Nucleic Acids Research Advance Access published August 24, 2007

Nucleic Acids Research, 2007, 1–12
doi:10.1093/nar/gkm624

Altered target site specificity variants of the I-PpoI His-Cys box homing endonuclease

Jennifer L. Eklund[1], Umut Y. Ulge[3], Jennifer Eastberg[3,4] and Raymond J. Monnat, Jr[1,2,*]

[1]Department of Genome Sciences, [2]Department of Pathology and [3]the Molecular and Cellular Biology Program, University of Washington, Seattle, WA and [4]Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Received June 24, 2007; Revised July 27, 2007; Accepted July 30, 2007

ABSTRACT

We used a yeast one-hybrid assay to isolate and characterize variants of the eukaryotic homing endonuclease I-PpoI that were able to bind a mutant, cleavage-resistant I-PpoI target or 'homing' site DNA *in vivo*. Native I-PpoI recognizes and cleaves a semi-palindromic 15-bp target site with high specificity *in vivo* and *in vitro*. This target site is present in the 28S or equivalent large subunit rDNA genes of all eukaryotes. I-PpoI variants able to bind mutant target site DNA had from 1 to 8 amino acid substitutions in the DNA–protein interface. Biochemical characterization of these proteins revealed a wide range of site–binding affinities and site discrimination. One-third of variants were able to cleave target site DNA, but there was no systematic relationship between site-binding affinity and site cleavage. Computational modeling of several variants provided mechanistic insight into how amino acid substitutions that contact, or are adjacent to, specific target site DNA base pairs determine I-PpoI site-binding affinity and site discrimination, and may affect cleavage efficiency.

INTRODUCTION

Proteins that bind DNA play a critical role in regulating gene structure, replication and expression in all organisms. Biochemical and structural analyses of proteins that bind specific DNA sequences have begun to provide insight into the molecular basis of both DNA binding and sequence-specific DNA recognition (1–3). These analyses have identified protein folds for DNA binding, together with a few general rules for the protein-mediated recognition of specific DNA bases (1,4,5). This work has also begun to suggest ways to modify the recognition specificity of existing, sequence-specific DNA-binding proteins. Two classes of site-specific DNA-binding proteins that have been the focus for efforts to engineer new DNA recognition specificities are the Type II restriction endonucleases (6), and sequence-specific transcription factors (1,7).

The homing endonucleases, a group of highly sequence-specific DNA-binding proteins, are also being investigated for recognition specificity engineering. Four different families of homing endonuclease proteins have been identified on the basis of protein sequence comparisons, and one or more families have been identified in all Kingdoms of life (8,9). The physiologic role of homing endonucleases is to target the lateral transfer of parasitic DNA elements known as mobile introns by making a highly sequence-specific DNA double-strand break in an intron-less recipient allele (8,9).

The high site specificity of many homing endonucleases reflects a combination of long (15–40 bp) DNA target or 'homing' sites, together with a high degree of sequence specificity at most target site base-pair positions. A second intrinsic property of many homing endonucleases is tight coupling of site recognition to catalysis (10). This is a particularly attractive feature of homing endonucleases, in contrast to other potential genome engineering reagents such as zinc finger nucleases (11). High site specificity and tight coupling of site binding to catalysis may reflect the evolutionary history of many homing endonucleases: these two properties in concert permit the continued lateral transfer—and thus the persistence—of endonuclease-encoding mobile introns to related target sites, while minimizing spurious chromosome cleavage events (12,13).

Our aim was to determine whether structure-guided protein engineering could be used to alter the DNA recognition specificity of the eukaryotic homing endonuclease I-PpoI, a member of the His-Cys box family of homing endonucleases. I-PpoI was originally identified as an open reading frame in a self-splicing mobile intron found in extrachromosomal copies of the 28S rRNA genes of Physarum polycephalum, a Plasmodial myxomycete slime mold (14). It is the best characterized member of the His-Cys box family of homing endonucleases, one family

*To whom correspondence should be addressed. Tel: 206 616 7392; Fax: 206 543 3967; Email: monnat@u.washington.edu
Present address:
Jennifer L. Eklund, University of Michigan School of Education, Ann Arbor, MI, USA.

in the ααβ-Me or His-Me endonuclease superfamily (15,16), and has not thus far been a focus for structure-guided design.

The active form of I-PpoI is a 36 kDa homodimer that cleaves a 15-bp semi-palindromic DNA target site in the 28S Physarum rDNA locus, and in the corresponding large subunit rRNA genes of all eukaryotes [(17–19); Figure 1]. Rare target sites may also exist outside the rDNA repeats (20). We had previously determined high-resolution apo- and co-crystal structures of native and mutant forms of I-PpoI that allowed us to identify the molecular basis for high affinity site binding and cleavage (21,22). We had also identified amino acid changes that interfere with I-PpoI catalysis or site binding, together with DNA base-pair changes that disrupted target site cleavage by native I-PpoI protein (23–26). These data were used to construct a yeast one-hybrid (Y1H) screening assay to identify I-PpoI protein variants able to bind a specific mutant target site in vivo. Biochemical characterization together with computational modeling were used to gain insight into the molecular basis for mutant site recognition and cleavage by variant I-PpoI proteins.

MATERIALS AND METHODS

Plasmids and yeast strains

Yeast reporter plasmids were constructed using the YEp24 two-micron plasmid vector or the integrating plasmid vector pRS404 (27,28). An I-PpoI-specific YEp24-lacZ reporter was constructed by inserting I-PpoI target sites into the SalI site upstream of a cyc5 promoter and lacZ gene. Target site inserts were prepared by annealing phosphorylated oligonucleotides (PPOSITES5 and PPOSITES6; all oligonucleotide sequences are given in Supplementary Data Table 1), or oligonucleotides that when annealed created three oriented copies of native or mutant I-PpoI target sites (oligonucleotides PPOX3_WT and _6 #1-4). Yep24-HIS3Ppo was constructed from the resulting plasmids by replacing the lacZ gene with a PCR fragment containing the budding yeast HIS3 gene.

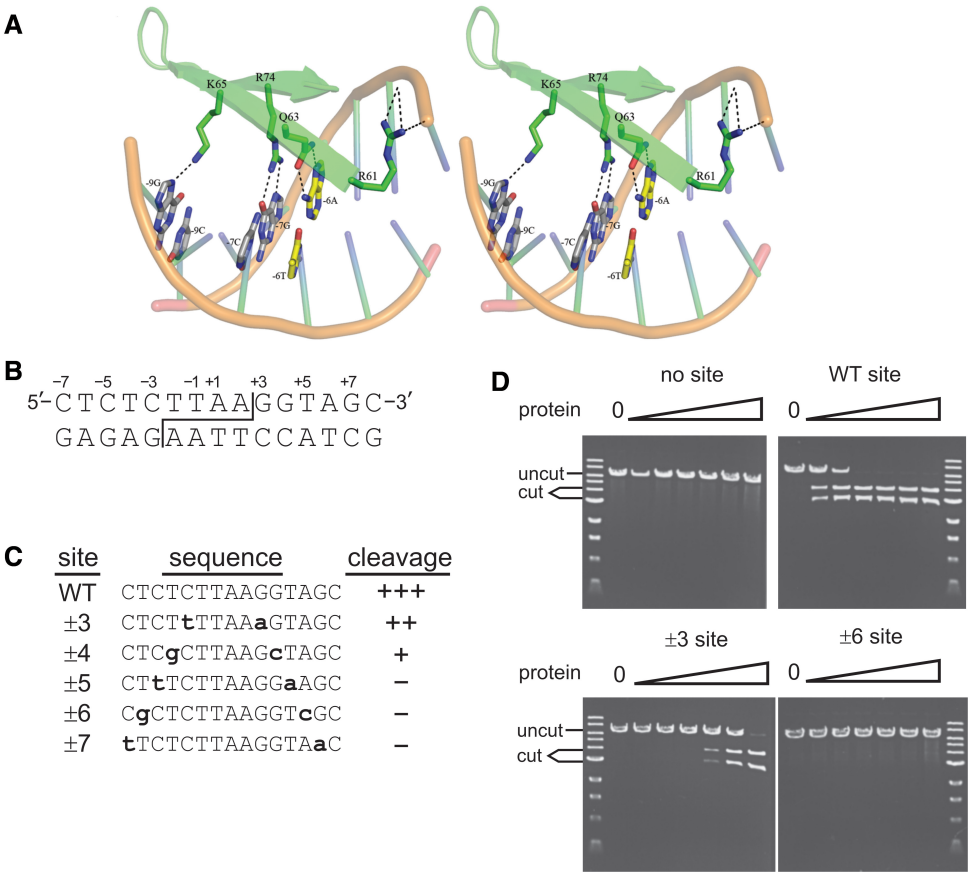

Figure 1. Amino acid geometries were determined from the I-PpoI co-crystal structure (21). (**A**) Stereo representation of the native I-PpoI DNA–protein interface. Canonical contacts are present between Q63:±6A and R74:±7G. The native ±6A/T base pair is shown in yellow. Amino acid geometries were modeled using RosettaDesign. Sequence-specific hydrogen bonds are shown as black dotted lines. (**B**) Sequence of the native I-PpoI target or 'homing' site. Site cleavage across the minor groove (staggered line) generates complementary 4 base, 3′ OH-extended single-stranded ends. The convention for numbering target site base-pair positions is shown above the top strand. (**C**) Sequence and cleavage sensitivity of native and five variant I-PpoI target sites with palindromic base substitutions (shown lower case bold) at positions ±3 to ±7 assayed for native I-PpoI cleavage sensitivity. Cleavage sensitivity is indicated to the right of each sequence ('+ + +', fully sensitive; '–', no cleavage). (**D**) Representative cleavage assays of native and mutant I-PpoI target sites by native I-PpoI protein. Linearized plasmid DNA (10 μM) was digested for 1 h with 10 pM to 1 μM native I-PpoI prior to agarose gel electrophoresis. ('uncut', plasmid substrates; 'cut', site-specific cleavage products).

An integrating version of Yep24-HIS3Ppo was constructed by transferring an EcoRI fragment containing a multiple-cloning site, minimal promoter and HIS3 gene into the yeast TRP1 integration vector pRS404, and then inserting three copies of the native or +6C/–6G mutant I-PpoI target site as described earlier Yeast containing an integrated copy of pRS404-HIS3Ppo were constructed by transforming MfeI-cleaved linear plasmid DNA into yeast strain W1588-4C (MATa ade2-1 can1-100 his3-11, 15 leu2-3, 112trp1-1 ura3-1 RAD5), followed by selection on SD media lacking tryptophan (29,30). Selection for HIS3 reporter expression was performed by growth on SD medium lacking histidine, followed by a demonstration of growth suppression by 3 mM 3-aminotriazole. All reporter plasmids and strains were verified by DNA sequencing of the I-PpoI target site region.

Generation of I-PpoI protein variants

CR library. Expression plasmids for Ppo-activation domain (AD) fusion proteins were constructed by inserting I-PpoI open reading frame cassettes in-frame into the GAL4 activation domain plasmid pOAD (kindly provided by Stan Fields, University of WA, USA). The corrected map and sequence of this plasmid can be found at: depts.washington.edu/sfields/protocols/pOAD.html. After removing flanking BamHI sites, the open reading frame of catalytically inactive H98A I-PpoI was inserted into the NcoI and PstI sites of pOAD to form pOAD-Ppo. A 3 kb insert or 'dummy' fragment was inserted into the NdeI–BamHI site in the I-PpoI ORF to create pOAD-Ppo + d. This 'dummy' fragment was subsequently replaced with an oligonucleotide insert encoding 25 codons of the I-PpoI ORF to restore the open reading frame and randomize I-PpoI residues N57, R61, Q63, K65 and R74. This insert was constructed by the annealing of three oligonucleotides, PPO(-)5, PPO+ NDEI and PPO NDEI SPLINT and ligation, at 500-fold molar excess, into NdeI-cut pOAD-Ppo + d. The PPO(-)5 oligonucleotide region containing randomized codons was converted to double-stranded DNA by primer extension with Klenow DNA polymerase prior to cleaving the resulting insert and plasmid with BamHI (31), ligation and electroporation into E. coli TB1 cells. The number of independent transformants was determined by plating a small portion of the pooled transformation, and the library was amplified by growth overnight in 1l of L-broth followed by plasmid DNA isolation [(32); Qiagen MegaPrep].

RD1 and RD2 libraries. The RD1 and RD2 libraries were constructed by replacing the region encompassing I-PpoI residues 55–76 with an insert generated by PCR-mediated assembly of degenerate oligonucleotides. Five oligonucleotides (the 4 +6G_B oligonucleotides and PPOOADN were used to assemble the N-terminal portion of the I-PpoI H98A ORF. Two +6G_E oligonucleotides and PPOOADC (RD1) or PPOOADCdeg (RD2) were used to assemble the C-terminal portion of the I-PpoI ORF. PPOADCdeg included a 9 bp randomized 'molecular bar tag' sequence to allow different starting plasmids to be distinguished. Both PCR assembly products were gel purified by polyacrylamide gel electrophoresis (PAGE). The N-terminal assembly product was extended in a second PCR reaction using a pool of 27 (RD1) or 18 (RD2) +6_M oligonucleotides together with PPOOADN. The resulting product was gel-purified and amplified with recombination cloning primers AD70F and AD70R (kindly provided by Stan Fields, University of WA; http://depts.washington.edu/sfields/protocols/protocols.html) to generate a full-length I-PpoI open reading frame insert, and then used for in vivo gap repair of PvuII + NcoI-cleaved pOAD-Ppo plasmid DNA.

Ppo-AD libraries were characterized by sequencing plasmid DNA isolated from independent colonies. Seventy-four CR library plasmids were sequenced, 11 from plasmid preps and 63 from PCR products generated using PADSCREENF and PADSCREENB oligonucleotides. RD1 and RD2 were characterized by sequencing yeast colony PCR products. PCR and the sequencing were performed with PADSEQF and 882 oligonucleotide primers. The resulting forward and reverse sequences were aligned using Sequencher (www.sequencher.com).

Site-directed mutagenesis. The I-PpoI open reading frame was subcloned from Ppo-AD fusion plasmids into a pET11c derivative containing an N-terminal hexahistidine affinity purification tag prior to expression in *E. coli* and purification by Ni-NTA affinity chromatography (Qiagen). The pET11c-histagPpo expression vector was constructed by annealing oligonucleotides pET-histagf and pET-histagr containing the affinity tag, followed by ligation into the NdeI site of pET-Ppo (23), to generate pET11c-histagPpo. Site-directed mutagenesis using a QuikChange protocol (Stratagene) was then used to generate pET vectors expressing variants A, B, C, F, P, Q, R, S, T, Y and H98A/L116A. Variants U, V and W were made by site-directed mutagenesis of the variant C expression vector. Variants D, E, G, H, J, K, L, M and N were constructed by inserting an NcoI–BamHI fragment from RD2 Ppo-AD plasmids into the open reading frame of pET11c-histagPpo.

Library screening and β-galactosidase activity assays

CR or RD library screening was performed by transforming CR or RD plasmid libraries, together with the +6C/-6G *lacZ* reporter plasmid Yep24-*lacZPpo*, into cells possessing an integrated copy of the pRS404-*HIS3Ppo* reporter plasmid. The resulting cells were then selected for growth on histidine-minus media followed by a visual screen for β-galactosidase activity. The number of independent co-transformants was determined by plating a small known fraction of the transformed yeast on SD–leucine–uracil plates. Library screening was performed by dense plating of transformed cells on SD–leucine–uracil–histidine plates supplemented with 3 mM 3-aminotriazole, followed by growth at 30°C for 7 days. Large colonies were then picked into 1 ml 96-well deep blocks containing SD–leucine–uracil media and grown for 7 days without shaking at 30°C.

A colorometric screen for β-galactosidase reporter gene induction was used to identify Ppo-AD plasmids with

in vivo site-binding activity. In brief, cells were pelleted, then resuspended and permeabilized in Z-buffer using SDS and chloroform (33). The β-galactosidase substrate ONPG was added to 0.6 mg/ml, and wells were incubated at 30°C to monitor color change. Incubations were stopped between 2 and 4 h by adding Na_2CO_3 prior to visual screening to identify positive wells by comparison against positive control (H98A Ppo-AD and native I-PpoI site lacZ reporter) or negative control (H98A Ppo-AD and mutant +6C/–6G site lacZ reporter) cells.

Ppo-AD coding plasmids were recovered from activity-positive cells by use of either a Qiagen or Sigma miniprep kit, amplified by transformation and growth in E. coli, and then transformed into yeast strain W1588-4C. We assayed at least eight transformants from each original positive colony for β-galactosidase induction by ONPG assay prior to streaking cells on SD–leucine plates supplemented with 5-fluoroorotic acid (5-FOA) to eliminate the URA3 reporter plasmid. In order to verify site-specific β-galactosidase induction, plasmid DNAs from activity-positive cultures were retransformed as described above with a +6C/−6G or native site reporter plasmid. The DNA sequence of the Ppo-AD open reading frame was determined from the same plasmid preparation by sequencing using a PADSCREENF sequencing primer.

Quantitative β-galactosidase activity assays were performed to further characterize several I-PpoI variants by using a modification of the screening assay as described above (33). Cells were grown overnight at 30°C in SD–leucine–uracil media. Cell density was determined by measuring absorbance at 600 nm, and 2 ml of each culture was pelleted and resuspended in 0.8 ml of Z-buffer followed by the addition of 50 µl 0.1% SDS and 50 µl chloroform with vortexing for 30 s. Activity was measured by adding 160 µl of 4 mg/ml ONPG in Z-buffer followed by incubation at 30°C, and stopped by adding 0.4 ml Na_2CO_3 after the development of yellow color. Cell debris was spun out of reactions prior to measuring absorbance at 420 nm and calculating units of β-galactosidase activity. All reported values represent an average of three independent determinations.

Protein expression and purification

Variant I-PpoI proteins subcloned into pET11c-histagPpo were expressed in E. coli host strain BL21(DE3) (Novagen). Single colonies were picked into 2–3 ml of L-broth supplemented with 100 µg/ml carbenicillin and 0.2% glucose, then grown overnight at 37°C. Overnight cultures (300 µl each) were then diluted into 100 ml of L-broth supplemented with carbenicillin and 1 mM zinc acetate and grown for 2–4 h at 37°C. Protein production was induced by adding 1 mM IPTG. After induction for 3–5 h at 37°C, cells were pelleted and stored overnight at −80°C.

Protein was purified from frozen cell pellets by thawing and resuspending cells in 1 ml of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl and 10 mM imidazole) containing 1 mg/ml lysozyme on ice, followed by the addition of protease inhibitors (1 mM PMSF, 1 µg/ml leupeptin and 1 µg/ml pepstatin) and sonication. RNAse A (10 µg/ml) and DNAse I (5 µg/ml) were added to cell lysates followed by incubation on ice for 15 min. Cell debris was pelleted at 4°C, and Ni-NTA resin (100 µl; Qiagen) was added to the cleared supernatant prior to incubation at 4°C on a roller for 1–2 h. Resin was pelleted from the supernatant in a benchtop centrifuge at 4°C, washed once with 1 ml of 20 mM imidazole buffer (50 mM NaH_2PO_4, 300 mM NaCl and 20 mM imidazole), and then washed twice with 1 ml 50 mM imidazole buffer (50 mM NaH_2PO_4, 300 mM NaCl and 50 mM imidazole). Protein was eluted with 3 × 50 µl washes of 50 mM NaH_2PO_4, 300 mM NaCl and 250 mM imidazole. Eluates were pooled to determine protein concentration by Bradford assay prior to the addition of glycerol to 50% (v/v) and storage at −20°C. Polyacrylamide gel electrophoresis of protein samples followed by Coommassie staining indicated that I-PpoI protein preparations were typically >75% pure with no other major contaminating bands.

Site-binding affinity and cleavage assays

The binding affinity (K_d) and cleavage of I-PpoI target sites were determined as previously described (26). The substrates for binding affinity assays were formed by annealing gel-purified PPOx1 oligonucleotide pairs in which (+) strand oligonucleotides had been end-labeled with ^{32}P. K_d values were determined from a minimum of two, or in most cases three, independent assays. Cleavage assays were performed using linearized plasmids containing either three copies of native or mutant I-PpoI target sites inserted into the SalI site of *pRS404-HIS3Ppo*, or single site PPOx1_Sal oligonucleotides pairs that had been annealed and inserted into pBlueScriptII KS(+) plasmid DNA (Stratagene). pRS404 plasmids were linearized with SnaBI, and pBluescriptII plasmids with XmnI, prior to performing cleavage assays.

Computational modeling of I-PpoI variants

Modeling of the I-PpoI DNA interface of native and variant proteins was done using RosettaDesign (34,35). In brief, changes to the substrate DNA sequence and the amino acid sequence of the protein were simulated *in silico*, and the DNA–protein complex was allowed to relax according to an energy function that mimics protein folding. I-PpoI variants E, G and T were modeled and visualized for inspection using the PyMOL molecular viewer (DeLano Scientific LLC).

RESULTS

Cleavage sensitivity of I-PpoI target site variants

We identified I-PpoI target sites that were cleavage-resistant when challenged with native I-PpoI protein. We focused initially on 5-bp positions, ±3 to ±7, and on base-pair substitutions at these sites that were known to affect cleavage efficiency on the basis of previous homing site degeneracy, biochemical or functional analyses (24,25,36). These base-pair positions are in a well-ordered part of the I-PpoI DNA–protein interface (Figure 1A). Palindromic ±3 G>A/C>T and ±4 G>C/T>G base-pair substitutions

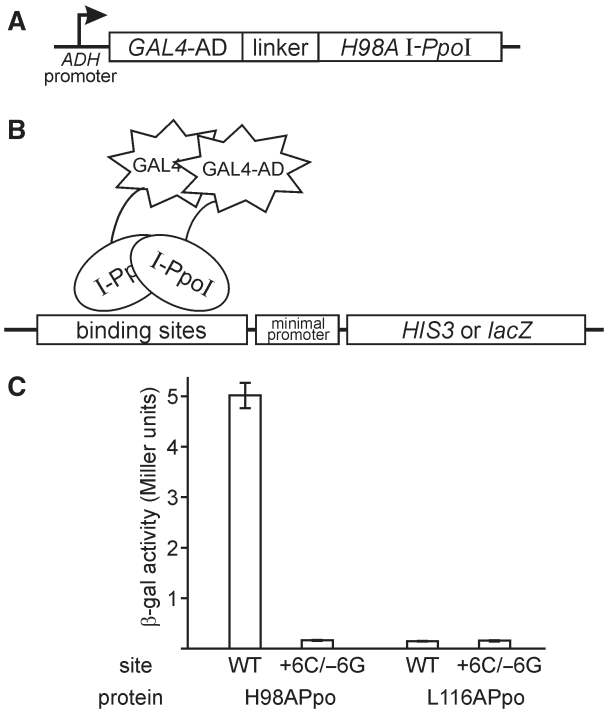

Figure 2. Yeast one-hybrid assay used to identify GAL4-I-PpoI fusion proteins that bind target site DNA *in vivo*. (**A**) GAL4-I-PpoI fusion protein generated by inserting the open reading frame of H98A I-PpoI downstream of the GAL4 transcriptional transactivation domain (GAL4-AD). The H98A substitution abolishes I-PpoI catalytic activity without affecting site-binding affinity. (**B**) Reporter genes to detect *in vivo* binding of GAL4-I-PpoI were constructed in the yeast plasmid YEp24 or in the *TRP1* integration vector pRS404 by inserting three direct repeats of native or mutant I-PpoI target site upstream of a minimal promoter and the budding yeast HIS3 or bacterial *lac*Z gene. (**C**) *In vivo* activity of H98A (positive control) and L116A (negative control) GAL4-I-PpoI fusion proteins in budding yeast, on native or +6C/-6G target site β-galactosidase reporter plasmids. Activity is measured by a quantitative β-galactosidase activity assay (see Methods section).

were permissive of cleavage only at high protein concentrations (10 nM for the ±3 position; 1 μM for the ±4 position). In contrast, base-pair substitutions at positions ±5 (T>A/C>T), ±6 (A>C/T>G) and ±7 (G>A/C>T) completely inhibited cleavage at all protein concentrations tested (Figure 1D; data not shown). Of the six possible base-pair changes from the native +6A/−6T target site, +6A>C/−6T>G substitutions most strongly inhibited site cleavage by native I-PpoI (Figure 1C and D; data not shown). We chose this target site for re-engineering of the site recognition specificity of I-PpoI.

A yeast one-hybrid screening assay

A yeast one-hybrid (Y1H) screen was used to identify I-PpoI protein variants that bound the +6C/−6G target site *in vivo* to induce reporter gene expression. To establish this assay, we fused a *Saccharomyces cerevisiae* Gal4p transcriptional activation domain to the N-terminus of catalytically inactive H98A I-PpoI to generate Ppo-AD. The H98A substitution was incorporated to abolish the catalytic activity of I-PpoI without disrupting high affinity site binding [Figure 2A; (22,23,26)]. A Ppo-AD negative control protein was constructed by fusing L116A I-PpoI to the Gal4p activation domain. L116A I-PpoI is catalytically inactive, and does not bind I-PpoI target site DNA (26). Reporter activity was determined using a budding yeast *HIS3* or plasmid-borne bacterial *lacZ* reporter genes located downstream of 1 or more I-PpoI target site and a minimal promoter (Figure 2B). Site binding was quantified by the induction of β-galactosidase activity, and was strongest when the *lacZ* reporter gene was flanked by 6 direct repeats of the native I-PpoI target site: 25-to-75-fold above the 'no site' or negative control background. The Gal4-L116A Ppo-AD negative control protein did not induce reporter expression above background on any target site or number of repeats (Figure 2C; data not shown). We performed all subsequent library screens using reporter genes flanked by three target sites, as these reporters could be readily constructed and gave reporter activity well above background.

Generation and screening of I-PpoI protein variant libraries

In order to generate protein variants to screen against the mutant +6C/−6G target site, we made substitutions in the I-PpoI DNA–protein interface at residues that contacted base–pair position 6 or adjacent base–pair positions. This 'contacting residue' (CR) library was designed to allow all 20 amino acid residues at positions N57, R61, Q63, K65 and R74, and had a maximum potential complexity of 3.2×10^6 (20^5) protein variants. The resulting CR library was difficult to construct, and thus not large enough to encompass all predicted variants. Upon experimental verification, the CR library was found to encode 1.6×105 different protein variants that each had an average of 4.7 amino acid substitutions (CR library, Table 1).

The CR library was screened by Y1H in two steps. We isolated yeast transformants that grew on histidine-minus plates containing 3-aminotriazole (3-AT), and then determined whether the same Ppo-AD plasmids could induce *in vivo* β-galactosidase expression. The Ppo-AD open reading frame of active variants was then sequenced to identify amino acid substitution(s) that conferred *in vivo* activity. Among the 2×10^6 colony transformants and 4600 His+ colonies screened, only one CR library transformant had *in vivo* activity on both reporter genes.

This variant (variant A, Table 2) had four amino acid substitutions at residue positions randomized during CR library construction.

Variant A served as a starting point for the construction of two additional, rationally designed (RD) libraries that were used to further explore determinants of I-PpoI recognition and *in vivo* activity on native and +6C/−6G target sites. The first RD library allowed +6C/−6G contacting residue 63 to vary among seven amino acid residues: the native Q, or D, E, K, N, R or Y. Neighboring β-strand residues A55, N57, W62, Y64, R74 and G76 were substituted with residues that differed slightly in size and/or polarity from native residues found at these positions (RD1; Table 1). The predicted maximum complexity of the RD1 library was 3.5×10^4. Sequencing indicated that

Table 1. Residue substitutions and library statistics for three I-PpoI yeast one hybrid libraries and screens

Residue	I-PpoI variant library		
	CR	RD1	RD2
A55		A,G,V,L,I	A,G,V,L
N57	All	N,C	Y,E,D,R,K,Q,N,C
R61	All		Y,E,D,R,K,Q,N,C
W62		W,F,L,I,Y	F,W
Q63	All	Y,E,D,R,K,Q,N	R
Y64		Y,F,S,T,W	
K65	All	K,R	Y,E,D,R,K,Q,N,C
R74	All	R,T	Y,E,D,R,K,Q,N,C
G76		A,G,V,L,I	
Maximum complexity	3.2×10^6	3.5×10^4	3.2×10^4
Library size	1.6×10^5	1.8×10^5	1.4×10^5
Colonies His-screened	2×10^6	2.5×10^5	2.1×10^5
His+lacZ+ plasmids recovered	1	6	9

The nine I-PpoI amino acid residues targeted for substitution during library constructions are indicated in the left column, together with substitutions made at each residue in the contacting residue (CR) or two rationally designed (RD1 and RD2) protein variant libraries. Maximum complexity is the predicted complexity of the library. Library size is the experimental complexity of the library, determined from the number of independent colonies that contained inact open reading frames.

72% of RD1 Ppo-AD plasmids had intact open reading frames. Thus the completed RD1 library of 1.8×10^5 members was sufficiently large to include all RD1 design variants (RD1, Table 1).

In vivo screening of RD1 against +6C/−6G target site reporter genes yielded 352 His+ colonies, and 15 plasmids from this pool were verified by re-transformation and quantitative *lacZ* activity prior to sequencing. Five active Ppo-AD plasmids had the same Ppo-AD amino acid substitutions found in variant A (Table 2). Six different RD1 starting plasmids harbored the same Q63R and K65R substitutions (Table 2, variant E), as indicated by plasmid DNA sequence differences. Four other plasmids had the variant E Q63R and K65R substitutions together with 2 to 4 other substitutions at residues targeted during library construction (variants B, C, D and F; Table 2).

A second rationally designed library was constructed to determine whether additional residue substitutions could augment the *in vivo* activity of Ppo-AD proteins identified in the CR and RD1 libraries. The RD2 library design features were: (i) three positions were fixed as residues found in active variants of RD1 (63R, 64Y and 76G); (ii) residue 55 was an A, G, L or V, and residue 62 was an F or W, again as previously identified in active variants; (iii) contacting residues 57, 61, 65 and 74 were varied

Table 2. Biochemical properties and residue substitutions in I-PpoI variants

Variant	Source	Binding K_d (nM)		Site cleavage		Residue at position								
		WT	+6C/−6G	WT	+6C/−6G	55	57	61	62	63	64	65	74	76
H98A	SDM	0.11 ± 0.02	≥1000	++++	No	A	N	R	W	Q	Y	K	R	G
L116A	SDM	350 ± 60	≥1000	No	No									
T	SDM	0.034 ± 0.01	6 ± 1.5	++++	+/−							R		
B	RD1	110 ± 10	130 ± 30	No	No	G			F	R		R	T	
U	SDM	130 ± 40	≥1000	+++	No	V	C					R	T	
Y	SDM	180 ± 60	≥1000	+	No		C			R		R		
D	RD1, RD2	230 ± 40	240 ± 80	No	No	L				R		R		
E	RD1, RD2	280 ± 90	140 ± 40	++	No					R		R		
H	RD2	460 ± 140	310 ± 90	No	No				F	R		R		
P	SDM	480 ± 30	570 ± 150	++++	No	V								
R	SDM	500 ± 20	≥1000	No	No					K				
Q	SDM	620 ± 90	570 ± 160	++	No					A				
W	SDM	620 ± 160	480 ± 100	No	No	V	C			R		R		
A	CR, RD1	630 ± 190	≥1000	No	No		C			R		R	T	
J	RD2	760 ± 40	430 ± 50	No	No	I				R		R		
V	SDM	810 ± 120	540 ± 140	No	No	V	C			R			T	
S	SDM	820 ± 170	≥1000	++	No					R				
G	RD2	≥1000	190 ± 40	No	No			K		R		R		
N	RD2	≥1000	300 ± 10	No	No	V		K		R		R		
C	RD1	≥1000	440 ± 80	No	No	V	C			R		R	T	
L	RD2	≥1000	500 ± 100	No	No	L		K		R		R		
M	RD2	≥1000	≥1000	No	No	L			F	R		R		
K	RD2	≥1000	≥1000	No	No	L		K	F	R		R		
F	RD1	≥1000	≥1000	No	No	I	C		I	R	W	R	T	I

The designation and source of each I-PpoI variant protein are listed in the first two columns. Positive (H98A) and negative (L116A) control proteins are listed in the first two rows for comparison. Binding affinities for native and +6C/–6G target site DNAs, determined by gel shift assay, are given in columns 3 and 4 and are arranged in order of decreasing affinity for the native I-PpoI target site. The ability of each I-PpoI variant to cleave native or +6C/–6G target site DNA was determined after reverting the catalytically inactivating H98A substitution that was required to establish the yeast one-hybrid assay. The final columns detail amino acid substitutions at nine positions in the DNA-protein interface. A blank box indicates presence of the native residue (shown top in single letter code). Key: CR, contacting residue library; RD1, RD2, rationally designed protein variant libraries RD1 and RD2; SDM, site-directed mutagenesis. Note: This Table has been reformatted in Supplementary table 2 to show variants rank-ordered on the basis of increasing number of residue substitutions at nine positions in the DNA–protein interface.

among eight amino acid residues commonly found in homing endonuclease DNA–protein interfaces (Y, E, D, R, K, Q, N, C; Table 1, RD2) and (iv) inclusion of a 9 bp post-C-terminal 'molecular bar code' to allow different RD2 plasmid molecules to be identified and distinguished. The RD2 library had a predicted maximum complexity of 3.2×10^4. Sequencing indicated that 67% of RD2 library plasmids had intact Ppo-AD open reading frames. Thus the completed RD2 library of 1.4×10^5 members was sufficiently large to include all designed RD2 variants.

Thirty-five RD2 plasmids displayed *in vivo* activity against +6C/–6G reporter genes. These plasmids encoded nine different Ppo-AD proteins: seven Ppo-AD variants were unique to the RD2 library (variants G, H, J, K, L, M and N), whereas two variants had been previously identified in RD1 (variants D and E; Table 2). Ppo-AD variants D, H, K and L were independently isolated, from 3 to 6 times each from RD2, as indicated by plasmid molecular bar code sequences. In addition to these library variants, we constructed nine I-PpoI variants (variants P to Y, Table 2) by site-directed mutagenesis. This was done to determine the contribution of specific amino acid residues identified in the CR, RD1 and RD2 libraries to target site binding and cleavage. We reasoned that specific residue substitutions in library isolates might promote or interfere with site binding and cleavage, and thus wanted to assay individual substitutions in defined sequence contexts. These site-directed mutants were generated, in consequence, in either the native or variant A or C I-PpoI sequence contexts.

Biochemical characterization of I-PpoI variant proteins

In order to characterize the site-binding properties of Ppo-AD variants, we expressed and purified variant proteins from *E. coli and* then determined their dissociation constants (K_d's) on native and on +6C/–6G target site DNAs. We also determined the ability of selected variants to induce reporter gene expression *in vivo* to allow a comparison of *in vitro* and *in vivo* site-binding properties (Figure 3). The I-PpoI portion of each Ppo-AD fusion protein was subcloned into a pET11c expression vector containing an N-terminal six-histidine tag to facilitate purification. In order to assay site cleavage as well as binding by specific variant proteins, we used site-directed mutagenesis to revert the catalytically inactivating H98A substitution required to establish our Y1H assay.

Dissociation constants for variant proteins were determined by electrophoretic gel mobility shift analysis using +6C/−6G and native target site DNAs. Fourteen of 22 variants displayed a higher affinity for +6C/−6G homing site DNA than did native I-PpoI, although this site-binding preference was modest (typically <5-fold; Table 2). Seven of the remaining variants displayed no site-binding preference. One variant, T, displayed high binding affinity for both the native and +6C/−6G target sites and a 175-fold higher affinity for native target site DNA (Table 2). We also determined the binding affinity of native I-PpoI and four variant proteins on target site DNAs with base-pair substitutions at the ±6 position, the ±7 position and in one instance ±6/±7 position substitutions. The variants had amino acid substitutions at position 63 contacting the ±6 bp; at residue 74 contacting the ±7 bp; and at residues 55, 57 and 65 in adjacent regions of the DNA–protein interface. None of these variant proteins bound any mutant target site DNA with high site affinity or specificity (Table 2, and Supplementary Table 2).

Seven of 22 variant I-PpoI proteins cleaved native homing site DNA *in vitro*. Cleavage activity was determined on linear plasmid substrates that contained a single I-PpoI target site. Both native I-PpoI and variant T bound and cleaved native target site DNA, though did not cleave +6C/–6G target site DNA in the presence of Mg^{2+} or Mn^{2+} (Figure 3A; data not shown). Other variant proteins with activity cleaved only native target site DNA (variants E, P, Q, S, U and Y; Table 2, Figure 3A). In order to better understand the relationship between biochemical properties and *in vivo* activity of variant proteins, we quantified the ability of six variants to induce a *lacZ* reporter gene in budding yeast. Variants A, C, E, G, J and T were chosen for analysis as they represent a range of *in vitro* binding affinities. Expression plasmids for each variant as a Ppo-AD fusion were transformed into yeast containing a native or +6C/–6G target site *LacZ* reporter gene plasmid. β-Galactosidase activity was quantified for three independent colonies grown in liquid media and assayed on the same day. Protein variants C, G and J had higher activity on +6C/–6G than on native target site reporters. This paralleled the 1.8-to-5.3-fold higher binding affinity of these variant proteins for +6C/–6G sites (Table 2 and Figure 3B). A graphical summary of *in vitro* site binding and site cleavage analyses, for all variants, is shown in Figure 3 (panels C and D).

Computational modeling and analysis of I-PpoI protein variants

Structural modeling of base pair and residue substitutions in the I-PpoI DNA–protein interface was used to gain mechanistic insight into the properties of several I-PpoI variant proteins. When I-PpoI is bound to native target site DNA, Q63 makes an energetically favorable, glutamine:adenine contact with two hydrogen bonds to adenine 6. R74 makes a canonical contact with guanine 7. K65 makes an H-bond contact to guanine 9, whereas R61 makes a series of backbone contacts between base pairs 3 and 4 (Figure 1A). Modeling indicates the +6C/–6G base-pair substitution inhibits site binding and cleavage by disrupting the glutamine:adenine 6 contact, and by forcing the Q63 residue to rotate 135° into and in part disrupt the adjacent R74:guanine 7 bp contact (Figure 4A and B).

Several residue substitutions found in variant G could in part restore mutant +6C/–6G target site binding. These were Q63R at the +6C/–6G base-pair-contacting position (present in 17 variants) K65R (present in 17 variants) and R61K (present in 4 variants; Table 2 and Figure 4, compare panels B and C). Modeling indicates that Q63R substitutions can in part restore mutant site-binding affinity by making a new contact at base-pair position 5, and by restoring an R74:guanine 7 canonical base-pair contact (Figure 4C). Modeling also provides an

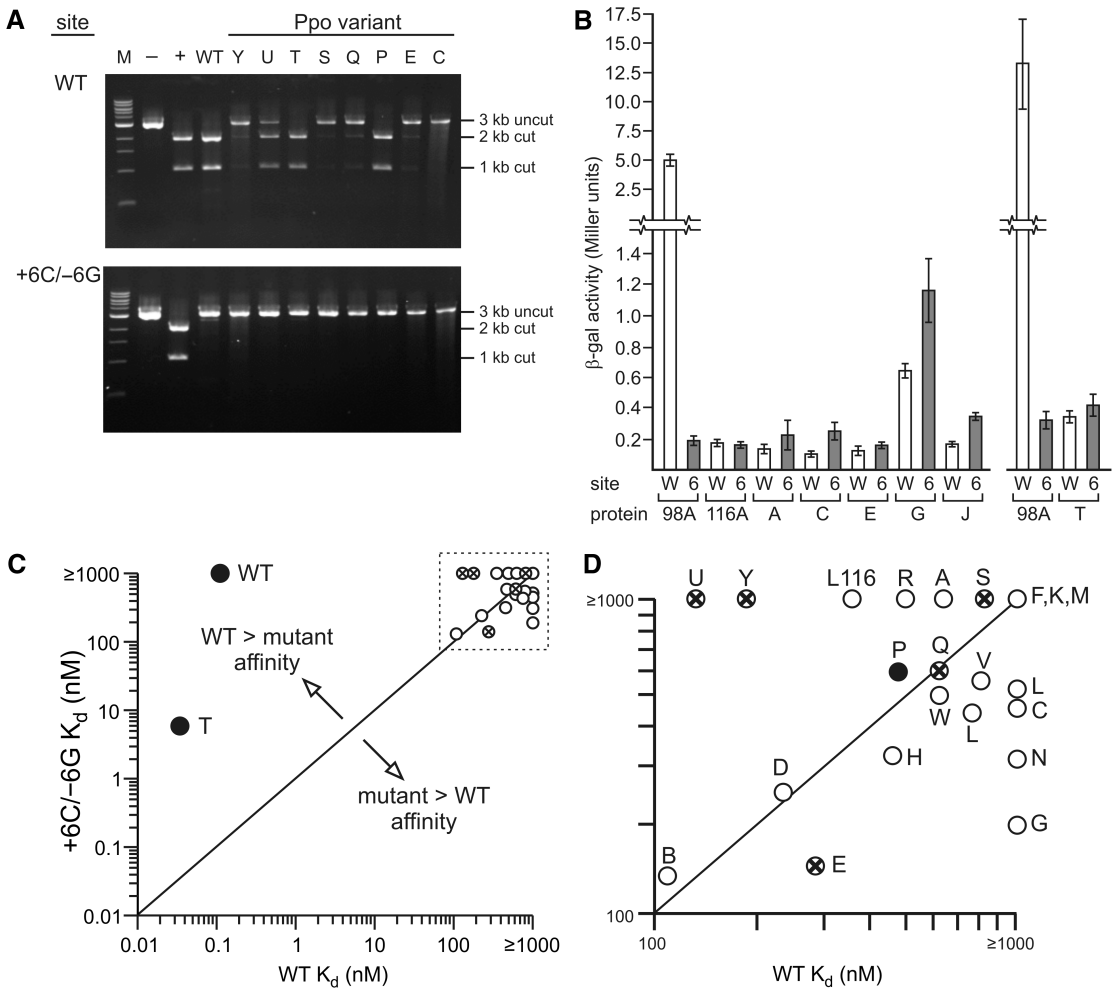

Figure 3. Target site cleavage and site binding by I-PpoI variant proteins. (**A**) *In vitro* cleavage of native and +6C/–6G target site DNAs by native and 8 variant I-PpoI proteins. Plasmid DNA (10 nM) containing a native (WT, top panel) or +6C/–6G variant (bottom panel) target site was cut with XmnI to generate a 3 kb linear DNA substrate ('uncut') that was cleaved with 0.1 μM native (WT lanes) or variant I-PpoI protein (rightmost 8 lanes, Y to C) for 30 min at 37°C. I-PpoI target site-specific cleavage generated 2 and 1 kb product fragments ('cut'). *M* = 1 kb DNA size marker (New England Biolabs). Controls include plasmid cleaved with XmnI alone (– lanes) or plasmid cleaved with XmnI, and with AflII that cleaves the central 6 bp of the I-PpoI target site (+ lanes). (**B**) *In vivo* yeast one-hybrid reporter gene activity of selected I-PpoI protein variants. *LacZ* reporter gene activity was determined by expressing specific proteins in yeast in the presence of native (site W) or +6C/–6G (site 6) I-PpoI target site reporter plasmids, then measuring β-galactosidase activity in permeabilized cells using ONPG as a substrate. Proteins included as controls were H98A Ppo-AD (98A; positive control) and L116A (116A; negative control). Variant T was assayed independently of the other five variants shown, and thus is displayed with a simultaneously performed H98A control (rightmost 4 bars). All activities represent the mean ± SD of three independent colonies grown and assayed on the same day. (**C and D**) Graphical summary of I-PpoI variant protein site binding and cleavage properties. Site-binding affinities (nM) for native (*X* axis) or +6C/–6G (*Y* axis) target site DNAs. Variant proteins with partial or full cleavage activity on native homing site DNA are indicated, respectively, by circled X's or filled circles. Only one I-PpoI variant protein, T, had any detectable cleavage activity on +6C/–6G target site DNA. The identities of the variants in the dashed line box in the upper right of panel C are shown in expanded panel D.

explanation for why selected variants with Q63R substitutions, e.g. variant E, show preferential binding of the mutant +6C/−6G target site: when R63 packs in the native DNA interface, it leaves a small cavity adjacent to the native ±6 A/T base pairs. This cavity is partially filled by the cytosine in the +6C/–6G substrate (Figures 1A and 4B; data not shown). These results indicate that the native contacting residue, Q63, acts as a gatekeeper at the ±6 A:T base pair by controlling contacts at this and adjacent base-pair positions.

Variant T, which contains a K65R substitution, showed a marked increase in binding affinity for both native and +6C/–6C target sites. Modeling of this substitution revealed that K65 forms a single H-bond to the guanine base at base pair 9, whereas K65R substitution make two H-bonds to base pairs at positions 8 *and* 9 (compare Figure 4 panels B and C). These contacts are energetically more favorable than the single contact made by the native K65 residue (−1.22 units for R65 versus −1.09 units for K65; RosettaDesign analyses not shown). Neither K nor R65 residues contact target site position 6, and thus do not discriminate between native or mutant +6C/–6G target site DNAs.

R61K substitutions decrease the DNA-binding affinity of I-PpoI, while increasing target site selectivity to favor the binding of mutant, +6C/–6G target site DNA

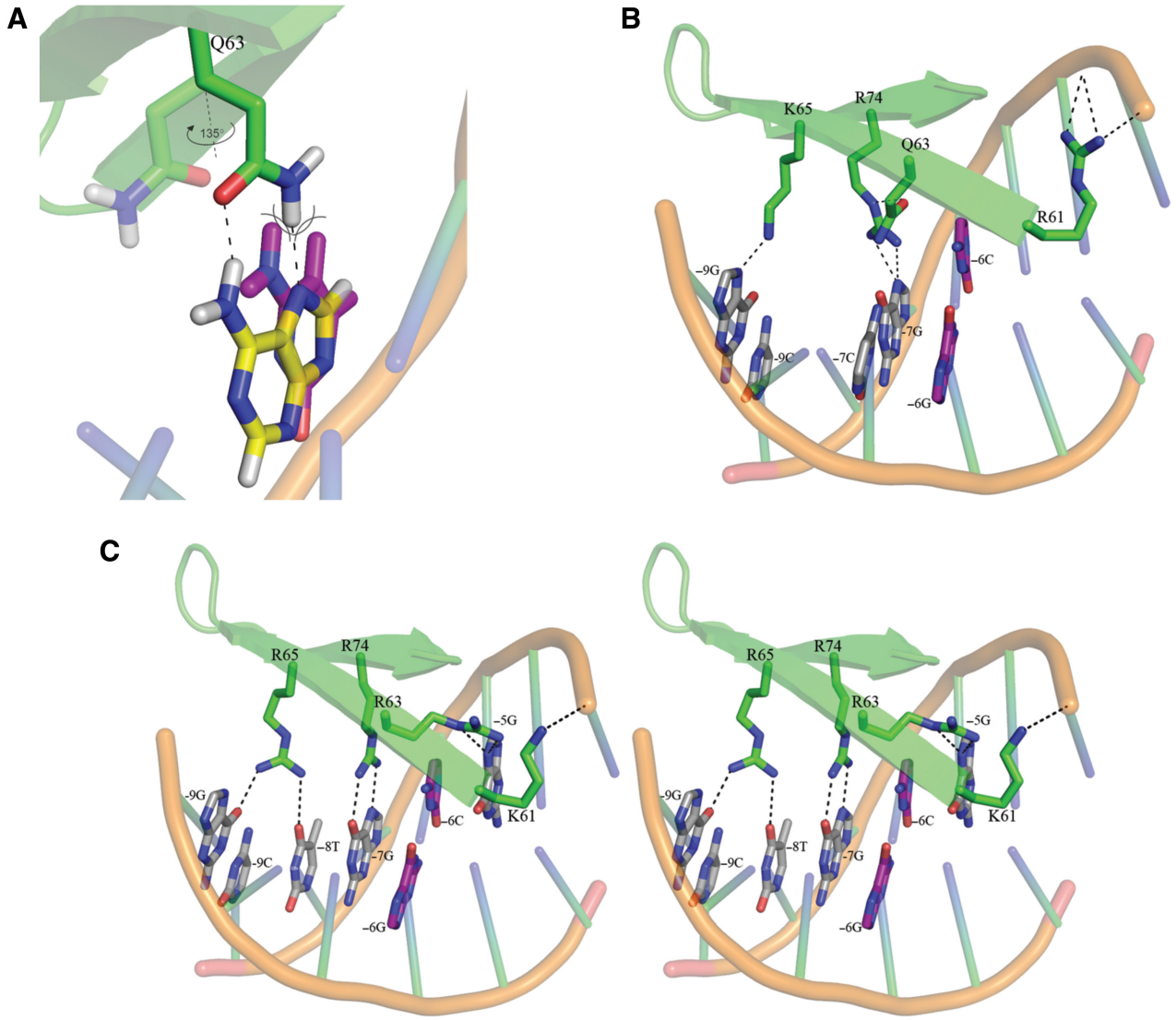

Figure 4. Molecular modeling of I-PpoI variant protein on native and mutant target site DNAs. The native ±6A/T base pair is shown in yellow, while the mutant ±6C/G base pair is shown in purple. Amino acid geometries pictured with the mutant target site were modeled using RosettaDesign. Relevant sequence-specific hydrogen bonds are shown as black dotted lines. (**A**) Geometry of position 6 contacting residue Q63 on native A:T and mutant C:G target sites. The mutant −6C base (purple) forces Q63 to rotate 1358 to alleviate a stearic clash (shown as crossed arc lines). (**B**) Rotation of Q63 forced by the ±6C/G base-pair substitution disrupts the canonical R74:guanine contacts made at adjacent base-pair position 7. (**C**) Stereo representation of I-PpoI variant G modeled with the ±6C/G mutant DNA target. R63 interacts with ±5G, and allows R74 to again contact base pair position 7. R65 makes two H-bonds to target site base-pairs 8 and 9, which increases the overall affinity of I-PpoI for native and mutant DNA target sites. R61K substitutions alter DNA backbone contacts near the scissile phosphate between base pairs 2 and 3 (compare panel C with B or Figure 1A).

(compare, e.g. variants E and G, Table 2). Modeling indicates that the R61K substitution modifies backbone contacts in the native interface between base pair positions 3 and 4 (Figure 4, compare panels B and C). These backbone positions are immediately adjacent to the scissile phosphate, and form part of the most deformed region of the I-PpoI:DNA substrate complex (21). The selectivity of K61 variants for mutant target site DNA is likely explained by sequence-dependent conformation changes in the DNA–protein interface, as no new contacts are established with mutant target site DNA.

DISCUSSION

We used a yeast one-hybrid (Y1H) assay to isolate and characterize variants of the I-PpoI homing endonuclease with altered DNA target recognition specificity. Variants were isolated using a mutant binding site target with symmetrical ±6 bp position substitutions that abolished site binding and cleavage by native I-PpoI. We reasoned that amino acid substitutions in the DNA–protein interface adjacent to this target site position, at contacting residue 63 alone or in conjunction with other substitutions, might restore high site-binding affinity and

specificity *in vivo*. This strategy resembles Y1H screens that have been used to identify and characterize site-specific DNA-binding proteins from yeast and other organisms. One similar precedent reported for homing endonucleases was the use of a bacterial two-hybrid screen to explore the recognition specificity of the LAGLIDADG homing endonuclease PI-SceI (37).

The variant I-PpoI proteins we generated contained from 1 to 8 amino acid substitutions in the DNA–protein interface (Tables 1 and 2). Most variants had low site-binding affinities (100–1000 nM K_d's), and a modest (2-to-10-fold) preference to bind native or mutant target site DNA. One exception, variant T with only a K65R substitution, had high binding affinities for native and mutant target site DNAs that included a clear (<175-fold) preference for native site binding. Cleavage competence was assayed by reverting the catalytically inactivating H98A substitution required to establish the Y1H screen, and then assaying proteins for the ability to cleave native or mutant target site DNAs. We found little correlation among site binding, discrimination and cleavage properties of individual variant proteins (Table 2 and Figure 3, panels C and D). One likely explanation for this is that I-PpoI must bind and bend substrate DNA in order to generate a productive substrate complex (21,22,26). Our hypothesis is that active variants (e.g. variants U, S or P, Figure 3A and Table 2) retain the ability to both bind and bend site DNA to permit cleavage.

Structure-based molecular modeling of I-PpoI variant proteins on both native and mutant target site DNAs was used to gain insight into the contribution of specific residue substitutions to the biochemical behavior of variant proteins. Modeling revealed that Q63R substitutions fail to confer ±6 mutant target site specificity by virtue of loss of a canonical glutamine:adenine contact at position 6, together with partial disruption of a canonical arginine:guanine contact at the adjacent base pair position 7. Q63R variants were nonetheless able to discriminate in favor of mutant +6C/–6G target sites by packing more favorably with mutant 6C than with the native 6A target site base, while making a high quality contact at the adjacent position 5 base pair (Figure 4). K65R substitutions, in contrast, increase affinity for native *and* ±6 mutant target sites by making an additional H-bond to bridge base pair positions 8 and 9 [Figures 1 and 4 (panels B and C)].

R61K substitutions, in contrast, appear to alter site binding by modifying DNA backbone contacts immediately adjacent to the scissile phosphate located between base pairs 2 and 3 [(21); Figure 4]. This type of indirect readout may be strongly influenced by local, sequence-dependent DNA conformation (38,39). These sequence-dependent effects may be further amplified by DNA substrate deformation in this region of the I-PpoI substrate complex, and thus interfere with correct positioning of the scissile phosphate (Figures 1 and 4) (21,26).

The Y1H screening assay we used employed the canonical two-hybrid reporters *HIS3* and *lacZ* that, respectively, confer a growth advantage or have readily detectable activity at low expression levels. The sensitivity of these reporters, in retrospect, permitted the identification of I-PpoI variants with modest *in vivo* activity and site-binding affinity (Table 2 and Figure 3). One explanation for the failure to recover I-PpoI variants with high affinity and specificity for +6C/–6G mutant target site DNA is that they were not present in our starting libraries. This may be the case for the initial randomized contacting residue (CR) library, which was too large to be exhaustively screened. The smaller rationally designed libraries (RD1 and RD2), in contrast, were exhaustively screened but may have been too small to encode a high affinity binding variant.

The experimental and computational analyses described above indicate one productive approach to alter the DNA recognition specificity of I-PpoI and other homing endonuclease proteins. Computational DNA–protein interface design can be used to predict different residue substitutions that may confer high binding affinity and specificity for a mutant target site DNA. The resulting protein variants can then be rank-ordered on the basis of predicted DNA-binding energies, and further evaluated for structural quality by molecular modeling. This general approach has already been shown to work for single base-pair positions in I-MsoI, a member of the LAGLIDADG homing endonuclease family (40), and should be applicable to His-Cys box proteins such as I-PpoI. A prerequisite for this engineering approach is a high resolution co-crystal structure.

In addition to protein computational design, it should also be possible to improve the experimental selection or screening assays to identify variant proteins with high site-binding affinity and specificity. For example, the Y1H assay could be adapted to use reporter genes or selections that would require comparatively high levels of expression to identify active variants. Alternatively, a combination of positive and negative selection could be used to recover variants with high site-binding affinity and specificity (41).

Homing endonucleases remain the most attractive starting point for the generation of new, highly sequence-specific proteins for biology and medicine. They encode a wide range of different DNA recognition specificities, and display high site-binding specificity that is tightly linked to DNA cleavage. Homing endonucleases including I-PpoI have already been successfully expressed in human cells. They cleave their target sites with high site specificity (19,42–44), and thus are being used to promote site-specific recombination (45) and high resolution DNA double-strand break repair analyses (20). The precedents outlined above indicate that it should be possible in the near future to generate many highly sequence-specific homing endonuclease variants useful for genome engineering, disease therapy or disease prevention.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Mike Moser, Meggen S. Chadsey and Alden Hackmann for help, respectively, with experimental

design, library construction and graphics support. Stan Fields and his laboratory provided generous help in establishing the yeast one-hybrid assay, and Barry L. Stoddard for I-PpoI structure determinations, initial structural modeling and continuous discussion. U.S. National Institutes of Health RO1 and T32 funding (RO1 CA88942 to R.J.M., Jr; T32 GM07735 to J.L.E. T32 GM07270 to J.E. and T32 GM007266 and T32 CA077116 to U. U.). U. U. is also the recipient of a Poncin Fund Award. Funding to pay the Open Access publication charges for this article was provided by the U.S. National Institutes of Health.

Conflict of interest statement. None declared.

REFERENCES

1. Pabo,C.O. and Sauer,R.T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.*, **61**, 1053–1095.
2. Branden,C. and Tooze,J. (1999) *Introduction to Protein Structure* Garland Publishing, Inc., New York, USA.
3. Luscombe,N.M., Austin,S.M., Berman,H.M. and Thornton,J.M. (2000) An overview of the structures of protein-DNA complexes. *Genome Biol.*, **1**, 1–37.
4. Luscombe,N.M., Laskowski,R.A. and Thornton,J.M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860–2874.
5. Sarai,A. and Kono,H. (2005) Protein-DNA recognition patterns and predictions. *Annu. Rev. Biophys. Biomol. Struct.*, **34**, 379–398.
6. Pingoud,A., Fuxreiter,M., Pingoud,V. and Wende,W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell. Mol. Life Sci.*, **62**, 685–707.
7. Latchman,D.S. (2004) *Eukaryotic Transcription Factors* Elsevier Academic Press, London.
8. Belfort,M. and Roberts,R. (1997) Homing endonucleases – keeping the house in order. *Nucleic Acids Res.*, **25**, 3379–3388.
9. Belfort,M., Derbyshire,V., Stoddard,B.L. and Wood,D.W. (2005) *Homing Endonucleases and Inteins* Springer, Berlin.
10. Stoddard,B.L. (2005) Homing endonuclease structure and function. *Quarterly Rev. Biophys.*, **38**, 1–47.
11. Porteus,M.H. and Carroll,D. (2005) Gene targeting using zinc finger nucleases. *Nat. Biotechnol.*, **23**, 967–973.
12. Gimble,F.S. (2000) Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiol. Lett.*, **185**, 99–107.
13. Burt,A. and Koufopanou,V. (2004) Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr. Opin. Genet. Dev*, **14**, 609–615.
14. Muscarella,D.E. and Vogt,V.M. (1989) A mobile Group I intron in the nuclear rDNA of Physarum polycephalum. *Cell*, **56**, 443–454.
15. 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.
16. Galburt,E.A. and Jurica,M.S. (2005) His-Cys box homing endonucleases. In Belfort,M., Derbyshire,V., Stoddard,B.L. and Wood,D.W. (eds), *Homing Endonucleases and Inteins*, Springer-Verlag, Berlin, Vol. 16, pp. 85–102.
17. Muscarella,D.E., Ellison,E.L., Ruoff,B.M. and Vogt,V.M. (1990) Characterization of I-Ppo, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of Physarum polycephalum. *Mol. Cell. Biol.*, **10**, 3386–3396.
18. Ellison,E.L. and Vogt,V.M. (1993) Interaction of the intron-encoded mobility endonuclease I-PpoI with its target site. *Mol. Cell. Biol.*, **13**, 7531–7539.
19. Monnat,R.J., Hackmann,A.F.M. and Cantrell,M.A. (1999) Generation of highly site-specific DNA double-strand breaks in human cells by the homing endonucleases I-PpoI and I-CreI. *Biochem. Biophys. Res. Commun.*, **255**, 88–93.
20. Berkovich,E., Monnat,R.J. and Kastan,M.B. (2007) Roles of ATM and NBS1 in chromatin structure modulation and DNA double-strand break repair. *Nat. Cell Biol.*, **9**, 683–690.
21. Flick,K.E., Jurica,M.S., Monnat,R.J., Jr and Stoddard,B.L. (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature*, **394**, 96–101.
22. Galburt,E.A., Chevalier,B., Tang,W., Jurica,M.S., Flick,K.E., Monnat,R.J., Jr and Stoddard,B.L. (1999) A novel endonuclease mechanism directly visualized for I-PpoI. *Nat. Struct. Biol.*, **6**, 1096–1099.
23. 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**, 1–4.
24. Argast,G.M., Stephens,K.M., Emond,M.J. and Monnat,R.J. (1998) I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. *J. Mol. Biol.*, **280**, 345–353.
25. Wittmayer,P.K., McKenzie,J.L. and Raines,R.T. (1998) Degenerate DNA recognition by I-PpoI endonuclease. *Gene*, **206**, 11–21.
26. Galburt,E.A., Chadsey,M.S., Jurica,M.S., Chevalier,B.S., Ehro,D., Tang,W., Monnat,R.J, Jr and Stoddard,B.L. (2000) Conformational changes and cleavage by the homing endonuclease I-PpoI: a critical role for a leucine residue in the active site. *J. Mol. Biol.*, **300**, 877–887.
27. Botstein,D., Falco,S.C., Stewart,S.E., Brennan,M., Scherer,S., Stinchcomb,D.T., Struhl,K. and Davis,R.W. (1979) Sterile host yeasts (SHY): a eukaryotic system of biological containment for recombinant DNA experiments. *Gene*, **8**, 17–24.
28. Sikorski,R.S. and Hieter,P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics*, **122**, 19–27.
29. Gietz,R.D. and Woods,R.A. (2006) Yeast transformation by the LiAc/SS carrier DNA/PEG method. *Methods Mol. Biol.*, **313**, 107–120.
30. Sherman,F. (2002) Getting started with yeast. *Meth. Enzymol.*, **350**, 3–41.
31. Worthington,M.T., Pelo,J. and Lo,R.Q. (2001) Cloning of random oligonucleotides to create single-insert plasmid libraries. *Anal. Biochem.*, **294**, 169–175.
32. Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1987) *Current Protocols in Molecular Biology* John Wiley & Sons, New York, USA.
33. Miller,J. (1992) *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.
34. Havranek,J.J., Duarte,C.M. and Baker,D. (2004) A simple physical model for the prediction and design of protein-DNA interactions. *J. Mol. Biol.*, **344**, 59–70.
35. Morozov,A.V., Havranek,J.J., Baker,D. and Siggia,E.D. (2005) Protein-DNA binding specificity predictions with structural models. *Nucleic Acids Res.*, **33**, 5781–5798.
36. Muscarella,D.E. and Vogt,V.M. (1993) A mobile Group I intron from Physarum polycephalum can insert itself and induce point mutations in the nuclear ribosomal DNA of Saccharomyces cerevisiae. *Mol. Cell. Biol.*, **13**, 1023–1033.
37. Gimble,F.S., Moure,C.M. and Posey,K.L. (2003) Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system. *J. Mol. Biol.*, **334**, 993–1008.
38. Lavery,R. (2005) Recognizing DNA. *Quarterly Rev. Biophys.*, **38**, 339–344.
39. Koudelka,G.B., Mauro,S.A. and Ciubotaru,M. (2006) Indirect readout of DNA sequences by proteins: the roles of DNA sequence-dependent intrinsic and extrinsic forces. *Prog. Nucleic Acid Res. Mol. Biol.*, **81**, 143–177.
40. Ashworth,J., Havranek,J.J., Duarte,C.M., Sussman,D., Monnat,R.J., Stoddard,B.L. and Baker,D. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, **441**, 656–659.
41. 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.

42. Puchta,H., Dujon,B. and Hohn,B. (1993) Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. *Nucleic Acids Res.*, **21**, 5034–5040.
43. Rouet,P., Smih,F. and Jasin,M. (1994) Introduction of double--strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.*, **14**, 8096–8106.
44. Lukacsovich,T., Yang,D. and Waldman,A.S. (1994) Repair of a specific double-strand break generated within a mammalian chromosome by yeast endonuclease I-SceI. *Nucleic Acids Res.*, **22**, 5649–5657.
45. Pâques,F. and Duchateau,P. (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives on gene therapy. *Curr. Gene Ther.*, **7**, 49–66.