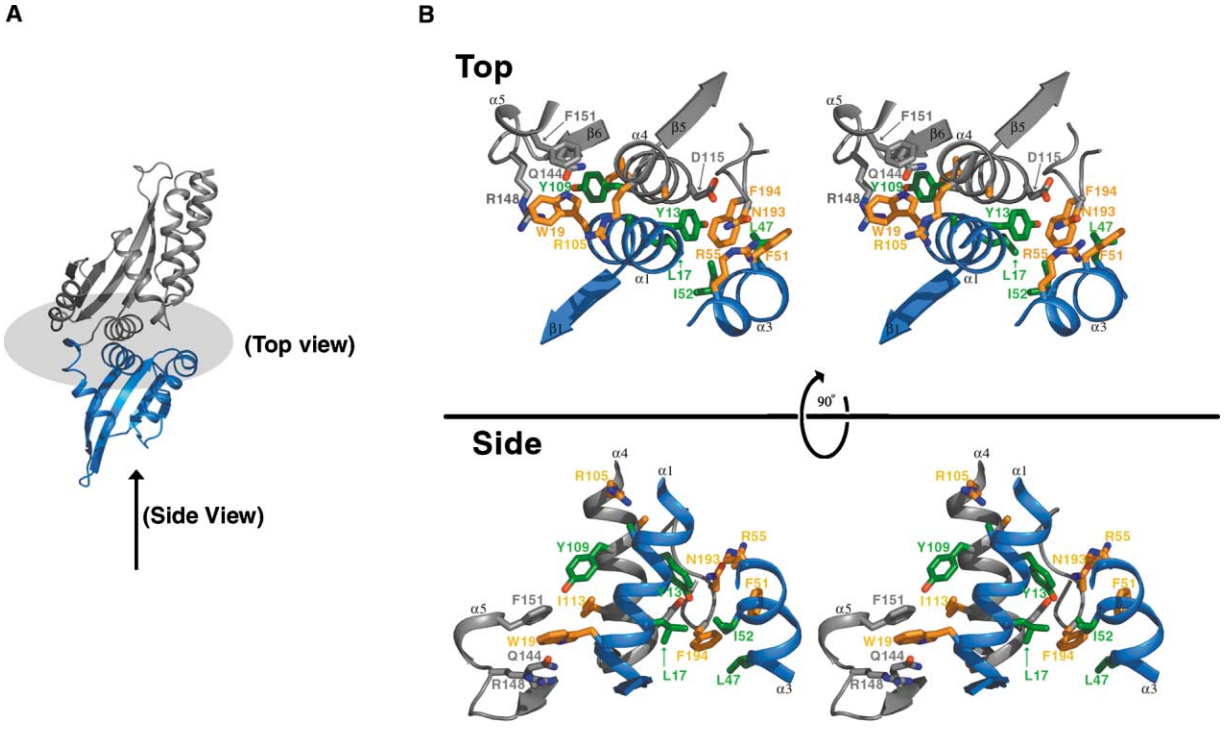

Figure 3. The E-DreI Domain Interface

The blue backbone denotes N-terminal domain from I-DmoI; the gray backbone denotes C-terminal domain from I-CreI.

(A) Location of the interface in the overall structure (gray oval).

(B) Detailed view of the interface highlighting the packing of interacting side chains. Fourteen residues were computationally randomized; of these, eight residues were finally changed (orange: I19W, H51F, L55R, E105R, L108A [not labeled], F113I [labeled in the side view figure], K193N, and L194F). The remaining six residues (green) are found in the native sequence (A12 [not labeled], Y13, L17, L47, I52, and Y109). Other residues that interact with the computationally screened residues are shown in their backbone color (top view: D115, Q144, R148, and F151; side view: Q144, R148, and F151).

an average B of 35 Å^2, while the remaining two complexes were poorly ordered and have been modeled as polyalanine/DNA (average B ~115 Å^2). In the well-ordered complexes, density is present for all residues except 1–4 and 253–260 (which are similarly disordered in the I-DmoI and I-CreI structures, respectively [Chevalier et al., 2001; Jurica et al., 1998; Silva et al., 1999]).

The general topology of the E-DreI structure and its domain interface was similar to those found in the previously determined structures of the parental endonucleases. The conserved core LAGLIDADG helices pack tightly against one another, and the redesigned residues cluster to either side of these helices to form a well-packed domain interface (Figure 3). The buried surface area of the interface in the E-DreI structure (1460 Å^2) is comparable to that in I-DmoI (1430 Å^2) and I-CreI (1870 Å^2) and is in the size range of typical protein-protein interfaces (Conte et al., 1999).

The crystal structure of E-DreI correlated well with the initial structural model (Figure 4A); the Cα rmsd between predicted and actual structures is 0.8 Å. The greatest divergence is in the DNA binding loops between β1/β2 and β3/β4 in the domain originating from I-DmoI. However, the model of these loops was derived from the I-DmoI structure lacking a DNA substrate, and upon DNA binding there are conformational changes at this part of the DNA interface. Slight differences are also noticeable at the C-terminal end of α3 and N-terminal end of α4 (connected by the linker between the two domains), where the ends of each of these helices are closer to one another (~0.4 Å) compared to the model. Prediction of the exact position of the top of α4 was complicated by slight divergence in the backbone positions of the tops of the LAGLIDADG helices of I-DmoI and I-CreI (about 1.4 Å). In E-DreI, helix α4 assumes an intermediate position compared to the same helix in native I-DmoI and I-CreI; the slight movement of α3 aids to accommodate this fit. The linker sequence of -$_{102}$NGN$_{104}$- is packed against the top of the protein, and side chains are easily seen.

The computationally redesigned model accurately predicted the actual E-DreI interface structure (Figures 4B–4E). Overall, the side chains in the designed and experimental interfaces superimpose well including both conserved and substantially altered residues. All residues substituted in the design procedure were predicted computationally to contribute significantly to the interface free energy of the E-DreI crystal structure; a subsequent computational packing analysis on the actual structure of E-DreI confirmed this prediction quantitatively for most of these sites (data not shown). These analyses indicate that residues Y13, W19, and F194 are vital energetic hot spots for interface stabilization. Y13 forms hydrogen bonds across the interface to both D110

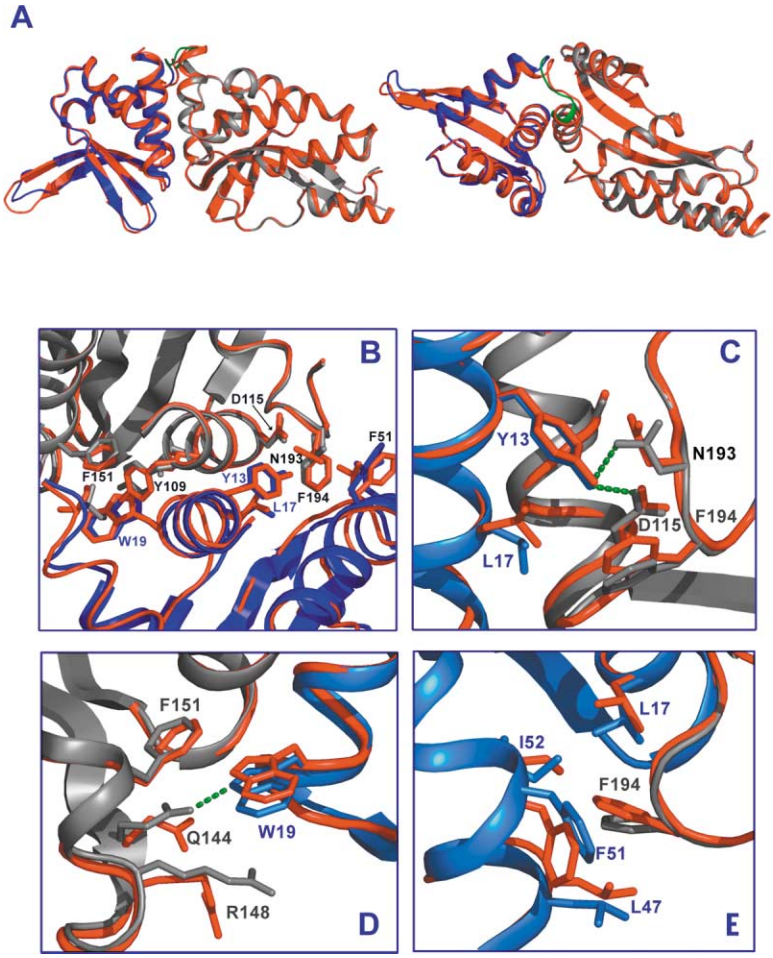

Figure 4. Comparison of E-DreI X-Ray Structure and Computational Model

Domains from I-DmoI and I-CreI in the X-ray structure are blue and gray, respectively; the designed model is red.

(A) Overall superposition of the backbone template used for computational design and the E-DreI crystal structure.

(B) Top view of the domain interface showing superposition of the structure and computational model, and detailed side chain packing interactions including three interface residues identified as hot spots in the computational analysis: Y13, W19, and F194.

(C) Y13 forms two hydrogen bonds across the interface to D115 and N193.

(D) W19 stacks across the interface against F151 and forms an unanticipated hydrogen bond to Q144 and a loose cation-π interaction with R148.

(E) F194 is buried across the interface in a hydrophobic pocket lined by residues L14, L47, F52, and I52.

and N193 (Figure 4C). W19 interacts with three residues across the domain interface; its main interaction, as designed, is stacking with F151. In addition, it forms an H bond with Q144 and a loose cation-π interaction with R148 (Figure 4D). These last two interactions were not predicted since Q144 and R148 are residues adjacent to an active site and, as such, were excluded from repacking calculations (see below). Finally, F194 reaches across the interface into a hydrophobic pocket created by L17, L47, F51, and I52 (Figure 4E).

A fundamental assumption of our engineering strategy (Figure 1) was that the two independent DNA binding domains in E-DreI would continue to recognize and bind their respective native DNA half-sites. The validity of this assumption was borne out by the crystal structure: all substrate contacts made by E-DreI originate from β sheets in the major groove and contacts made across the cre half-site contained in E-DreI closely resembled those previously documented in I-CreI/DNA cocrystal structures (Chevalier et al., 2001; Jurica et al., 1998) (Figure 5). Residues making contacts in the cre half-site of E-DreI include two arginines, three glutamines, an asparagine, and a tyrosine. The dmo half-site interface, which had not been previously visualized, includes direct contacts to DNA bases by four arginines, two acidic residues (asp and glu), a tyrosine, and a threonine. The I-DmoI domain also makes two base-specific contacts from the protein backbone and stacking between a thymine methyl group and a tyrosine ring. As has been observed in all other homing endonuclease structures, the number of H bonds in the DNA-protein interface of E-DreI is undersaturated; of 92 potential H bonds that could be made in the major grove of the 23 base pair interface, only 32 direct and 16 water-mediated contacts were observed (Figure 5). These two DNA-protein interfaces illustrate the diversity of DNA-protein contacts employed by LAGLIDADG endonucleases.

It was not as clear at the outset of the project that successful redesign of the E-DreI protein interface would generate an active enzymatic catalyst because the predicted active sites of E-DreI are located directly at the bottom of the redesigned protein interface (Chevalier et al., 2001; Jurica et al., 1998) and might be structurally perturbed during the engineering process. In addition, in both I-CreI and I-DmoI at least one residue in each active site is domain swapped and is not recapitulated in the E-DreI enzyme. For example, R51 in the I-CreI enzyme, which is located proximal to a catalytic metal and the nucleophilic solvent molecule, is absent from one active site of E-DreI. However, all of the soluble E-DreI variants retained catalytic activities comparable to the parent endonucleases. Activity appears to be retained because the E-DreI active sites still contain the conserved LAGLIDADG active site architecture (Figure

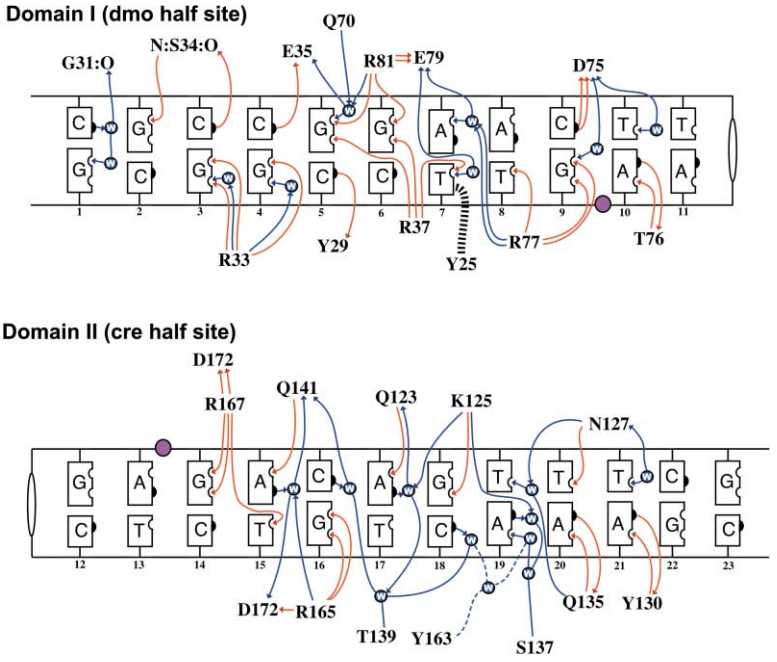

Figure 5. Base-Specific Contacts Made by E-DreI to Target Site DNA

Contacts to the dmo half-site are shown in the top, and contacts to the cre half-site are shown in the bottom. Both strands of each half-site are shown. Direct hydrogen bonds are indicated in red, and water-mediated hydrogen bonds are blue where arrowheads indicate inferred hydrogen bond acceptors. The crosshatch line indicates stacking between a tyrosine ring and methyl group on thymine. The DNA interface of the domain from I-CreI is identical to that made by native I-CreI; the dmo half-site complex has not been previously visualized.

6), which consists of a strictly conserved acidic residue at the base of each LAGLIDADG helix, three catalytic metals bound between the two active sites by these acidic residues, and extended pockets in each endonuclease domain that accommodate and order solvent molecules around each scissile phosphate. The ordering of water molecules appears to be important for catalysis in I-CreI; protein side chains make no direct contact with catalytic nucleophiles, the scissile phosphates, or the leaving groups during catalysis (Chevalier et al., 2001).

The two active sites of E-DreI contain three Mg^{2+} ions with the central metal being shared by both active sites. These metal ions are coordinated by D21 and D117, each of which is located at the base of a LAGLIDADG interface helix, similar to what had been previously observed for I-CreI bound to target site DNA (Chevalier et al., 2001). During the redesign process we avoided perturbing catalytically important residues such as the active site, metal-coordinating aspartates, and residues at the periphery of the active sites (e.g., Q42, Q144, R148, and K195). However, the cocrystal structure of E-DreI bound to target DNA revealed that W19, which was designed to stack against F151 in the protein interface, made two important interactions with active site residues. These unanticipated contacts included an H bond across the protein interface with Q144 and a loose cation-π interaction with R148 (Figure 4D). While these interactions were not anticipated, they did not impede metal binding or catalysis.

Conclusions

We have used a combination of domain fusion and computational protein interface redesign strategies to generate a new, highly sequence-specific endonuclease, E-DreI, from portions of two existing homing endonucleases. Our interest in this approach was stimulated by the observation that nature has repeatedly used protein domain fusion and interface optimization to generate new homing endonucleases with different target specificities. We reasoned that if this process could be replicated experimentally, then it might be possible to generate many new, highly sequence-specific DNA binding proteins from existing homing endonuclease proteins.

In contrast to the frequently observed immutability of restriction endonucleases when subjected to selection or structure-based redesign for altered specificity, homing endonucleases (and in particular the LAGLIDADG enzyme family) appear to be remarkably tolerant to similar efforts. Such studies include both selection of altered DNA binding residues for novel sequence specificity (Seligman et al., 2002; R.J.M., unpublished data) and recombination and fusion of unrelated enzyme domains (this work). While the basis for this redesign tolerance is not absolutely clear, it is probably a function of both the natural biological function of homing endonucleases as catalysts of lateral DNA transfer events (rather than the highly regulated defense of a bacterial genome in concert with a protective methylase partner) and of the biophysical characteristics of the homing endonuclease-DNA recognition interface. DNA recognition by homing endonucleases consistently utilizes a mechanism whereby a large number of protein side chains make subsaturating hydrogen bond contacts across a long target site, with the majority of contacts located a significant distance from the scissile phosphate groups. Therefore, the relative energetic cost of a single structural mismatch in the DNA interface may be lower than for many nucleic acid binding proteins, and the structure and alignment of the active site catalytic groups may be less intimately coupled to the position and interactions formed by many of the DNA binding residues within that interface.

It should be possible to further diversify the site specificity of newly generated, highly sequence-specific chimeric DNA binding proteins such as E-DreI by directed

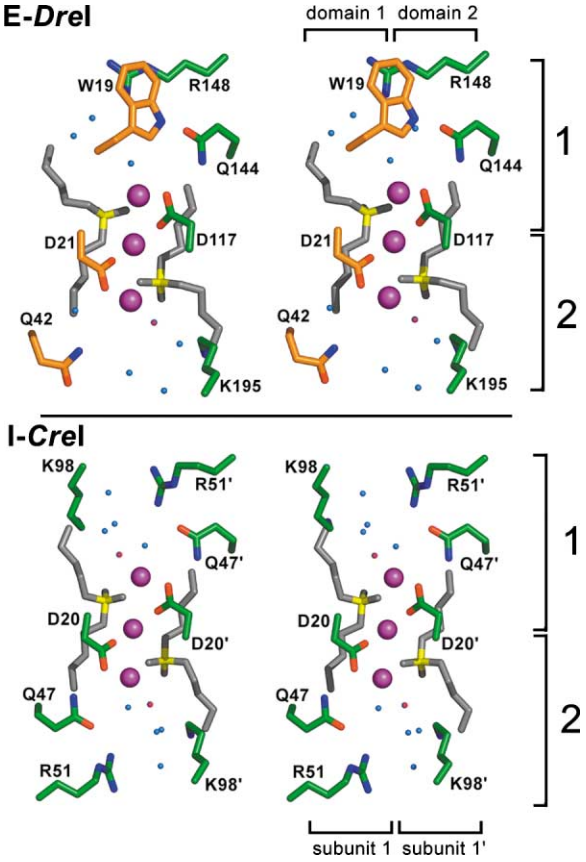

Figure 6. Stereo Views of the E-DreI and I-CreI Active Sites

The substrate DNA in the E-DreI is trapped as an uncleaved substrate complex by a combination of low temperature and low pH during crystal growth. The three catalytic metals are purple, the DNA backbone is gray, the scissile phosphates are yellow, waters are blue, and putative nucleophilic waters are red. Each structure contains two active sites that share a central metal; each active site is chimeric and composed of residues from both protein domains.

protein evolution and further computational redesign strategies. A critical question for future studies is the extent to which computational interface redesign, using strategies similar to those discussed in this article, can be used to successfully redesign protein-DNA interfaces for this enzyme family. Such methods have generally been unsuccessful for other DNA binding protein systems. It is worth noting, however, that there appears to be lower structural and functional interdependence between DNA binding residues in homing endonucleases than for many analogous systems. Additionally, even a partially successful computational redesign effort targeting the protein-DNA interface may greatly facilitate associated attempts at directed evolution of new DNA sequence specificities.

An interesting question to consider with respect to the engineering of homing endonucleases into gene-specific reagents is the generalizability of such efforts as compared to similar efforts to create DNA binding proteins that utilize zinc finger domains (see Introduction). Those systems possess the obvious advantages of: (1) being easily permutable (by altering the order and presence of the highly modular, individual fingers) and (2) of permitting the straightforward design of constructs that can recognize many long sites (because zinc finger domains have been identified that recognize many of the 64 possible nucleotide codons). However, it is not clear at this time whether such constructs will prove to be consistently more generalizable and specific to any given DNA sequence than engineered homing endonucleases. This is because the use of flexible linkers in such constructs causes nonspecific catalytic activity, because zinc fingers display promiscuity (wobble) in base recognition that might lower their overall specificity and because zinc finger constructs have only been described that are specific for approximately half of the 64 possible codons of the genetic code (Beerli and Barbas, 2002). Thus, it seems reasonable that both types of DNA binding motifs should be further developed with the intent of creating gene-specific reagents for a variety of purposes.

Finally, it should be possible to broaden the range of biochemical functions possessed by new, highly site-specific DNA binding proteins using homing endonucleases as prototypes. For example, I-CreI, one of E-DreI fusion partners, and I-PpoI, a member of the His-Cys box homing endonuclease family, can be converted with single amino acid substitutions into catalytically inactive DNA binding proteins and/or site-specific repressors (Galburt et al., 1999; Seligman et al., 1997). The covalent attachment of small molecular reagents to engineered surface cysteine residues and the genetic tethering of independent protein domains to the enzyme termini could further increase the biological repertoire of such designed molecules. Such highly sequence-specific reagents should have broad utility in research, diagnosis, and perhaps therapy, where the ability to target specific biochemical activities to individual genes would be advantageous.

Experimental Procedures

Initial Construction of E-DreI

E-DreI was initially modeled by superimposing the LAGLIDADG helix (residues 9–21) from the N-terminal domain of I-DmoI (RCSB1B24) onto the same helix (residues 8–20) of the first subunit of the I-CreI structure (1G9Z). Based on this model, we created a 260 amino acid E-DreI coding sequence consisting of 101 codons (Met_{start} to Phe_{101}) from the N-terminal domain of I-DmoI fused to the last 156 codons of I-CreI (Glu_9 to Pro_{163}), separated by a three amino acid linker that mimicked the native I-DmoI linker in length (consisting of residues -NGN- to encourage β-turn formation [Hutchinson and Thornton, 1994]) within the parental pI-CreI vector (Thompson et al., 1992). The graphical model revealed potential side chain clashes across the new domain interface involving residues L47, H51, L55, L108, K193, and L194. Eight initial constructs, in which these side chains were serially mutated to alanine, were created and overexpressed in BL21[DE3] *E. coli* cells. These constructs were insoluble (data not shown).

Computational Interface Redesign

The graphical model created above was subjected to a computational interface redesign procedure which models amino acid side chains in an all-atom representation (all heavy atoms and polar hydrogens) onto a fixed polypeptide backbone template. Using this template with original I-DmoI and I-CreI side chains, all positions making significant side chain-mediated interactions in the interface were identified. For each of these positions, the program created rotamers for 19 amino acid types (excluding cysteine)—on average about 500 rotamers per sequence position. The rotamers were taken

from a backbone-dependent library (Dunbrack and Cohen, 1997) with additional rotamers added by rotations around the χ_1 and χ_2 angles by 5°–20° and extra rotamers for χ_3 and χ_4 angles (Dahiyat and Mayo, 1997). All amino acid side chains not considered in the design procedure were left in their native conformations obtained from the parent I-DmoI and I-CreI crystal structures, in particular, the active site residues D20 and D117 and residues in the active site vicinity (Q44, Q144, R148, and K195). Energies were computed for each rotamer with the constant part of the protein complex (the template backbone and all side chains not subjected to design), and for all pairwise rotamer-rotamer combinations using the free energy function described below. Selection of the best amino acid in its best conformation was performed using a Monte Carlo-simulated annealing procedure as described previously (Kuhlman and Baker, 2000). As the Monte Carlo procedure is not guaranteed to find the global free energy minimum, different independent Monte Carlo runs were performed and yielded slightly different sequences with similar energies. All combinations of the most frequently obtained amino acids at each position were then exhaustively enumerated and scored. The 16 top scoring sequences with the lowest interface free energies were then chosen for further analysis.

Our free energy function consisted of the attractive and repulsive parts of a 6-12 Lennard-Jones potential, an explicit hydrogen bonding potential, an implicit solvation model (Lazaridis and Karplus, 1999), and statistical terms representing the backbone-dependent internal free energies of amino acid rotamers as described previously (Kuhlman and Baker, 2000). The relative weights of the different contributions were determined as described elsewhere (Kortemme and Baker, 2002).

The relative contribution of side chains to the interface free energy was evaluated by computational alanine scanning as described elsewhere (Kortemme and Baker, 2002). In brief, each amino acid in the interface was separately substituted in silico by alanine, and the effect of the replacement was computed both for the whole protein (ΔG_{dre}) as well as the two domains separately (ΔG_{cre_domain} and ΔG_{dmo_domain}), to yield the change in binding energy $\Delta\Delta G_{int}$ ($\Delta\Delta G_{int} = \Delta G_{dre} - \Delta G_{cre\_domain} - \Delta G_{dmo\_domain}$).

In Vivo lacZ-Based Solubility Assay

The lacZ complementation-based in vivo solubility screen was essentially as previously described (Wigley et al., 2001). E-DreI variants tagged at their C termini with lacα were overexpressed in *E. coli* containing lacω. Soluble constructs were able to complement lacω and form blue colonies in the presence of Xgal; insoluble constructs failed to complement lacω, and colonies remained white. Twenty-one different E-DreI constructs (8 from initial interface truncations and 12 from interface redesign) were assayed for blue colony formation after transformation into XL1-Blue *E. coli* cells (Stratagene) and spread on LB plates supplemented with 0.1 mM IPTG. After 17 hr, colonies were lifted on nylon membranes (Magnalift NLOHY08250) and lysed by flash freezing in liquid nitrogen. Membranes were developed in 100 mM sodium phosphate-buffered solution (pH 7) containing 1 mg/ml X-gal for 30 min, then dried. Relative color development was assessed visually.

Protein Expression and Purification

Three constructs that yielded the bluest colonies were subcloned back into the pI-CreI parental vector lacking the lacα tag and induced with 0.5 mM IPTG in BL21[DE3] *E. coli* cells overnight at 15°C. Cells were harvested by centrifugation and lysed by sonication in 50 mM Tris (pH 8.0), 100 mM NaCl, and 1 mM CaCl2. Cell debris was removed by centrifugation at 40,000 g for 45 min at 4°C. The supernatant was forced through a 0.2 μm syringe filter, and applied to a heparin column (Pharmacia). E-DreI was eluted with an increasing salt gradient. Collected fractions from the single peak were diluted with an equal volume solution of 50 mM Tris (pH 8.0), and loaded back over the heparin column. After the second elution, E-DreI was >95% pure (SDS-PAGE); it was then dialyzed overnight into 125 mM NaCl, 50 mM Tris (pH 8.0), 1 mM CaCl2, and 5% glycerol, concentrated to ~4 mg/ml by centrifugation, (Centriprep, Millipore), and stored at −80°C.

Biochemical Analysis

The digestion of labeled oligonucleotide substrates was performed in I-CreI buffer (20 mM Tris 9.0, 10 mM $MgCl_2$, 1 mM DTT, 50 ug/ml BSA) at 65°C (I-CreI digests), or in New England Biolabs Buffer 4 containing 50 ug/ml BSA at 37°C (I-DmoI digests) or 65°C (E-DreI digests). 5′ end-labeled primers and template DNA (a dre3 target site cloned into pBSIISK$^+$), were used to generate both the sequencing ladders and the dsDNA substrates for E-DreI cleavage. Denatured E-DreI-digested substrates (labeled as "X" in Figure 2C) were run alongside their corresponding sequencing reactions to map cleavage positions. Gels were imaged on a Storm Phosphorimager 840 (Molecular Dynamics, Sunnyvale, CA).

Crystallization and Data Collection

Crystals were grown at room temperature in hanging drops with 1 μl reservoir solution (2.1–2.4 M $(NH_4)_2SO_4$, 100 mM MES [pH 6.5], 10 mM $MgCl_2$), and 1 μl protein mixed with 0.1 μl solution containing blunt-ended DNA corresponding to the dre4 target site (15× molar excess to protein; oligos used were 5′-CCAAACTGTCTCAAGT TCCGGCG-3′ and 5′-CGCCGGAACTTGAGACAGTTTGG-3′). Crystals (200 μm × 200 μm × 50 μm; space group $P3_1$, a = b = 131.76 Å, c = 120.91 Å) grew in 4–6 weeks at room temperature. They were transferred up to reservoir solution containing 20% glycerol in 5% steps and flash frozen in liquid nitrogen. Diffraction data were recorded to 2.4 Å resolution at the Advanced Light Source beamline 5.0.1 (305673 measured reflections, 91688 unique reflections, 99% complete; R_{merge} = 8.1). Intensities were integrated and scaled using DENZO and SCALEPACK (Otwinowski and Minor, 1997).

Structure Determination and Refinement

The structure was solved via molecular replacement using EPMR (Kissinger and Gehlhaar, 1997) with an I-CreI subunit bound to its DNA half-site as an initial search model. Four E-DreI/DNA complexes were found in each asymmetric unit, two of which lacked well-defined density due to low occupancy and/or disorder; these were modeled solely as polyalanine/DNA chains. The structure was modeled in XtalView (McRee, 1999) and refined using CNS (Brunger et al., 1998) with 5% of the data set aside for crossvalidation. The final refinement statistics were R_{work}/R_{free} = 0.236/0.256; rmsd bond distance = 0.005 Å; rmsd bond angles 1.07°. Per PROCHECK (Laskowski et al., 1993), 91.6% of the backbone dihedral angles are in the most favorable region of the Ramachandron plot; 8.4% are within the favorable region. Average B factors for protein, DNA, Mg^{2+}, and solvent in the two well-ordered complexes are 35.1 Å^2, 38.0 Å^2, 26.4 Å^2, and 48.6 Å^2, respectively. Average B factors for protein and DNA in the disorderd polyalanine/DNA complexes are 100.3 Å^2 and 134.7 Å^2, respectively.

Acknowledgments

We thank Alexis Kaushansky for assistance with molecular biology experiments and Michael Moser for insightful suggestions and discussion early in this work. Marlene Belfort and Victoria Derbyshire provided an initial clone for I-DmoI and purified protein for biochemical comparisons. Gerry McDermott, Keith Henderson, and Thomas Earnest at the ALS beamline 5.0.2 provided help during X-ray data collection. The work described here was funded by grants from the NIH to B.S.C. (GM49857), R.J.M. (CA88942), and D.B. (GM59224). B.S.C. was supported by an Interdisciplinary Training Grant in Cancer Research predoctoral fellowship (T32 CA80416), and T.K. was supported by the Human Frontier Science Program Organization and EMBO.

Received: July 8, 2001
Revised: September 11, 2002

References

Argast, G.M., Stephens, K.M., Emond, M.J., and Monnat, R.J.J. (1998). I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential *in vitro* enrichment. J. Mol. Biol. *280*, 345–353.

Beerli, R.R., and Barbas, C.F. (2002). Engineering polydactyl zinc-finger transcription factors. Nat. Biotechnol. *20*, 135–141.

Belfort, M., and Perlman, P.S. (1995). Mechanisms of intron mobility. J. Biol. Chem. *270*, 30237–30240.

Belfort, M., and Roberts, R.J. (1997). Homing endonucleases—keeping the house in order. Nucleic Acids Res. *25*, 3379–3388.

Belfort, M., Reaban, M.E., Coetzee, T., and Dalgaard, J.Z. (1995). Prokaryotic introns and inteins: a panoply of form and function. J. Bacteriol. *177*, 3897–3903.

Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.G., and Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol. Cell. Biol. *21*, 289–297.

Bolon, D.N., and Mayo, S.L. (2001). Enzyme-like proteins by computational design. Proc. Natl. Acad. Sci. USA *98*, 14274–14279.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Bryk, M., Belisle, M., Mueller, J.E., and Belfort, M. (1995). Selection of a remote cleavage site by I-*Tev*I, the *td* intron-encoded endonuclease. J. Mol. Biol. *247*, 197–210.

Carroll, S.B., Genier, J.K., and Weatherbee, S.D. (2001). From DNA to Diversity (Malden, MA: Blackwell Science Inc.).

Chevalier, B.S., and Stoddard, B.L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res. *29*, 3757–3774.

Chevalier, B., Monnatt, R., and Stoddard, B.L. (2001). The LAGLIDADG homing endonuclease I-CreI shares three divalent cations between two active sites. Nat. Struct. Biol. *8*, 312–316.

Conte, L.L., Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition sites. J. Mol. Biol. *285*, 2177–2198.

Dahiyat, B.I., and Mayo, S.L. (1997). De novo protein design: fully automated sequence selection. Science *278*, 80–81.

Dalgaard, J.Z., Klar, A.J., Moser, M.J., Holley, W.R., Chatterjee, A., and Mian, I.S. (1997). Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. Nucleic Acids Res. *25*, 4626–4638.

Duan, X., Gimble, F.S., and Quiocho, F.A. (1997). Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. Cell *89*, 555–564.

Dunbrack, R., Jr., and Cohen, F. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci. *6*, 1661–1681.

Farinas, E.T., Bulter, T., and Arnold, F.H. (2001). Directed enzyme evolution. Curr. Opin. Biotechnol. *12*, 545–551.

Galburt, E.A., Chevalier, B., Tang, W., Jurica, M.S., Flick, K.E., Monnat, R.J., and Stoddard, B.L. (1999). A novel endonuclease mechanism directly visualized for I-PpoI. Nat. Struct. Biol. *6*, 1096–1099.

Gruen, M., Chang, K., Serbanescu, I., and Liu, D.R. (2002). An in vivo selection system for homing endonuclease activity. Nucleic Acids Res. *30*, 29–34.

Guo, H., Karberg, M., Long, M., Jones, J.P., Sullenger, B., and Lambowitz, A. (2000). Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. Science *289*, 452–457.

Halford, S.E., Johnson, N.P., and Grinsted, J. (1980). The EcoRI restriction endonuclease with bacteriophage λ DNA. Biochem. J. *191*, 581–592.

Harbury, P.B., Plecs, J.J., Tidor, B., Alber, T., and Kim, P.S. (1998). High resolution protein design with backbone freedom. Science *282*, 1462–1467.

Heath, P.J., Stephens, K.M., Monnat, R.J., and Stoddard, B.L. (1997). The structure of I-CreI, a group I intron-encoded homing endonuclease. Nat. Struct. Biol. *4*, 468–476.

Hutchinson, E.G., and Thornton, J.M. (1994). Revised set of potentials for β-turn formation in proteins. Protein Sci. *3*, 2207–2216.

Ichiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K., and Morikawa, K. (2000). Crystal structure of an archaeal intein-encoded homing endonuclease PI-PfuI. J. Mol. Biol. *300*, 889–901.

Jacob, F. (1977). Evolution and tinkering. Science *196*, 1161–1166.

Jurica, M.S., Monnat, R.J., and Stoddard, B.L. (1998). DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. Mol. Cell *2*, 469–476.

Kissinger, C.R., and Gehlhaar, D.K. (1997). EPMR: A Program for Crystallographic Molecular Replacement by Evolutionary Search (La Jolla, CA: Agouron Pharmaceuticals).

Kortemme, T., and Baker, D. (2002). A simple physical model for binding energy hotspots in protein-protein complexes. Proc. Natl. Acad. Sci. USA, in press.

Kuhlman, B., and Baker, D. (2000). Native protein sequences are close to optimal for their structures. Proc. Natl. Acad. Sci. USA *97*, 10383–10388.

Lanio, T., Jeltsch, A., and Pingoud, A. (2000). On the possibilities and limitations of rational protein design to expand the specificity of restriction endonucleases. Protein Eng. *13*, 275–281.

Laskowski, R.J., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–290.

Lazaridis, T., and Karplus, M. (1999). Effective energy function for proteins in solution. Proteins *35*, 133–152.

McRee, D.E. (1999). A versatile program for manipulating atomic coordinates and electron density. J. Struct. Biol. *125*, 156–165.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Pokala, N., and Handel, T.M. (2001). Protein design—where we were, where we are, where we're going. J. Struct. Biol. *134*, 269–281.

Schmidt-Dannert, C. (2001). Directed evolution of single proteins, metabolic pathways and viruses. Biochemistry *40*, 13125–13135.

Seligman, L.M., Stephens, K.M., Savage, J.H., and Monnat, R.J. (1997). Genetic analysis of the *Chlamydomonas reinhardtii* I-CreI mobile intron homing system in *Escherichia coli*. Genetics *147*, 1653–1664.

Seligman, L.M., Chisholm, K.M., Chevalier, B.S., Chadsey, M.S., Edwards, S.T., Savage, J.H., and Veillet, A.L. (2002). Mutations altering the cleavage specificity of a homing endonuclease. Nucleic Acids Res. *30*, 3870–3879.

Silva, G.H., Dalgaard, J.Z., Belfort, M., and Roey, P.V. (1999). Crystal structure of the thermostable archaeal intron-encoded endonuclease I-*Dmo*I. J. Mol. Biol. *286*, 1123–1136.

Smith, J., Berg, J.M., and Chandrasegaran, S. (1999). A detailed study of the substrate specificity of a chimeric restriction enzyme. Nucleic Acids Res. *27*, 274–281.

Smith, J., Bibikova, M., Whitby, F.G., Reddy, A.R., Chandrasegaran, S., and Carroll, D. (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic Acids Res. *28*, 3361–3369.

Thompson, A.J., Yuan, X., Kudlicki, W., and Herrin, D.L. (1992). Cleavage and recognition pattern of a double-strand-specific endonuclease (I-CreI) encoded by the chloroplast 23S rRNA intron of *Chlamydomonas reinhardtii*. Gene *119*, 247–251.

Vriend, G. (1990). WHATIF: a molecular modeling and drug design program. J. Mol. Graph. *8*, 52–56.

Wang, J., Kim, H.-H., Yuan, X., and Herrin, D.L. (1997). Purification, biochemical characterization and protein-DNA interactions of the I-CreI endonuclease produced in *Escherichia coli*. Nucleic Acids Res. *25*, 3767–3776.

Wigley, W.C., Stidham, R.D., Smith, N.M., Hunt, J.F., and Thomas, P.J. (2001). Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein. Nat. Biotechnol. *19*, 131–136.

Accession Numbers

The E-DreI sequence has been deposited in the Protein Data Bank under ID code 1MOW.