LETTERS

Exploitation of binding energy for catalysis and design

Summer B. Thyme[1,3], Jordan Jarjour[2,5], Ryo Takeuchi[6], James J. Havranek[7], Justin Ashworth[1,3], Andrew M. Scharenberg[2,5], Barry L. Stoddard[3,6] & David Baker[1,3,4]

Enzymes use substrate-binding energy both to promote ground-state association and to stabilize the reaction transition state selectively[1]. The monomeric homing endonuclease I-AniI cleaves with high sequence specificity in the centre of a 20-base-pair (bp) DNA target site, with the amino (N)-terminal domain of the enzyme making extensive binding interactions with the left (−) side of the target site and the similarly structured carboxy (C)-terminal domain interacting with the right (+) side[2]. Here we show that, despite the approximate twofold symmetry of the enzyme–DNA complex, there is almost complete segregation of interactions responsible for substrate binding to the (−) side of the interface and interactions responsible for transition-state stabilization to the (+) side. Although single base-pair substitutions throughout the entire DNA target site reduce catalytic efficiency, mutations in the (−) DNA half-site almost exclusively increase the dissociation constant (K_D) and the Michaelis constant under single-turnover conditions (K_M*), and those in the (+) half-site primarily decrease the turnover number (k_{cat}*). The reduction of activity produced by mutations on the (−) side, but not mutations on the (+) side, can be suppressed by tethering the substrate to the endonuclease displayed on the surface of yeast. This dramatic asymmetry in the use of enzyme–substrate binding energy for catalysis has direct relevance to the redesign of endonucleases to cleave genomic target sites for gene therapy and other applications. Computationally redesigned enzymes that achieve new specificities on the (−) side do so by modulating K_M*, whereas redesigns with altered specificities on the (+) side modulate k_{cat}*. Our results illustrate how classical enzymology and modern protein design can each inform the other.

Enzymes use interactions with the substrate to promote catalysis both by bringing the substrate into close proximity and proper alignment with catalytic groups on the enzyme and by selectively stabilizing the transition state for the chemical reaction[3–5]. Dissection of the contributions to enzyme catalysis has taken on renewed importance with the advent of computational and directed evolution approaches for engineering novel enzymatic activities for applications ranging from synthetic chemistry to therapeutics[6,7]. Reprogramming the specificity of the LAGLIDADG family of homing endonucleases for genome engineering and biotechnology purposes is one such application[8,9].

Control experiments probing the binding specificity of the I-AniI homing endonuclease, in preparation for computational redesign of specificity, revealed a striking asymmetry in the effect of base substitutions on binding affinity (Fig. 1a). DNA cleavage and DNA binding by Y2 I-AniI endonuclease[10] were assayed for 60 different target sites, each containing a single base-pair substitution from the wild-type recognition sequence. Consistent with previous observations[11], enzyme activity assays showed that many nucleotide substitutions throughout the extended 20-bp recognition site abrogated or reduced cleavage, reflecting the high sequence specificity of the endonuclease (Fig. 1b). Fluorescence-binding experiments showed that for mutations between −10 and −3 on the (−) side of the interface, this loss of cleavage activity is associated with a loss of binding affinity. In sharp contrast, mutations in the −2 to +10 region of the recognition site, which also eliminated or reduced cleavage, had a minimal affect on substrate binding (Fig. 1c).

To determine whether the differences between the (−) and (+) side substitutions reflected differential contributions to ground-state association versus transition-state stabilization, the extent of cleavage of a linear double-stranded template as a function of time was determined for all 60 singly substituted sites under single-turnover conditions, and pseudo-Michaelis–Menten parameters[12] K_M* and k_{cat}* were obtained from these data (Supplementary Figs 1–3). Comparison of k_{cat}*/K_M* for related substrates highlights the high sequence specificity of the enzyme: for example, at position −4 k_{cat}*/K_M* for the wild-type G:C base pair is more than 2,000-fold greater than for A:T and more than 400-fold greater than for C:G (Fig. 1b and Supplementary Table 1; because specificity is determined by the differences in k_{cat}*/K_M* for different substrates[13], these results provide perhaps the most rigorous quantification of homing endonuclease specificity so far). The contribution of target-site interactions to ground-state stabilization (K_M*, Fig. 1d) versus transition-state stabilization (k_{cat}*, Fig. 1e) was found to be skewed: substitutions on the (−) side increased K_M* significantly without reducing k_{cat}*, whereas substitutions on the (+) side decreased k_{cat}* with little effect on K_M*. The overall segregation of the kinetic contributions to specificity is shown graphically in Fig. 1f and in the structural schematic in Fig. 1a: most single-base substitutions in the target affect k_{cat}* (blue, (+) side) or K_M* (red, (−) side) but not both. The striking feature of our results is that the apparent symmetry of the binding interface is completely broken during catalysis: chemically, very similar protein–DNA contacts are used for substrate association on the left side and selective transition-state stabilization on the right side.

Surface display methods are widely used to engineer proteins with new binding specificities[14]. The sequence specificity profile obtained for singly substituted target sites binding to I-AniI displayed on the surface of yeast closely parallels the profile observed in the solution fluorescence experiments (Fig. 1c) (J.J., B.L.S. and A.M.S., unpublished observations). We reasoned that cleavage of mutated target sites with increased K_M* should be suppressible by tethering the DNA duplex containing the target site adjacent to the displayed enzyme on the yeast surface; the increase in local substrate concentration should compensate for the decrease in ground-state binding affinity (Fig. 2b). Indeed, mutations between positions −10 and −3 (red) that greatly reduced binding in solution do not slow tethered cleavage on the yeast cell surface (Fig 2c). In contrast, substitutions on the right side of the target

[1]Department of Biochemistry, [2]Department of Immunology, [3]Graduate Program in Biomolecular Structure and Design, [4]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. [5]Seattle Children's Hospital Research Institute, 1900 9th Ave M/S C9S-7, Seattle, Washington 98177, USA. [6]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue, Seattle, Washington 98109, USA. [7]Department of Genetics, Campus Box 8232, Washington University School of Medicine, 4566 Scott Avenue, St Louis, Missouri 63110, USA.

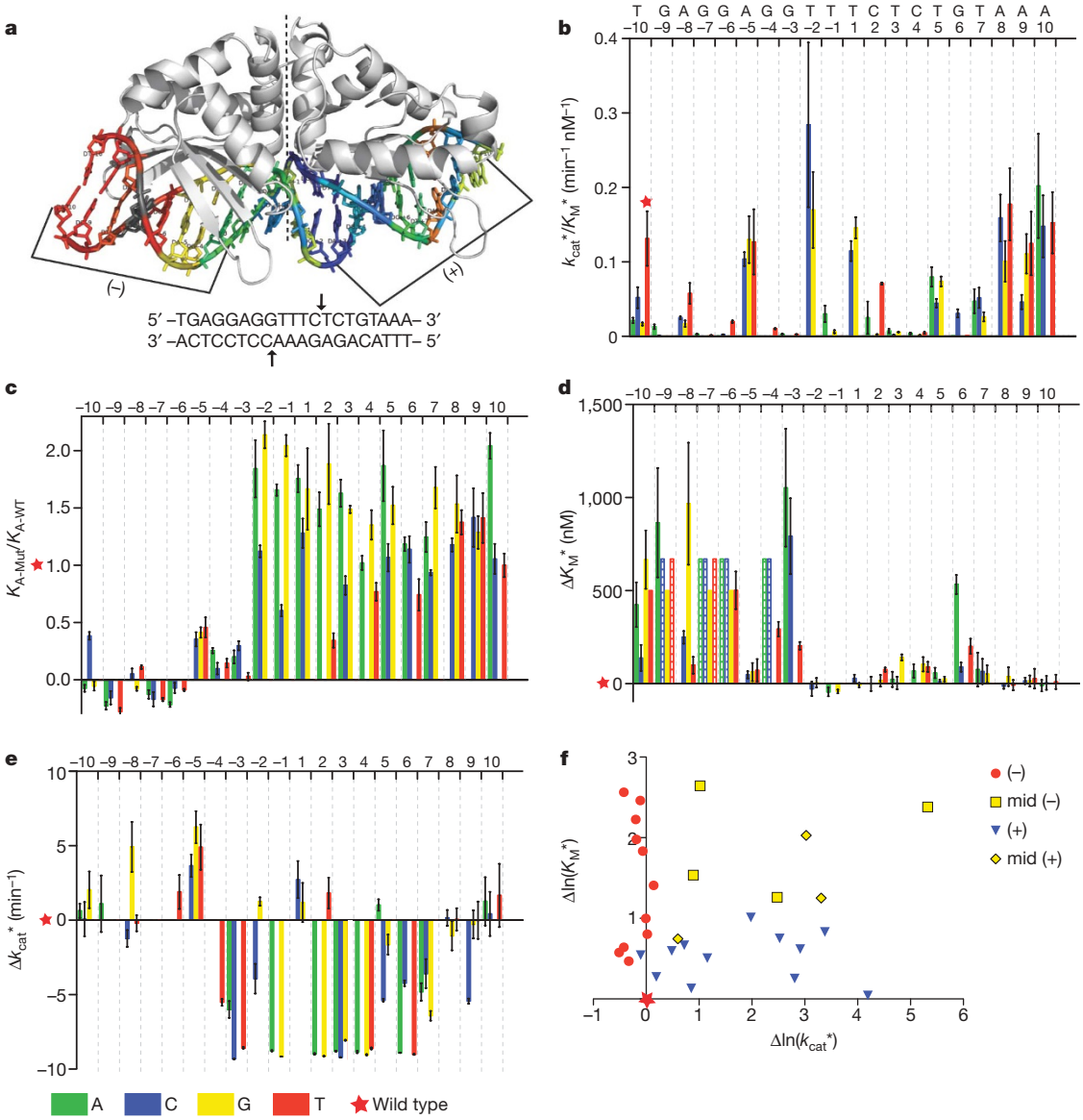

Figure 1 | Segregation of contributions to binding and catalysis. **a**, Ribbon diagram of the I-AniI enzyme in complex with the wild-type target site (2QOJ[11]). Target site and positions of DNA cleavage are shown below: (−) side cleavage site is cut before (+) side site[10]. **b–e**, Colour scheme: A, green; C, blue; G, yellow; T, red; error bars, s.e.m. **b**, k_{cat}^*/K_M^* values for the wild-type target site (red star) and each of the 60 singly substituted target sites (vertical bars). Substitutions throughout the length of the target site abrogate enzyme activity, demonstrating the high sequence specificity of the enzyme. **c**, Relative binding affinities determined for each singly substituted target site using fluorescence competition assays. Substitutions on the left side, but not the right side, significantly reduce binding affinity. **d**, K_M^* values for each singly substituted target site relative to the wild type. As in **c**, substitutions on the left but not the right display significantly different values from wild type. **e**, k_{cat}^* values for each singly substituted target site relative to the wild-type site. In contrast to **c** and **d**, substitutions between positions −4 and +9 have significant effects. Substitutions for which K_M^* was too high (>750 nM) to allow separate determination of k_{cat}^* and K_M^* are indicated by bars with dashed lines in **d**, and are left blank in **e**. **f**, Asymmetry of the contributions to k_{cat}^* and K_M^*. Positions shown in red are on the left (−) side of the target site from −10 to −5 and almost exclusively contribute to K_M^*. Positions shown in blue are on the right (+) side of the target site from positions +3 to +7. The boundary positions, −4, −3 and +6, contribute to both k_{cat}^* and K_M^* and are shown in yellow. To portray the structural context of these positions, the target site in **a** is coloured based on the effect of the mutation on k_{cat}^*, normalized by the sum of the effects on k_{cat}^* and K_M^* ($[|\Delta\ln(k_{cat}^*)|/|(\Delta\ln(K_M^*)| + |\Delta\ln(k_{cat}^*)|)]$): close to 1.0, blue; close to 0.0, red; intermediate, yellow; position where K_M^* and k_{cat}^* could not be separately determined; grey).

site (blue) that reduced cleavage in solution also reduced enzyme activity in the tethered cleavage assay, consistent with their reduction of k_{cat}^*. Substitutions that disrupt interactions involved in selective transition-state stabilization cannot be overcome by increasing the local concentration of substrate.

Assuming the simple free-energy diagram in Fig. 2a, we can make inferences from the kinetic data in solution and on the yeast surface about the structures of the Michaelis and transition-state complexes. Side-chain–base-pair interactions from positions −10 to −5 are present in both the Michaelis complex and the transition state (base substitutions increase K_M^* and K_D in solution and do not affect k_{cat}^* or the rate when tethered). Sequence-specific base-pair interactions from +3 to +8 are formed only in the transition state (substitutions have no effect on K_M^* or K_D, reduce k_{cat}^*, and slow the rate when tethered). A third class of interactions (at −5 and +7 for example) appear to be formed in the Michaelis complex but not the transition state (substitutions increase or decrease both k_{cat}^* and K_M^*/K_D).

Importantly for the design calculations described in the next section, three observations suggest that the crystal structure of the complex likely resembles the transition state more than the Michaelis complex: (1) specific interactions on the (+) side of the DNA target present in the crystal structure appear to be formed in the transition

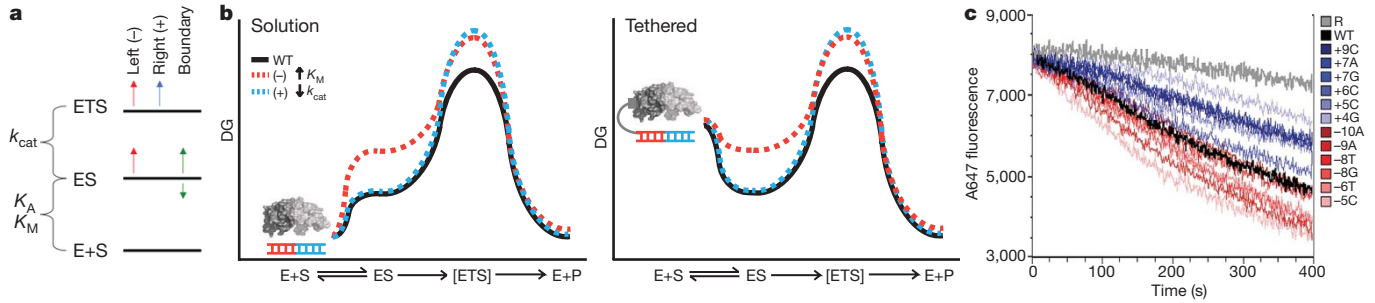

Figure 2 | Contributions to catalysis. **a**, Free-energy diagram showing the effect of target-site substitutions on the free energies of substrate binding and transition-state stabilization. ETS, enzyme–transition-state complex; ES, enzyme–substrate complex; E+S, free enzyme and substrate; E+P, free enzyme and product; DG, free energy difference. Most substitutions on the left side increase K_M*, and hence the free energy difference between ES and E+S, but leave K_{cat}*, and hence the free energy difference between ETS and ES, unchanged. These substitutions thus seem to disrupt interactions made in both states. Most substitutions on the right side instead leave K_M* unchanged but decrease K_{cat}*, and hence raise only ETS, suggesting they remove interactions present in ETS but not ES. A small subset of positions (labelled boundary) appears to stabilize or destabilize ES selectively while not affecting ETS; these substitutions may disrupt interactions present in ES but not in ETS. **b**, Free-energy profiles for a free (left) and tethered (right) system. Red profile: substitutions that remove interactions present in both ES and ETS; blue line, substitutions that remove interactions present in ETS but not ES. Tethering increases the free energy of free E + S to the point that the rate depends only on the free-energy difference between ES and ETS. Because this free-energy difference is unchanged by substitutions that remove interactions made in both ES and ETS (red profile), they do not affect the rate in the tethered case. **c**, Cleavage kinetics of endonuclease tethered to yeast cells. Surface displayed enzyme cleaves a tethered fluorescently labelled oligonucleotide, which then diffuses away from the yeast surface resulting in loss of fluorescence. Black, wild-type target site; random DNA, grey; shades of red, left-side target-site substitutions; shades of blue, right-side target-site substitutions. Tethering suppresses decreases in cleavage rate produced by (−) side but not (+) side mutations. The rate increases observed for some (−) side variants such as −5c were consistent across multiple independent experiments and with the *in vitro* results in Fig. 1e.

state but not the Michaelis complex; (2) the third class of substitutions mentioned above that appear to stabilize only the Michaelis complex make few interactions in the crystal structure (Supplementary Fig. 4); and (3) calculations of Rosetta specificity based on the crystal structure correlate better with catalytic efficiency than with binding affinity (Supplementary Fig. 5).

Monomeric LAGLIDADG homing endonucleases, which recognize non-palindromic targets, are attractive scaffolds for genome engineering applications[15]. An important challenge is to reprogram the substrate specificity of these enzymes towards desired target sequences[9]. To redesign I-AniI specificity using Rosetta[16], the target site in the crystal structure of the I-AniI protein–DNA complex is mutated *in silico* and the program searches for combinations of amino-acid substitutions that allow the formation of energetically favourable interactions with the new base pairs, but not with the wild-type base pairs[16]. Design calculations were performed for six target-site variants bearing single base-pair substitutions, genes encoding the amino-acid sequences of eight redesigned enzymes were constructed and the enzymes purified. DNA cleavage assays revealed that the designed specificity changes were for the most part achieved (Supplementary Table 2). These results demonstrate that I-AniI cleavage specificity can be reprogrammed by computational protein design, thereby providing starting points for the larger-scale specificity changes required to cleave physiological target sites.

An enzyme redesigned for a new target site could achieve altered specificity either by changing k_{cat}*, changing K_M* or changing both. To determine whether the designed changes in specificity were a result of changes in K_M* or k_{cat}*, for each of eight designed endonucleases we measured the single-turnover cleavage kinetics for target substrates containing each of the four possible base pairs at the redesign position (Supplementary Table 2). A design aimed at specific recognition of a DNA target site containing base pair −8G:C (Fig. 3a) achieved specificity exclusively by modulating K_M*: K_M* decreased for G:C, and increased for A:T, T:A and C:G. In contrast, a design aimed at specific recognition of +8C:G (Fig. 3b) achieved specificity entirely through k_{cat}*: k_{cat}* decreased for A:T, G:C and T:A, but was unchanged for +8C:G. Both of these designed enzymes have high specificity at neighbouring base pairs, and overall specificities that are higher than the wild-type enzyme in the targeted regions (Supplementary Fig. 6). A design aimed at specific recognition of the −3C:G substitution (Fig. 3c), at the boundary between K_M*- and k_{cat}*-influencing positions (Fig. 1), displayed changes in both k_{cat}* and K_M*, consistent with the results with the wild-type enzyme at this position. These trends hold for the remaining designs as well (Supplementary Table 2 and Supplementary Fig. 7): we find generally that the left-side designs achieve specificity primarily by modulating ground-state binding affinity, whereas the right-side designs achieve specificity by modulating the stability of the transition state.

Our results suggest that initial binding of I-AniI to its target site involves formation of base-specific interactions on the (−) side and lower-affinity non-specific interactions on the (+) side to form the Michaelis complex (the latter are suggested by yeast display experiments which show that the enzyme binds less tightly to the (−) half-site than to the full site (J.J., B.L.S. and A.M.S., unpublished observations)). Catalysis then requires bending of the DNA (note bend in Fig. 1a), which is stabilized at the transition state by newly formed specific interactions between the (+) side and the enzyme. Such a two-stage mechanism (see Supplementary Information section C) may be a general solution to the problem of specific target-site recognition by enzymes that act on distorted DNA substrates. If the enzyme only bound to the distorted site, binding would require enzyme to be at the site (which may occur only once in the genome) simultaneous with fluctuation of the DNA into the distorted conformation; because both are rare events, the net rate of binding, the product of two small numbers, would be very slow. If, instead, the enzyme can bind with some sequence specificity to undistorted target sites, the probability of being close enough to capture (and perhaps promote) fluctuations that distort the DNA will be very much higher. In I-AniI the total transition-state binding energy appears to be roughly divided between the two steps: the N-terminal domain guides the enzyme to potential target sites that match on the (−) side, and the C-terminal domain specifically stabilizes the transition state if there is also a match on the (+) side.

There is considerable synergy between classical enzymology and modern computational design. Design should be informed by detailed analyses of the wild-type enzyme because, depending on the enzyme and substrate concentrations in the application the designed enzymes are to be used for, it may be necessary to re-engineer K_M, k_{cat} and/or k_{cat}/K_M. Conversely, computational design can provide insight into the basis for transition-state stabilization. Our kinetic dissection of I-AniI cleavage activity also has implications for endonuclease

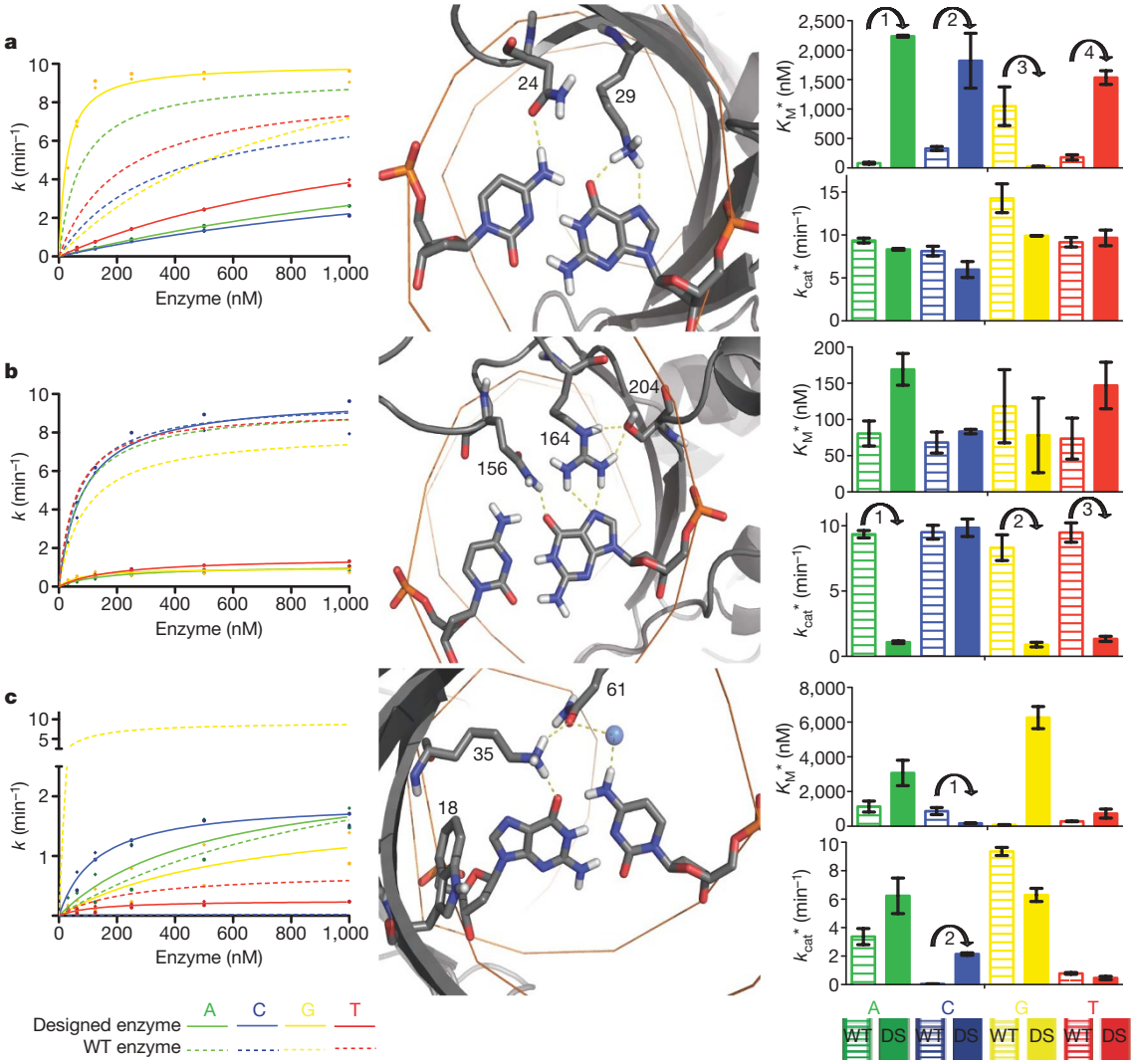

Figure 3 | Computational redesign of specificity. Colour scheme: A, green; C, blue; G, yellow; T, red; error bars in right panels, s.e.m. **a**, Design for −8A:T to −8G:C substitution (K24N, T29K). Middle panel: the designed residues N24 and K29 make direct hydrogen bonds to −8G and −8C, respectively. Left panel: the concentration dependence of the cleavage activity for the designed enzyme (solid lines) for different base pairs at the −8 position differs considerably from the wild-type enzyme (dashed lines). Right panel: the k_{cat}* values remain approximately the same for both the wild-type and designed enzymes against all target sites, but the K_M* values are decreased for the target G base pair (arrow 3) and increased significantly for the other three substitutions (arrows 1, 2 and 4). **b**, Design for +8A:T to +8C:G substitution (L156Q, I164R, T204S). Middle panel: designed residues R164 and Q156 make direct hydrogen bonds to +8G. Designed residue S204 holds R164 in position. The kinetic traces (left panel) and bar graphs (right panel) show this design achieves altered specificity through changing k_{cat}*. The K_M* values remain approximately the same for both the wild-type and designed enzymes against all target sites, but the k_{cat}* values are significantly decreased for all of the competitor target sites (arrows 1, 2 and 3). **c**, Design for −3G:C to −3C:G substitution (Y18W, E35K, R61Q). Middle panel: designed residues K35 and Q61 make a direct hydrogen bond to −3G and a water-mediated hydrogen bond to −3C, respectively. Q61 and K35 also hydrogen bond with each other, and designed residue W18 further helps position K35 through packing interactions. The kinetic traces (left panel) and bar graphs (right panel) show this design achieves altered specificity through changing both k_{cat}* and K_M*. The designed enzyme has an increased k_{cat}* (arrow 2) and decreased K_M* for the −3C (arrow 1).

re-engineering using yeast display: selection based on binding may be sub-optimal because substrate binding could be optimized at the expense of transition-state stabilization, whereas selection for cleavage in the tethered substrate system could yield variants with decreased solution cleavage due to increased K_M*. These pitfalls could potentially be overcome by selecting both for k_{cat}* and K_M*, perhaps by alternating between the two selection procedures. More generally, the union of classical enzymology with modern computational design and selection technology, as illustrated here, provides a powerful approach to revealing the mechanistic basis for, and subsequently reprogramming of, sequence-dependent molecular recognition.

METHODS SUMMARY

Experimental preparation and kinetic analysis. I-AniI was expressed in *Escherichia coli* BL21(DE3) using a standard auto-induction protocol and purified over a His-trap column. Linearized plasmid substrates were prepared for each of the 60 singly substituted target sites. Kinetic assays were performed over a 20-fold range of enzyme concentrations (from 30 to 1,500 nM, depending on the substrate) with 5 nM DNA substrate, and analysed by agarose gel electrophoresis followed by integration of product and substrate band densities. The velocity versus enzyme concentration profiles were determined two to four independent times; reported k_{cat}* and K_M* values are the average of those determined from the independent experiments. Fluorescence competition-binding assays were performed as previously described[17].

Computational design. New target sequences were mapped on to the I-AniI-DNA crystal structure (2QOJ[11]) and the Rosetta computational design methodology was used to optimize the amino-acid sequence of the protein to maximize affinity for the new site[16]. The predicted specificity of the resulting protein models for the desired target sequence was computed using Rosetta, and designs that were predicted to bind tightly and specifically were subjected to further optimization using flexible backbone protein design (Supplementary Methods). The tightest binding and most specific designs were again selected, and the designed amino-acid substitutions were removed one at a time. If no

significant loss was predicted in either specificity or binding energy, the substitution was removed from the design. The '-8G:C_A' (K24N/T29K) and '-8G:C_B' (K24N/T29Q) designs were generated instead using a genetic algorithm to optimize binding affinity and specificity simultaneously (Supplementary Methods). Genes encoding the designed proteins were assembled from oligonucleotides, and the designed proteins were expressed, purified and assayed as described above.

Tethered cleavage on yeast surface. PCR-generated DNA substrates, labelled with biotin and Alexa 647, were tethered by an antibody–streptavidin–phycoerythrin bridge to the haemagglutinin epitope of I-AniI expressed on the surface of *S. cerevisiae* in conditions that prohibited catalysis. Samples were then spiked with 10 mM $MgCl_2$ and placed in a pre-warmed chamber at 37 °C and fluorescence measurements were acquired on a flow cytometer. The Alexa 647 signal from a phycoerythrin-normalized population of each sample was then plotted against time to generate the curves shown.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 June; accepted 15 September 2009.

1. Jencks, W. P. Mechanism of enzyme action. *Annu. Rev. Biochem.* **32,** 639–676 (1963).
2. Bolduc, J. M. *et al.* 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 (2003).
3. Wells, T. N., &. Fersht, A. R. Use of binding energy in catalysis measured by mutagenesis of tyrosyl-tRNA synthetase. *Biochemistry* **25,** 1881–1886 (1986).
4. Fersht, A. R. Relationships between apparent binding energies measured in site-directed mutagenesis experiments and energetics of binding and catalysis. *Biochemistry* **27,** 1577–1580 (1988).
5. Benkovic, S. J. & Hammes-Schiffer, S. A perspective on enzyme catalysis. *Science* **301,** 1196–1202 (2003).
6. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453,** 190–195 (2008).
7. Collins, C. H., Yokobayashi, Y., Umeno, D. & Arnold, F. H. Engineering proteins that bind, move, make, and break DNA. *Curr. Opin. Biotechnol.* **14,** 371–378 (2003).
8. Smith, J. *et al.* A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.* **34,** e149 (2006).
9. Redondo, P. *et al.* Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature* **456,** 107–111 (2008).
10. Takeuchi, R., Certo, M., Caprara, M. G., Scharenberg, A. M. & Stoddard, B. L. Optimization of *in vivo* activity of a bifunctional homing endonuclease and maturase reverses evolutionary degradation. *Nucleic Acids Res.* **37,** 877–890 (2008).
11. Scalley-Kim, M., McConnell-Smith, A. & Stoddard, B. L. Coevolution of a homing endonuclease and its host target sequence. *J. Mol. Biol.* **372,** 1305–1319 (2007).
12. Halford, S. E., Johnson, N. P. & Grinsted, J. The EcoRI restriction endonuclease with bacteriophage lambda DNA. Kinetic studies. *Biochem. J.* **191,** 581–592 (1980).
13. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Analysis and Protein Folding* (W. H. Freeman, 1998).
14. Gai, S. A. & Wittrup K. D.. Yeast surface display for protein engineering and characterization. *Curr. Opin. Struct. Biol.* **17,** 467–473 (2007).
15. Perez, E. E. *et al.* Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nature Biotechnol.* **26,** 808–816 (2008).
16. Ashworth, J. *et al.* Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* **441,** 656–659 (2006).
17. Zhao, L., Pellenz, S. & Stoddard, B. L. Activity and specificity of the bacterial PD-(D/E)XK homing endonuclease I-*Ssp6803*I. *J. Mol. Biol.* **385,** 1498–1510 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by a National Science Foundation graduate research fellowship to S.B.T., the US National Institutes of Health (GM084433 and RL1CA133832), the Foundation for the National Institutes of Health through the Gates Foundation Grand Challenges in Global Health Initiative, and the Howard Hughes Medical Institute. We thank A. Quadri for help with plasmid substrate preparation and M. Scalley-Kim for I-AniI cleavage data collected in the presence of Mn^{2+}.

Author Contributions S.B.T and J.J.H. performed computational design calculations and S.B.T. performed kinetic characterization of all designed and wild-type enzymes. R.T. performed the fluorescence competition binding experiment. J.J. performed the surface-expressed tethered cleavage assay. J.A. and J.J.H. developed computational design procedures. S.B.T. and D.B. wrote the paper. Multiple discussions of shared data among all authors at Northwest Genome Engineering Consortium (http://research.seattlechildrens.org/centers/immunity_vaccines/ngec/) group meetings contributed to the recognition of binding/catalysis asymmetry in I-AniI Y2 and the conceptual development of this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.B.T. (sthyme@u.washington.edu) or D.B. (dabaker@u.washington.edu).

METHODS

Protein expression and purification. Genes encoding Y2 I-AniI[10,18] designs were assembled from oligonucleotides[19], cloned into a variant of the pet15 expression vector and sequence-verified plasmids were transformed into BL21 Star (Invitrogen). A 1-l culture of auto-induction media[20] was inoculated with several colonies, grown at 37 °C for about 12 h (to approximately saturation) and expression at 18 °C continued for about 24 h. Cells were harvested, resuspended in Tris 20 mM pH 7.5, 1.0 M NaCl and 30 mM imidazole, and lysed by sonication and lysozyme. The soluble fraction was loaded onto a 1 mL HisTrap FF crude column (GE Healthcare) and I-AniI variants were purified by imidazole gradient elution on an AKTA express (GE Healthcare). The proteins were concentrated and the buffer was exchanged to Tris 20 mM pH 7.5, 500 mM NaCl and 50% (v/v) glycerol for storage. Purity of the proteins was assessed by SDS–polyacrylamide gel electrophoresis and the concentration of samples with about >95% purity was determined by measuring the absorbance at 280 nm using the calculated extinction coefficient[21]. The concentration of enzyme in the <95% pure samples was determined by generating a standard curve with a pure I-AniI protein, correlating protein concentration with band density (calculated with ImageJ (http://rsbweb.nih.gov/ij/)) and comparing the band density of the I-AniI protein in impure samples run on the same gel as the standard curve.

Plasmid substrate construction. All single base-pair variants from the wild-type target site in pBluescript were individually constructed by site-directed mutagenesis as described[22]. Sequence-verified plasmids were linearized with ScaI before the kinetic assays to facilitate product identification.

Endonuclease activity assays. Kinetic assays. Previous work[23] has confirmed that I-AniI, like other LAGLIDADG endonucleases, is a single-turnover enzyme, and the conditions for single-turnover kinetics[12] were met in all experiments. The ionic strength of the enzyme reaction buffer was optimized for enzyme activity and stability to a final solution of 170 mM KCl, 10 mM $MgCl_2$ and 20 mM Tris, pH 9.0. Enzyme was diluted in 1.25× reaction buffer to working concentrations, serial twofold dilutions were made, and both substrate plasmid and diluted enzyme were incubated separately at 37 °C for 1 min. The appropriate amount of plasmid (one-fifth of the reaction volume) was added to each reaction for a final 1× reaction buffer and final plasmid concentration of about 5 nM (the lowest concentration still readily visible on agarose gel). The plasmid (one-fifth of reaction volume) was added to the enzyme (four-fifths of reaction volume) to minimize heat loss during the transfer (which was found to add significant noise to the data). Reactions were halted with 200 mM EDTA, 30% glycerol and bromophenol blue. DNA fragments were separated on 1.2% agarose TBE gels, which were then stained in a standard ethidium bromide solution and subsequently destained in water for maximum contrast between DNA and background. All data were collected by integrating the density of the substrate (2,959 bp) and product bands (1,801 bp and 1,158 bp) using ImageJ (http://rsbweb.nih.gov/ij/). The percentage of product formed is equal to the sum of the density of the two product bands divided by the total sum of the densities of the three bands. The progress curves fit to single exponentials for all enzyme concentrations (Supplementary Fig. 1) and for all target sites except for several substitutions in the central four base pairs between the cleavage sites on the two DNA strands (Supplementary Fig. 2).

Assays for specificity positions adjacent to designed nucleotide. Twofold serial dilutions of enzyme from 1500 to 11 nM were made in 1.25× reaction buffer and the enzyme was reacted with about 5 nM substrate (in 1× reaction buffer) for half an hour at 37 °C. Reactions were halted and data were analysed as described in the 'Kinetic assays' section.

Fluorescence competition binding assay[17]**.** Unlabelled DNA oligonucleotides with each of the 60 single base-pair substitutions in the I-AniI target site (wild-type I-AniI site, 5′-TGAGGAGGTTTCTCTGTAAG-3′), a negative control sequence (5′-CTCTTCTTGCATATATCTCC-3′), an unlabelled wild-type site oligonucleotide and a wild-type site oligonucleotide labelled with 5′ Cy3 were synthesized with six consecutive 'A' flanking on each end (Integrated DNA Technology, 100-nmol scale, salt-free). Complementary oligonucleotides were ordered for all 63 sites and double-stranded target DNA was prepared by annealing equal amounts of complementary strands.

His-tagged I-AniI was immobilized by incubating 200 µl of 100 nM I-AniI in TBS/BSA buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.2% BSA) in wells of Nickel-NTA coated HisSorb Plates (Qiagen) for 2 h at room temperature. Unbound protein was removed and the plates were washed four times with TBS/Tween-20 (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.05% Tween-20). The immobilized I-AniI in the microtitre plate was incubated for about 4 h with both 100 nM labelled target DNA duplex and 3 µM (30-fold excess) of one unlabelled duplex per well in 200 µl of binding buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.02 mg ml^{-1} poly(dI-dC), 10 mM $CaCl_2$). The plates were washed four times with TBS (50 mM Tris-HCl (pH 7.5), 150 mM NaCl), and the fluorescent signal retained in each well was quantified using a SpectraMax M5/M5e Microplate Reader (Molecular Devices) (excitation, 510 nm; emission, 565 nm; cutoff, 550 nm). Additional negative control experiments performed in the absence of the enzyme indicated that no significant detectable fluorescent signal was retained after the protocol described above was completed. Relative binding affinities were calculated using the following equation: relative binding affinity $= [(F_n - F_x) \times F_t]/[(F_n - F_t) \times F_x]$, where F_x, F_t and F_n indicate fluorescent intensities obtained from wells in which the immobilized protein was incubated with the unlabelled singly substituted target sites, wild-type target site and negative control sequence, respectively.

Cleavage kinetics of endonuclease tethered to yeast cells. Surface display of I-AniI on *S. cerevisiae* was performed using standard methods[14]. For each sample, 5×10^6 cells were stained with biotinylated anti-HA11, followed by secondary staining with streptavidin-DNA substrate conjugates (1:3 molar ratio) on ice and in the absence of divalent cations. DNA substrates were generated by PCR using biotinylated and Alexa 647-conjugated primers complementary to the 5′ and 3′ sequences flanking the I-AniI target (or indicated target-site variants). PCR products were purified by Exo1 digestion followed by size exclusion chromatography on a G-100 column (GE Healthcare) before conjugation with streptavidin–phycoerythrin. Samples were then spiked with 10 mM $MgCl_2$ and placed in a pre-warmed 37 °C chamber and acquired at an approximate event rate of 3,000 s^{-1} for 400 s on a BD FACSAria II flow cytometer. Processing was performed using FloJo software (Treestar). Live cells were gated by forward- and side-scatter properties, and doublets and clumped cells were excluded on the basis of forward scatter area versus height linearity. The Alexa 647 signal from a phycoerythrin-normalized population of each sample was then plotted against time to generate the curves shown.

Computational design methods. Single-state design (designs −9C, −8G_C, −6C, +5C and +8C). The computational design of homing endonuclease–DNA specificity was performed using the Rosetta design software in a manner specifically designed to predict new protein sequences that would bind with high affinity to novel DNA sequences[24]. The prediction of designed proteins with novel interactions to substituted base pairs in the I-AniI recognition sequence was performed by mutation and Monte Carlo repacking of amino-acid side chains as described in Ashworth *et al.* 2006 (ref. 16). The template for the design calculations was the crystal structure of the I-AniI-DNA complex (Protein Data Bank accession number 2QOJ[11]). Additionally, minor shifts of the protein backbone were modelled only in the vicinity of the designed region using a loop-rebuilding algorithm[25,26]. The specificity of each hypothetical new protein sequence for the intended new DNA recognition sequence was calculated as the Boltzmann probability of the intended complex versus a partition function consisting of each base-pair possibility at the redesigned DNA base pair[27]. After design, predicted protein sequences with the most favourable binding energy and highest predicted specificity were reverted position by position to the wild-type amino-acid sequence to identify (and revert) designed mutations that did not significantly contribute to the energy or specificity of the designed complex.

Multi-state design (designs −8G_A and −8G_B). Two base-pair positions in the structure were computationally mutated to generate a partial match to a recognition site in the *IL-2Rγ* gene in a mouse model of severe combined immunodeficiency disease. Specifically, positions −9G:C and −8A:T were modelled as −9A:T and −8G:C. A multistate design calculation[28] was performed to select amino acids at positions 24Z, 26Z, 27Z and 29Z. Three states were included in the design. The first state was the target state, which was modelled using the altered DNA structure. The second state was the original structure with the wild-type DNA sequence, which served as a competitor to enforce binding specificity of the selected proteins for the altered recognition site (negative design state). The third state was the modelled structure of the best single-state design for the target state with the modified DNA sequence; the energy associated with this state is a constant during the multi-state design procedure. It represents the best scoring protein-altered DNA complex as assessed with the Rosetta energy potential, and it is therefore impossible for the energy associated with the target state to be lower than this value. As a result, multiple calculations were performed which differed from each other only in an artificial offset applied to the third state. Progressively larger offsets bias the calculations to select sequences that achieve higher specificity for the first state over the second state at the expense of achieving Rosetta scores that are allowed to be progressively worse than the third state.

A genetic algorithm was used to evolve a population of sequences that prefer the target state to the two competitors. An initial population of 2,000 sequences was generated by selecting random amino acids at the four design positions. The side-chain conformations of these four residues (with the rest of the protein and DNA structure held fixed) were predicted for the first and second states using a Monte Carlo algorithm, and the Rosetta score recorded. As noted above, the energy of the third state is a constant. A 'fitness' score for each sequence i in the population is calculated: $\text{fitness}_i = E_{\text{target}} - \langle E_{\text{competitors}} \rangle$, where E_{target} is the energy of the

target state and the angled brackets denote an ensemble (Boltzmann-weighted) average over the energies of the competitors. Conceptually, the fitness corresponds to the transfer free energy of the protein from the ensemble of competitors to the target state. Subsequent generations were constructed using the following procedure. First, the sequence with the best (lowest) fitness was promoted automatically. Next, 1,980 sequences were created by recombining two members of the population using uniform crossover of two parents chosen by tournament selection.[29] Finally, the remaining 19 sequences were generated by mutating a single parent chosen by tournament selection with a 25% chance of randomizing each position in turn. A fitness value was calculated for each new sequence, and the population was propagated for 30 generations.

18. Bolduc, J. M. *et al.* 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 (2003).
19. Stemmer, W. P. C., Crameri, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164,** 49–53 (1995).
20. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).
21. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4,** 2411–2423 (1995).
22. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154,** 367–382 (1987).
23. Geese, W. J., Kwon, Y. K. & Waring, R. B. *In vitro* analysis of the relationship between endonuclease and maturase activities in the bi-functional group I intron-encoded protein, I-AniI. *Eur. J. Biochem.* **270,** 1543–1554 (2003).
24. Havranek, J. J., Duarte, C. M. & Baker, D. A simple physical model for the prediction and design of protein–DNA interactions. *J. Mol. Biol.* **344,** 59–70 (2004).
25. Canutescu, A. A. & Dunbrack, R. L. Jr. Cyclic coordinate descent: robotic algorithm for protein loop closure. *Protein Sci.* **12,** 963–972 (2003).
26. Das, R. *et al.* Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home. *Proteins* **69,** 118–128 (2007).
27. Ashworth, J. & Baker, D. Assessment of optimization of affinity and specificity at protein-DNA interfaces. *Nucleic Acids Res.* **37,** e73 (2009).
28. Havranek, J. J. & Harbury, P. B. Automated design of specificity in molecular recognition. *Nature Struct. Biol.* **10,** 45–52 (2003).
29. Mitchell, M. *An Introduction to Genetic Algorithms* (MIT Press, 1996).