doi:10.1016/j.jmb.2004.07.031 J. Mol. Biol. (2004) 342, 31–41

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Isolation and Characterization of New Homing Endonuclease Specificities at Individual Target Site Positions

Django Sussman[1], Meg Chadsey[2], Steve Fauce[3], Alex Engel[3] Anna Bruett[3], Ray Monnat Jr[2], Barry L. Stoddard[1]* and Lenny M. Seligman[3]

[1]*Division of Basic Sciences Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N. A3-025 Seattle, WA 98109, USA*

[2]*Department of Pathology and Genome Sciences, University of Washington, Seattle, WA 98195 USA*

[3]*Department of Biology and Program in Molecular Biology Pomona College, 609 North College Avenue, Claremont CA 91711, USA*

**Corresponding author*

Homing endonucleases are highly specific DNA endonucleases, encoded within mobile introns or inteins, that induce targeted recombination, double-strand repair and gene conversion of their cognate target sites. Due to their biological function and high level of target specificity, these enzymes are under intense investigation as tools for gene targeting. These studies require that naturally occurring enzymes be redesigned to recognize novel target sites. Here, we report studies in which the homodimeric LAGLIDADG homing endonuclease I-*Cre*I is altered at individual side-chains corresponding to contact points to distinct base-pairs in its target site. The resulting enzyme constructs drive specific elimination of selected DNA targets *in vivo* and display shifted specificities of DNA binding and cleavage *in vitro*. Crystal structures of two of these constructs demonstrate that substitution of individual side-chain/DNA contact patterns can occur with almost no structural deformation or rearrangement of the surrounding complex, facilitating an isolated, modular redesign strategy for homing endonuclease activity and specificity.

Keywords: engineering; endonuclease; specificity; crystal structure

Introduction

Homing endonucleases are catalysts of gene-specific transpositions in which mobile intervening sequences are duplicated into homologous host alleles.[1–3] These enzymes are being studied as potential tools for the creation of novel gene-specific reagents. Such uniquely tailored proteins could be useful for many applications, including targeted allele replacement (i.e. gene therapy), detection of rare genetic sequences and genome mapping. These applications require that homing endonucleases be subjected to selection and re-engineering to recognize and cleave novel DNA target sites with affinities and specificities that rival their natural activities.

Homing endonucleases are found in all biological super-kingdoms. On the basis of primary sequence homology, four homing enzyme families have been identified: the LAGLIDADG, GIY-YIG, HNH, and His-Cys Box endonucleases.[1,2] The latter two of these families (HNH and His-Cys Box) are diverged from a common ancestor and share similar active sites and cleavage mechanisms.[4,5] The largest family, LAGLIDADG, contains several hundred identified members, many of which have been shown to be functional endonucleases.[6,7] The conserved LAGLIDADG sequence motif forms the core of a structural interface between the endonuclease domains or subunits and contributes conserved acidic residues to the enzyme active sites.[8,9] All LAGLIDADG nucleases recognize long DNA target sites (19–30 base-pairs), cleaving their target sites to generate cohesive four base, 3′ overhangs.[10–13] The enzymes typically bind their

Present addresses: S. Fauce, Department of Pathology and Laboratory Medicine, UCLA Medical School, 10833 Le Conte Ave, Los Angeles, CA 90095-7300, USA; A. Engel, Department of Biochemistry and Biophysics, UCSF, 600 16th Street N316, San Francisco, CA 94143, USA.

Abbreviations used: SC-GSRs, single chain gene-specific reagents; BSA, bovine serum albumin.

E-mail address of the corresponding author: bstoddar@fhcrc.org

physiological target sites with dissociation constants ranging from 0.5 nM to 5 nM.

Endonucleases that contain a single LAGLIDADG motif per polypeptide chain form homodimers that recognize palindromic DNA target sites and their pseudopalindromic variants. The enzymes with two motifs form pseudo-symmetric monomers capable of recognizing asymmetric DNA target sites.[14] Several independent studies have demonstrated that domains from unrelated LAGLIDADG enzymes can be structurally fused to create fully active, chimaeric homing endonucleases that recognize corresponding chimaeric target sites.[15–17] This technology requires extensive repacking of the domain interface, but allows the creation of new protein scaffolds with novel specificities, in addition to those encoded by naturally occurring enzymes.

Based on their modular, compact structures, their straightforward mechanisms of DNA recognition, and their tolerance of domain exchanges, the LAGLIDADG homing endonucleases are an attractive protein family for the creation of single chain gene-specific reagents (SC-GSRs). Beyond the wholesale domain fusions noted above, the complete redesign of homing endonucleases for recognition of a desired DNA target must involve the additional selection and combination of enzyme mutations that alter specificity towards individual base-pairs. However, the resulting behavior of such altered enzyme constructs, in terms of their discrimination between cognate and non-cognate target sites and their structural response to engineering, is unknown. Here, we report the results of such studies for the enzyme I-*Cre*I.

Of the LAGLIDADG homing endonucleases, the I-*Cre*I enzyme (along with the other founding members of the family, I-*Sce*I and PI-*Sce*I) has been the best characterized in terms of recognition specificity and flexibility. The native DNA target site (or "homing" site) for the enzyme is a pseudopalindromic 22 base-pair site in which symmetry is broken at base-pairs ±1, 2, 6 and 7 between the target half sites (Figure 1).[11,18] Palindromic variants of this site, consisting of inverted repeats of the left or right half-sites from the native target, are recognized and cleaved with affinities and activities similar to those of the physiological homing site. *In vitro* site selection experiments, in which variant DNA sequences that can be cleaved by the wild-type enzyme were recovered and sequenced, indicate that several nucleotide positions in the site may be mutated to at least one alternative base-pair without loss of binding or cleavage sensitivity.[19] The positions of polymorphisms that are tolerated most readily by the enzyme generally correspond to base-pairs that are not conserved palindromically between native

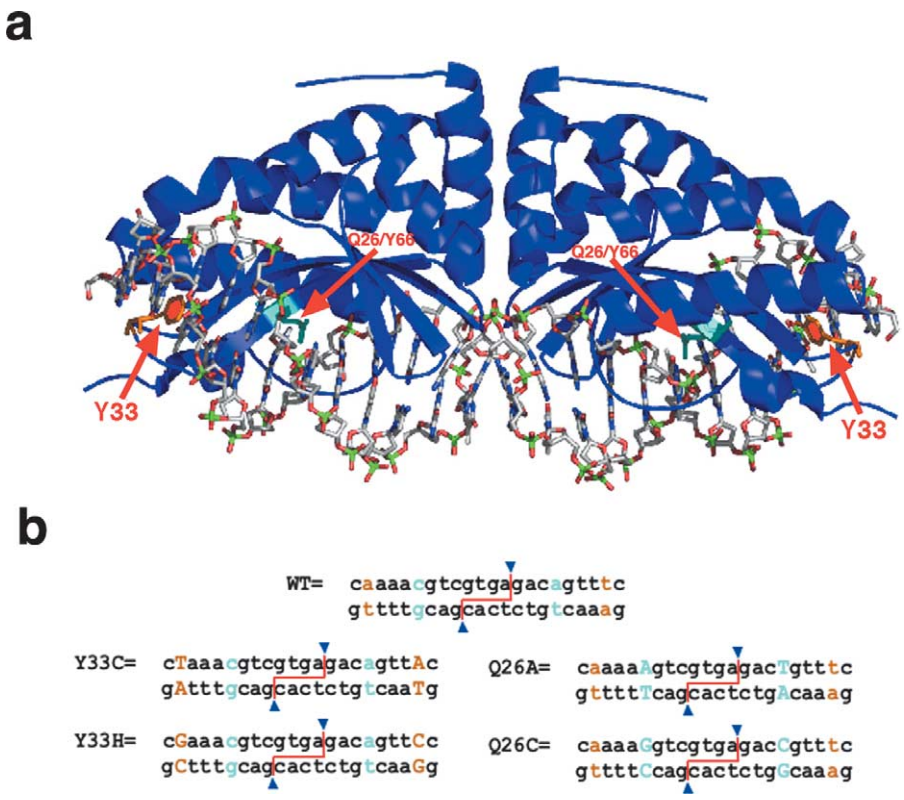

Figure 1. Structure of I-*Cre*I and DNA target sites used in this study. (a) Structure of wild-type I-*Cre*I bound to DNA. The positions of residues that are targeted for selection are indicated in the homodimer by red labels and arrows. (b) Wild-type and mutant enzyme DNA-binding sites. Base-pairs ±6 and ±10, that interact with Q26/Y66 and Y33, respectively, are colored to correspond to the scheme in the top panel. Points of cleavage are noted with blue triangles and red cleavage patterns.

half-sites. These, in turn, generally correspond to base-pairs that display fewer direct contacts with enzyme side-chains in the protein–DNA complex.[20] The information content (specificity of recognition) of individual base-pairs in the I-*Cre*I target site have been calculated and correlated with their contacts to individual protein side-chains and solvent molecules,[21] allowing us to begin studying the effect of modifying those contacts on enzymatic activity and structure. Similar studies of site recognition have been reported for PI-*Sce*I;[22] the trends of that study are quite similar to those described for I-*Cre*I.

Here, we report the identification, *in vivo* and *in vitro* activities, and DNA-bound cocrystal structures of several selected variants I-*Cre*I bound to their new targets. These altered enzyme constructs recognize DNA targets that have been altered at base-pairs that correspond to both high and moderate levels of specificity. The results demonstrate that, unlike restriction enzymes[23] and many other DNA-binding proteins, homing endonucleases recognize DNA through a moderately independent set of protein–DNA contacts that can be isolated and optimized individually. Selection of enzymes with altered specificities, and with minimal effects on protein structure and nearby DNA contacts, is possible.

Results

Isolation of enzyme mutants and genetic phenotypes

A strategy for isolating I-*Cre*I derivatives with increased affinities for altered target sites has been described.[24] Endonuclease mutants with single amino acid substitutions at positions predicted to make base-specific DNA contacts were assayed against appropriate DNA target site mutants in an *Escherichia coli*-based system. In this system, cleavage of F′ borne target sites results in cells being converted from lac^+ to lac^-. Three levels of activity can thus be identified on medium containing the β-galactosidase indicator Xgal: white colonies indicate efficient site cleavage, blue colonies indicate no cleavage, and sectored colonies indicate an intermediate level of cleavage activity. Out of screens directed towards target sites altered symmetrically at positions ±10, endonucleases with Y33C and Y33H substitutions were each identified as having increased cleavage activities toward specific DNA targets (Figure 1).[24] The structures of these novel protein–DNA complexes are described below.

Within the I-*Cre*I DNA interface at positions +6 *versus* −6, and the other positions that differ between the target half-sites, base-specific contacts occur between the same amino acid residues and different nucleotide bases. For example, Gln26 in one I-*Cre*I subunit is within hydrogen bonding distance of an adenine nucleotide at base-pair +6, while Gln26 in the other subunit is within hydrogen bonding distance of a guanine nucleotide at base-pair −6.[25] Each interaction contributes to DNA recognition of the asymmetric physiological target site.[21] To test the relative importance of these interactions, the four symmetric target site mutants altered at bases ±6 were introduced onto F′ lac for *in vivo* assays as described.[24] When assayed against wild-type I-*Cre*I, each of the two mutant sites that share a base with the native site (a C:G base-pair at both +6 and −6, and T:A base-pair at the same positions) gave rise to white colonies, indicating efficient cleavage. However, each of the two completely novel sites (an A:T base-pair or a G:C base-pair at both positions) gave rise to blue colonies, indicating no site cleavage.

Having identified DNA target sites symmetrically altered at bases ±6 that are resistant to

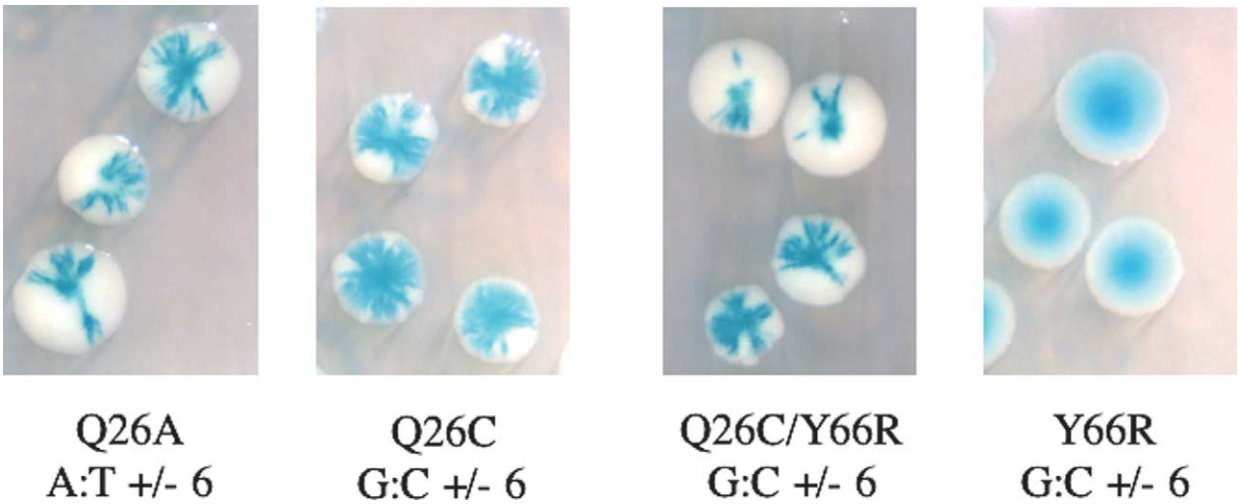

Figure 2. *In vivo* activity of cognate and non-cognate pairs directed against base-pair 6. Colony sectoring and color formation for enzyme mutants selected against novel target sites. A non-cognate pair (Y66R *versus* G:C base-pair at position ±6) is shown on the far right; various selections for mutants of altered specificity at the same base-pair are shown in the first three panels.

cleavage by wild-type I-*Cre*I, we next screened for I-*Cre*I derivatives with increased activities against these sites. Endonuclease mutants with each of the 19 single amino acid substitutions altering residue 26 were generated and assayed against these sites. A single endonuclease mutant with increased activity against each site was identified: a Q26A substitution resulted in increased activity against the A:T at ±6 sites, while a Q26C substitution resulted in increased activity against the G:C at ±6 sites (Figure 1). In each case, *in vivo* assays resulted in sectored colonies (Figure 2), indicating a lower cleavage activity than wild-type I-*Cre*I displays toward its cognate site. Neither mutant displayed cleavage of the wild-type target site *in vivo*, indicating that each single amino acid substitution had caused a decreased recognition of the native site along with the increased recognition of its cognate mutant site.

The intermediate phenotypes displayed by the Q26 substitutions implied that it might be possible to optimize these interactions with further mutation(s). In the wild-type complex, a tyrosine at position 66 is predicted to participate in additional, water-mediated hydrogen bonding with the bases present at positions ±6 and ±7.[21] We reasoned that alterations at this position, in combination with those at position 26, might yield increased affinities for particular mutant sites. Eighteen single amino acid substitution mutants altered at position 66 were isolated (all but Y66E), and examined for activity. Of these, 17 (all but Y66P) retained full or partial endonuclease activity toward the wild-type target site *in vivo*. All 18 substitutions were subsequently combined with the Q26C substitution, and assayed against the G:C ±6 target sites *in vivo*. One double mutant, Q26C Y66R, resulted in an increased cleavage activity towards these sites, as evidenced by an increased fraction of white cells within sectored colonies (Figure 2). When the Y66R single mutant alone was assayed against the G:C ±6 target sites *in vivo*, no cleavage was observed (Figure 2). A Q26A Y66R double mutant was made and assayed against the A:T ±6 target site *in vivo*. Here, the double mutant displayed decreased affinity relative to the single mutant, indicating that the enhancement provided by the Y66R substitution is specific for a G:C base-pair at positions ±6.

Binding affinities

The affinity of the wild-type enzyme was determined against all the sites described here (five in all; the original site and four variants). The affinities of the individual mutant enzyme constructs were each determined against the wild-type I-*Cre*I site and their appropriate, individual novel cognate site. All of the constructs yielded clean patterns of gel retardation with K_d values ranging from 0.4 nM to 120 nM. All measured K_d values, with standard deviations from triplicate measurements, are shown in Table 1.

The overall result of the selection and characterization of enzyme point mutants against individual target site variants is both a shift and a broadening in binding specificity. Each mutant displays a higher dissociation constant (lower affinity) against the original wild-type target site than does the wild-type enzyme, and each mutant displays a lower dissociation constant (higher affinity) against its novel target than does the wild-type enzyme. The ratios of K_d values of the mutant enzymes against its non-cognate *versus* cognate sites are 1/3–1/60 of the ratios of the wild-type enzyme K_d values against the same sites.

Overall, mutations of the I-*Cre*I target site at base-pairs ±6 or ±10 causes a decrease in the wild-type enzyme affinity. The ratio of K_d values for the wild-type enzyme against the four separate non-cognate

Table 1. Substrate binding affinities

	Enzyme constructs					
	WT	Q26A	Q26C	Q26C/Y66R	Y33C	Y33H
Wild-type	0.4±0.6	2.4±0.5	1.5±0.2	4.4±0.6	25.0±9.0	2.0±0.06
Target sites						
A:T±6	3.0±0.2	1.0±0.08				
G:C±6	120±10		0.3±0.08	0.6±0.3		
T:A±10	63±5.0				5.9±0.5	
G:C±10	4.8±0.7					1.2±0.04
Non-cognate/ cognate ratio	7.5–300	2.4	5.0	7.3	4.2	1.7
Specificity shift[a]		18	1500	2200	670	20
Specificity broadening[b]		0.3	0.03	0.02	0.03	0.14

K_d values in nM enzyme concentrations.

[a] Product of wild-type and mutant non-cognate/cognate binding affinity ratios for a given site. A shift from a tenfold cognate/non-cognate preference for wild-type enzyme to an altered, tenfold cognate/non-cognate preference for an enzyme mutant gives a 100-fold shift in site preference.

[b] Ratio of mutant cognate/non-cognate binding affinity to wild-type cognate/non-cognate binding affinity. If wild-type and mutant display the same overall preference (example: each tenfold for cognate over non-cognate), then the specificity broadening is 1.0, implying no overall change in the breadth of the specificity profile for the mutant enzyme. Values less than 1.0 correspond to broader specificity, values greater than 1.0 correspond to tighter specificity.

sites, relative to the original wild-type site, range from 7.5 to 300. This range of values represents the binding preference of the wild-type enzyme for its original site *versus* the altered targets.

All five of the enzyme mutants recovered and analyzed from the selection experiments display decreased affinities against the wild-type target site, with K_d values elevated fourfold to 62-fold. All of these mutants bind more tightly to their individual, novel cognate sites than does the wild-type enzyme, with individual K_d values ranging from 0.3 nM to 5.9 nM. The mutant enzymes with the most improved binding to a novel site relative to the wild-type enzyme are Q26C and Q26C/Y66R, which binds its site (G:C at positions ± 6) with a K_d of 0.3 nM to 0.6 nM (*versus* 120 nM for the wild-type enzyme).

Whereas the wild-type enzyme displays a preference or specificity of binding of 75-fold to 300-fold when comparing wild-type *versus* non-cognate targets, the mutants display shifted but reduced preferences, with ratios of non-cognate *versus* cognate K_d values of 1.7–7.3 nM. Pairwise comparisons of the non-cognate to cognate preference of each mutant enzyme to the wild-type enzyme indicates that the overall binding specificities of all the mutants are reduced (Table 1, final row).

The extent of the shift in binding specificity of individual mutants corresponds to the position of the corresponding base-pair change, and it is a function of both the position and the identity of each individual base-pair at those positions, and the initial magnitude of binding preference for the wild-type enzyme. For example, a site containing an A:T base-pair at position ± 6 causes a small, 7.5-fold increase in K_d of the wild-type enzyme; this preference is nearly matched by a single point mutation (Q26A) at a residue in direct contact with that base-pair. In contrast, incorporation of a G:C at that same base-pair causes a 300-fold increase in the wild-type enzyme's K_d. The best individual mutant combination at this position (Q26C/Y66R) displays only a 7.3-fold difference in cognate/non-cognate K_d values, corresponding to a reduction in preference to 1/40 the wild-type value for that site polymorphism. A similar result is evident for two separate site mutations at position ± 10 (Table 1).

Cleavage efficiencies

The relative cleavage efficiencies of the wild-type and mutant enzymes against their corresponding cognate and non-cognate sites were determined as described here and elsewhere.[24] In these experiments, both sites are present in a single digest reaction, and the extent of cleavage under standardized digest conditions of each site across a broad range of enzyme concentrations is compared (Figure 3b and Table 2). The overall result of this analysis is that the enzyme mutants display similar kinetics of substrate cleavage, with shifts and broadening in substrate preferences similar to those described for binding affinities. However, the sites that yield significant reduction of cleavage preference for their corresponding selected enzyme mutants are different from the sites that yield the most significant reductions in binding specificities described above.

Each mutant cleaves its new cognate site more efficiently than the original site. For the two site variants at base-pairs ± 6, the mutants (Q26A, Q26C and Q26C/Y66R) display reduced cleavage preferences. In contrast, the site variants at base-pairs ± 10 yield single enzyme point mutants (Y33C and Y33H) that display cleavage preferences that are reversed, but similar in magnitude to those of the wild-type enzyme (Figure 4).

Structural consequences of individual contact reselection

Crystallization experiments were attempted with each reselected enzyme construct against its corresponding cognate target site. No suitable specimen of Q26/Y66 mutants in complex with target site variants at base-pairs ± 6 was recovered, despite extensive crystallization trials and seeding attempts. In contrast, diffracting single crystals were obtained for both Y33C and Y33H in complex with their corresponding target sites variants containing base-pair substitutions at positions ± 10.

In the wild-type complex, the phenolic oxygen atom of Y33 forms bifurcated hydrogen bonds to the extracyclic amino nitrogen and the N7 nitrogen atoms of the adenosine base at positions ± 10 (Figure 1). A neighboring asparagine residue (N30) is in close proximity to Y33 and makes contacts to adjacent DNA bases, including the thymidine base-paired partner of adenine 10 and neighboring thymidine 9. The structure of Cre-Y33C bound to its selected cognate target site (in which a T:A base-pair at positions ± 10 is inverted into an A:T base-pair) was determined at 2.5 Å resolution and compared to the wild-type structure. The hydrogen bonds between Y33 and adenosine 10 are replaced by a single 3.8 Å van der Waals contact between the sulfur atom of C33 and the C5 methyl group of the newly substituted thymidine base (Figure 1). The DNA sequence change cannot be accommodated by the wild-type enzyme without significant local structural perturbations, due to predicted steric clash between the aromatic side-chain of Y33 and the thymidine methyl group.

The structure of Y33H bound to its cognate site, in which the same original T:A base-pair at positions ± 10 is altered into a G:C base-pair, was determined at 2.9 Å resolution. In this structure, the same bifurcated hydrogen bonds between Y33 and adenosine 10 in the wild-type structure appear to be replaced by two separate hydrogen-bonds from the ε and δ nitrogen atoms of H33 to the N7 nitrogen atoms of Gua± 10 and Gua± 11. This new interaction thus maintains a contact to base-pair 10 that is similar to the wild-type complex, while introducing an additional contact to the adjacent

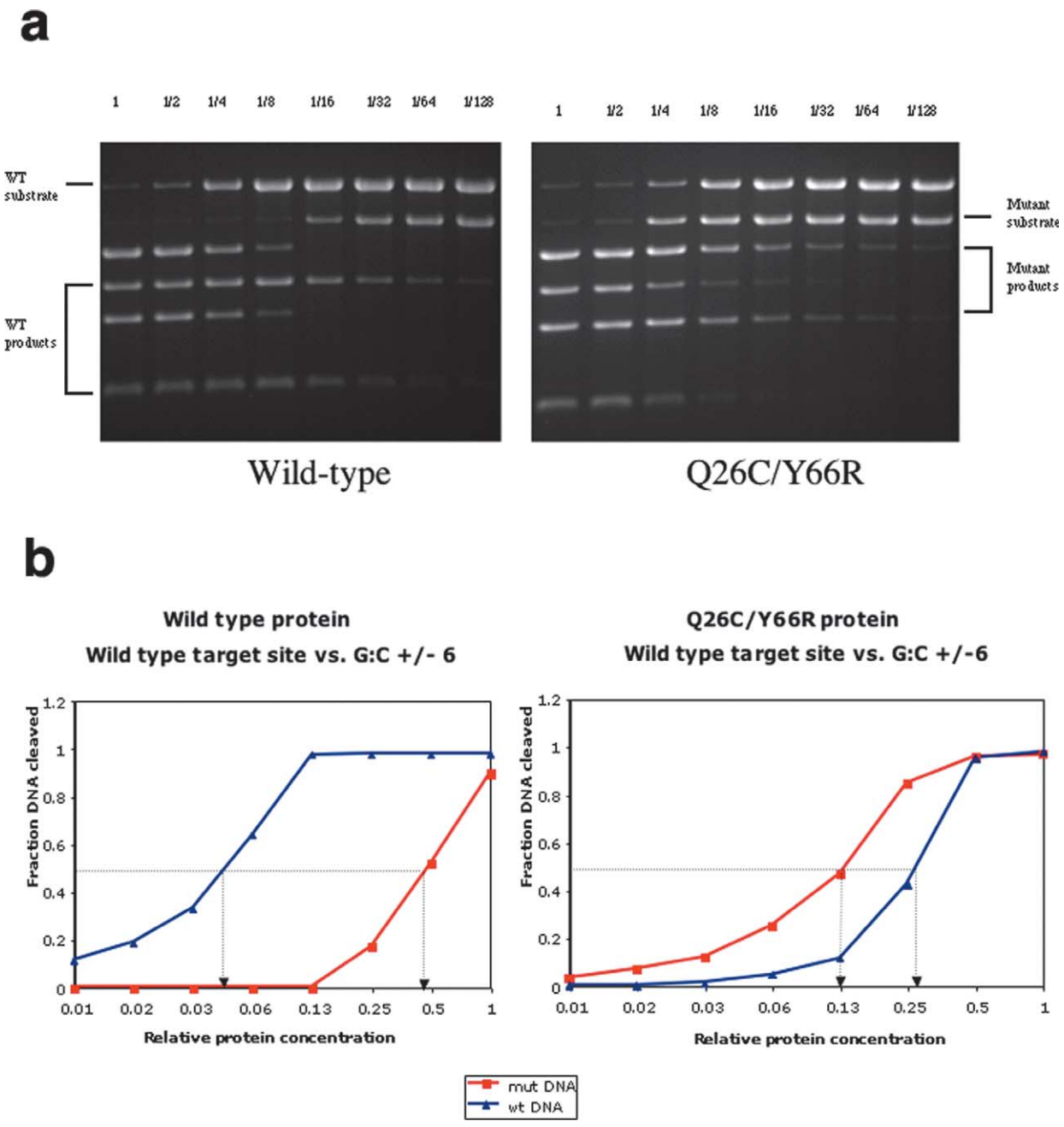

Figure 3. Relative cleavage of cognate and non-cognate targets. (a) *In vitro* competitive cleavage assay. Wild-type I-*Cre*I and the Q26CY66R mutant were exposed to linearized plasmids containing wild-type and mutant homing sites. The numbers above each gel indicate relative amounts of endonuclease, with 1 corresponding to the minimal enzyme concentration sufficient to completely digest each plasmid and subsequent fractions reflecting serial twofold enzyme dilutions. Homing site identities are indicated at the sides of each photo and endonuclease identities below. (b) Representative plot of competitive cleavage assay. The density of agarose gel bands was determined using program ImageJ; the fraction cleaved was calculated by dividing the density of product bands by the density of product plus reactant bands.

base-pair at the extreme end of the protein–DNA complex.

In both structures, the nearest neighbors to residue 33 exhibit little structural rearrangement as a result of the protein and DNA site alterations. Residues 29–36 (corresponding to a β-strand and adjacent turn that flank residue 33) superpose with an rmsd of less than 0.5 Å across all atoms. The two nearest neighbors, N30 and S32, maintain rotameric conformations nearly identical with those observed in the wild-type complex, and exhibit the same distances between their atoms and those of neighboring DNA bases. Other than the altered DNA base-pair and Y33C/H mutations, the only detectable structural difference is the disappearance and/or positional shifts of single water molecules

Table 2. Cognate *versus* non-cognate cleavage efficiencies

	WT	Q26A	Q26C	Q26C/Y66R	Y33C	Y33H
Wild-type	–					
Target sites						
A:T ±6	>50	3.0				
G:C ±6	8.0		1.35	2.4		
T:A ±10	3.8				4.0	
G:C ±10	4.0					3.3
Specificity shift[a]		>150	10.8	19.2	15.2	13.2
Specificity broadening[a]		<0.06	0.17	0.3	1.05	0.83

Ratio of enzyme concentrations yielding 50% cleavage of cognate site *versus* 50% cleavage of non-cognate target sites at equivalent reaction time-points.

[a] See the legend to Table 1 for a full explanation. The same calculations were performed, substituting the cognate/non-cognate enzyme concentration ratios for 50% cleavage, in place of K_d values for binding measurements.

that form a bridging interaction between S32 and guanine 11 in the wild-type structure. This difference may contribute to the lower affinity of Cre-Y33C to its cognate site than that observed for the wild-type complex.

Discussion

In contrast to the relative ease with which homing endonucleases with new (albeit broadened) specificities can be created, both by selection of individual enzyme mutants and by creation of artificial chimaeric enzymes, other DNA-binding enzymes have proven resistant to engineering. In particular, creation of altered restriction endonucleases with new specificities has proven extremely difficult,[26] although some recent successes have been reported.[27–29] The difficulty of such experiments is ascribed to tight coupling of protein structure, DNA-binding affinity and catalysis, and to the difficulties inherent in cloning and manipulating toxic restriction endonuclease genes. In contrast, homing endonucleases offer a high level of site specificity that reduces host toxicity during selection experiments, a structurally forgiving DNA-binding motif, and relatively straightforward mutational strategies for uncoupling binding and catalytic functions. Accordingly, methods for *in vivo* selection of homing endonuclease variants have been developed successfully.[20,22,23,30]

In genetic assays, a variety of additional mutations at residue 33 (leucine, serine and threonine) appear to block cleavage of the wild-type target while also conferring activity against the altered target in genetic assays, albeit at a lower level of efficiency than Y33C.[24] While the activity of the Y33S mutant is not surprising, based on the structure of its similar cysteine counterpart, the interactions formed to the target base-pair by a non-polar leucine residue and an aliphatic threonine residue are far more difficult to model without postulating additional structural rearrangements that cannot be predicted *a priori*. This illustrates the difficulties inherent in modeling and designing biomolecular interfaces in the absence of experimentally determined activities and structures. A similar statement can be made for the mutations at residues 26 and 66 (Q26A/C; Y66R) reported here,

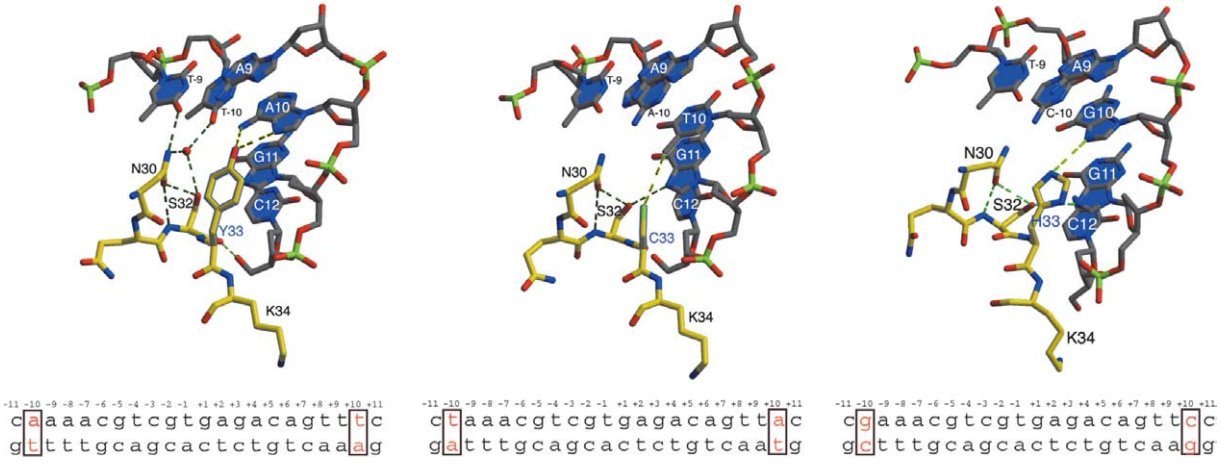

Figure 4. Structures and interactions of cognate pairs at residue 33 and base-pair 10. Protein–DNA contacts in the vicinity of base-pair ±10, in bound complexes containing either wild-type enzyme and DNA target site (left), or Y33C and Y33H mutant enzymes bound to their cognate target site (middle and right, respectively). The sequence of wild-type I-*Cre*I target sequence (left) and alternate target sequences (middle and right) targeted for selection are shown below their corresponding structures.

where again the structural basis of altered specificity (at base-pairs ±6) is difficult to model.

The three experimental measures of endonuclease specificity employed here, *in vivo* cleavage assays, *in vitro* competitive cleavage assays, and *in vitro* binding assays, give somewhat disparate results. While all the enzyme mutants do in fact display reversed specificities in both *in vitro* assays, the most pronounced shifts in cognate/non-cognate preferences correspond to different enzyme-target site pairs when site binding is compared to site cleavage (Tables 1 and 2). Furthermore, in general the enzyme mutants appear to maintain tighter specificity for site cleavage than for site binding, perhaps reflecting the fact that the *in vivo* screen that generated the mutants reports on relative cleavage and elimination of targets, and indicating that the enzymes can partly uncouple their relative binding affinities and cleavage activities against specific sites as a result of *in vivo* selection.

It is important to consider differences in the assays reported here. Binding assays provide an accurate measure of an enzyme's affinity for a particular site. *In vivo*, homing endonucleases must find and cleave specific target sites. In so doing, they must discriminate between optimal target sites and other sites that may be related closely. For example, the *E. coli* genome contains seven copies of a sequence in its 23 S rDNA genes that differs from the native 22 base-pairs I-*Cre*I target site at only three base-pairs.[31] Mutations that either increase an endonuclease's affinity for a particular desired target, or decrease its affinity for non-desired target(s), such as the 23 S site, may display increased signal in our *in vivo* assays. The *in vitro* cleavage competition assays seek to mimic this discrimination problem, albeit at a greatly reduced scale of complexity.

The results from the study reported here indicate that the simplest form of single-site reselection, in which individual contact points are mutated, can produce desired dramatic shifts in binding and cleavage preferences, generally with variable amounts of broadening of those preferences. *In vitro* binding assays revealed that the single amino acid substitutions analyzed here resulted in 18 to 2000-fold shifts in substrate specificity, calculated as the product of wild-type endonuclease *versus* mutant endonuclease site preference (Table 1; see the footnotes for calculations). A similar pattern emerged in the cleavage assays, where an 11-fold to greater than 150-fold effect was observed (Table 2). Further, the Q26C Y66R double mutant displayed greater substrate specificity shifts than the Q26C mutant in each assay. Thus, we have shown that single amino acid substitutions are capable of producing dramatic shifts in substrate specificity, and that additional mutations can be found that result in further increases.

A variety of genomic sequence analyses conducted in our laboratories against targets of interest from genetic disease loci and essential genes from microbial pathogens indicate that many potentially useful targets for the development of gene-specific reagents contain sequences that differ from naturally occurring homing endonuclease targets at five to eight positions (out of a typical recognition site spanning 20–22 base-pairs) (R.M. Jr & B.L.S., unpublished results). While domain fusions from panels of these endonucleases can help to reduce the gap between existing substrates and desired target specificities, the ability to completely exchange target preferences at individual base-pairs while maintaining overall levels of site discrimination and specificity is the central challenge facing the field of homing endonuclease engineering. The structures of a pair of mutants from this study in complex with their new DNA targets indicate that individual contacts can be altered with minimal, local effects on protein architecture and surrounding contact positions. This, in turn, implies that a great deal of homing endonuclease engineering selection can be accomplished by conducting simultaneous, parallel selection experiments at multiple positions in the protein–DNA target interface, followed by final rounds of higher-stringency selections to maximize discrimination between cognate and potential undesirable, non-cognate sites. Experiments of the type reported here can serve as a model for achieving this goal.

In addition to homing endonucleases, molecular systems involving the tethering of tandem arrays of DNA-binding zinc fingers to catalytic domains of DNA-acting enzymes (such as the endonuclease domain of FokI) also offer great promise for the design of novel DNA-binding proteins.[32,33] Such artificial biomolecules have been generated and characterized *in vitro*, and recently have been shown to drive specific, predictable recombinational events *in vivo*.[33–35] Finally, engineered group II homing endonucleases (which form ribonucleoprotein complexes in which binding specificity is influenced by the RNA sequence) are under development.[36] Taken together, these results indicate that the ability to create gene-specific reagents is near.

Methods

Isolation of endonuclease mutants

I-*Cre*I mutants, individually mutated at either of two residues (Q26 and Y66) that contact base-pairs ±6, were constructed in an arabinose-inducible vector (pA-E) and passed through screens for recognition and cleavage of altered target sites as described.[24] In this strategy, wild-type and symmetrically mutated I-*Cre*I target sites are transferred by reciprocal recombination to the F′ factor F128 of *E. coli* strain CC136 to create F′o-*Cre* alleles. Host strains containing these constructs are subsequently used to identify endonuclease mutants with altered site recognition and cleavage activity (which leads to conversion of $LacZ^+$ cells to $LacZ^-$, giving rise to white and sectored colonies in the presence of 5-bromo-4-chloro-3-indolyl-β,D-galactopyranoside (Xgal). The relative

activities of various mutants can be estimated qualitatively by visualization of colony sectoring in this assay.

The mutants that gave the best indication of altered specificity based on these phenotypic screens and assays were tested for their relative binding and cleavage of wild-type and mutant sites, as described below. In order to purify the enzyme for biochemical assays, the genes were subcloned into a C-terminal His-tagged version of the arabinose-inducible pB-E vector.[37] Mid-exponential phase cells were induced for two hours with 0.2% (w/v) arabinose, and I-*Cre*I derivatives purified by nickel-affinity chromatography using TALON® resin.

Gel mobility-shift assays

Gel mobility-shift assays were based on retardation of the electrophoretic mobility of ^{32}P kinase-labeled DNA when bound by I-*Cre*I. Appropriate 47 base oligonucleotides were annealed and end-labeled with ^{32}P. Endonuclease and 2.5 fmol of labeled double-stranded (ds) DNA was incubated for 30 minutes at room temperature in 20 mM Tris–HCl (pH 9.0), 10 mM $CaCl_2$, 1 mM DTT, 50 μg/ml of non-specific competitor DNA and 3% (v/v) glycerol (Mg^{2+} is required for I-*Cre*I cleavage activity; substitution of Ca^{2+} for Mg^{2+} permits DNA binding but not cleavage). Samples were electrophoresed on non-denaturing 10% polyacrylamide gels containing 1 mM $CaCl_2$ at 200 V at 4 °C. Gels were imaged using a Storm Phosphorimager 840 (Molecular Dynamics, Sunnyvale, CA) and the intensity of the free and bound DNA bands were quantified using ImageQuant software (Molecular Dynamics). The K_d values of the I-*Cre*I–DNA complexes were defined as the concentration of I-*Cre*I at which 50% of the DNA was shifted into a complex with slower mobility and represent averages of three experiments.

Competitive cleavage assays

The relative ability of wild-type and mutant I-*Cre*I sites to serve as substrates across a range of concentrations of purified I-*Cre*I constructs were determined as described,[24] using linearized plasmid constructs, one containing a wild-type site and the other a mutant site, as substrate. Assays were performed on 100 ng of each linearized plasmid in 10 μl of 20 mM Tris–HCl (pH 9.0), 10 mM $MgCl_2$, 1 mM DTT and 50 mg/ml of bovine serum albumin (BSA). Minimal concentrations of enzyme sufficient to achieve complete digestion of each substrate were determined empirically and used to generate a series of twofold dilutions of enzymes. Standardized digests of substrates were carried out for 30 minutes at 37 °C and were terminated by placing digestions on ice, followed by addition of loading buffer containing 0.5% (w/v) SDS and electrophoresis through 1.2% (w/v) agarose in TBE buffer (Figure 3a). To measure relative cleavage efficiency of individual enzyme constructs against cognate and non-cognate target sites, digest progression curves were generated by scanning and quantifying DNA product band intensities, and determining the relative concentrations of enzyme required for 50% cleavage of the two sites under identical digest conditions (Figure 3b).

Crystallographic analyses

The enzyme mutants with most significantly altered site specificity, identified by screening and biochemical analyses described above, were overexpressed and purified for crystallization trials in complex with their cognate DNA target sites. These constructs were generated by point mutagenesis in a catalytically inactive Q47E I-*Cre*I mutant background, to facilitate overexpression necessary for crystallization experiments. The background mutation is located in the enzyme active site and does not affect protein structure or DNA-binding contacts or affinity (our unpublished results). Protein was expressed, purified and crystallized as described,[21] using two DNA oligonucleotides to form a pseudopalindromic

Table 3. Data processing and refinement statistics

Protein	Y33C	Y33H
Space group	$P2_1$	$P1$
Cell parameters		
	a=43.1 Å	a=43.8 Å
	b=68.0 Å	b=70.3 Å
	c=87.2 Å	c=88.2 Å
		α=90.3°
		β=89.9°
	γ=92.0°	γ=98.8°
Resolution (Å)	2.5	2.9
Redundancy	4.7	4.1
Completeness[a] (%)	95.6 (93.9)	94.8 (45.6)
Average $I/\sigma(I)$[a]	20.4 (2.8)	24.6(6.9)
R_{sym}[a] (%)	5.7 (31.9)	3.1 (12.6)
R_{work} (%)	21.4	21.7
R_{free} (%)	27.7	25.6
Ramachandran plot		
Most-favored (%)	87.8	87.6
Additionally allowed (%)	12.2	12.4
Generousy allowed (%)	0.0	0.0
Dissallowed (%)	0.0	0.0
rms deviation from ideality		
Bond lengths (Å)	0.006	0.007
Bond angles (deg.)	1.2	1.2
Average B (Å^2) (protein, DNA)	34.5	34.4
PDB ID code	1UOC	1UOD

[a] Outer resolution bin 2.59–2.50 Å in parentheses.

duplex corresponding to the mutant binding site (Figure 1).

Whereas mutants at residue 33 (Y33C and Y33H, which recognize sites with altered base-pairs at positions ±10) crystallized readily in complex with their DNA targets, constructs with mutations at residues 26 and 66 (isolated in selections against site variants at positions ±6) did not yield data-quality specimens. It is most likely that the failure of complexes with mutations at base-pairs ±6 is due to the increased propensity of those DNA constructs (which possess additional palindromic symmetry relative to the wild-type DNA site) to form a dynamic equilibrium between double-stranded duplexes and individual hairpin structures; this aspect of homing endonuclease/DNA cocrystallization has been observed and reported for I-*Cre*I and other similar enzyme systems.[21,38] In contrast, the substitutions at base-pairs ±10 simply maintain existing wild-type symmetry between the DNA half-sites.

For Y33C and Y33H, crystals grew in two different space groups. Y33C crystals are isomorphous with the space group $P2_1$ previously described for wild-type I-*Cre*I/DNA,[21] while Y33H crystals grow in the *P*1 space group. Data from crystals of both mutants were collected at the Advanced Light Source beamline 5.0.2 to 2.5 Å and 2.9 Å resolution. Data were processed using DENZO/SCALEPACK.[39] Refinement was carried out using CNS,[40] with 5% of the reflections withheld for an R_{free} calculation.[41] Rigid body refinement was used to position an initial model of an I-*Cre*I dimer bound to cognate DNA.[38] Residues 29–36 from each I-*Cre*I monomer and base-pairs ±8–10 were deleted from the model to prevent bias during subsequent map calculations. Omitted model atoms including protein and DNA mutations were placed easily in the subsequent maps. Table 3 provides data and refinement statistics.

Protein Data Bank accession codes

The structures of the I-*Cre*I Y33C–DNA and the Y33H–DNA complex have been deposited in the RCSB Protein Data Bank with ID codes 1UOC and 1UOD, respectively.

Acknowledgements

The authors thank Karen Chisholm for the initial isolation and characterization of Y33C and Y33H, and Brendan Springstubb, Emilie Vander Haar and Aaron Korkegian for experimental contributions. Funding was provided by NIH GM49857 (to B.L.S.), NIH CA88942 (to R.J.M.), NSF RUI-9870817 and NIH R15 GM069485-01 (to L.M.S.).

References

1. Belfort, M. & Roberts, R. J. (1997). Homing endonucleases—keeping the house in order. *Nucl. Acids Res.* **25**, 3379–3388.
2. Chevalier, B. S. & Stoddard, B. L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucl. Acids Res.* **29**, 3757–3774.
3. Lambowitz, A., Caprara, M. G., Zimmerly, S. & Perlman, P. S. (1999). Group I and group II ribozymes as RNPs: clues to the past and guides to the future. In *The RNA World II* (Gesteland, R. F., Atkins, J. F. & Cech, T. R., eds), pp. 451–485, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
4. Kuhlmann, U. C., Moore, G. R., James, R., Kleanthous, C. & Hemmings, A. M. (1999). Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? *FEBS Letters*, **463**, 1–2.
5. Shen, B. W., Landthaler, M., Shub, D. A., & Stoddard, B. L. (2004). DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.*
6. Dalgaard, J. Z., Klar, A. J., Moser, M. J., Holley, W. R., Chatterjee, A. & 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. *Nucl. Acids Res.* **25**, 4626–4638.
7. Chevalier, B., Monnatt, R. J. & Stoddard, B. L. (2001). The LAGLIDADG homing endonuclease I-*Cre*I shares three divalent cations between two active sites. *Nature Struct. Biol.* **8**, 312–316.
8. Heath, P. J., Stephens, K. M., Monnat, R. J. & Stoddard, B. L. (1997). The structure of I-*Cre*I, a group I intron-encoded homing endonuclease. *Nature Struct. Biol.* **4**, 468–476.
9. Duan, X., Gimble, F. S. & Quiocho, F. A. (1997). Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell*, **89**, 555–564.
10. Colleaux, L., D'Auriol, L., Galibert, F. & Dujon, B. (1988). Recognition and cleavage site of the intron-encoded omega transposase. *Proc. Natl Acad. Sci. USA*, **85**, 6022–6026.
11. Thompson, A. J., Yuan, X., Kudlicki, W. & Herrin, D. L. (1992). Cleavage and recognition pattern of a double-strand-specific endonuclease (I-*Cre*I) encoded by the chloroplast 23S rRNA intron of *Chlamydomonas reinhardtii*. *Gene*, **119**, 247–251.
12. Marshall, P. & Lemieux, C. (1991). Cleavage pattern of the homing endonuclease encoded by the fifth intron in the chloroplast large subunit rRNA-encoding gene of *Chlamydomonas eugametos*. *Gene*, **104**, 241–245.
13. Dalgaard, J. Z., Garrett, R. A. & Belfort, M. (1993). A site-specific endonuclease encoded by a typical archaeal intron. *Proc. Natl Acad. Sci. USA*, **90**, 5414–5417.
14. Chevalier, B. S. & Stoddard, B. L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucl. Acids Res.* **29**, 3757–3774.
15. Steuer, S., Pingoud, V., Pingoud, A. & Wende, W. (2004). Chimeras of the homing endonuclease PI-SceI and the homologous *Candida tropicalis* intein: a study to explore the possibility of exchanging DNA-binding modules to obtain highly speciric endonucleases with altered specificity. *ChemBiochem.* **5**, 206–213.
16. Epinat, J. C., *et al*. (2003). A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucl. Acids Res.* **31**, 2952–2962.
17. Chevalier, B. S., Kortemme, T., Chadsey, M. S., Baker, D., Monnat, R. J. & Stoddard, B. L. (2002). Design, activity and structure of a highly specific artificial endonuclease. *Mol. Cell*, **10**, 895–905.
18. Durrenberger, F. & Rochaix, J.-D. (1993). Characterization of the cleavage site and the recognition sequence of the I-*Cre*I DNA endonuclease encoded by the chloroplast ribosomal intron of *Chlamydomonas reinhardtii*. *Mol. Gen. Genet.* **236**, 409–414.
19. Argast, G. M., Stephens, K. M., Emond, M. J. &

Monnat, R. J. (1998). I-*Ppo*I and I-*Cre*I homing site sequence degeneracy determined by random mutagenesis and sequential *in vitro* enrichment. *J. Mol. Biol.* **280**, 345–353.
20. Jurica, M. S., Monnat, R. J. & Stoddard, B. L. (1998). DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-*Cre*I. *Mol. Cell*, **2**, 469–476.
21. Chevalier, B., Turmel, M., Lemieux, C., Monnat, R. J. & Stoddard, B. L. (2003). Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-*Cre*I and I-MsoI. *J. Mol. Biol.* **329**, 253–269.
22. Gimble, F. S., Moure, C. M. & Posey, K. L. (2003). Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system. *J. Mol. Biol.* **334**, 993–1008.
23. Pingoud, A. & Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Nucl. Acids Res.* 2001;, 3705–3727.
24. Seligman, L., Chisholm, K. M., Chevalier, B. S., Chadsey, M. S., Edwards, S. T., Savage, J. H. & Veillet, A. L. (2002). Mutations altering the cleavage specificity of a homing endonuclease. *Nucl. Acids Res.* **30**, 3870–3879.
25. Jurica, M. S., Monnat, R. J., Jr & Stoddard, B. L. (1998). DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-*Cre*I. *Mol. Cell*, **2**, 469–476.
26. Lanio, T., Jeltsch, A. & Pingoud, A. (2000). On the possibilities and limitations of rational protein design to expand the specificity of restriction endonucleases. *Protein Eng.* **13**, 275–281.
27. Rimseliene, R., Maneliene, Z., Lubys, A. & Janulaitis, A. (2003). Engineering of restriction endonucleases: using methylation activity of the bifunctional endonuclease Eco571 to select the mutant with a novel sequence specificity. *J. Mol. Biol.* **327**, 383–391.
28. Schottler, S., Wenz, C., Lanio, T., Jeltsch, A. & Pingoud, A. (1998). Protein engineering of the restriction endonuclease EcoRV–structure- guided design of enzyme variants that recognize the base-pairs flanking the recognition site. *Eur. J. Biochem.* **258**, 184–191.
29. Samuelson, J. C. & Xu, S.-Y. (2002). Directed evolution of restriction endonuclease BstYI to achieve increase substrate specificity. *J. Mol. Biol.* **319**, 673–683.
30. Gruen, M., Chang, K., Serbanescu, I. & Liu, D. R. (2002). An *in vivo* selection system for homing endonuclease activity. *Nucl. Acids Res.* **30**, 29–34.
31. Blattner, F. R., *et al*. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453–1474.
32. Smith, J., Bibikova, M., Witby, F. G., Reddy, A. R., Chandrasegaran, S. & Caroll, D. (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucl. Acids Res.* **28**, 3361–3369.
33. Bibikova, M., Caroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y. G. & Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell. Biol.* **21**, 289–297.
34. Porteus, M. H. & Baltimore, D. (2003). Chimeric nucleases stimulate gene targeting in human cells. *Science*, **300**, 763.
35. Bibikova, M., Beumer, K., Trautman, J. K. & Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. *Science*, **300**, 764.
36. Guo, H., Karherg, M., Long, M., Jones, J. P., 3rd., Sullenger, B. & Lamhowitz, A. M. (2000). Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science*, **289**, 452–457.
37. Seligman, L. M., Stephens, K. M., Savage, J. H. & Monnat, R. J., Jr (1997). Genetic analysis of the *Chlamydomonas reinhardtii* I-*Cre*I mobile intron homing system in *Escherichia coli*. *Genetics*, **147**, 1653–1664.
38. Chevalier, B. S., Monnat, R. J., Jr & Stoddard, B. L. (2001). The homing endonuclease I-*Cre*I uses three metals, one of which is shared between the two active sites. *Nature Struct. Biol.* **8**, 312–316.
39. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
40. Brunger, A. T. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
41. Brunger, A. (1993). Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Crystallog. sect. D*, **49**, 24–36.

Edited by K. Morikawa

(Received 2 June 2004; received in revised form 13 July 2004; accepted 14 July 2004)