Published online 18 October 2007 *Nucleic Acids Research, 2007, Vol. 35, No. 21* ***7209–7221***
doi:10.1093/nar/gkm867

Thermodynamics of DNA target site recognition by homing endonucleases

Jennifer H. Eastberg[1,2], Audrey McConnell Smith[1], Lei Zhao[1,3], Justin Ashworth[3], Betty W. Shen[1] and Barry L. Stoddard[1,*]

[1]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, A3-025 Seattle, WA 98109, [2]Graduate Program in Molecular and Cellular Biology and [3]Graduate Program in Molecular Biophysics, Structure and Design, University of Washington, Seattle, WA 98195, USA

Received August 23, 2007; Revised September 11, 2007; Accepted September 24, 2007

ABSTRACT

The thermodynamic profiles of target site recognition have been surveyed for homing endonucleases from various structural families. Similar to DNA-binding proteins that recognize shorter target sites, homing endonucleases display a narrow range of binding free energies and affinities, mediated by structural interactions that balance the magnitude of enthalpic and entropic forces. While the balance of ΔH and $T\Delta S$ are not strongly correlated with the overall extent of DNA bending, unfavorable $\Delta H_{binding}$ is associated with unstacking of individual base steps in the target site. The effects of deleterious basepair substitutions in the optimal target sites of two LAGLIDADG homing endonucleases, and the subsequent effect of redesigning one of those endonucleases to accommodate that DNA sequence change, were also measured. The substitution of base-specific hydrogen bonds in a wild-type endonuclease/DNA complex with hydrophobic van der Waals contacts in a redesigned complex reduced the ability to discriminate between sites, due to nonspecific $\Delta S_{binding}$.

INTRODUCTION

Site-specific DNA-binding proteins typically display tight binding to their cognate target sites, and can discriminate between those sites and potential noncognate targets under demanding physiogical conditions—including high salt, the presence of many competing DNA-binding proteins in the cell and a vast background of potential nonspecific target sites within genomic DNA (1,2). Site-specific protein–DNA recognition is driven by structural interactions and conformational changes in both molecules that occur upon binding (3,4). This includes: (i) the overall complementarity of shape and charge between the two macromolecules, (ii) generic interactions with the DNA phosphoribosyl backbone (which attenuates overall affinity), (iii) interactions between chemical groups on the protein and on the nucleotide bases and (iv) binding-induced distortions of the DNA duplex.

Interactions with the DNA bases include both nonpolar contacts that are dominated by the van der Waals (Lennard–Jones) potential and hydrogen bonds formed between individual polarizable groups (5). These latter contacts can be made directly between DNA and protein atoms, or can be mediated by water molecules trapped in the molecular interface (6). Distortion of the DNA also contributes substantially to overall specificity, through the influence of sequence-specific DNA conformational preferences, because the energetic cost of disrupting the DNA double helix (particularly by base unstacking) can be strongly dependent on the sequence of the DNA target (7).

Historical studies of specific protein–DNA-binding energetics demonstrate that the free energy change and affinity of DNA binding displayed by most proteins fall within a relatively narrow range, by balancing inversely proportional changes of enthalpy (ΔH) and entropy ($T\Delta S$) for any given binding event (1). This balance has been termed 'isothermal enthalpy–entropy compensation' (1,8), and allows DNA-binding proteins that must perform extensive DNA remodeling or distortion (such as the bacterial CAP protein or TATA binding protein) to function under the same physiological conditions and at similar molecular concentrations as proteins that induce almost no distortion of their DNA targets (such as the GCN4 transcriptional activator). The absolute magnitude of these opposing thermodynamic forces can vary dramatically between individual proteins (each by up to $+/-30$ kcal/mol).

The majority of DNA-binding proteins that have been subjected to detailed thermodynamic analyses recognize relatively short (less than 10 basepair) target sites (1). Most of those proteins display high fidelity, exhibiting

*To whom correspondence should be addressed. Tel: 1 206 667 4031; Fax: 1 206 667 6877; Email: bstoddar@fhcrc.org

a marked decrease in affinity and activity as a result of basepair substitutions in their cognate target site. For those systems, a strong correlation has been noted between the extent of DNA distortion and the relative balance of ΔH and ΔS upon protein binding. Proteins that strongly distort their DNA targets are generally observed to display unfavorable ΔH as a result of base unstacking and molecular strain (1,8,9), which is balanced by a net favorable $T\Delta S$. This can be accomplished by minimizing protein folding upon binding and by extensive release of solvent molecules and counterions when the molecules form a complex. In contrast, proteins that bind to their DNA targets with minimal distortion often display a favorable ΔH, balanced by unfavorable $T\Delta S$. This can be accomplished by coupling the binding event to additional protein folding, as well as by quenching of protein dynamic flexibility (10).

Homing endonucleases that act as mobile genetic elements (usually in concert with microbial introns and inteins) differ greatly from the previously studied DNA-binding proteins mentioned above. They are small proteins (typically less than 250 amino acids) and are among the most specific DNA-binding proteins found in nature, recognizing long target sites of 14–30 bp (11). The most specific of the homing endonucleases are under intense study as potential enzymatic reagents for various applications that require highly specific targeting of DNA sequences, including corrective gene therapy and genomic engineering (12). All homing endonucleases are able to tolerate individual polymorphisms at some target site positions, providing a mechanism to increase their overall genetic mobility. For the most specific of homing endonucleases, the length of their DNA target sites assures that any one target sequence can occur by chance at a very low frequency (for example, a single 22 bp site occurs by chance at a frequency of 1 in 10^{13}). Even if several thousand closely related target site variants of that length can be recognized by the protein, the chance that any one of them might occur in random genomic DNA is equal to the sum of their individual probabilities—in some cases as low as 1 in 10^9, making them capable of acting as 'gene specific reagents' (13).

Homing endonucleases can be classified into at least five distinct structural families (Figure 1), based on their structural similarities and conservation of active site geometries(11). These families are the LAGLIDADG proteins (the most specific, found in archaebacteria and in mitochondrial and chloroplast genomes of fungi and algae), His-Cys box proteins (found in protist nuclear genomes), HNH and GIY-YIG proteins (both found in primarily in phage) and PD-(D/E)XK proteins (found in bacteria). With the exception of the His-Cys box and HNH endonucleases that share a common $\beta\beta\alpha$-metal active site motif (14), each of these families has arisen from unique ancestors. The extraordinary DNA-binding specificity displayed by these proteins requires neither large tertiary structures nor the assistance of additional protein factors.

Crystallographic structures are available for representatives of all homing endonuclease families (Figure 1), showing that their mechanisms of DNA recognition and association are widely divergent. The phage endonucleases (HNH and GIY-YIG families) are completely asymmetric and are composed of separate N- and C-terminal catalytic and DNA-binding domains that are tethered together by a variety of additional domains and linkers (15–17). In nature these proteins often exchange or swap either their catalytic or their DNA-binding domains in a highly combinatorial manner. Their isolated nuclease domains typically display considerable non-specific DNA cleavage activity.

In contrast, the remaining homing endonuclease families display significant symmetry in their tertiary and/or quaternary structures, ranging from pseudosymmetric, gene-fused monomers (such as the LAGLIDADG endonucleases I-SceI and I-AniI) (18,19) to homodimers (found in many lineages of LAGLIDADG and His-Cys box proteins) (20,21) and tetramers (most recently visualized for the bacterial PD-D/E-XK endonuclease I-Ssp8603I) (22). In homing endonuclease systems involving protein dimers or tetramers, the corresponding cognate target site in the biological host is often a pseudopalindrome. In those cases, palindromic derivatives of the DNA sites are also viable substrates for their endonucleases.

In this study, we survey the free energy and DNA-binding affinity, and the balance of enthalpy and entropy, for representatives of most known families of homing endonucleases using the method of isothermal titration calorimetry (ITC) (23). In addition to a global survey of DNA-binding energetics by these proteins, we also characterize three particularly significant aspects of homing endonuclease DNA recognition: the ability of symmetric homodimers to recognize asymmetric DNA targets, the energetic costs of encorporating deleterious basepair mismatches in their cognate target sites and the effect of subsequent efforts to redesign protein contact points to accommodate those same basepair substitutions.

MATERIALS AND METHODS

Materials

All homing endonucleases used in this study were overexpressed in *Escherichia coli* and purified as described previously (methods found in the following citations). The proteins surveyed were: I-MsoI (a LAGLIDADG homodimer from the green algae *Monomastix*) (20), I-AniI (a LAGLIDADG monomer from the fungi *Aspergillus nidulans)* (18), I-PpoI (a His-Cys box protein from the slime mold *Physarum polycephalum)* (24), I-HmuI (an HNH protein from bacteriophage SPO1) (17) and I-SspI (a PD-D/E-XK protein from the cyanobacteria *Synechocystis*) (22). All endonucleases were wild-type constructs, with the exception of I-SspI (which was expressed and purified as a catalytically inactive point mutant, because the wild-type protein cannot be expressed in *E. coli* at sufficient yield to permit ITC experiments).

All DNA oligonucleotides (1 μmol scale, salt-free syntheses; see Table 1 for sequences) were purchased

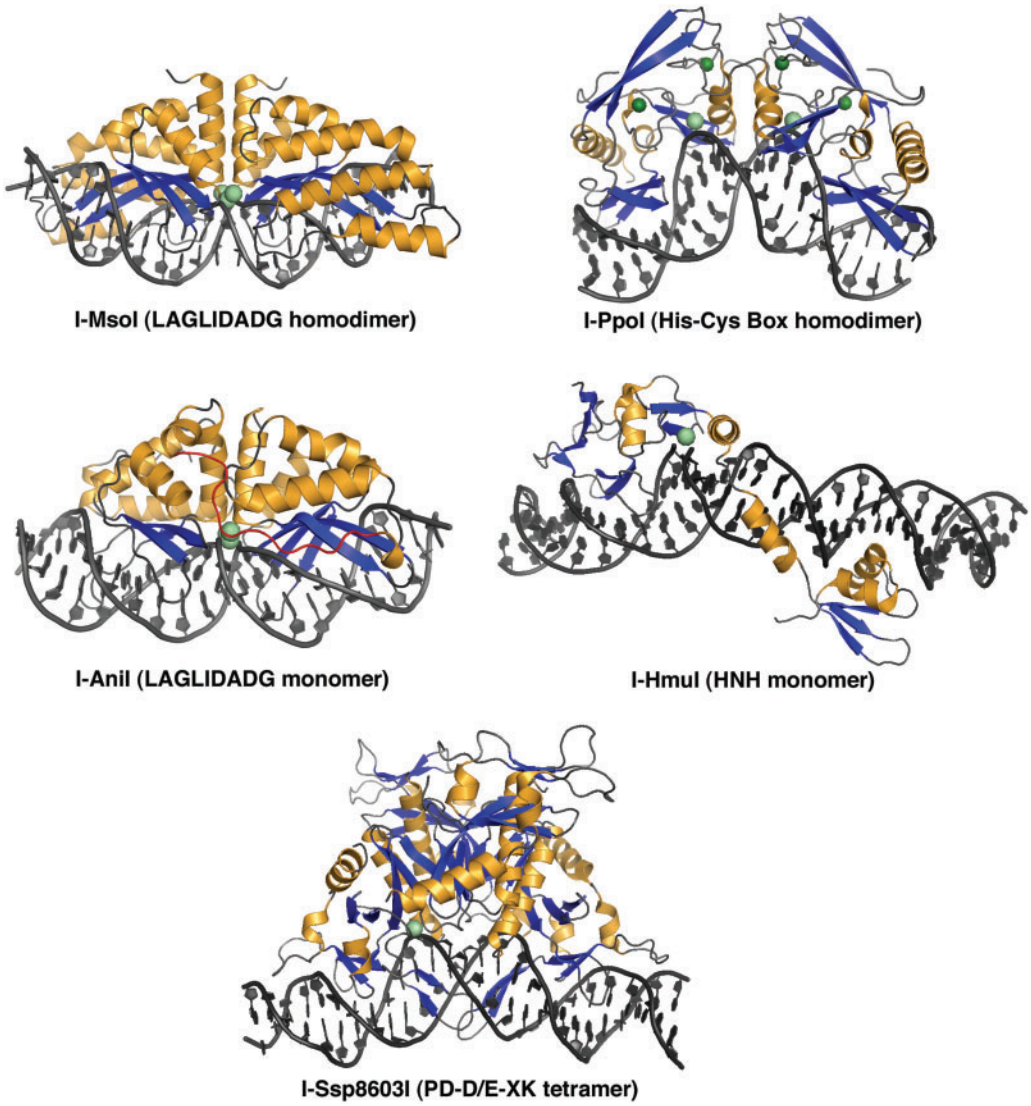

Figure 1. Structures of homing endonuclease–DNA complexes in this study. The structures correspond to a homodimeric LAGLIDADG enzyme (I-MsoI; 2.5 Å resolution), a monomeric LAGLIDADG enzyme (I-AniI; 2.4 Å), a homodimeric His-Cys box enzyme (I-PpoI, 1.7 Å), an HNH enzyme (I-HmuI, 3.1 Å) and a tetrameric PD-D/E-XK enzyme (I-SspI; 3.1 Å). All structures shown are in complex with wild-type physiological target sites from the biological host of the corresponding homing endonuclease. Bound metal ions are shown as spheres.

from Operon Biotechnologies, Inc. (Huntsville, AL, USA). Complementary strands were annealed by heating to 95°C followed by slow-cooling to room temperature to generate the double-stranded DNA target site for binding analysis by ITC. Oligonucleotide target site sequences were chosen based on previous identification of their cognate target sites and additional biochemical analyses. Asymmetric G-C anchors were added to each end of palindromic and pseudo-palindromic sites to prevent undesired hairpin formation and thus promote formation of stable heterodimers. The homogeneity of duplex constructs was verified by electrophoresis in 4% agarose.

Isothermal titration calorimetry

Appropriate pairs of protein and DNA samples were dialyzed overnight into identical buffers, prior to calorimetric measurements, using Tube-O-Dialyzers (G Biosciences, St Louis, MO, USA). Individual buffer compositions, chosen based on previous published studies of optimal binding conditions for each endonuclease (while preventing DNA cleavage by withholding magnesium), were as follows:

I-PpoI: 20 mM sodium phosphate pH 8.0, 50 mM NaCl and 10 mM EDTA (25,26).

I-MsoI: 50 mM Tris pH 8.0, 75 mM NaCl and 10 mM $CaCl_2$ (20).

I-AniI: 50 mM Tris, pH 7.5, 50–100 mM NaCl, 10 mM $CaCl_2$ (27).

I-SspI: 50 mM Tris pH 7.5 and 200 mM NaCl (28).

I-HmuI: 20 mM HEPES pH 8.0, 50 mM NaCl and 1 mM EDTA (29).

Following dialysis, sample concentrations were quantified by UV spectroscopy using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE, USA). The extinction coefficients

Table 1. DNA target site constructs

Protein	Structural family	Site name and site type	Sequence of duplex (target site in bold)
I-PpoI	His-Cys box homodimer	WT: wild-type cognate pseudopalindrome	5′ GGGGGAGTAACTATGA**CTCTCTTAAGGTAG**CCAAAGGGGG 3′ 3′ CCCCCTCATTGATACT**GAGAGAATTCCATC**GGTTTCCCCC 5′
I-PpoI	His-Cys box homodimer	LL: left–left palindrome	5′ GGGGGTGCAGTGA**CTCTCTTAAGAGAGT**CAGACGTGGGGG 3′ 3′ CCCCCACGTCACT**GAGAGAATTCTCTC**AGTCTGCACCCCC 5′
I-PpoI	His-Cys box homodimer	RR: right–right palindrome	5′ GGGGGTGCAGTGG**CTACCTTAAGGTAG**CCAGACGTGGGGG 3′ 3′ CCCCCACGTCACC**GATGGAATTCCATC**GGTCTGCACCCCC 5′
I-MsoI	LAGLIDADG homodimer	WT: wild–type cognate pseudopalindrome	5′ GGGGGG**CAGAACGTCGTGAGACAGTTCG**GGGGGG 3′ 3′ CCCCCC**GTCTTGCAGCACTCTGTCAAGC**CCCCCC 5′
I-MsoI	LAGLIDADG homodimer	LL: left–left palindrome	5′ GGGGGG**CAGAACGTCGTACGACGTTCTG**GGGGGG 3′ 3′ CCCCCC**GTCTTGCAGCATGCTGCAAGAC**CCCCCC 5′
I-MsoI	LAGLIDADG homodimer	RR: right–right palindrome	5′ GGGGGG**CGAACTGTCTCGAGACAGTTCG**GGGGGG 3′ 3′ CCCCCC**GCTTGACAGAGCTCTGTCAAGC**CCCCCC 5′
I-MsoI	LAGLIDADG homodimer	MIS: miscognate pseudopalindrome	5′ GGGGGG**CAGAAGGTCGTGAGACCGTTCG**GGGGGG 3′ 3′ CCCCCC**GTCTTCCAGCACTCTGGCAAGC**CCCCCC 5′
I-AniI	LAGLIDADG monomer	WT: wild–type cognate asymmetric	5′ CCTTCCC**TGAGGAGGTTTCTCTGTAA**ACCCTTCC 3′ 3′ GGAAGGG**ACTCCTCCAAAGAGACATT**TGGGAAGG 5′
I-AniI	LAGLIDADG monomer	WT-OPT: selected optimal target asymmetric	5′ CCTTCCC**TGAGGAGGTTACTCTGTTA**ACCCTTCC 3′ 3′ GGAAGGG**ACTCCTCCAATGAGACAAT**TGGGAAGG 5′
I-SspI	PD-D/E-XK tetramer	WT: wild-type cognate pseudopalindrome	5′ GGG**CCTTCGGGCTCATAACCCGAAGGGA**CG 3′ 3′ CCC**GGAAGCCCGAGTATTGGGCTTCCCT**GC 5′
I-HmuI	HNH monomer	WT: wild-type cognate asymmetric	5′GGGGGGTAAT**GAGCCTAACGCTCAGCAATTCCCAC**GTAAGGGGGG3′ 3′CCCCCCATTA**CTCGGATTGCGAGTCGTTAAGGGTG**CATTCCCCCC5′

Underlined bases indicate positions in target sites for homodimeric (I-PpoI, I-MsoI) or tetrameric (I-SspI) homing endonucleases that are not conserved between DNA half-sites (i.e. they contribute to asymmetry in otherwise palindromic targets).

(ϵ) for the proteins were calculated using ProtParam (30). The annealed DNA oligonucleotide ϵ-values were determined experimentally by digesting the dsDNA oligonucleotides to completion with DNAse and using known nucleotide-monophosphate ϵ-values (31) to calculate ϵ for the dsDNA complex (32). Sample concentrations ranged from 1.5 to 4.5 μM protein and from 30 to 70 μM dsDNA target.

ITC experiments were conducted on a VP-ITC MicroCalorimeter (MicroCal, LLC, Northampton, MA, USA) following manufacturer guidelines. The protein sample was placed in the cell and the DNA target in the auto-pipette. Individual runs consisted of 25–35 injections of 5–9 μl each, depending on sample concentrations, and were conducted at a temperature of 30°C and a stirring speed of 329 r.p.m. A full experiment was composed of three experimental runs with both DNA and protein components in the instrument, as well as a control run of DNA into buffer. Following integration and normalization of peaks, data from the DNA into buffer control run was subtracted to minimize the effect of DNA dilution on measured heat of mixing. Data were fit using Origin 7 SR2 software (OriginLab Corporation, Northampton, MA, USA). The data were well modeled by the one-site fitting algorithm in Origin and the resulting fit parameters were used to calculate K_D.

For a ligand X binding to a single set of 'n' identical sites on a macromolecule, the single-site binding constant 'K_D' = (filled sites)/(empty sites) and $\Delta G_o = -R\ T\ \ln K = \Delta H_o - T\ \Delta S_o$, where ΔG_o, $\Delta\ H_o$ and ΔS_o are the free energy, enthalpy and entropy change for single site binding. By nonlinear least squares fit of calorimetric titration data, the parameters K_D, ΔH_0, and n are determined directly in a single experiment and ΔG_0 and ΔS_0 may then be calculated. Titration calorimetry is capable of defining all of these parameters in a single experiment, resulting in nearly complete thermodynamic characterization of the interaction (33).

DNA bending analyses

The extent of DNA bending and parameters of individual basepair steps were analyzed for the crystal structures of individual homing endonuclease-DNA complexes using program 'ReadOut' (34) (via a web-based server located at: http://gibk26.bse.kyutech.ac.jp/jouhou/readout/).

RESULTS AND DISCUSSION

Affinity, binding free energy and thermodynamic compensation across homing endonuclease families

The binding of 11 separate combinations of homing endonucleases and various DNA target sites, representing

Table 2. Affinities and thermodynamic values of homing endonuclease–DNA-binding events

Protein	Site	K_D *nM*	ΔH *kcal/mol*	ΔS *cal/mol/deg*	$-T\Delta S$ *kcal/mol*	ΔG *kcal/mol*
I-PpoI	WT	17 ± 1.7	−35.5 ± 0.29	−81.6 ± 0.87	24.7	−10.8
I-PpoI	LL	87 ± 8.8	−35.1 ± 0.57	−83 ± 3.0	25.3	−9.8
I-PpoI	RR	23 ± 1.6	−38.8 ± 0.26	−93.1 ± 2.4	28.2	−10.6
I-AniI	WT	96 ± 10	11.2 ± 1.1	69.3 ± 3	−21.0	−9.8
I-AniI	WT-OPT	8 ± 1	6.6 ± 0.7	59.0 ± 3	−17.9	−11.3
I-MsoI	WT	21 ± 5.0	12.7 ± 0.27	77 ± 1.1	−23.3	−10.6
I-MsoI	LL	6 ± 1.4	16.4 ± 0.20	91.7 ± 0.36	−27.8	−11.4
I-MsoI	RR	17 ± 6.0	14.4 ± 0.40	83.0 ± 0.64	−25.2	−10.8
I-MsoI	MIS	760 ± 53	21.2 ± 0.42	98 ± 2.7	−29.7	−8.4
I-MsoI redesigned	MIS	46 ± 6.5	34.2 ± 0.42	146.5 ± 0.7	−44.4	−10.2
I-MsoI redesigned	WT	243 ± 20	37.1 ± 4	152.7 ± 6	−46.3	−9.2
I-SspI (E11Q)[a]	WT	350 ± 30	−23.3 ± 3	−47.3 ± 4	14.3	−9.0
I-HmuI[b]	WT	(30)	Endothermic	–	–	–

[a]'I-SspI' corresponds to a catalytically inactive mutant (E11Q) of the I-Ssp8603I homing endonuclease, containing an alteration of an active-site putative metal-binding residue. In addition, the enzyme contains a second mutation on its surface far from the DNA-binding interface (F55K) to improve its solubility. The resulting affinity and $\Delta G_{binding}$ therefore may not reflect the wild-type enzyme.
[b]I-HmuI displays complex behavior in ITC experiments that is not reliably modeled mathematically (Figure 2). The reaction is strongly endothermic, and the K_d for that binding event has been estimated at 30 nM.

five major structural classes of these proteins, was analyzed by ITC. The DNA constructs used for these studies are shown in Table 1; the structure of each wild-type protein/DNA complex is shown in Figure 1. The affinities and thermodynamic signatures of these interactions are summarized in Table 2 and Figures 2 and 3.

The dissociation constants of the wild-type homing endonuclease families against their wild-type, physiological target sites range from 17 (I-PpoI) to 96 nM (I-AniI). The precise affinity and thermodynamic values for one wild-type endonuclease (I-HmuI, of the phage HNH family) could not be estimated directly from ITC data, due to multiphasic behavior in its binding isotherm (described in detail subsequently). Gel shift experiments for that endonuclease give K_D values of ~30 nM to its cognate DNA targets (B.L.S., unpublished data). Finally, the dissociation constant of a catalytically inactive point mutant (E11Q) of I-Ssp8603I (a bacterial PD-D/E-XK homing endonuclease) was 350 nM; this probably represents a compromised binding affinity due to the loss of a bound metal ion in the protein–DNA interface. For the wild-type endonucleases that could be analyzed in detail, the range of binding free energies (ΔG) to their physiological DNA targets was −10.8 kcal/mol (I-PpoI) to −9.8 kcal/mol (I-AniI).

With the exception of I-HmuI, the stoichiometries of protein–DNA interactions measured in these experiments were within 25% of 1:1 (protein functional assemblies:DNA target sites), in agreement with previously published biochemical studies and crystallographic analyses (Figure 2). The lower stoichiometry observed for I-HmuI binding may reflect a higher percentage of aggregated or misfolded protein (that may also contribute to the unusual isotherm for that endonuclease).

For two endonucleases, their physiological binding site is not the optimal substrate for that protein. I-AniI (a LAGLIDADG monomer) binds a variant target site containing two basepair substitutions with 12-fold lower K_D (8 nM) than its target in *Aspergillus*. Similarly, I-MsoI (a LAGLIDADG homodimer) binds a palindromic DNA target that consists of its two left half-sites with a 3-fold lower K_D (6 nM) than its physiological target site. Both of these observations are discussed in more detail in the following sections.

The thermodynamic DNA-binding profiles of these proteins are widely distributed across the continuum of enthalpy/entropy compensation, and obey the same linear pattern of thermodynamic compensation noted previously for a wide variety of DNA-binding proteins (1) (Figures 2 and 3). Two of the endonucleases (the His-Cys box protein I-PpoI and the PD-D/E-XK protein I-Ssp8603I) display strongly exothermic DNA binding, whereas the monomeric and homodimeric LAGLIDADG endonucleases (I-AniI and I-MsoI) and the HNH endonuclease (I-HmuI) display strongly endothermic DNA binding (Figure 2). The range of values for ΔH (from −35 to +13 kcal/mol) and $T\Delta S$ (−23 kcal/mol to +25 kcal/mol) that separate these individual protein–DNA interactions are as broad as previously observed for other DNA-binding proteins (Figure 3).

DNA bending and thermodynamic signatures

All available crystal structures of homing endonuclease/DNA complexes demonstrate significant DNA bending that is associated with the mechanism by which each accomplishes recognition of long target sequences (11) (Figures 1 and 4). In the case of dimeric (LAGLIDADG and His-Cys-Box) and tetrameric (PD-D/E-XK) homing endonucleases, cleavage is executed across the minor groove of the target site (generating 3′ overhangs), thus allowing the protein to contact and 'read out' contiguous sets of nucleotide basepairs within the major groove of each DNA half-site, at positions that flank the cleavage sites. This strategy

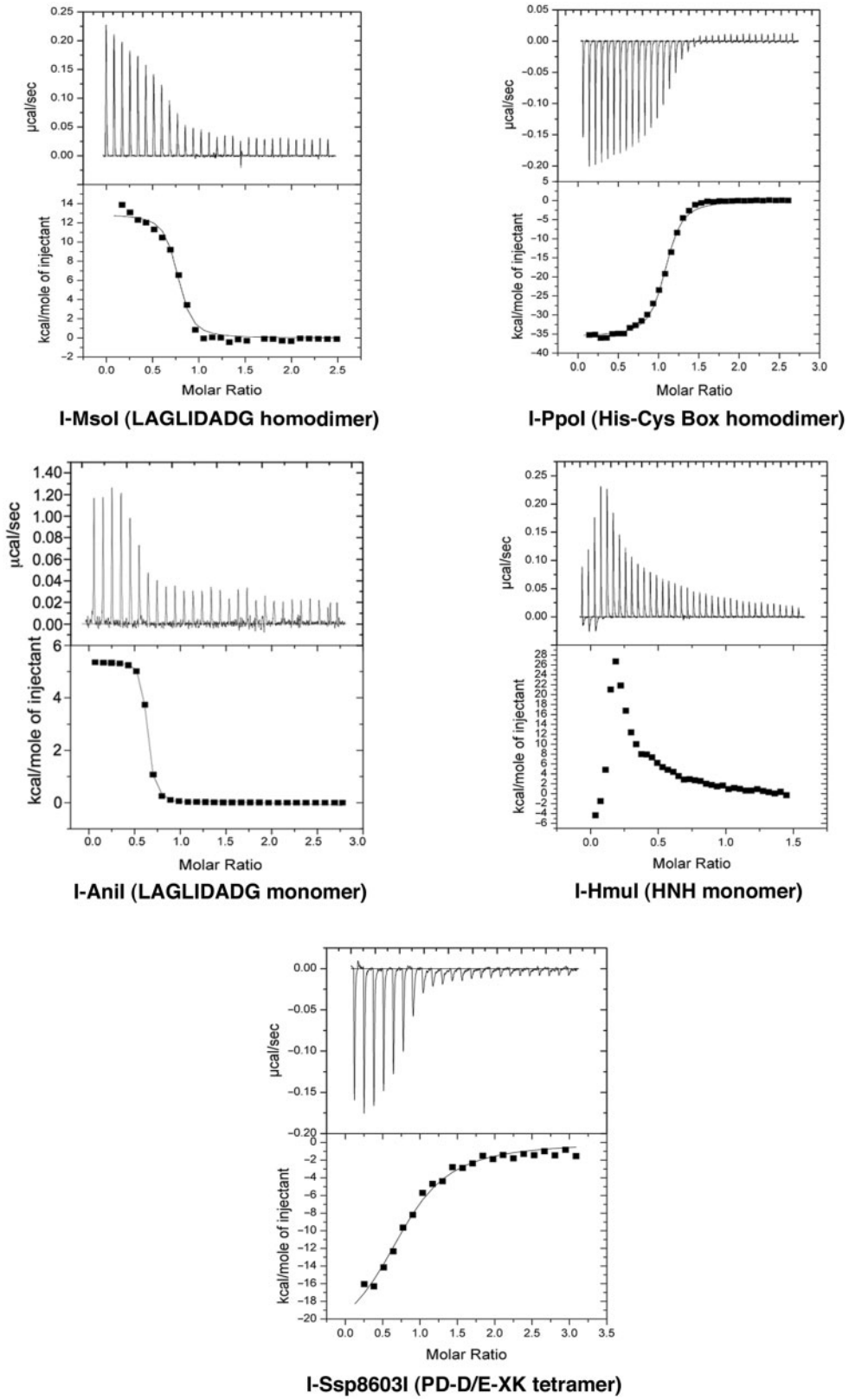

Figure 2. Thermodynamic profiles of representative homing endonuclease binding events. Heat profiles of sequential injections of DNA against relevant endonucleases are shown. All endonucleases are wild-type except for a catalytically inactivated mutant of I-Ssp8603I; all DNA target sites are the physiological sequences form the corresponding biological hosts. Three of the five representative homing endonucleases (the monomeric and homodimeric LAGLIDADG enzymes and the HNH enzyme) display endothermic binding; the other two (the His-Cys box and PD-D/E-XK enzymes) display exothermic binding. All of the enzymes studied could be fit to standard saturation curves, except for I-HmuI (described in detail in the text), which displays complex multiphasic binding behavior.

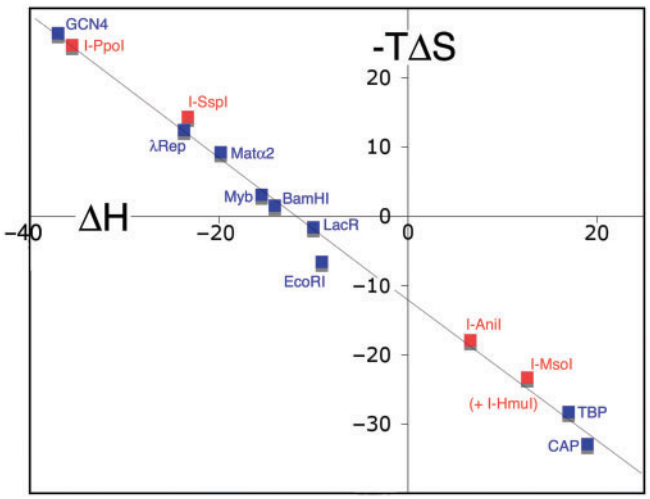

Figure 3. Isothermal enthalpy–entropy compensation by homing endonucleases and other DNA-binding proteins. The enthalpic (ΔH) and entropic ($-T\Delta S$) contributions to site-specific DNA recognition of different protein–DNA complexes, including several representative wild-type homing endonucleases (highlighted in red) are shown. The thermodynamic values for previously studied DNA-binding proteins are shown in blue, and are taken from previous analyses by Jen-Jacobson *et al.* (1) and references therein.

requires that the center of the DNA target sites be significantly distorted near the scissile phosphates, either by narrowing the minor groove and using two closely juxtaposed active sites (as seen for the LAGLIDADG enzymes) or by widening the minor groove and using two physically separated active sites (as seen for the His-Cys box and PD-D/E-XK enzymes). In contrast, the monomeric phage endonucleases (such as the HNH enzyme I-HmuI) bind even longer targets (>25 bp), and use a tandem series of protein domains to contact intermittent stretches of DNA bases within both the major and minor groove of the target site. For those complexes, DNA bending is again necessary in order to access DNA bases in the minor groove and to straddle the DNA backbone at various positions in the complex.

For previously examined DNA-binding proteins, unfavorable enthalpic changes upon DNA binding (corresponding to an endothermic reaction event) have been observed to correlate with increased distortion of the DNA target, which in turn corresponds to basepair unstacking and molecular strain (1). As discussed subsequently, a comparison of the binding thermodynamics of the representative homing endonuclease complexes (Figure 2, Table 2) to the distortion of their DNA target sites (Figure 4) also demonstrates the dominant role of base unstacking in promoting unfavorable changes in enthalpy during DNA binding, while conversely also demonstrating that significant DNA bending can be accomplished while still maintaining base stacking and producing a strongly exothermic-binding event.

The three homing endonucleases that display strongly 'endothermic' DNA binding (the two LAGLIDADG endonucleases and the I-HmuI HNH endonuclease) all display basepair roll angles and unstacking at individual basepair steps that significantly depart from undistorted B-form DNA (Figure 4). The LAGLIDADG enzymes both display a single significant negative roll angle at their central $-1/+1$ bp step of $\sim -20^\circ$, corresponding to pronounced narrowing of the minor groove at the center of the target cleavage region. This DNA bend and unstacking is accompanied by overwinding of the central basepair step ('twist' rising to 50° and 60° for the I-MsoI homodimer and I-AniI monomer, respectively). In contrast, the HNH endonuclease I-HmuI displays strongly 'positive' roll angles at two separate basepair steps ($+20^\circ$ at step $+9/+10$ and then $+37^\circ$ at step $+13/+14$). At each position the protein 'hurdles' the phosphate backbone and inserts side chains into the minor groove. The latter of these bends is accompanied by widening of the minor groove and significant underwinding of the DNA ('twist' being reduced to 22°).

The two endonucleases that exhibit strongly 'exothermic' DNA binding (the His-Cys box I-PpoI endonuclease, and the PD-D/E-XK endonuclease I-SspI) also display significant DNA bending (overall bends of $\sim 75^\circ$ and 25° across the central 8 bp of their target sites) and corresponding widening of the minor groove at the site of cleavage. However, both proteins accomplish binding and distortion of their DNA targets with roll angles that never exceed $+/-12^\circ$ at any single basepair step, and relatively small distortions of helical winding ('twist') values across their target sites (Figure 4). In the case of I-PpoI, the severe bend imparted to the DNA target is accomplished through a mixture of smaller cumulative roll angles (which act in concert to widen the minor groove at the cleavage sites) and mutually opposing tilts in a single DNA basepair step in each half-site.

Multidomain DNA recognition: I-HmuI

Whereas the eukaryotic and bacterial homing endonucleases display relatively compact structures containing single DNA-binding domains arranged in various oligomeric symmetries, the phage endonucleases (HNH and GIY-YIG families) contain separate DNA interaction domains arranged in a sequential tandem array along a single peptide chain (15,16,35). These proteins are prone to significant exchange and shuffling of their domains during evolution, and the occasional insertion of additional structural elements such as zinc fingers (36).

I-HmuI displays such a multidomain structure (Figure 1) (17). Its N-terminal, antiparallel β–sheet (contacting the 5′ end of the target site) is associated with the HNH nuclease core, which is then followed by two α-helices that intercalate in the minor groove in the center of the site, and finally a helix–turn–helix domain that generates the majority of base-specific contacts at the 3′ end of the target site. The DNA binding isotherm displayed by I-HmuI (Figure 2) is unique among the measurements described here, in that it displays complex multiphasic behavior as the total concentration of protein increases with sequential injections. Each of the first three injections (generating DNA concentrations

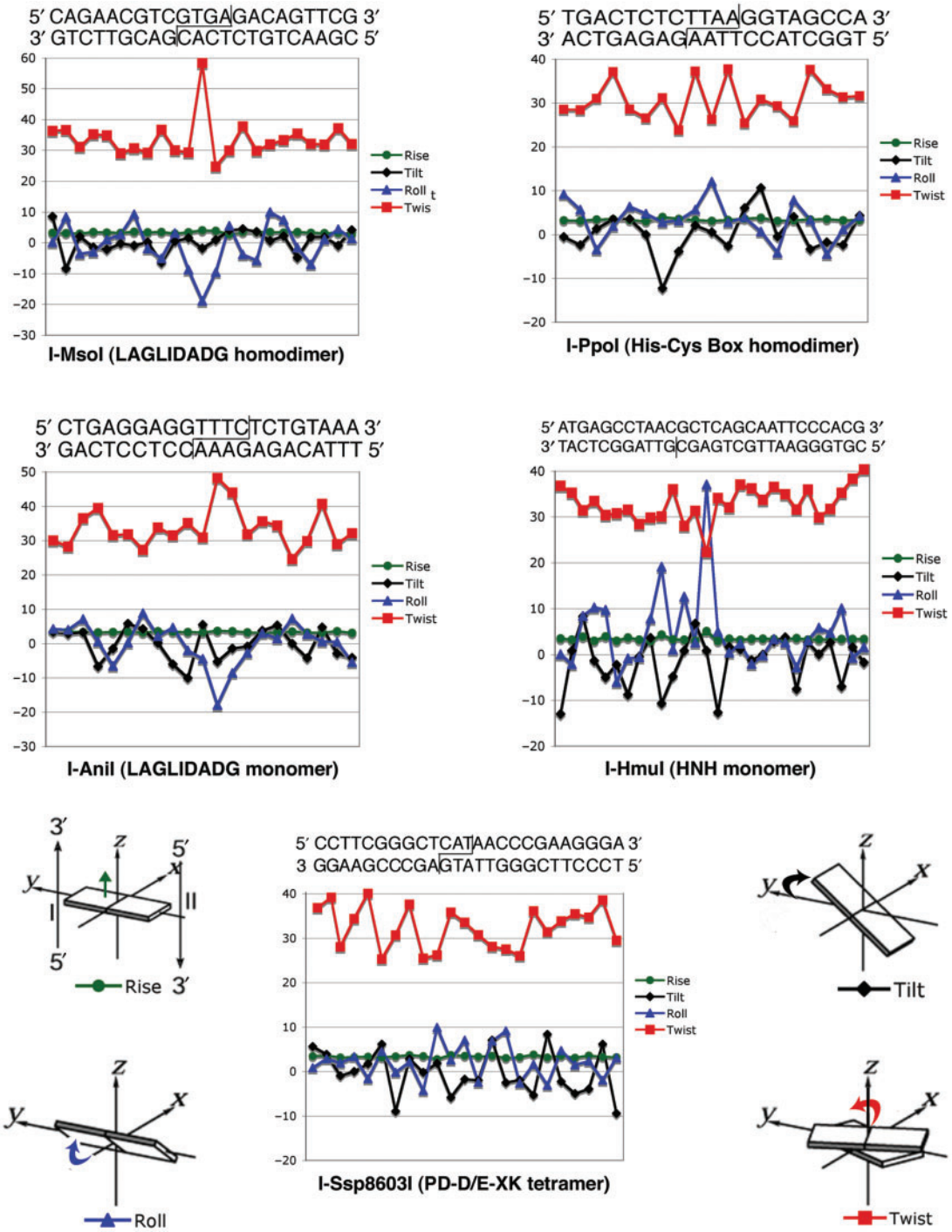

Figure 4. DNA distortion induced by homing endonuclease binding. DNA bend parameters were quantitated using program 'Readout' (34), via a web-based server located at: http://gibk26.bse.kyutech.ac.jp/jouhou/readout/). The cognate target sequence, used in the individual crystal structures from which the bend parameters were calculated, are shown; cleavage sites are indicated. All endonucleases in this study except for I-HmuI cleave both strands to produce 3' overhangs; I-HmuI nicks the lower strand only (vertical line). The individual features of basepair steps distortion shown in the graphs (roll, tilt, twist and rise) are illustrated relative to a standard coordinate frame for double-stranded DNA.

in the sample cell rising from ~10 to 30 nM) results in a rapid absorption of heat ($+\Delta H$) followed by a slower, more modest release of heat ($-\Delta H$). The overall heat signature becomes stronger with each of these early injections. Beginning with the fourth injection (resulting in a DNA concentration of ~40 nM), the isotherm demonstrates only a single feature corresponding to heat absorption. This signal then displays a hyperbolic reduction in strength with each ensuing injection as saturation of the DNA target sites is achieved. The transition from the multiphasic injection isotherms to single peaks and subsequent return to baseline occurs

at a protein concentration near the previously measured K_D for the I-HmuI–DNA interaction.

While the pattern of heat absorption and release demonstrated by I-HmuI described above and shown in Figure 2 is difficult to explain to complete satisfaction, there are two likely explanations. The first is that that a small percentage of the protein is aggregated, a behavior that is eliminated by early injections of DNA (perhaps as free protein is titrated into DNA-bound complexes). The second is that the early injection isotherms might reflect to a biphasic binding event consisting of a relatively fast endothermic association of one protein domain (possibly the C-terminal helix–turn–helix domain) followed by a slower exothermic step of intramolecular association. It is possible that the N-terminal nuclease domain, which displays fewer contacts with the DNA and is known to display reduced DNA specificity on its own (37), may require a slow step of DNA distortion at the center of the target site before docking to the remaining target sequence. Recent studies of GIY-YIG endonucleases, which possess similar structural organization, have also demonstrated DNA binding and cleavage behaviors that involve initial rapid association and subsequent slower conformational changes (38).

Symmetry recognizing asymmetry

Many multimeric homing endonucleases are encoded within introns that interrupt highly conserved rDNA and tRNA genes (39–43). The preponderance of base-paired stem–loop elements in folded RNA structures increases the frequency of palindromic sequences within their coding regions, encouraging the persistence of multimeric homing endonucleases, including both homodimers (such as the His-Cys box endonuclease I-PpoI and the LAGLIDADG endonuclease I-MsoI, encoded by introns within algral rDNA genes) and tetramers (such as I-Ssp8603I, encoded by an intron within a bacterial tRNA gene). While such homing endonucleases possess the advantage of being encoded by particularly short reading frames, which are tolerated well by their intron hosts, they are presented with the challenge of maintaining activity when target site symmetry is broken. In many cases, a homodimeric protein is found to display cleavage activity against a physiological DNA target site that displays surprisingly little palindromic symmetry between left and right half-sites. For example, the LAGLIDADG homodimeric endonuclease I-CeuI acts at a target site that displays only 36% sequence identity between left and right half sites (21).

In such cases, one would anticipate that a homodimeric homing endonuclease would display binding and cleavage activity towards palindromic variants of its natural target site, and that many such proteins would display energetically superior (more favorable) contacts to one half-site over the other. This was examined for homodimeric endonucleases from both the His-Cys Box (I-PpoI) and LAGLIDADG (I-MsoI) families, by measuring their binding affinity and thermodynamic profiles against palindromic DNA sequences ('Left-Left', or 'LL' and 'Right-Right', or 'RR') derived from individual half-sites of their host target sites (Table 1). The two palindromic sequences derived from the I-PpoI target site differ from one another at 4 bp out of 14 total. In contrast, the two palindromic sites derived from the I-MsoI target (which is substantially more asymmetric) differ at 12 out of 22 bp. Each of the palindromic sites differs from the wild-type target at half of those basepairs (all within a single half-site in each case).

I-PpoI and I-MsoI both display noticeable differences in affinity towards the two palindromic DNA sequences derived from their wild-type target (Table 2). I-PpoI binds its 'RR' target 4-fold more tightly than its 'LL' target; I-MsoI binds its 'LL' target 3-fold more tightly than 'RR'. The corresponding differences in $\Delta G_{binding}$ between the palindromic targets for each enzyme ($\Delta\Delta G$) are $\sim$0.6 and 0.8 kcal/mol, respectively. In both cases, these differences in binding affinity are attributable to more favorable $\Delta H_{binding}$ for the superior palindromic target (-3.7 and -2.0 kcal/mol, respectively), which is partially offset by smaller, unfavorable changes in $T\Delta S$. The differences in the binding affinities and $\Delta H_{binding}$ to the LL and RR palindromes is correlated to the total number of contacts made to nucleotide bases in each unique half-site. For example, I-MsoI displays 20 contacts to atoms on DNA bases in the left half-site (10 water mediated, 10 direct to side chains) versus 15 similar contacts to DNA bases in the right half-site (again, half of the contacts are water mediated). The number of contacts to DNA backbone atoms are identical in both.

Interestingly, the two enzymes behave differently towards their wild-type asymmetric target sites from the host as compared to the palindromic variants. I-PpoI binds its asymmetric cognate target site with an affinity (K_D $\sim$17 nM) that is comparable to the more 'tightly' bound of the two palindromes (K_D $\sim$23 nM). In contrast, I-MsoI binds its asymmetric cognate site with an affinity (K_D $\sim$21 nM) that is comparable to the more 'loosely' bound palindromic target (K_D $\sim$17 nM). The ability of a homodimeric homing endonuclease to recognize an asymmetric target site with an affinity that rivals a palindromic repeat of one half-site, or to even prefer the asymmetric cognate target over either palindromic variant (rather than displaying an affinity that is obviously an intermediate between the two palindromes) has also been observed for the LAGLIDADG homing endonuclease I-CeuI (21).

The LAGLIDADG family: specificity and engineering

Homing endonucleases are under intense scrutiny as potential reagents for targeted genetic applications, in which genomes of target organisms or cells are manipulated *in vivo*, using site-specific recombination to alter or add desired traits. This field requires the development of highly specific DNA-binding proteins that can stimulate gene conversion events at unique sites within complex genomes. Over the past several years, two different approaches to create enzymes capable of inducing site-specific DNA double-strand breaks have been developed: zinc finger nucleases (44) and engineered

homing endonucleases (12,45). LAGLIDADG homing endonucleases are particularly attractive systems for the development of gene-specific reagents: they are the most specific of all known homing endonucleases, possess relatively small and highly modularized structures, and have tightly coupled mechanisms of specific DNA recognition and cleavage. A variety of studies have demonstrated that the DNA-binding specificity of LAGLIDADG endonucleases can be altered in predictable ways, ranging from the creation of artificial chimeric enzymes (46,47) to endonucleases that harbor individual amino acid substitutions in the protein–DNA interface (48–55).

Unlike many DNA-binding proteins, engineering and selection of homing endonuclease constructs that display altered DNA-binding specificity (particularly the LAGLIDADG family) is facilitated by two properties: they display relatively modest reductions in affinity when individual contacts to nucleotide bases are eliminated, and they utilize highly modularized and separable contacts between individual amino acid side chains and DNA bases. While this strategy of DNA recognition does not reduce the rules for binding specificity down to the simplicity of a one-to-one 'code' between protein residues and DNA bases, it does greatly simplify the complexity of DNA recognition and its redesign.

In order to investigate the exact effect of deleterious single base substitutions on binding specificity and affinity of LAGLIDADG endonucleases, we characterized the differences in binding thermodynamics for two separate LAGLIDADG homing endonucleases (monomeric I-AniI and homodimeric I-MsoI) when complexed to optimal target sites, versus nonoptimal miscognate target sites. For each study, the effect of simultaneously altering two basepairs was measured (sequences of sites shown in Table 1). We then went on to further characterize the thermodynamic effects of a 'redesign cycle' for I-MsoI, where amino acid substitutions in the enzyme were engineered with the goal of reacquiring high affinity, specific recognition of the formerly miscognate target sequence.

In the case of I-AniI, the optimal target site for the endonuclease differs from the wild-type, physiologic target site at two basepair positions (Table 1). Both mutations (at positions +1 and +8) consist of an inversion of an A:T basepair and were identified in an *in vitro* screen for hypercleavable target site variants (13). The endonuclease displays a 12-fold difference in affinity towards these two targets ($K_{D(\text{wild-type})} = 96\,\text{nM}$; $K_{D(\text{optimal})} = 8\,\text{nM}$), corresponding to a difference in $\Delta G_{\text{binding}}$ of −1.5 kcal/mol. The less favorable $\Delta G_{\text{binding}}$ to the wild-type site is caused by a substantial increase in the heat absorbed during binding to that site (ΔH is increased by 4.6 kcal/mol), which is partially offset by a more favorable entropic change for the formation of the miscognate complex.

In contrast, the homodimeric I-MsoI displays near optimal binding affinity to its physiological cognate target site. Two simultaneous alterations of that target sequence, consisting of a substitution of −6 C:G to −6G:C in the 'left' DNA half-site, and a similar change from +6 T:A to +6 G:C in the symmetry-related 'right' half-site, result in significant reduction of cleavage activity under standard reaction protocols. These substitutions were chosen based on structure-based computational predictions of DNA mutations that would cause a significant reduction in binding affinity of the wild-type enzyme (51). At both of these DNA positions, Lys 28 is engaged in a hydrogen bond to the purine ring and Thr 83 makes a water-mediated contact to the pyrimidine (Figure 5A). Converting either base pair to a G:C was predicted to disrupt binding by the loss of the direct hydrogen-bonding interactions and by desolvation of Lys28.

I-MsoI displays a 36-fold increase in the dissociation constant K_D against the miscognate target site relative to the wild-type target site (from 21 nM up to 760 nM) (Figure 5B). This reduction in affinity corresponds to a 2.1 kcal/mol loss of favorable $\Delta G_{\text{binding}}$ (from −10.6 kcal/mol for the wild-type complex to −8.5 kcal/mol for the miscognate site). Similar to I-AniI, the change in $\Delta G_{\text{binding}}$ to the miscognate site is caused by a significant increase in the heat absorbed during binding (ΔH rises from 12.7 to 21.2 kcal/mol), which is partially offset by a more favorable entropic change ($T\Delta S$ decreases from −23.3 to −29.7 kcal/mol).

For both enzymes, the deleterious basepair mismatches are presumed to lead to the loss of hydrogen bonds between protein side chains and DNA bases, and possibly the introduction of interatomic steric clashes in the interface. The total magnitude of unfavorable thermodynamic changes induced by these mismatches is reduced, however, by increased entropy in the miscognate complex: presumably the complexes between the endonucleases and their miscognate target sites display increased torsional and vibrational disorder in both the protein and the DNA at the site of nonoptimal contacts.

The I-MsoI endonuclease was subsequently engineered to compensate for the basepair substitutions at positions +/−6 in the DNA site, by redesign of the surrounding amino acids (Figure 5D). In these studies, a double point mutation in the enzyme, consisting of K28L and T83R, was predicted to reestablish energetically favorable interactions with the DNA bases (51). This computational redesign prediction was validated by crystallographic analyses, and by functional assays for cleavage activity against the novel target site. In the structure of the redesigned cognate enzyme/DNA complex, Leu28 makes a non-polar contact with the C5 of the cytosine rings at the altered basepair positions and Arg83 makes two hydrogen bonds to the corresponding guanine base of the same basepairs.

Armed with crystal structures of the WT:WT and Mutant:Mutant complexes of I-MsoI bound to DNA (Figure 5A and D), the thermodynamic profiles of the redesigned enzyme to its corresponding 'cognate' target and to the original wild-type DNA sequence were determined. As described above, the wild-type enzyme displays dissociation constants of 21 and 760 nM to its cognate and noncognate target sites (a 36-fold difference in affinity). In comparison, the redesigned enzyme displays

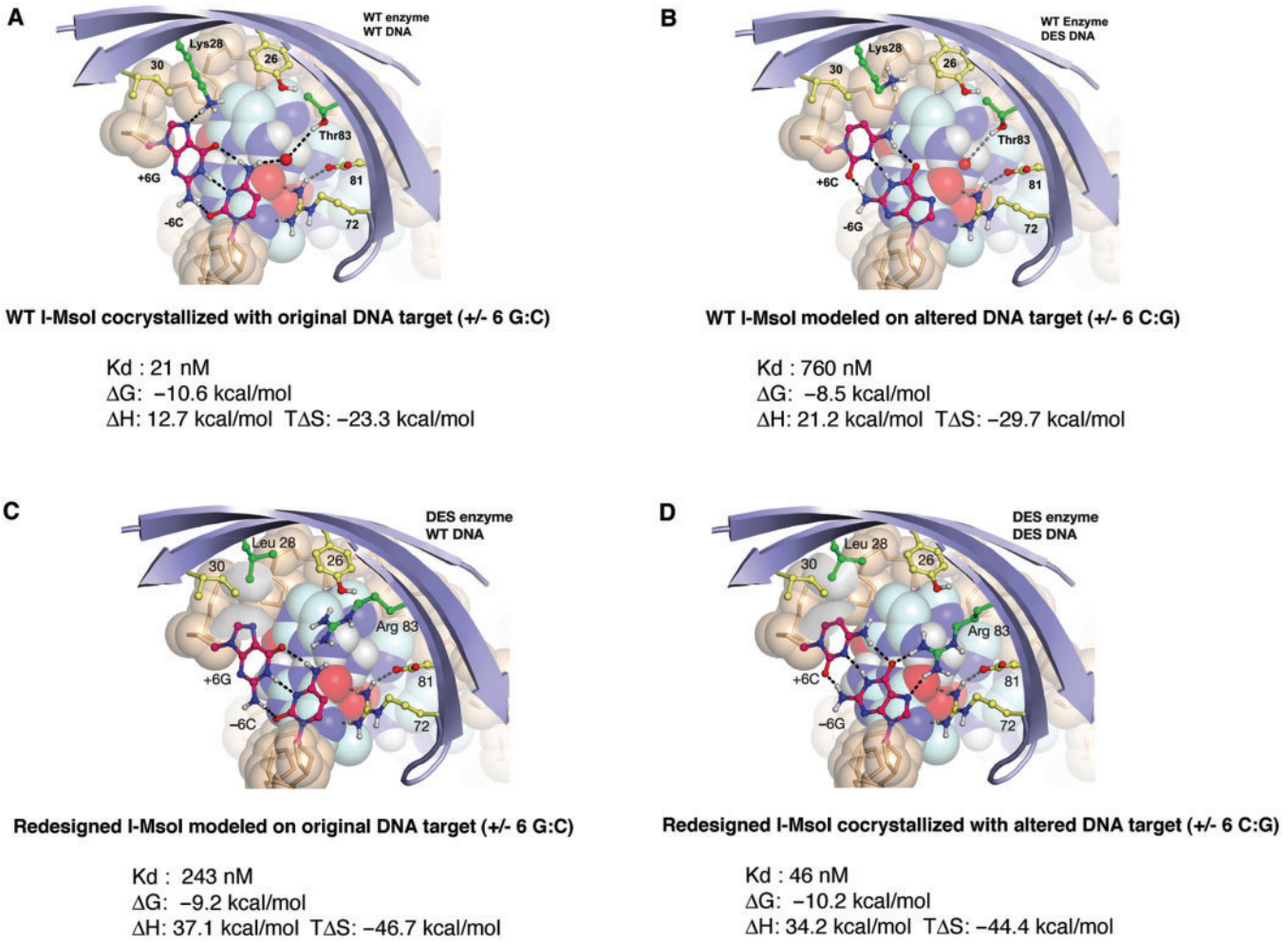

Figure 5. Thermodynamic signature of a homing endonuclease redesign cycle. The wild-type I-MsoI endonuclease (top left; **A**) binds to its cognate target site with a K_d of 21 nM, driven by a favorable entropic change ($T\Delta S$) upon binding. Simultaneous alteration of a base pair in each half site (position +/− 6) results in a 30-fold increase in K_d, correlated with an unfavorable increase of 2.1 kcal/mol in the free energy of binding ($\Delta\Delta G$) (top right; **B**). This energetic penalty is caused by a large unfavorable increase in the enthalpy of binding. Redesign of the enzyme, via two point mutations in the protein/DNA interface (bottom right; **D**), almost entirely restores affinity and free energy of binding (K_D = 46 nM). Finally, analysis of the redesigned enzyme against the original target site (bottom left; **C**) indicates that the newly created LHE, while displaying a specificity switch, only displays about a 5-fold to 10-fold increase in K_D.

a binding affinity to the mutated DNA target that is almost entirely restored as compared to the original wild-type complex (K_D = 46 nM) and a 5.3-fold higher dissociation constant for the original wild-type site (K_D = 243 nM). The amount of discrimination between cognate and noncognate sites displayed by the redesigned enzyme is therefore somewhat lower than the wild-type protein, corresponding to a $\Delta\Delta G_{binding}$ (cognate versus noncognate) of ~1 kcal/mol.

The near wild-type affinity and $\Delta G_{binding}$ between the redesigned I-MsoI and the altered DNA target site is driven entirely by a significant improvement in the already favorable entropic change upon binding. $T\Delta S_{binding}$ is increased by over 20 kcal/mol as compared to the original wild-type binding event, compensating for a significant unfavorable increase in heat absorption ($+\Delta H$) upon binding. This dramatic alteration in the thermodynamic profile of the redesigned protein–DNA pair is probably a result of the introduction of a hydrophobic leucine side chain at residues 28 and 28′ in place of lysine. In both enzyme constructs, residue 28 is solvent-exposed in the unbound enzyme and buried in the DNA complex. In the case of the wild-type protein, the lysine side chain simply exchanges hydrogen bonds with solvent for similar interactions with the DNA. In contrast, binding of the redesigned endonuclease is presumably accompanied by significant desolvation of the solvent-exposed leucine side chain. When the original DNA target sequence is bound by the same redesigned enzyme construct, the less favorable enthalpy of binding is partially negated by the still strongly favorable entropic changes that accompany burial of the hydrophobic leucine side chains. Therefore, the enzyme does not display as large a difference in affinity between cognate and miscognate DNA targets as is observed for the wild-type enzyme. An important take-home

message of this analysis is that the replacement of hydrogen-bond contacts with hydrophobic van der Waals contacts in the protein–DNA interface, during protein engineering, may lead to an undesirable loss of specificity and discrimination.

CONCLUDING REMARKS

Homing endonucleases are unique in that they all must arrive at a successful balance of two somewhat contradictory properties: high-enough specificity to avoid toxicity to their biological host, and low-enough fidelity to maximize opportunities for evolutionary transfer and mobility. Each unique structural family of homing endonuclease is largely constrained to a certain biological host range, and has arrived at the optimal balance of these properties relative to the average genomic size of those hosts. The comparative analysis of multiple lineages of these proteins offers many insights into the constraints placed on DNA recognition by such proteins. Additionally, the characterization of thermodynamic profiles of engineered homing endonucleases promises to add detail to this field for future applications in genomic engineering and medicine.

ACKNOWLEDGEMENTS

Funding to pay the Open Access publication charges for this article was provided by NIH.

Conflict of interest statement. None declared.

REFERENCES

1. Jen-Jacobson,L., Engler,L.E. and Jacobson,L.A. (2000) Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure*, **8**, 1015–1023.
2. vonHippel,P.H. (2007) From 'Simple' DNA-Protein interactions to the macromolecular machines of gene expression. *Ann. Rev. Biophys. Biomol. Struct*., **36**, 79–105.
3. Arauzo-Bravo,M.J., Fujii,S., Kono,H., Ahmad,S. and Sarai,A. (2005) Sequence-dependent conformational energy of DNA derived from molecular dynamics simulations: toward understanding the indirect readout mechanism in protein-DNA recognition. *J. Am. Chem. Soc*., **127**, 16074–16089.
4. Michael Gromiha,M., Siebers,J.G., Selvaraj,S., Kono,H. and Sarai,A. (2004) Intermolecular and intramolecular readout mechanisms in protein-DNA recognition. *J. Mol. Biol*., **337**, 285–294.
5. Jones,S., Heyningen,P.V., Berman,H.M. and Thornton,J.M. (1999) Protein-DNA interactions: a structural analysis. *J. Mol. Biol.,*287.
6. Jayarum,B. and Jain,T. (2004) The role of water in protein-DNA recognition. *Ann. Rev. Biophys. Biomol. Struct*., **33**, 343–361.
7. Koudelka,G.B., Mauro,S.A. and Ciubotaru,M. (2006) Indirect readout of DNA sequence by proteins: the roles of DNA sequence-dependent intrinsic and extrinsic forces. *Prog. Nucleic Acid Res. Mol. Biol*., **81**, 143–177.
8. Searle,M.S. and Williams,D.H. (1993) On the stability of nucleic acid structures in solution: enthalpy-entropy compensations, internal rotations and reversibility. *Nucleic Acids Res*., **21**, 2051–2056.
9. Jayaram,B., McConnell,K.J., Dixit,S.B. and Beveridge,D.L. (1999) Free energy analysis of protein-DNA binding: the EcoRI endonuclease-DNA complex. *J. Comp. Phys*., **151**, 333–357.
10. Spolar,R.S. and Record,M.T., Jr. (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263**, 777–784.
11. Stoddard,B.L. (2005) Homing endonuclease structure and function. *Q. Rev. Biophys*., **38**, 49–95.
12. Paques,F. and Duchateau,P. (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr. Gene Ther*., **7**, 49–66.
13. Scalley-Kim,M., McConnell-Smith,A. and Stoddard,B.L. (2007) Coevolution of homing endonuclease specificity and its host target sequence. *J. Mol. Biol*., **372**, 1305–1319.
14. Kuhlmann,U.C., Moore,G.R., James,R., Kleanthous,C. and Hemmings,A.M. (1999) Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? *FEBS Lett*., **463**, 1–2.
15. VanRoey,P., Waddling,C.A., Fox,K.M., Belfort,M. and Derbyshire,V. (2001) Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *EMBO J*., **20**, 3631–3637.
16. VanRoey,P., Meehan,L., Kowalski,J.C., Belfort,M. and Derbyshire,V. (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nat. Struct. Biol*., **9**, 806–811.
17. Shen,B.W., Landthaler,M., Shub,D.A. and Stoddard,B.L. (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol*., **342**, 43–56.
18. Bolduc,J.M., Spiegel,P.C., Chatterjee,P., Brady,K.L., Downing,M.E., Caprara,M.G., Waring,R.B. and Stoddard,B.L. (2003) Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. *Genes Dev*., **17**, 2875–2888.
19. Moure,C.M., Gimble,F.S. and Quiocho,F.A. (2003) The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *J. Mol. Biol*., **334**, 685–696.
20. Chevalier,B., Turmel,M., Lemieux,C., Monnat,R.J. and Stoddard,B.L. (2003) Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. *J. Mol. Biol*., **329**, 253–269.
21. Spiegel,P.C., Chevalier,B., Sussman,D., Turmel,M., Lemieux,C. and Stoddard,B.L. (2006) The structure of I-CeuI homing endonuclease: Evolving asymmetric DNA recognition from a symmetric protein scaffold. *Structure*, **14**, 869–880.
22. Zhao,L., Bonocora,R.P., Shub,D.A. and Stoddard,B.L. (2007) The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. *EMBO J*., **26**, 2432–2442.
23. Ladbury,J.E. and Chowdhry,B.Z. (1996) Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chem. Biol*., **3**, 791–801.
24. Flick,K.E., McHugh,D., Heath,J.D., Stephens,K.M., Monnat,R.J.Jr. and Stoddard,B.L. (1997) Crystallization and preliminary X-ray studies of I-PpoI: a nuclear, intron-encoded homing endonuclease from Physarum polycephalum. *Protein Sci*., **6**, 2677–2680.
25. Ellison,E.L. and Vogt,V.M. (1993) Interaction of the intron-encoded mobility endonuclease I-*Ppo*I with its target site. *Mol. Cell. Biol*., **13**, 7531–7539.
26. Wittmayer,P.K. and Raines,R.T. (1996) Substrate binding and turnover by the highly specific I-PpoI endonuclease. *Biochemistry*, **35**, 1076–1083.
27. Ho,Y., Kim,S.J. and Waring,R.B. (1997) A protein encoded by a group I intron in Aspergillus nidulans directly assists RNA splicing and is a DNA endonuclease. *PNAS USA*, **94**, 8994–8999.
28. Bonocora,R. and Shub,D.A. (2001) A novel group I intron-encoded endonuclease specific for the anticodon region of tRNA-fMet genes. *Mol. Microbiol*., **39**, 1299–1306.
29. Landthaler,M., Lau,N.C. and Shub,D.A. (2004) Group I intron homing in Bacillus phages SP01 and SP82: a gene conversion event initiated by a nicking homing endonuclease. *J. Bacteriol*., **186**, 4307–4314.
30. Gasteiger,E., Hoogland,C., Gattiker,A., Duvaud,S., Wilkins,M.R., Appel,R.D. and Bairoch,A. (2005) In Walker,J.M. (ed.), *The Proteomics Protocols Handbook*. Humana Press, Totowa, NJ, pp. 571–607.
31. Cavaluzzi,M.J. and Borer,P.N. (2004) Revised UV extinction coefficients for nucleoside-5′-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res*., **32**, e13.

32. Kallansrud,G. and Ward,B. (1996) A comparison of measured and calculated single- and double-stranded oligodeoxynucleotide extinction coefficients. *Anal. Biochem.*, **236**, 134–138.
33. Pierce,M.M., Raman,C.S. and Nall,B.T. (1999) Isothermal titration calorimetry of protein-protein interactions. *Methods*, **19**, 213–221.
34. Ahmad,S., Kono,H., Arauzo-Bravo,M.J. and Sarai,A. (2006) ReadOut: structure-based calculation of direct and indirect readout energies and specificities for protein-DNA recognition. *Nucleic Acids Res.*, **34**, W124–127.
35. Shen,B.W., Dyer,D.H., Huang,J.Y., D'Ari,L., Rabinowitz,J. and Stoddard,B.L. (1999) The crystal structure of a bacterial, bifunctional 5,10 methylene- tetrahydrofolate dehydrogenase/ cyclohydrolase. *Protein Sci.*, **8**, 1342–1349.
36. Dean,A.B., Stanger,M.J., Dansereau,J.T., VanRoey,P., Derbyshire,V. and Belfort,M. (2002) Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *Proc. Natl Acad. Sci.*, **99**, 8554–8561.
37. Landthaler,M., Shen,B.W., Stoddard,B.L. and Shub,D.A. (2006) I-BasI and I-HmuI: two phage intron-encoded endonucleases with homologous DNA recognition sequences but distinct DNA specificities. *J. Mol. Biol.*, **358**, 1137–1151.
38. Mueller,J., Smith,D., Bryk,M. and Belfort,M. (1995) Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the td homing site. *EMBO J.*, **14**, 5724–5735.
39. Lemieux,B., Turmel,M. and Lemieux,C. (1988) Unidirectional gene conversions in the chloroplast of Chlamydomonas inter-specific hybrids. *Mol. Gen. Genet.*, **212**, 48–55.
40. Turmel,M., Boulanger,J., Schnare,M.N., Gray,M.W. and Lemieux,C. (1991) Six group I introns and three internal transcribed spacers in the chloroplast large subunit ribosomal RNA gene of the green alga Chlamydomonas eugametos. *J. Mol. Biol.*, **218**, 293–311.
41. Turmel,M., Gutell,R.R., Mercier,J.-P., Otis,C. and Lemieux,C. (1993) Analysis of the chloroplast large subunit ribosomal RNA gene from 17 Chlamydomonas taxa: three internal transcribed spacers and 12 group I intron insertion sites. *J. Mol. Biol.*, **232**, 446–467.
42. Biniszkiewicz,D., Cesnaviciene,E. and Shub,D.A. (1994) Self-splicing group I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. *EMBO J.*, **13**, 4629–4635.
43. Turmel,M., Cote,V., Otis,C., Mercier,J.P., Gray,M.W., Lonergan,K.M. and Lemieux,C. (1995) Evolutionary transfer of ORF-containing group I introns between different subcellular compartments (chloroplast and mitochondrion). *Mol. Biol. Evol.*, **12**, 533–545.
44. Porteus,M.H. (2006) Mammalian gene targeting with designed zinc finger nucleases. *Mol Ther*, **13**, 438–446.
45. Stoddard,B.L., Monnat,R.J., J and Scharenberg,A.M. (2007) Advances in engineering homing endonucleases for gene targeting: ten years after structures. *Prog. Gene Ther.*, in press.
46. Epinat,J.C., Arnould,S., Chames,P., Rochaix,P., Desfontaines,D., Puzin,C., Patin,A., Zanghellini,A., Paques,F. *et al.* (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.*, **31**, 2952–2962.
47. Chevalier,B.S., Kortemme,T., Chadsey,M.S., Baker,D., Monnat,R.J.Jr and Stoddard,B.L. (2002) Design, activity and structure of a highly specific artificial endonuclease. *Mol. Cell.*, **10**, 895–905.
48. Arnould,S., Perez,C., Cabaniols,J.-P., Smith,J., Gouble,A., Grizot,S., Epinat,J.-C., Duclert,A., Duchateau,P. and Paques,F. (2007) Engineered I-CreIderivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J. Mol. Biol.*, **371**, 49–65.
49. Rosen,L.E., Morrison,H.A., Masri,S., Brown,M.J., Springstubb,B., Sussman,D., Stoddard,B.L. and Seligman,L.M. (2006) Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Res.*, **34**, 4791–4800.
50. Doyon,J.B., Pattanayak,V., Meyer,C.B. and Liu,D.R. (2006) Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J. Am. Chem. Soc.*, **128**, 2477–2484.
51. Ashworth,J., Havranek,J.J., Duarte,C.M., Sussman,D., Monnat,R.J.Jr., Stoddard,B.L. and Baker,D. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, **441**, 656–659.
52. Arnould,S., Chames,P., Perez,C., Lacroix,E., Duclert,A., Epinat,J.C., Stricher,F., Petit,A.S., Patin,A. *et al.* (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J. Mol. Biol.*, **355**, 443–458.
53. Chames,P., Epinat,J.C., Guillier,S., Patin,A., Lacroix,E. and Paques,F. (2005) In vivo selection of engineered homing endonucleases using double-strand break induced homologous recombination. *Nucleic Acids Res.*, **33**, e178.
54. Sussman,D.J., Chadsey,M., Fauce,S., Engel,A., Bruett,A.R.J., Monnat,J., Stoddard,B.L. and Seligman,L.M. (2004) Isolation and characterization of new homing endonuclease specificities at individual target site positions. *J. Mol. Biol.*, **342**, 31–41.
55. 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.