Published online 22 December 2008

Nucleic Acids Research, 2009, Vol. 37, No. 3 ***877–890***
doi:10.1093/nar/gkn1007

Optimization of *in vivo* activity of a bifunctional homing endonuclease and maturase reverses evolutionary degradation

Ryo Takeuchi[1,2], Michael Certo[2,3], Mark G. Caprara[4], Andrew M. Scharenberg[2,3] and Barry L. Stoddard[1,2,*]

[1]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, [2]Northwest Genome Engineering Consortium, Seattle, WA 98101, [3]Seattle Children's Hospital Research Institute and the Graduate Program in Molecular and Cellular Biology, M/S C9S, Seattle, WA 98101 and [4]Center for RNA Molecular Biology, School of Medicine, Case Western Reserve University, Cleveland, OH 44106-4960, USA

Received November 4, 2008; Revised November 26, 2008; Accepted December 1, 2008

ABSTRACT

The LAGLIDADG homing endonuclease (LHE) I-AniI has adopted an extremely efficient secondary RNA splicing activity that is beneficial to its host, balanced against inefficient DNA cleavage. A selection experiment identified point mutations in the enzyme that act synergistically to improve endonuclease activity. The amino-acid substitutions increase target affinity, alter the thermal cleavage profile and significantly increase targeted recombination in transfected cells. The RNA splicing activity is not affected by these mutations. The improvement in DNA cleavage activity is largely focused on one of the enzyme's two active sites, corresponding to a rearrangement of a lysine residue hypothesized to act as a general base. Most of the constructs isolated in the screen contain one or more mutations that revert an amino-acid identity to a residue found in one or more close homologues of I-AniI. This implies that mutations that have previously reduced the endonuclease activity of I-AniI are identified and reversed, sometimes in combination with additional 'artificial' mutations, to optimize its *in vivo* activity.

INTRODUCTION

Homing endonucleases (HEs) are highly specific DNA-binding and cleavage enzymes that are encoded by open reading frames that are embedded within intervening sequences (introns and inteins) in all microbial forms of life, including phage, eubacteria, archaea and single cell eukaryotes (1,2). First discovered in the 1980s as the drivers of mobile introns in yeast (3,4), HEs promote the mobility of the intervening sequences (and their own reading frames) by generating double strand breaks in homologous alleles that lack the intron or intein (5). Break repair leads to transfer of the element via homologous recombination, using the allele that contains the HE gene as a template. Thus, HE genes are selfish DNA sequences that are inherited in a dominant, non-Mendelian manner.

Based on primary sequence homology, five homing enzyme families have been identified, each primarily associated with a unique size and distribution of host genome (2): the LAGLIDADG endonucleases (usually found in archaea and fungal and algal organellar genomes), His-Cys Box family (protist nuclear genomes), the HNH and GIY-YIG endonucleases (usually found in bacteriophage, but with additional rare exceptions in bacteria) and the PD-(D/E)XK family (found in bacterial genomes). The His-Cys Box and HNH family appear to be descended from a common ancestral nuclease (which also gave rise to bacterial colicins and many additional DNA modifying enzymes), while the remaining three families each contain a unique catalytic scaffold and appear to have arisen independently of one another.

The structures of several LHEs bound to their DNA targets have been determined, as summarized in (2). All of these endonucleases display highly modularized structures and DNA-binding scaffolds that are amenable to engineering and selection experiments. Monomeric LHEs (that contain two active sites and two DNA recognition regions on a single peptide chain) are particularly compact proteins (~200–250 amino acids total) that can recognize very asymmetric DNA target sites (6).

LHEs are the most specific of all known naturally occurring DNA-binding enzymes, recognizing DNA

*To whom correspondence should be addressed. Tel: +1 206 667 4031; Fax: +1 206 667 3331; Email: bstoddar@fhcrc.org

target sites ranging in length from 20 to 24 basepairs with low nanomolar dissociation constants. Their specificity of cognate DNA recognition and cleavage, and the structural basis for this behavior, has been evaluated both for the homodimeric I-*Cre*I (7,8) and for the monomeric I-AniI (9) endonucleases. In both studies, the estimated cleavage specificity of the wild-type (WT) endonuclease under ideal *in vitro* cleavage conditions approaches 1 in 10^9. Extensive experimentation by a wide variety of laboratories, mostly with the I-*Sce*I endonuclease, indicate little or no toxicity when the enzyme is transfected and expressed in a variety of cellular experimental contexts (10).

The collaboration between HE genes and their surrounding introns or inteins represent an efficient symbiotic relationship (11,12). The extra-Mendelian inheritance of these elements is driven by the endonuclease, assuring that they rapidly become fixed in a population. In return, the self-splicing intron or intein serves as a refuge for the endonuclease gene, allowing it to invade but not inactivate the host gene. However, after invasion and fixation occurs there is little selective pressure for the maintenance of a functional HE, leading to the gradual accumulation of mutations that reduce its activity. This evolutionary degradation eventually leads to loss of the endonuclease gene and the surrounding intron, followed by subsequent re-invasion. This HE life cycle has been especially well illustrated in a comparative study of analogous HEs in yeast (13).

A small but significant fraction of HE genes has acquired a secondary activity that provides a benefit to the host. As a result of this activity, these proteins are therefore subject to purifying selection that helps maintain their fold and stability. For example, a subset of HEs also facilitates the splicing of their host intron (a behavior termed 'maturase' activity) (14–18). While the presence of the intron itself might not benefit the host *per se*, its splicing is necessary for proper function of the surrounding gene. A HE that is required for intron splicing is therefore under 'partial' selection for its DNA cleavage activity. The overall fold of the protein must be maintained in order to maintain splicing activity (which obviously also helps maintain its endonuclease capabilities) but individual mutations that only reduce DNA cleavage are still under minimal purifying selection pressure.

Those naturally-occurring HEs that moonlight as splicing cofactors (or as any other factor that benefits the host) are more persistent in the genome, but often display a chronic disability in their DNA cleavage activity (9,19). Here we study these points using a model HE (I-AniI) that is a very efficient intron splicing cofactor, but a quite inefficient endonuclease. We conducted an *in vivo* selection experiment for mutated variants of I-AniI that display enhanced cleavage of its DNA target. This experiment yields several reoptimized endonuclease constructs, harboring between one and four point mutations, which display significantly enhanced activity in the screen. Most of these mutations correspond to residues found in homologues of I-AniI, while some are unique. These amino-acid substitutions act synergistically to restore DNA cleavage activity. The construct containing the fewest point mutations of the WT enzyme sequence (and that also displays near 100% recovery in the cleavage activity screen) binds the WT DNA target more tightly, exhibits a catalytic rate enhancement for one of its two active sites, is more active at low physiological temperatures, and is much more efficient at stimulating recombination than its WT parent.

MATERIALS AND METHODS

Proteins

All I-AniI constructs used for *in vitro* experiments were expressed and purified from *Escherichia coli* strain BL21-CodonPlus (DE3)-RIL (Stratagene), as previously described (20), with the exception that expression was induced at 15°C for ~20 h after the culture had achieved early log growth phase ($OD_{600} \sim 0.6$). For all experiments, two point mutations (F80K and L232K) were introduced that improve the solubility of the enzyme and facilitate biochemical studies as previously described (9). Both mutations are far removed from the DNA-binding surface of the enzyme, and are on exposed surfaces of α-helices.

Bacteria-based *in vivo* cleavage assay and selection

This cleavage assay was performed based on the previous report by Doyon *et al.* (21). Two copies each of a WT target site or the closely related 'LIB4' site (that harbors two basepair substitutions and is bound and cleaved more readily by the endonuclease) were inserted between *Afl*III and *Bgl*II and between *Nhe*I and *Sac*II in the reporter (pCcdB) plasmid (four copies total). This plasmid encodes 'control of cell death B', a toxic protein to bacteria, and the expression of this gene is induced by isopropyl-β-thio-galactopyranoside (IPTG). The open reading frame (ORF) of the I-AniI construct described above was amplified by PCR, in order to incorporate an *Nco*I site and *Not*I site at the 5′ and 3′ termini, and was cloned into the expression (pEndo) plasmid. The HE gene is tightly regulated by pBAD promoter, and addition of L-arabinose promotes the gene transcription. The pEndo plasmid was transformed into DH12S competent cells harboring pCcdB plasmid (containing four copies of WT target site or LIB4 site) by electroporation. The transformants were grown in 2× YT medium at 37°C for 30 min, and were 5-fold diluted with 2× YT medium supplemented with 125 mg/ml carbenicillin and 0.025% L-arabinose. After the culture was grown at 30°C for 4 h, the cells were harvested, resuspended in sterilized water, and spread on both the non-selective plates (1× M9 salt, 1% glycerol, 0.8% tryptone, 1 mM $MgSO_4$, 1 mM $CaCl_2$, 2 μg/ml thiamine and 100 μg/ml carbenicillin) and selective plates (the non-selective plates supplemented with 0.02% L-arabinose and 0.4 mM IPTG). The plates were incubated at 30°C for two days, and the cleavage activity was evaluated by comparing the number of the colonies between the non-selective and the selective plates.

Selection of I-AniI variants with efficient cleavage in bacteria by random mutagenesis

Random mutations were induced in I-AniI ORF using GeneMorph II Random Mutagenesis Kit (Stratagene), by following the manufacture's instruction, and the amplified fragments with *Nco*I/*Not*I sites were inserted between *Nco*I and *Not*I sites in pEndo plasmid. The library with random mutations were transformed into DH12S competent cells harboring pCcdB including WT target site, and the transformants were selected on the selective plates as described above. The survival colonies were streaked on both the selective plates and the negative control plates (the non-selective plates supplemented with 0.02% L-arabinose and 35 μg/ml chloramphenicol), and the mutated endonuclease genes were amplified by PCR from clones that could survive only on the selective plates. The PCR fragments were cloned into pEndo plasmid, and were selected with the same procedures as for the first round selection.

Far-UV circular dichroism

CD spectra were recorded at 25°C between 190 and 260 nm using 0.1 cm path length quartz cuvettes (Hellma) and 6 μM I-AniI variants in 10 mM potassium phosphate buffer (pH 7.0). The spectra were analyzed in the continuous mode with 1 nm bandwidth, 8 s response, and a scan speed of 10 nm/min. Thermal denaturation curves were obtained by measuring the ellipticity of 10 nM protein at 222 nm from 5 to 80°C at 2°C/min. These experiments were independently repeated three times.

Isothermal titration calorimetry

The dissociation constant and thermodynamic values of Y2 I-AniI binding WT target or LIB4 target DNA sequences were analyzed as previously described (9). Briefly, Y2 I-AniI and the target DNA dialyzed in ITC buffer (50 mM *Tris*-HCl (pH 7.5), 50 mM NaCl, 10 mM $CaCl_2$) were quantified by UV spectroscopy, using extinction coefficients to calculate concentrations. Sample concentrations ranged from 5.5 to 6.0 μM I-AniI and 38 to 60 μM dsDNA target. A VP-ITC MicroCalorimeter (μCal) was used for all experiments, with the protein sample in the cell and the DNA target in the auto-pipette. Individual runs consisted of 30 injections of 5-8 μL each, depending on sample concentrations, and were conducted at 30°C with a 351 r.p.m. stirring speed. Data were fit using Origin 7 SR2 software.

In vitro cleavage assays

DNA substrates were amplified by PCR from pBluescript plasmid containing either WT target site or LIB4 target site, with primers designed to generate approximately 700 basepair products. The reaction mixtures (20 μl) contained 50 mM *Tris*-HCl (pH 7.5), 50 mM NaCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, 100 ng (~11 nM) of the DNA substrates, and the indicated amounts of I-AniI variants. The reactions ran at 37°C for 30 min, and were terminated by adding 5 μl of stop buffer (100 mM EDTA (pH 8.0), 2.5% SDS, 10% glycerol and 0.01% bromophenol blue). The products were separated on a 2% agarose gel with ethidium bromide, and quantified as described above.

For kinetic determination of the relative rates of individual DNA strand cleavage events, the 28 basepair DNA substrate was prepared from two oligonucleotides: 5′-ATT TGAGGAGGTTTCTCTGTAAATAATG and 5′- CATT ATTTACAGAGAAACCTCCTCAAAT. Oligonucleotides to be 5′ end-labelled were incubated in buffer with T4 polynucleotide kinase (New England Biolabs Beverly, Massachusetts) and [γ^{-32}P]-ATP (6000 Ci/mmole; Perkin Elmer Boston, Massachusetts). The reactions were incubated at 37°C for 45 min extracted with phenol:chloroform:isoamyl alcohol solution (25:24:1; PCI), and centrifuged through a Sephadex G-50 spun column (Sigma Chemical Co., St. Louis, M0).

The substrate duplex was formed by incubating the labelled oligo (1.25 μM, final) with the corresponding unlabelled partner oligo (5 μM, final) in 50 mM *Tris*-HCl (pH 7.5), 10 mM $MgCl_2$ and 50 mM NaCl (TMN buffer) at 90°C for 2 min and then cooled to 25°C over 20 min. DNA cleavage reactions were carried out at 37°C and contained substrate (6.25 nM) and TMN and the reactions initiated by the addition of WT I-AniI or Y2 I-AniI to a final concentration of 400 nM. Aliquots were removed at specific times (ranging from 0.5 to 40 min) and the reactions terminated by the addition of EDTA to a final concentration of 67 mM, followed by organic extraction. Samples were electrophoresed on 8% acrylamide/7 M urea gels, dried and exposed on a phosphorimager screen and visualized/quantified using a phosphorImager and ImageQuant software (Molecular Dynamics, Sunnyvale, CA). The imager counts in the substrate and product were summed and the fraction product plotted against time. The data fit to a first order rate equation: Fraction of 5′-labelled product = $A(1 - e^{-kt})$ where A is the amplitude of the product and k represents the pseudo-first order rate constant, k_{obs}. Note that the cleavage rate for the top strand in the small duplex reported here is greater than that reported previously for a larger (~200 bp) duplex (0.22 min^{-1}) while the rate for the bottom strand is the same (0.11 min^{-1}) (22). As noted in the text, the designation of 'top' and 'bottom' strand in these methods is opposite that previously described in (22) where the labeling of strands and cleavage products were inadvertantly reversed.

In vivo gene conversion assay

The DR-GFP reporter contains a GFP sequence interrupted by a HE target site and in frame stop codon as a recipient), followed by a 5′ and 3′-truncated GFP as a donor (23). Gene conversion event between these non-functional GFP genes, which is promoted by double strand breaks at the HE target site results in the expression of the intact GFP protein. 3×10^6 of 293T cells were plated 24 h prior to transfection in 12 well plates. The cells were transfected with 1.25 μg each of DR-GFP reporter and HE expression plasmid using calcium phosphate method. Media was replaced at 24 h post transfection, and the GFP positive cells were detected by flow cytometry

at 48 h post transfection. Western blots were probed with anti HA monoclonal antibody (Clonetech 6E2) and anti β-actin polyclonal antibody (cell signaling).

RNA splicing assay

The preparations of radiolabelled COBme-pre RNA as well as the kinetic analyses and data fits were carried as described previously (24). The imager counts in the precursor and splicing products were summed and the fraction precursor remaining plotted against time. The protein-dependent splicing data were fit to a first order equation with a double exponential: Fraction RNA bound = $A(e^{-kt}) + B(e^{-kt})$ where A and B are the amplitude of RNA in each phase and k represents the first order rate constants for each phase. Only the fast phase is reported in Supplementary Table S2. The slow phase, which presumably represents a population of RNA that undergoes a slow folding step required for protein binding or RNA catalysis, (24) was 0.025 (±0.004) and 0.024 (±0.009) min^{-1} for the WT and Y2 proteins.

Crystallographic analysis

The DNA oligonucleotides used for cocrystallization (5′-GCG CGC TGA GGA GGT TTC TCT GTA AAG CGC A-3′ and 5′-GCG CTT TAC AGA GAA ACC TCC TCA GCG CGC T-3′) were purchased from Integrated DNA Technologies (1 mM scale, HPLC purified). The oligonucleotides were dissolved in 10 mM *Tris*-EDTA (TE) buffer pH 8.0, and the complementary DNA strands were annealed by incubation at 95°C for 10 min and slow cooling to 4°C over a 6 h period. Purified Y2 I-AniI protein (at 10 mg/mL concentration) in 50 mM Hepes-NaOH (pH 7.5), 500 mM NaCl, 20 mM $CaCl_2$, and 5% (v/v) glycerol was mixed with 1.5-fold molar excess of the DNA substrate. The protein-DNA drops were mixed in a 1:1 volume ratio with a reservoir solution containing 80 mM Bis-*Tris* (pH 6.2), 50 mM $CaCl_2$, and 32% (v/v) polypropylene glycol P400 and equilibrated at 22°C. Crystals grew within two weeks, and were frozen by direct looping and submersion into liquid nitrogen. The crystals diffracted up to approximately 2.7Å resolution at the ALS beamline 5.0.1. Data sets were processed using the DENZO/SCALEPACK software package (25). The structure of Y2 I-AniI/DNA complex was solved by molecular replacement using the PDB coordinates of the WT I-AniI/DNA complex (PDB: 1P8K), and was refined using programs REFMAC5 (26) and CNS (27) (Supplementary Table S3).

RESULTS

Generation of I-AniI variants with optimal *in vivo* activity towards its WT target sequence

Compared to other well-characterized LHEs, WT I-AniI binds to its physiological, WT target site with reduced affinity (the dissociation constant of I-AniI to its that site is estimated to be ~100 nM) and displays relatively slow and incomplete cleavage of that site under optimized *in vitro* reaction conditions (9). The enzyme also displays a temperature optima (~50°C) that is far removed from the growth optima of its physiological host (this work). As a result, WT I-AniI displays poor cleavage activity in transfected cell lines, including both human cells and in bacteria.

In order to optimize the activity of I-AniI in transfected cells (and to subsequently study the basis for any observed improvement), we conducted a directed evolution experiment to improve cleavage activity of I-AniI against its WT target site. This experiment was conducted using the bacterial screen for HE cleavage activity developed by Doyon *et al*. (21), that reports on elimination of a reporter plasmid that encodes a cytotoxic protein (see Materials and Methods section). We induced random mutations in the open reading frame of I-AniI using low-fidelity PCR amplification, and screened for the variants that display efficient DNA cleavage using the bacterial selection system and an endonuclease library containing ~130 000 mutated transformants.

After two rounds of selection, we successfully isolated nine unique endonuclease variants which significantly improved cell growth and survival under selective conditions (Figure 1 and Table 1). These constructs each contained between one and four individual amino-acid substitutions. The variants of I-AniI displayed between 17-fold and 475-fold increases in survival rate and recovery as compared to the WT I-AniI endonuclease (between 3.5% and 95% survival, compared to less than 0.2% for WT I-AniI).

A total of 10 separate amino-acid positions were mutated (Figure 1), and a total of 13 separate point mutations were distributed across these positions. Two amino-acid positions (F13 and S111) yielded multiple mutations (F13 substituted with Y, L or V and S111 substituted with Y or F). Three of the thirteen mutations (23%) correspond exactly to residues found in one or more close homologues of I-AniI: I55V, S92T and S111Y (Figure 2). Two additional mutations (F13L and K49Q) correspond to similar amino acids also found in those homologues (an isoleucine is found at residue 13 and either an asparagine or an aspartate at residue 49). One construct contained a single mutation (F13Y); this mutation displayed a 180-fold increase in survival and recovery over WT I-AniI. Eight out of nine variants contained a mutation at this position (F13Y, L or V). A variant of I-AniI harboring only two mutations (F13Y/S111Y) yielded nearly 100% survival rates and formation of larger colonies than other variants (Figure 1A). This construct, termed 'Y2 I-AniI', represented the simplest mutational route to robust *in vivo* endonuclease activity and was chosen for detailed biochemical and structural studies.

Since most of the isolated variants possessed amino-acid substitutions at multiple positions, we deconvoluted several of the mutations via individual site-directed mutations into the WT protein scaffold, to determine their relative contribution of mutations that improve cell growth. The F13Y mutation (located in the first LAGLIDADG helix of the protein scaffold, positioned in the protein's hydrophobic core) was the single amino-acid substitution that contributed most significantly to an increase in the survival rate. The effect of this mutation was enhanced in

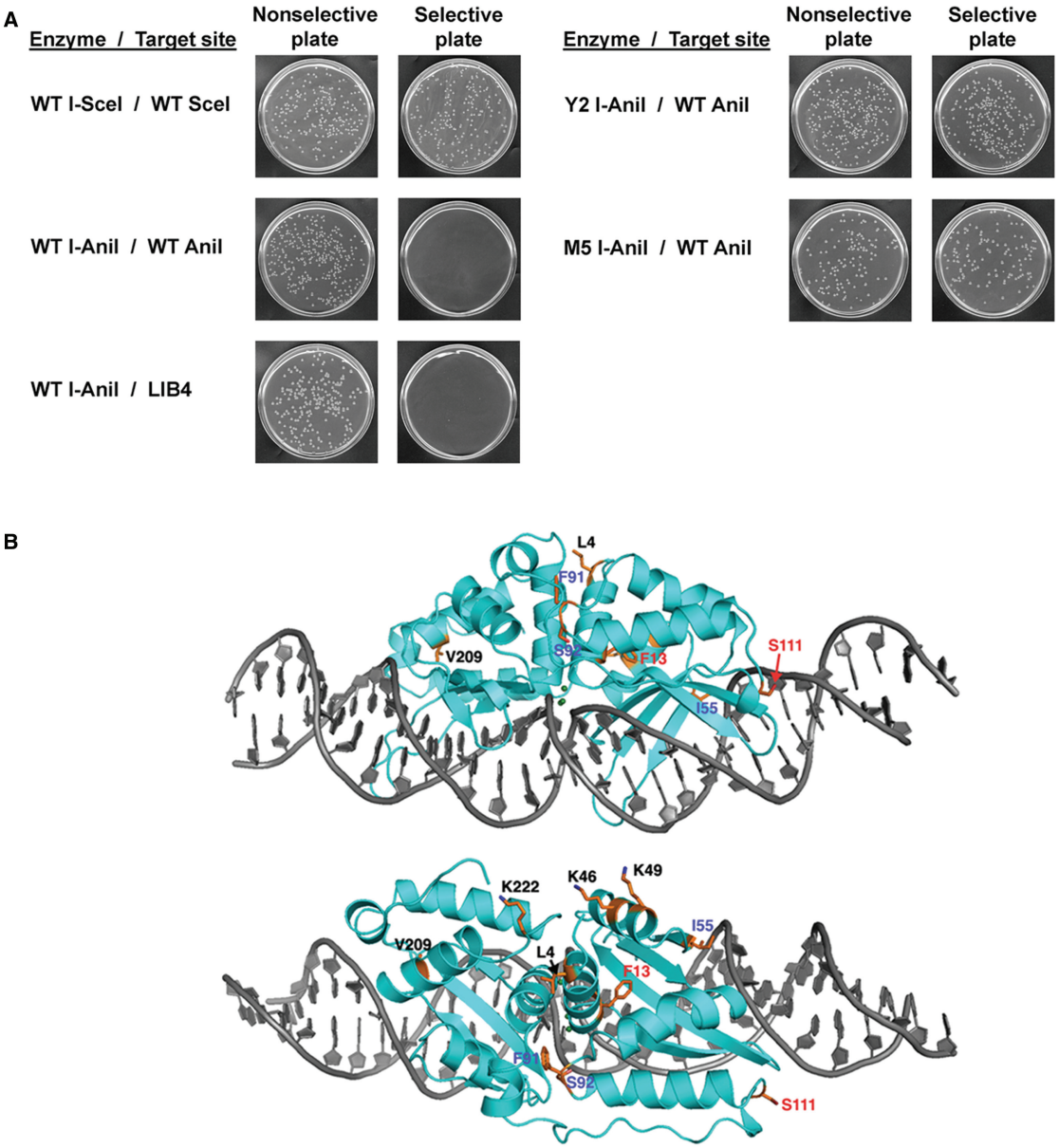

Figure 1. Isolation of optimized I-AniI variants by a bacteria based cleavage assay. (**A**) The expression of the toxic gene is restricted on the selective plates. WT I-SceI significantly cleaved its wild type target site in bacteria, but WT I-AniI inefficiently did against its physiological (WT I-AniI) and optimal (LIB4) sequences. The Y2 and M5 variants improved the cleavage activity. (**B**) Ribbon model of WT I-AniI structure. Gray sticks indicates the physiological target DNA. Mutated residues in Y2 construct are labeled in red; remaining mutated residues in M5 construct are labeled in blue.

the presence of several additional mutations, particularly the addition of S111Y (the additional mutation present in the 'Y2 I-AniI' constructs). This amino-acid substitution corresponds to the most strongly conserved alternative residue in homologues of I-AniI (Figure 2). On its own, the S111Y substitution increases survival by over 10-fold relative to WT I-AniI (2.7% recovery versus less than 0.2%) (Table 1).

Although the addition of the S111Y mutation to F13Y offered the most direct route to complete optimization of endonuclease activity, a variety of alternate combinations of mutations also increase bacterial survival (including those that echo conserved residues in I-AniI homologues, and others that are unique to the output of the screen). The mutations I55V and S92T, both of which are also recapitulated in one or more naturally occurring homologues of I-AniI, independently increase survival by ~2-fold and up to 30-fold, respectively when compared to analogous variants that lack those mutations. Similarly, the mutation F91I (which does not correspond to naturally occurring residues in homologues of I-AniI) also enhanced activity and survival.

Three additional mutations found in the survivors from the activity screen (K46Q, K49Q and K222R) alter a series of surface-exposed basic residues that are clustered together on a surface far from the bound DNA (Figure 1B). Various combinations of these three mutations in control experiments indicate that the contribution

Table 1. I-*Ani*I variants identified by a bacteria-based cleavage assay

I-*Ani*I variants	Survival rate (%)[a]
Wild type	<0.2
F13Y	**36 ± 3.8**
F13L, K46Q, S92T	75 ± 9.6
F13Y, S111Y (Y2)	**95 ± 30**
F13Y, I55V	85 ± 16
F13Y, K46Q, S92T	93 ± 4.7
F13V, K46Q, S92T	3.5 ± 1.2
F13Y, S111F, K222R	94 ± 18
F13L, K46Q, F91I	71 ± 11
L4I, K49Q, S111F, V209A	90 ± 17
F13L, S92T[b]	68 ± 11
F13L, K46Q[b]	2.7 ± 2.4
S111Y[b]	**2.6 ± 4.2**
S111F[b]	0.26 ± 0.28
F13Y, K222R[b]	37 ± 13
L4I, K49Q, V209A[b]	<0.2
F13Y, I55V, F91I, S92T, S111Y (M5)[b]	92 ± 10

[a]The survival rate was determined by dividing the number of colonies on the selective plates by that on the nonselective plates. Errors refer to mean ± S.D. of three replicate plating experiments.
[b]These variants were created by site-directed mutagenesis.
Boldface correspond to mutations (alone and in combination) found in the Y2 variant of I-AniI.

of those residues to activity and survival is relatively minimal. It is possible that these mutations block a secondary nucleic acid binding activity (perhaps related to its maturase function) in such a way as to reduce a competing source of toxicity caused by over-expression of the protein.

Taken together, it appears that five individual mutations (F13Y, I55V, F91I, S92T and S111Y/F) contribute most significantly, in various combinations, to observed improvements in cleavage of WT I-AniI site in bacteria. Simultaneous incorporation of all of these substitutions (F13Y/I55V/F91I/S92T/S111Y; termed 'M5 I-AniI') facilitated significant cell survival (Figure 1 and Table 1) at a level similar to Y2 I-AniI, suggesting that the combination of all five mutations neither impairs nor further improves the activity of the enzyme in transfected cells, relative to the simpler Y2 double mutation (F13Y/S111Y). Like the 'minimally mutated' Y2 construct described above, this 'maximally mutated' construct was also selected for additional biochemical analyses.

Thermal stability and thermal activity profiles

The location of the point mutations in the Y2 and M5 protein variants (three of which are distant from nucleotide bases, and partially or fully buried in the protein core) suggested that the stability and/or dynamic flexibility of the folded proteins might be altered. The circular dichroism spectra of WT, Y2 and M5 variants indicated that all three enzymes possess similar secondary structure content (Figure 3A), and is consistent with the previous observation, the mixed α/β topology of the I-AniI crystal structure. The WT enzyme and the Y2 variant display similar thermal denaturation curves, with observed transition midpoints (T_M) of 53.9 and 52.6°C, respectively (Figure 3B). In contrast, M5 variant was partially destabilized, with a reduction in the T_M value of approximately 10°C (43.5°C). Because the denaturation and/or aggregation of the protein at high temperature are irreversible, exact changes in the free energy of protein unfolding cannot be determined using this system.

The Y2 and M5 enzymes also display altered thermal reaction profiles, corresponding to enhanced cleavage of the DNA target at the physiological temperature (30°C) that was used in the selection experiment (Figure 3C). Whereas the WT enzyme displays a relatively sharp temperature optima for cleavage at approximately 50°C, the Y2 and M5 enzyme variants respectively display broadened (Y2) or reduced (M5) thermal activity profiles and optima.

Binding affinity

Because the Y2 construct displays the highest survival frequency in the bacterial screen for *in vivo* cleavage activity, contains fewer mutations and is physically more stable than the M5 mutant, its DNA-binding and endonuclease activity was examined further. The affinity and thermodynamic profile of cognate DNA target binding was measured using isothermal titration calorimetry. The Y2 variant displayed significantly tighter binding to its physiological target site than the WT enzyme, with a measured dissociation constant corresponding to 12 ± 3.0 nM (compared to a K_D value for the WT enzyme of approximately 96 nM) (Figure 4A and Supplementary Table S1). The tighter binding for Y2 I-AniI corresponds to a reduction in $\Delta G_{binding}$ (−1.2 kcal/mol) that results from a more favorable entropic change ($-T\Delta S$, −7.2 kcal/mol), which was partially offset by a more unfavorable enthalpic change (ΔH, +6.0 kcal/mol). In contrast, the affinity of the Y2 variant to the 'LIB4' target site [a sequence differing from the WT site at two basepairs, that is bound and cleaved more efficiently by the WT enzyme (9)] is nearly unchanged (Supplementary Figure S1 and Table S1).

In vitro cleavage activity

A series of *in vitro* cleavage assays were carried out using a linearized DNA construct containing a single copy of the WT target site as a substrate. WT I-AniI was unable to completely cleave this site under the digest conditions used, and the generation of cleavage products was decreased at higher concentrations of the protein (Figure 4B). In contrast, the Y2 I-AniI variant hydrolyzed the substrate more efficiently, at reduced concentrations of enzyme. Under the same reaction conditions, Y2 and WT enzyme constructs cleave the 'LIB4' substrate with similar efficiencies (data not shown). Therefore, the selected enzyme constructs display improved activity against the 'suboptimal', WT I-AniI site (which was used in the selection), but not against the closely related LIB4 site (which was not used in the selection). This implies that the specificity of the Y2 I-AniI construct is slightly reduced with respect to discrimination between these two target sequences; additional studies of binding and cleavage specifities across the entire target site are in progress.

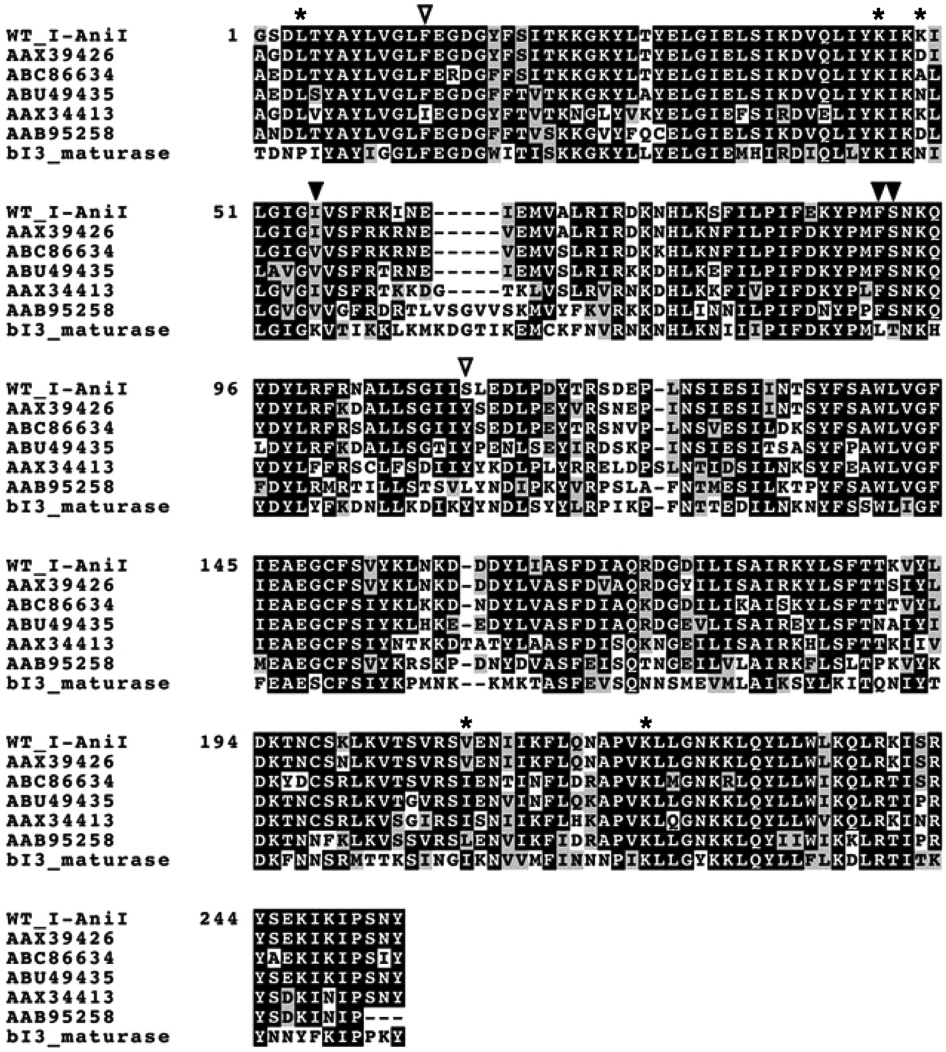

Figure 2. Sequence alignment of I-AniI homologs. The alignment was performed using CLUSTALW program. Open triangles indicate positions of F13Y and S111Y in the Y2 I-AniI construct. AAX39426, ABC86634, ABU49435, AAX34413, AAB95258 and bI3 maturase are LAGLIDADG proteins from *Neosartorya fischeri*, *Gibberella zeae*, *Phaeosphaeria nodorum SN15*, *Agrocybe chaxingu*, *Venturia inaequalis* and *Saccharomyces cerevisiae*, respectively. Closed triangles represent sites of additional mutations I55V, F91I and S92T in the M5 I-AniI construct. Asterisks indicate remaining positions of point mutations isolated in the cleavage activity screen as listed in Table 1.

I-AniI is an asymmetric LHE that displays a sequential mechanism of DNA cleavage. The C-terminal active site cleaves the top strand of its target site (the coding strand of *Aspergillus nidulans* mitochondrial apocytochrome *B* gene) significantly faster than the N-terminal active site cleaves the bottom (non-coding) strand (22). To test whether the Y2 variant changed the relative efficiencies of individual strand cleavage events, the rates of hydrolysis for each strand were measured. In these experiments, a small 28 nt duplex was prepared with either the top or bottom strand end-labeled with ^{32}P. Cleavage reactions were performed under single turnover conditions with saturating concentrations of WT I-AniI or the Y2 variant at 37°C. Under these conditions, the WT enzyme cleaved the top strand ~5-fold faster than the bottom strand [0.51 (±0.14) min^{-1} versus 0.10 (±0.01) min^{-1}; Figure 4C, Supplementary Table S2]. While cleavage of the top strand by the Y2 variant was not significantly different than the WT enzyme [0.45 (±0.10) min^{-1}, compared to 0.51 min^{-1}], the bottom strand was cleaved approximately 2-fold faster than the WT enzyme [0.21 (±0.05) min^{-1}, compared to 0.10 min^{-1}]. Thus, the Y2 I-AniI construct appears to specifically enhance the 'slow' cleavage by the N-terminal active site of I-AniI.

It should be noted that the designation of 'top' and 'bottom' strand in the paragraph above is opposite that previously described in (22) where the labeling of strands and cleavage products were inadvertently reversed.

Splicing activity

The effect of the F13Y and S111Y mutations on maturase activity was assessed by *in vitro* splicing assays. I-AniI stimulates splicing of the COB intron via the rapid formation of an initial RNA/protein complex which is then is

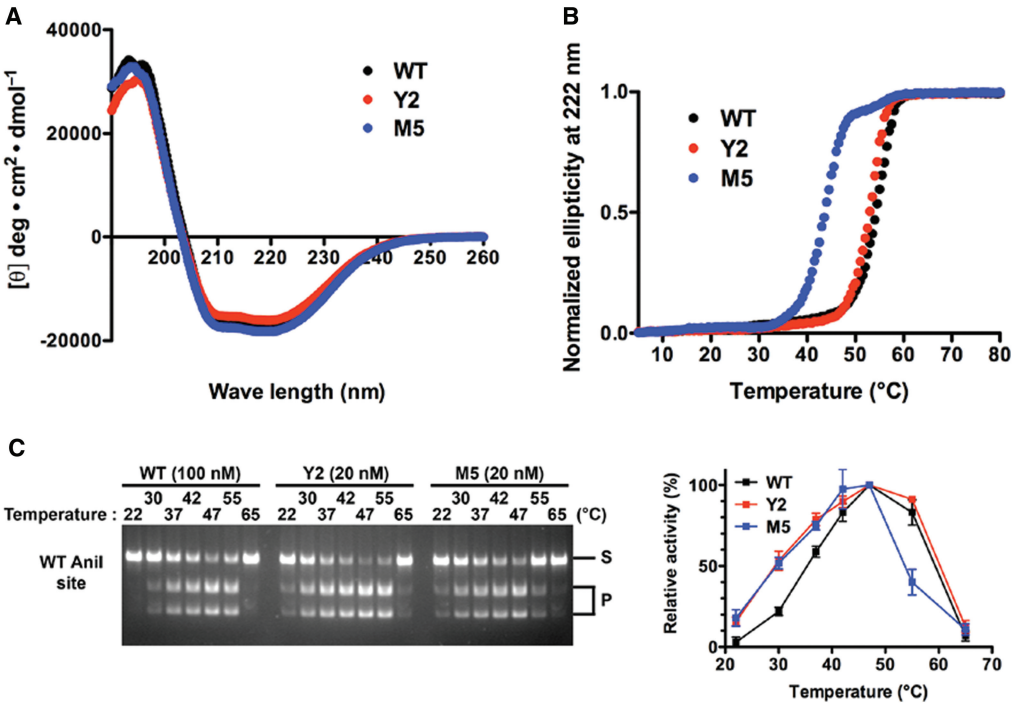

Figure 3. Secondary structure, thermal stability and thermal activity profile analyses of WT and mutant I-AniI variants. (**A**) Circular dichroism analysis. Far-UV CD spectra were measured between 190 and 260 nm at 25°C. (**B**) Temperature-dependent denaturation on the basis of changes in circular dichroism ellipticity at 222 nm. Temperature of 50% transition was 53.9, 52.6 and 43.5°C for WT, Y2 variant and M5 variant, respectively. (**C**) Thermal profiles of the cleavage activity. The reactions ran at 22, 30, 37, 42, 47, 55 and 65°C. Relative activity was shown by taking cleavage activity at 47°C as 100%. Errors mean ± S.D.

slowly resolved into a native, splicing-competent particle (24,28). Importantly, under non-saturating protein concentrations [~100 nM; (29)] the rate of splicing is dependent on both the bimolecular rate constant for formation of the initial complex and the resolution step (24). Therefore, the maturase activity was measured using trace amounts of COBme-preRNA and 20 nM protein. Under these conditions, the rate of splicing for the Y2 protein was the same within error to the WT protein (1.3 (±0.37) min^{-1} versus 1.7 (±0.66) min^{-1}; Figure 4D, Supplementary Table S2).

In vivo recombination activity

Homing endonucleases are under intense study as potential reagents for gene application, but to date only the I-SceI and I-CreI LAGLIDADG proteins have been reported to be capable of inducing homologous recombination in mammalian cells by generation of double strand breaks (10,30–32). To measure the relative ability of WT I-AniI and the Y2 variant to induce gene conversion in transfected human cells, we used an assay in which human cells are cotransfected with an expression vector for either protein, in combination with a 'DR-GFP' reporter plasmid that encodes two non-functional GFP genes (Figure 5A). One copy of the GFP gene is interrupted with WT I-AniI site and two stop codons in the middle of the open reading frame, while the other is prematurely truncated (23). Gene conversion between these two gene copies, either spontaneously or as a result of a double-strand break, leads to complementation of GFP and production of measurable fluorescent cells.

Human 293T cells were co-transfected with this reporter plasmid in the presence or absence of the expression construct encoding WT I-AniI or the Y2 variant, and GFP positive cells were detected by flow cytometry (Figure 5B). Transfection of the DR-GFP reporter plasmid alone leads to a measurable but relatively low background of GFP expression, caused by spontaneous conversion between the GFP genes on the reporter, whereas almost 100% of the cells were GFP positive when the GFP expression plasmid (as a positive control) was transfected in the same experimental conditions (data not shown). Co-transfection with WT I-AniI expression plasmid slightly increased the fraction of GFP positive cells over the background. In contrast, transfection with the expression construct for Y2 variant resulted in a dramatic increase in the GFP positive cells (Figure 5B and C).

In a separate control experiment to demonstrate the specificity of action of the Y2 variant, a DR-GFP reporter plasmid containing the target site for the I-SceI site instead of the WT I-AniI site was used, leading to gene conversion only when the gene for I-SceI was introduced in place of Y2 I-AniI (Figure 5D). Finally, the expression level of each HE in transfected human cells was verified by western blotting assay (Figure 5E). These results suggested that Y2 variant induced highly efficient gene conversion by generating double strand breaks at its physiological target site in 293T cells.

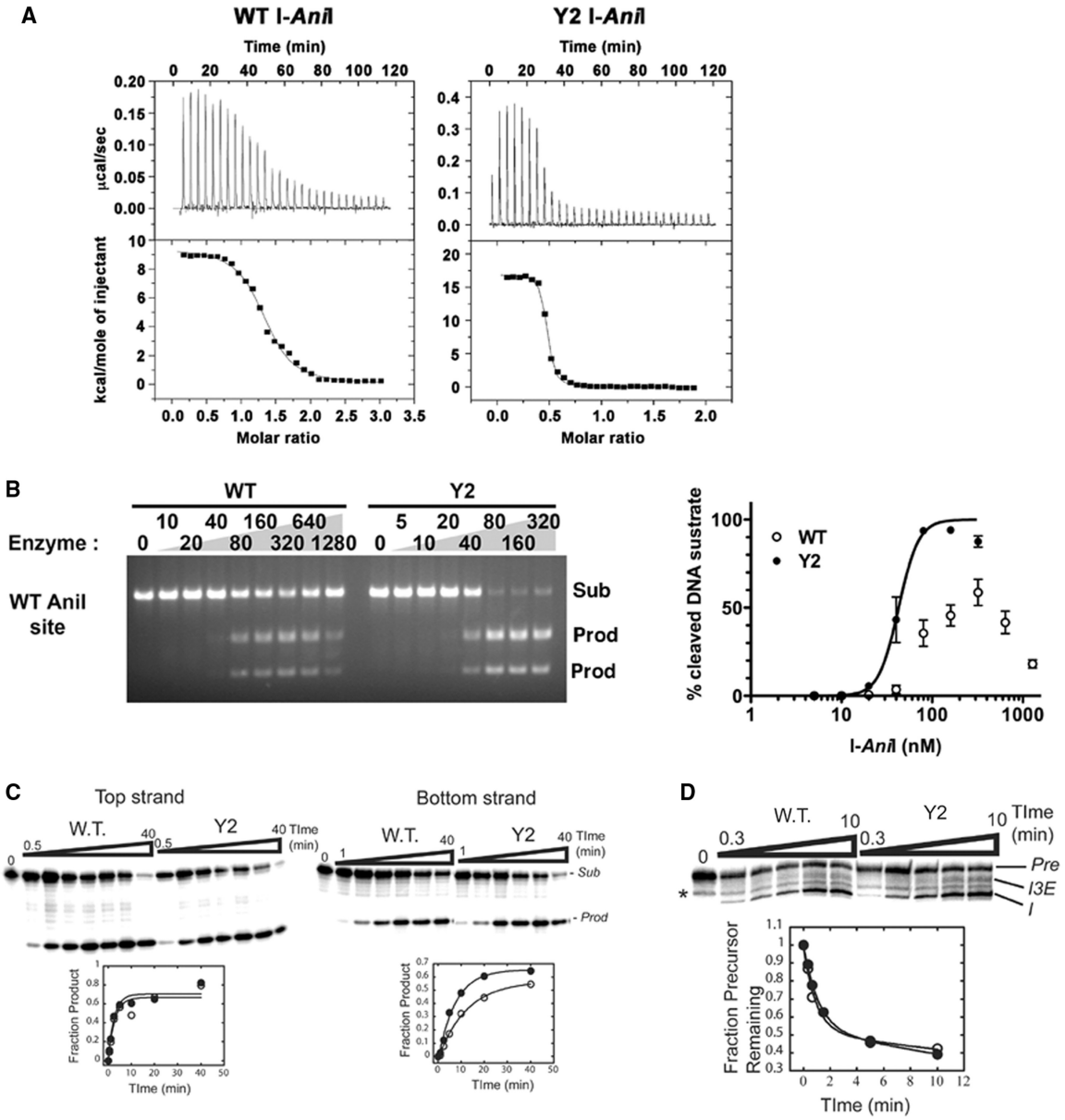

Figure 4. DNA binding affinity, DNA cleavage activity and intron splicing activity of WT and 'Y2' optimized I-AniI variant. (**A**) Binding assay with WT I-AniI (left) and Y2 variant (right) against WT I-AniI site by isothermal titration calorimetry. The left graph is taken from our previous report (9). Dissociation constants and thermodynamic values are summarized in Supplementary Table S1. (**B**) *In vitro* cleavage assay against the WT I-AniI target site. A serial dilution of each I-AniI variant was incubated with PCR fragment containing a single copy of the target site at 37°C. The reaction products were quantified using the NIH ImageJ program. Errors mean ± S.D. Sub: linearized plasmid substrate; prod: cleaved products. (**C**) Relative cleavage of top and bottom DNA strands by WT and Y2 I-AniI. The top strand includes the coding sequence of the mitochondrial apocytochrome *B* gene from *Aspergillus nidulans* (5′-TGAGGAGGATTTCTCTGAAA-3′). Sub = substrate; Prod = product. Open ovals = WT, filled ovals = Y2. The data are fit to an equation that describes a first order reaction (see Materials and Methods section). Note that the Y2 variant cleaves the bottom strand more rapidly that the WT protein. For values, see Supplementary Table S2. The designation of 'top' and 'bottom' strand in these methods is opposite that previously described in (22) where the labelling of strands and cleavage products were inadvertantly reversed. (**D**) Relative intron splicing activity of WT and Y2 I-AniI. Pre = precursor RNA (substrate); I3E = splicing intermediate after 5′-splice site cleavage with the 3′-exon still linked to the intron; I = free intron (product). The ligated exon product is not shown. The asterisk demarcates an RNA degradation product. Open ovals = WT; filled ovals = Y2. The data are fit to an equation that describes a first order reaction (see Materials and Methods section). For values, see Supplementary Table S2.

Crystal structure

We then solved the crystal structure of the Y2 I-AniI/DNA complex (Y2/DNA) complex to 2.7Å resolution (Figure 6 and Supplementary Figures S2 and S3). The data and refinement statistics are shown in Supplementary Table S3. The crystals were grown in the presence of calcium ions, the DNA substrate was uncleaved, and three independent complexes were present in the asymmetric unit.

Y2 I-AniI displays a similar overall structure to the structure of WT I-AniI (9,20), except for an

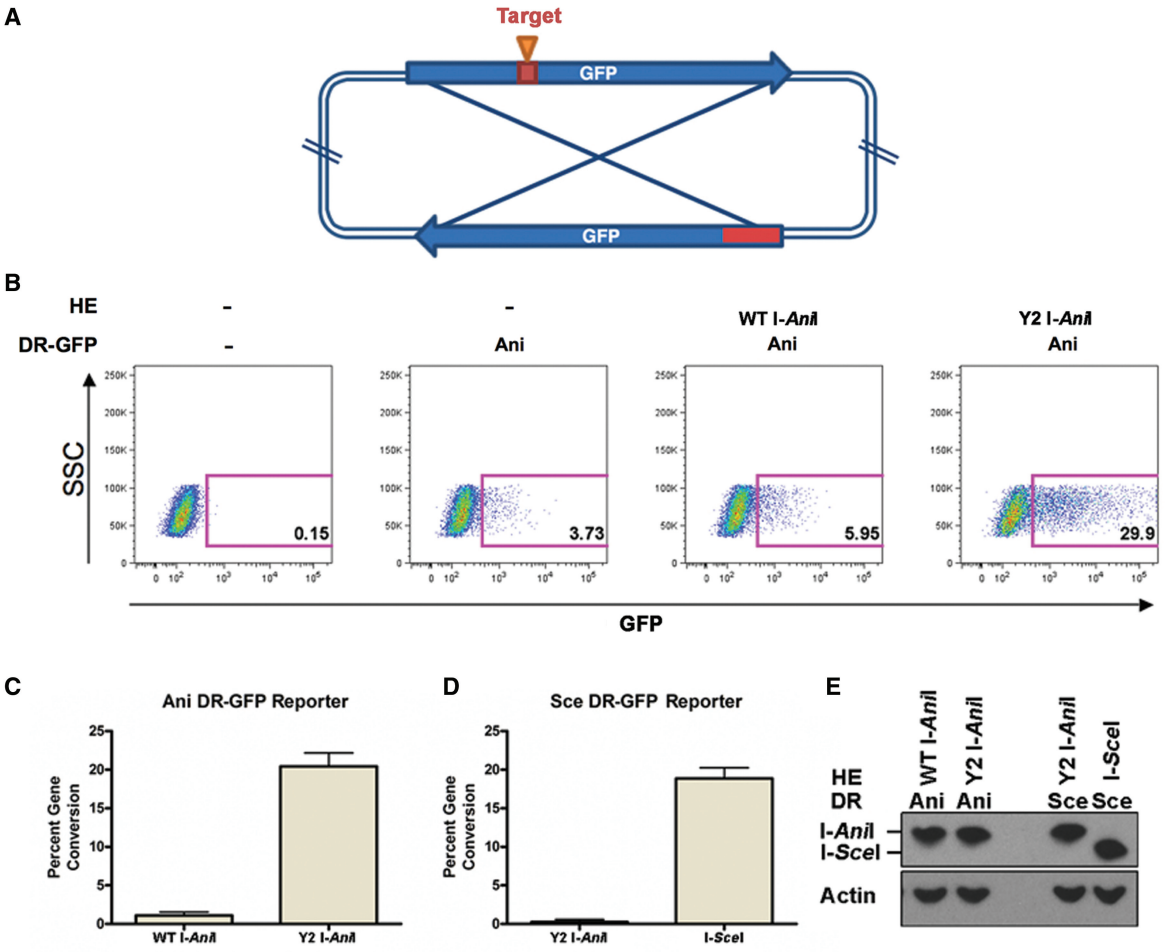

Figure 5. Comparative *in vivo* gene conversion activity of WT I-AniI and the Y2 variant. (**A**) Schematic representation of the DR-GFP reporter plasmid. The reporter plasmid harbors two non-functional copies of the GFP gene, one interrupted by two stop codons followed by a target site for either I-AniI or I-*Sce*I, the other encoding the 5′- and 3′-truncated gene. (**B**) Representative FACS plot of gene conversion on the DR-GFP reporter with WT AniI site (I-AniI DR-GFP). 293T cells were transiently co-transfected with the reporter and an expression plasmid for either WT I-AniI or Y2 variant, and analyzed by flow cytometry at 48 hours post transfection. In each plot, the fluorescent side scatter 'SSC' (as a standard of cell viability) is plotted on the *Y* axis and the GFP-associated fluorescent signal 'GFP' (a measurement of recombiantion of the reporter plasmid in the transfected cells) is plotted on the *X*-axis. (**C,D**) graph depicting percent gene conversion of I-AniI DR-GFP (C) and I-SceI DR-GFP (D) induced by either WT I-AniI, Y2 variant or I-SceI. Data represents three independent experiments performed in duplicate normalized to observed background by transfection of the DR-GFP reporter alone, Errors mean + SE. (**E**) Expression of HEs in the transfected cells. The expression level was analyzed by Western blotting using an antibody against hemagglutinin (HA)-epitope (upper) and β-actin as a control (lower).

extended peptide sequence (residues 108–127) linking the N-terminal and the C-terminal domains. This region contains one of the two amino-acid substitutions in the Y2 construct (S111Y) (Figure 6B and 6C). The overall bend angles of the DNA in the WT and Y2 complexes were also quite similar (Supplementary Figure S4), although with a local shift in the major groove dimensions spanning several basepairs, as described below. The structure indicates that mutation of serine 111 to tyrosine (S111Y) causes a rearrangement of several neighboring residues that leads to a repositioning of the N-terminal active site lysine residue (K94).

Residue Y111 in the Y2 variant makes a new, direct contact with the DNA backbone near base −11C (one position upstream from the protein's specifically recognized target site) (Figure 6C). This contact, and the motion of the surrounding residues, results in a corresponding shift in the neighboring side chain of R120 and a movement of the bottom DNA strand (at positions from −3 to −6) by up to 2.3Å, resulting in a localized narrowing of the DNA major groove (Figure 6D). These conformational changes are augmented by significant rotations of the side chains of N93 (which is a structural neighbor to R120) and its immediate neighbor, K94. The amino nitrogen of this latter side chain, which is hypothesized to act as a general base in one of the two endonuclease active sites, moves by 2.5Å toward the first LAGLIDADG motif, into a conformation that mirrors that of its K227 counterpart in the C-terminal active site (Figure 6E and Supplementary Figure S2).

The base-pairing interactions at positions −3 and −4 in the target site (which display the largest change in position relative to the WT complex, and are the

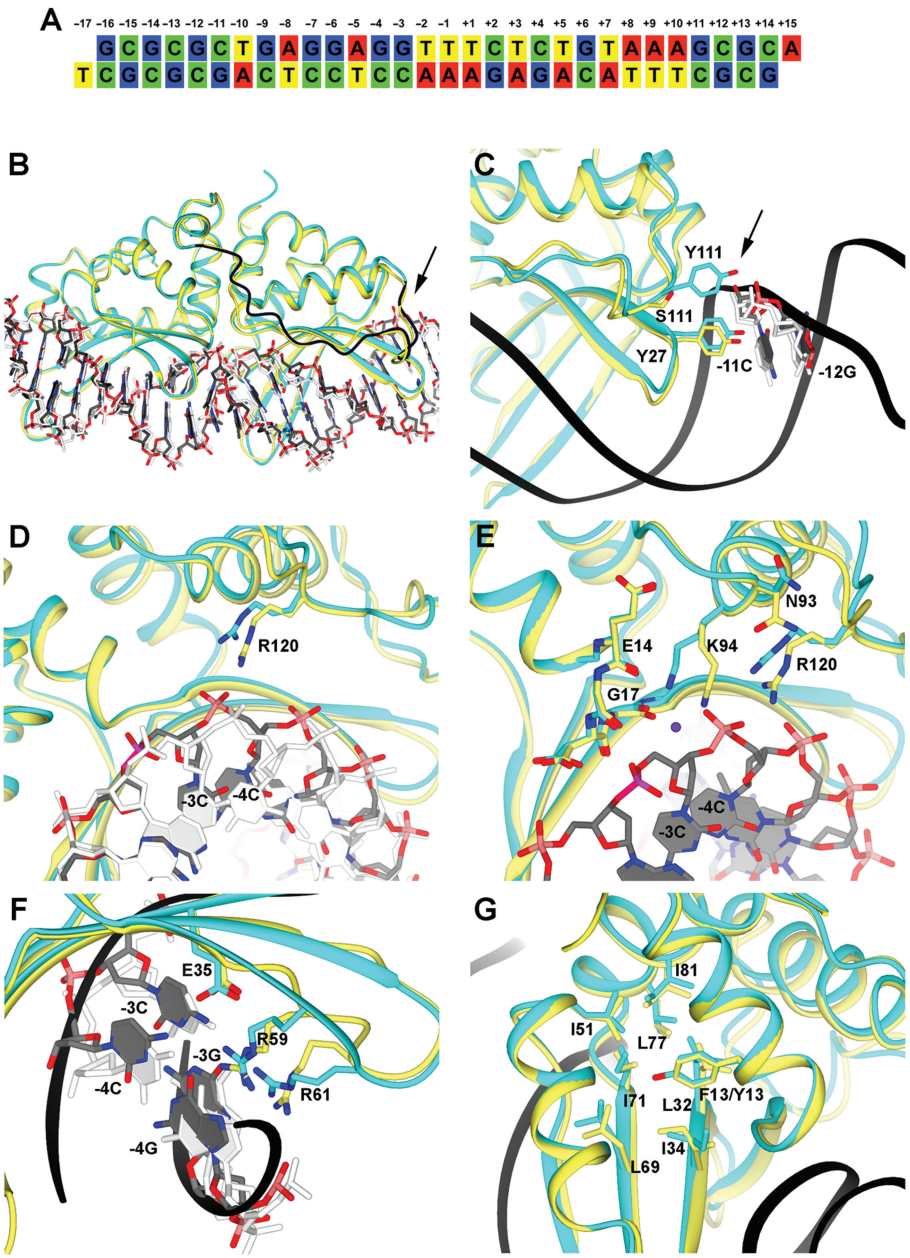

Figure 6. Structural analysis of the Y2 I-AniI construct. Electron density maps for several of the structural features shown here are provided in Supplementary Figure S3. For clarity of viewing, the orientation of the structure shown in panels B–G are rotated 180°C relative to the canonical orientation of the DNA substrate sequence shown in panel A. (**A**) Sequence of the DNA duplex used for the crystallization. The target site (20 basepairs) is numbered from positions −10 to +10, relative to the center of the four-base overhangs generated by DNA cleavage. The strand breaks are introduced between positions +2 and +3 in the top strand, and between −2 and −3 in the bottom strand. (**B**) Y2 I-AniI (cyan) in complex with the DNA duplex (grey) superimposed against WT I-AniI (light yellow) bound to the same DNA target sequence (white) (PDB: 2QOJ). The extended loop of Y2 I-AniI linking the N-terminal and the C-terminal domains is shown in dark blue. The arrow in panel B and C points towards S111Y. (**C**) Magnification of region surrounding residue 111 (S111Y mutation in the Y2 I-AniI construct). (**D**) Magnification of region surrounding residue 120. (**E**) Magnification of region surrounding residue 94 and its surrounding active site residues and DNA bases. (**F**) Magnification of region surrounding basepairs −3 and −4. (**G**) Magnification of region surrounding residue 13 (F13Y in the Y2 I-AniI construct).

sites of specific, high affinity interactions with several protein residues) are maintained by shift of the guanine bases of the top strand (Figure 6F). Contacts between these bases and the protein (R59 and −4G, R61 and −3G, and E35 and −4C) are also maintained by slight changes in rotameric conformations of those side chains.

In contrast to the direct influence of the S111Y mutation on the structure of the N-terminal active site, residue 13 in I-AniI (the site of a the F13Y substitution from the screen) is buried in the hydrophobic protein core of the protein's N-terminal domain, and does not seem to cause major changes in the hydrophobic residues around Y13, other than the addition of an additional hydroxyl group

that is in contact with the neighboring protein backbone (Figure 6G).

A schematic representation of the contacts exhibited by the WT endonuclease and the Y2 variant are shown in Supplementary Figure S5. The Y2 I-AniI/DNA complex is packed in a different space group relative to the WT I-AniI/DNA complex, and the Y2 variant structure was solved at relatively lower resolution (2.7Å). Nevertheless, the Y2 enzyme displays very similar protein-DNA interactions, especially with respect to direct contacts to DNA bases.

DISCUSSION

Laboratory reoptimization of homing endonuclease activity partially reverses evolutionary drift

It has been reported that LHEs can accomplish invasion of intron-deficient alleles with much higher frequency than expected from their measured *in vitro* cleavage efficiencies, suggesting that a significant amount of mutational deterioration can occur before the persistence and propagation of the endonuclease is irreversibly compromised (33). I-AniI clearly displays many of the suboptimal biophysical and enzymatic properties outlined above, that were reoptimized in the laboratory through the identification of amino-acid substitutions that were selected solely on the basis of improved cleavage activity in bacteria. These mutations, when considered separately and in combination, appear to affect the protein scaffold in three ways: by increasing target site affinity, by improving the catalytic efficiency of one active site, and by increasing cleavage activity at 30°C (a temperature well below the thermal optima of the WT enzyme).

The F13Y mutation, which is part of the LAGLIDADG sequence motif and is found in the protein core, clearly represents the most significant mutational 'hot spot' identified in the cleavage activity screen (with mutations to both leucine and valine also recovered in addition to tyrosine). This residue is conserved as a phenylalanine in five out of six of the closest homologues of I-AniI (with an isoleucine found in one sequence; Figure 2). When the sequence alignment to I-AniI is extended to more distant homologues (corresponding to BLAST alignment E values of 10^{-30}–10^{-6}) this residue is observed to be heavily diverged, with its position predominantly occupied by small aliphatic and β-branched amino acids (especially valine and threonine).

Given the location of this residue in the completely buried hydrophobic core of the protein, it seems likely that the effect of the mutations recovered at residue 13 (F13Y/L/V) is to increase the global dynamic flexibility of the protein, leading to the reduced thermal temperature profile exhibited by both the Y2 and the M5 enzyme constructs (and the increased survival at the *in vivo* screening temperature of 30°C). Such effects have been well-documented in both artificially engineered and naturally-occurring enzymes (34–36). If true, this may imply that the ancestral form of I-AniI and its immediate homologues originated in a more thermophilic host organism [prior to transfer to fungal genome(s) such as *Aspergillus*] and that the variation of this residue position strongly influences overall catalytic efficiency as a function of ambient temperature.

The additional mutation recovered in the Y2 I-AniI variant (S111Y) appears to convert the corresponding amino-acid identity to a residue (tyrosine) that is more strongly conserved in homologous endonucleases (Figure 2). This suggests that ancestors of the I-AniI endonuclease lineage may have possessed a tyrosine residue at this position, which is conserved in most homologues in a multi-sequence alignment. A crystal structure analysis of the Y2 variant, combined with the kinetics of individual strand cleavage events, indicates that this mutation is coupled to conformational rearrangements of both the protein and the DNA substrate that lead to a beneficial rearrangement, and catalytic enhancement, of the N-terminal active site. However, on its own this mutation has less effect on *in vivo* activity than does F13Y (Table 1).

Therefore, the poor *in vivo* activity of I-AniI in the bacterial screen appears to primarily result from inadequate activity at low temperature, and secondarily from a slightly malformed active site in one of the two protein domains. It may be that at significantly elevated temperatures, the relative importance of the two mutations in the Y2 I-AniI construct would be reversed, with the S111Y mutation contributing most strongly to the catalytic performance of the enzyme.

The position of two additional mutations that individually synergize with F13Y (I55V and S92T) appears to partially share structural and evolutionary features with the S111Y mutation described above. The side chain of V55 is in proximity to the DNA backbone (positioned to make non-specific contacts to the substrate) and S92T is also found in close proximity to K94 (Figure 2). Thus, the route to recovery of robust *in vivo* endonuclease activity seems to consistently involve a mutation that alters the thermal profile of the enzyme in combination with at least one additional mutation that alters the structure of the protein-DNA interface and/or the architecture of one of the two active sites.

Although almost all individual mutations described above are readily observed in homologues of I-AniI, their exact combination(s), as recovered in active clones from the screen, are not. This might be explained by two facts. First, the number of potential combinatorial interactions between distant residues in the protein scaffold, that act synergistically to optimize multiple properties and functions (thermal activity, binding affinity and catalytic efficiency) are virtually limitless. Second, the selection pressure applied at each amino-acid position in the screen (to improve DNA cleavage activity) is not a precise counterpart to the selection pressure imposed on those same residues in nature. I-AniI is a bifunctional enzyme that acts both as a highly specific endonuclease and also as an RNA maturase (18). This latter activity (in which the protein facilitates folding and splicing of its cognate intron) is required for proper expression of the mitochondrial apocytochrome b oxidase gene (18,37), and thus provides a benefit to the host that leads to selection pressure for maintenance of the LAGLIDADG protein-fold.

While acquisition of this secondary function can therefore indirectly stabilize endonuclease activity during the course of evolution, it would also suppress those mutations that are detrimental to RNA splicing but would otherwise optimize DNA cleavage activity.

If the selection pressures for the two activities of I-AniI and its homologues were strongly balanced, it is possible that certain amino-acid positions in I-AniI and its closest homologues might display elevated frequencies of asynonymous amino-acid substitutions caused by 'genetic conflict' – in this case, positions where the requirements for endonuclease and maturase functions produce closely balanced, antagonistic selection pressures in favor of different residues types. This possibility was examined by subjecting the sequences of I-AniI and its homologues (using a BLAST alignment E value cutoff of 10–6) to a 'Ka/Ks' analysis (searching for elevated levels of nonsynonymous substitutions, relative to silent or synonymous substitutions) (38). Absolutely no residues in the endonuclease were found to display elevated ratios of Ka/Ks; all positions appear to be under either neutral or purifying selection.

Asymmetry in monomeric LHEs active sites

I-AniI, which is a monomeric endonuclease with two separate structural domains, can cleave asymmetric DNA sequences in a highly sequence-specific manner and is an excellent model of an initial protein scaffold for development of a gene-specific reagent. The results described above indicate that during evolution this endonuclease or one of its immediate predecessors suffered a mutation (Y111S) that perturbed one of two active sites.

In addition to I-AniI, several additional homing monomeric LHEs have also been shown to display significant asymmetry in their DNA cleavage properties. The algal endonuclease I-CpaII displays specific DNA nicking behavior that varies as a function of metal concentration, preferentially nicking the bottom strand of its target site at low (<0.5 mM) magnesium concentrations (39). Similarly, the yeast HEs I-SceI recognizes its asymmetrical site with higher affinity binding to the 3′-DNA half-site, leading to accumulation of nicked intermediates during the cleavage reaction (40). Finally, the archaeal endonuclease I-DmoI cleaves the coding strand of its host gene with a slight preference over the non-coding strand (41); this asymmetry is enhanced in engineered variants of this enzyme in which the LAGLIDADG motif residues are exchanged for those of homologous endonucleases (42,43).

We believe the structural changes in the Y2 variant of I-AniI, leading to repositioning of lysine 94 into contact with a putative nucleophilic water molecule, explains the increase in catalytic efficiency in its corresponding active site. This conclusion is in agreement with the hypothesized reaction mechanism for this family of endonucleases (44), which require the presence of one or more positively charged groups to assist in deprotonation of the hydrolytic water and/or to stabilize the phosphoanion transition state. In these enzymes, the usual basic residue in each active site is a lysine residue (K98 and 98′ in I-CreI, K122 and 223 in I-SceI, K43 and 140 in I-DmoI, and K94 and K227 in I-AniI). In the very high resolution structures of the most active LAGLIDADG endonuclease, I-CreI, the positions of K98 and K98′ superimpose quite closely on the position of K94 in the Y2 variant of I-AniI.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Technical assistance and advice was provided by Jill Bolduc, Audrey McConnell-Smith, Lei Zhao and other members of the FHCRC structural biology program.

FUNDING

National Institutes of Health (R01 GM49857 and RL1 CA133833 to B.L.S., RL1 CA133832 to A.M.S.); the Gates Foundation Grand Challenge Program (to A.M.S. and B.L.S.); National Institutes of Health training grant (T32 GM07270 to M.C.); Japan Society for the Promotion of Science (to R.T.). Funding for open access charge: National Institutes of Health (GM49857 and RL1 CA133833).

Conflict of interest statement. None declared.

REFERENCES

1. Belfort,M. and Roberts,R.J. (1997) Homing endonucleases - keeping the house in order. *Nucleic Acids Res.*, **25**, 3379–3388.
2. Stoddard,B.L. (2005) Homing endonuclease structure and function. *Quarterly Reviews of Biophysics*, **38**, 49–95.
3. Jacquier,A. and Dujon,B. (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell*, **41**, 383–394.
4. Kostriken,R., Strathern,J.N., Klar,A.J., Hicks,J.B. and Heffron,F. (1983) A site-specific endonuclease essential for mating-type switching in Saccharomyces cerevisiae. *Cell*, **35**, 167–174.
5. Colleaux,L., d'Auriol,L., Betermier,M., Cottarel,G., Jacquier,A., Galibert,F. and Dujon,B. (1986) Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into E. coli as a specific double strand endonuclease. *Cell*, **44**, 521–533.
6. Dalgaard,J.Z., Klar,A.J., Moser,M.J., Holley,W.R., Chatterjee,A. and Mian,I.S. (1997) Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. *Nucleic Acids Res.*, **25**, 4626–4638.
7. Argast,G.M., Stephens,K.M., Emond,M.J. and Monnat,R.J. (1998) I-*Ppo*I and I-*Cre*I homing site sequence degeneracy determined by random mutagenesis and sequential *in vitro* enrichment. *J. Mol. Biol.*, **280**, 345–353.
8. Chevalier,B., Turmel,M., Lemieux,C., Monnat,R.J. and Stoddard,B.L. (2003) Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. *J. Mol. Biol.*, **329**, 253–269.
9. Scalley-Kim,M., McConnell-Smith,A. and Stoddard,B.L. (2007) Coevolution of homing endonuclease specificity and its host target sequence. *J. Mol. Biol.*, **372**, 1305–1319.
10. Paques,F. and Duchateau,P. (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Current Gene Therapy*, **7**, 49–66.
11. Lambowitz,A.M. and Belfort,M. (1993) Introns as mobile genetic elements. *Annu. Rev. Biochem.*, **62**, 587–622.

12. Goddard,M.R. and Burt,A. (1999) Recurrent invasion and extinction of a selfish gene. *Proc. Natl Acad. Sci. USA*, **96**, 13880–13885.
13. Posey,K.L., Koufopanou,V., Burt,A. and Gimble,F.S. (2004) Evolution of divergent DNA recognition specificities in VDE homing endonucleases from two yeast species. *Nucleic Acids Res.*, **32**, 3947–3956.
14. Delahodde,A., Goguel,V., Becam,A.M., Creusot,F., Perea,J., Banroques,J. and Jacq,C. (1989) Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. *Cell*, **56**, 431–441.
15. Wenzlau,J.M., Saldanha,R.J., Butow,R.A. and Perlman,P.S. (1989) A latent intron-encoded maturase is also an endonuclease needed for intron mobility. *Cell*, **56**, 421–430.
16. Mohr,G., Perlman,P.S. and Lambowitz,A.M. (1993) Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Res.*, **21**, 4991–4997.
17. Schafer,B., Wilde,B., Massardo,D.R., Manna,F., Del Giudice,L. and Wolf,K. (1994) A mitochondrial group-I intron in fission yeast encodes a maturase and is mobile in crosses. *Curr. Genet.*, **25**, 336–341.
18. Ho,Y., Kim,S.J. and Waring,R.B. (1997) A protein encoded by a group I intron in Aspergillus nidulans directly assists RNA splicing and is a DNA endonuclease. *Proc. Natl Acad. Sci. USA*, **94**, 8994–8999.
19. Szczepanek,T., Jamoussi,K. and Lazowska,J. (2000) Critical base substitutions that affect the splicing and/or homing activities of the group I intron bi2 of yeast mitochondria. *Mol. Gen. Genet.*, **264**, 137–144.
20. Bolduc,J.M., Spiegel,P.C., Chatterjee,P., Brady,K.L., Downing,M.E., Caprara,M.G., Waring,R.B. and Stoddard,B.L. (2003) Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. *Genes Dev.*, **17**, 2875–2888.
21. 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.
22. Chatterjee,P., Brady,K.L., Solem,A., Ho,Y. and Caprara,M.G. (2003) Functionally distinct nucleic acid binding sites for a group I intron-encoded RNA maturase/DNA homing endonuclease. *J. Mol. Biol.*, **329**, 239–251.
23. Pierce,A.J., Johnson,R.D., Thompson,L.H. and Jasin,M. (1999) XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev.*, **13**, 2633–2638.
24. Solem,A., Chatterjee,P. and Caprara,M.G. (2002) A novel mechanism for protein-assisted group I intron splicing. *RNA*, **8**, 412–425.
25. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, **276**, 307–326.
26. Winn,M.D., Murshudov,G.N. and Papiz,M.Z. (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.*, **374**, 300–321.
27. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D: Biol. Crystallogr.*, **54**, 905–921.
28. Caprara,M.G., Chatterjee,P., Solem,A., Brady-Passerini,K.L. and Kaspar,B.J. (2007) An allosteric-feedback mechanism for protein-assisted group I intron splicing. *RNA*, **13**, 211–222.
29. Downing,M.E., Brady,K.L. and Caprara,M.G. (2005) A C-terminal fragment of an intron-encoded maturase is sufficient for promoting group I intron splicing. *RNA*, **11**, 437–446.
30. Choulika,A., Perrin,A., Dujon,B. and Nicolas,J.F. (1995) Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Mol. Cell Biol.*, **15**, 1968–1973.
31. Epinat,J.C., Arnould,S., Chames,P., rochaix,P., Desfontaines,D., Puzin,C., Patin,A., Zanghellini,A., Paques,F. and Lacroix,E. (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.*, **31**, 2952–2962.
32. Rouet,P., Smih,F. and Jasin,M. (1994) Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl Acad. Sci. USA*, **91**, 6064–6068.
33. Kurokawa,S., Yamasaki,T., Komatsu,T., Watanabe,K.I. and Ohama,T. (2006) Degenerated recognition property of a mitochondrial homing enzyme in the unicellular green alga Chlamydomonas smithii. *Plant Mol. Biol.*, **62**, 141–150.
34. Prieto,J., Epinat,J.C., Redondo,P., Ramos,E., Padro,D., Cedrone,F., Montoya,G., Paques,F. and Blanco,F.J. (2008) Generation and analysis of mesophilic variants of the thermostable archaeal I-DmoI homing endonuclease. *J. Biol. Chem.*, **283**, 4364–4374.
35. Siddiqui,K.S. and Cavicchioli,R. (2006) Cold-adapted enzymes. *Annu. Rev. Biochem.*, **75**, 403–433.
36. Sriprapundh,D., Vieille,C. and Zeikus,J.G. (2003) Directed evolution of Thermotoga neapolitana xylose isomerase: high activity on glucose at low temperature and low pH. *Protein Eng.*, **16**, 683–690.
37. Ho,Y. and Waring,R.B. (1999) The maturase encoded by a group I intron from Aspergillus nidulans stabilizes RNA tertiary structure and promotes rapid splicing. *J. Mol. Biol.*, **292**, 987–1001.
38. Goldman,N. and Yang,Z. (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.*, **11**, 725–736.
39. Turmel,M., Mercier,J.P., Cote,V., Otis,C. and Lemieux,C. (1995) The site-specific DNA endonuclease encoded by a group I intron in the Chlamydomonas pallidostigmatica chloroplast small subunit rRNA gene introduces a single-strand break at low concentrations of Mg2+. *Nucleic Acids Res.*, **23**, 2519–2525.
40. Perrin,A., Buckle,M. and Dujon,B. (1993) Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. *EMBO Journal*, **12**, 2939–2947.
41. Agaard,C., Awayez,M.J. and Garrett,R.A. (1997) Profile of the DNA recognition site of the archaeal homing endonuclease I-DmoI. *Nucleic Acids Res.*, **25**, 1523–1530.
42. Silva,G.H. and Belfort,M. (2004) Analysis of the LAGLIDADG interface of the monomeric homing endonuclease I-DmoI. *Nucleic Acids Res.*, **32**, 3156–3168.
43. Silva,G.H., Belfort,M., Wende,W. and Pingoud,A. (2006) From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. *J. Mol. Biol.*, **361**, 744–754.
44. Chevalier,B., R. J. Monnat,J. and Stoddard,B.L. (2005) In al,M. B. e (ed.), *Homing Endonucleases and inteins*. Vol. 16, Springer Verlag, Berlin, pp. 34–47.