Mutability of an HNH Nuclease Imidazole General Base and Exchange of a Deprotonation Mechanism†

Jennifer H. Eastberg,‡ Jennifer Eklund,§,|| Raymond Monnat, Jr.,§ and Barry L. Stoddard*,‡

Fred Hutchinson Cancer Research Center and Graduate Program in Molecular and Cellular Biology, University of Washington, 1100 Fairview Avenue North, A3-025, Seattle, Washington 98109, and Departments of Pathology and Genome Sciences, University of Washington, Box 357705, Seattle, Washington 98195

Received March 1, 2007; Revised Manuscript Received April 6, 2007

ABSTRACT: Several unique protein folds that catalyze the hydrolysis of phosphodiester bonds have arisen independently in nature, including the PD(D/E)XK superfamily (typified by type II restriction endonucleases and many recombination and repair enzymes) and the HNH superfamily (found in an equally wide array of enzymes, including bacterial colicins and homing endonucleases). Whereas the identity and position of catalytic residues within the PD(D/E)XK superfamily are highly variable, the active sites of HNH nucleases are much more strongly conserved. In this study, the ability of an HNH nuclease to tolerate a mutation of its most conserved catalytic residue (its histidine general base), and the mechanism of the most active enzyme variant, were characterized. Conversion of this residue into several altered chemistries, glutamine, lysine, or glutamate, resulted in measurable activity. The histidine to glutamine mutant displays the highest residual activity and a pH profile similar to that of the wild-type enzyme. This activity is dependent on the presence of a neighboring imidazole ring, which has taken over as a less efficient general base for the reaction. This result implies that mutational pathways to alternative HNH-derived catalytic sites do exist but are not as extensively or successfully diverged or reoptimized in nature as variants of the PD(D/E)XK nuclease superfamily. This is possibly due to multiple steric constraints placed on the compact HNH motif, which is simultaneously involved in protein folding, DNA binding, and catalysis, as well as the use of a planar, aromatic imidazole group as a general base.

Protein structures and functions diverge from common ancestors over time, acquiring new activities and properties while maintaining similar core folds (*1*). A small number of protein fold superfamilies represent disproportionately large fractions of annotated proteomes and encompass a wide diversity of chemical functions (*2*, *3*). Highly successful protein folds include the TIM[1] (triosephosphate isomerase) barrel (*4*), the HAD (haloacid dehydrogenase) α-helical bundle (*5*), and the HUP (H/GH signature, UspA, and PP-ATPase) or "Rossman fold" sandwich (*6*). In many cases, proteins with similar folds display extreme divergence of their primary sequences, often to an extent that prevents confident assignment of homology in the absence of experimentally determined structures.

A critical determinant of the ability of a protein fold family to successfully expand and diversify during evolution is its ability to tolerate mutations (particularly at residues important for function) without a complete loss of either structural stability or core activity. This property may provide a corresponding ability to explore variations in sequence, structure, and function that can lead to the acquisition of altered (or even completely novel) activities and properties.

Enzymatic catalysis of phosphodiester bond hydrolysis and ligation is a fundamental requirement for virtually all forms of nucleic acid modification, rearrangement, and repair (*7*). This ubiquitous activity provides an excellent comparison of the ability of various protein folds to solve the same problem of maintaining a fundamental biochemical activity (phosphoryl transfer) while successfully diversifying their functional repertoires and expanding their presence in biological genomes.

In nature, a relatively small number of protein families are found to encompass the vast majority of enzymes that make and break phosphodiester bonds. In particular, two unrelated protein folds, the PD(D/E)XK and HNH domains, are each found in enzymes involved in similar processes, including restriction/modification, transposition and homing, Holliday junction resolution, and recombination. Thus, the comparison of these enzyme folds, and their ability to undergo mutational alteration and diversification, may shed light on properties that are important to the evolutionary success of structural protein superfamilies.

† This research was supported by NIH Grant GM49857 (B.L.S.) and NIH Training Grant T32GM07270 (J.H.E.).

* To whom correspondence should be addressed. Phone: (206) 667-4031. Fax: (206) 667-5894. E-mail: bstoddar@fhcrc.org.

‡ Fred Hutchinson Cancer Research Center and Graduate Program in Molecular and Cellular Biology.

§ Departments of Pathology and Genome Sciences.

|| Current address: University of Michigan Schools of Education and Public Health, Ann Arbor, MI 48109.

[1] Abbreviations: TIM, triosephosphate isomerase; HAD, haloacid dehydrogenase; HUP, H/GH signature, UspA, and PP-ATPase; DTT, dithiothreitol; Bis-Tris, bis(2-hydroxyethyl)iminotris(hydroxymethyl)-methane; CHES, cyclohexylaminoethanesulfonic acid; CAPS, 3-(cyclohexylamino)-1-propanesulfonic acid; BSA, bovine serum albumin; EDTA, ethylenediaminetetraacetic acid; TBE, Tris-borate-EDTA buffer; ITC, isothermal titration calorimetry; PEG, polyethylene glycol; PDB, Protein Data Bank.

10.1021/bi700418d CCC: $37.00
Published on Web 05/22/2007

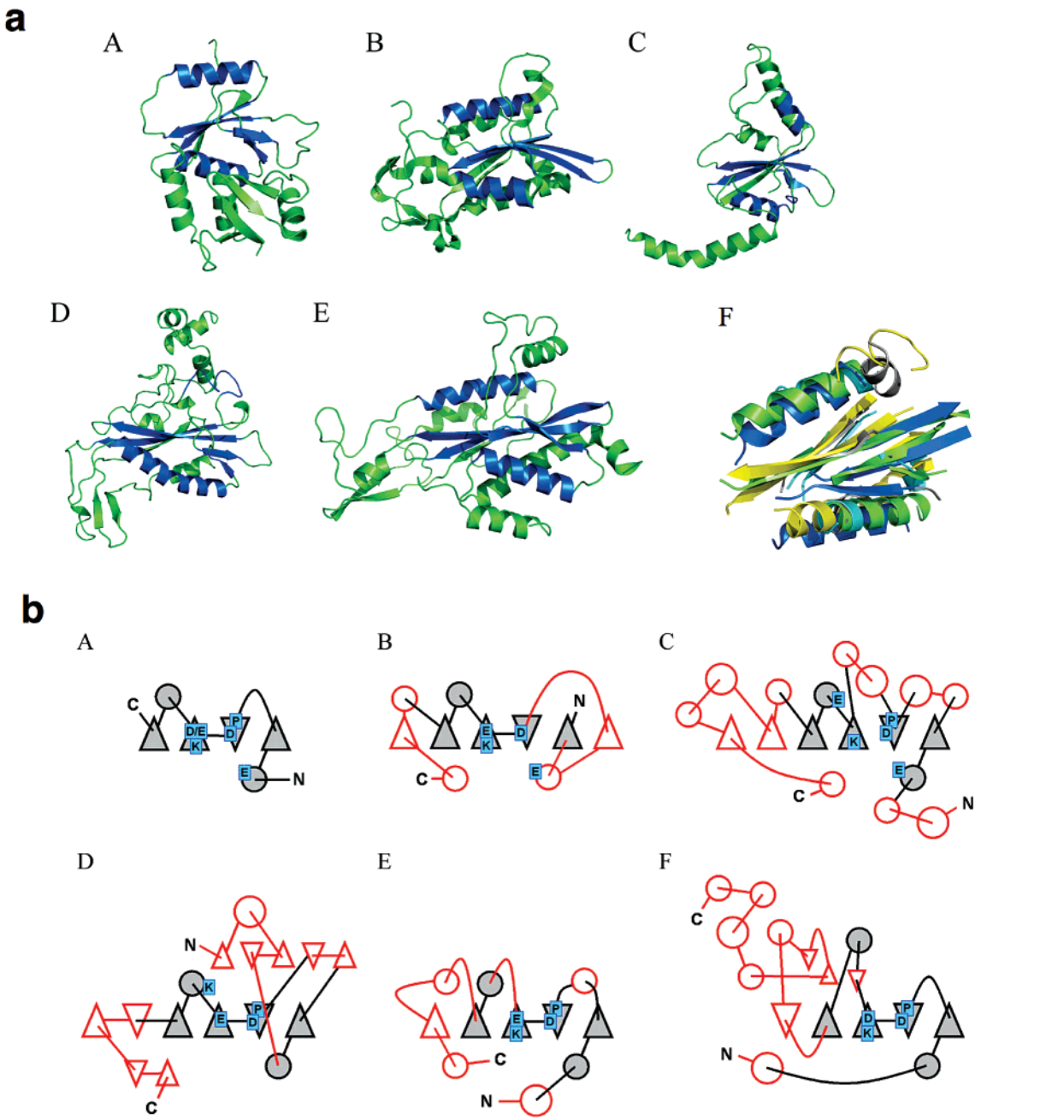

FIGURE 1: Comparison and superposition of PD(D/E)XK domain-containing enzymes. (a) PD(D/E)XK domains of EndA (A), BamHI (B), PvuII (C), EcoRV (D), and EcoRI (E) are colored blue; the remainder of each enzyme subunit is colored green. Only one monomer of each enzyme is shown. PDB entries 1A79, 1BHM, 1PVI, 1RVE, and 1ERI, respectively, were used. (F) Superposition of core PD(D/E)XK motifs. (b) Topology diagrams of the hypothetical ancestral PD(D/E)XK core (A), nuclease XPF (B), NgoMIV (C), Tt1018 (D), EcoRI (E), and EcoRV (F). The PD(D/E)XK core is colored gray, with elaborations in red. Catalytic residues are positioned as indicated.

The PD(D/E)XK domain was originally identified in type II restriction endonucleases (*8*) but has since been found in a variety of other DNA-interacting enzymes (*9*). Examples of other PD(D/E)XK domain-containing enzymes include phage λ exonuclease, archael resolvases, phage T7 endonuclease I, RecB nuclease, and the homing endonuclease I-Ssp6803I (*10*). The common core fold shared by PD(D/E)XK superfamily members is a four-strand mixed β-sheet flanked on both sides by α-helices (Figure 1). The overall structure of this fold is relatively loosely conserved, with significant differences in the length of both helices and strands, the curvature of the β-sheet, and positions of the α-helices relative to the β-sheet. Furthermore, the conserved core structure of PD(D/E)XK enzymes has functioned as a scaffold for fold elaboration. Thus, as the proteins have evolved, multiple insertions have been made between the conserved elements, resulting in altered topologies (*9*, *11*).

In addition to the variability of their core structures, PD(D/E)XK nucleases also exhibit strong divergence of their active site architectures. Most enzymes contain two or three acidic residues as well as one lysine residue as their hallmark catalytic motif. However, within the enzyme superfamily, these residues have migrated within the primary sequence but still maintain their spatial relationship to the substrate phosphate group (Figure 1b). Additionally, these residues display flexibility in their chemistries; for example, acidic

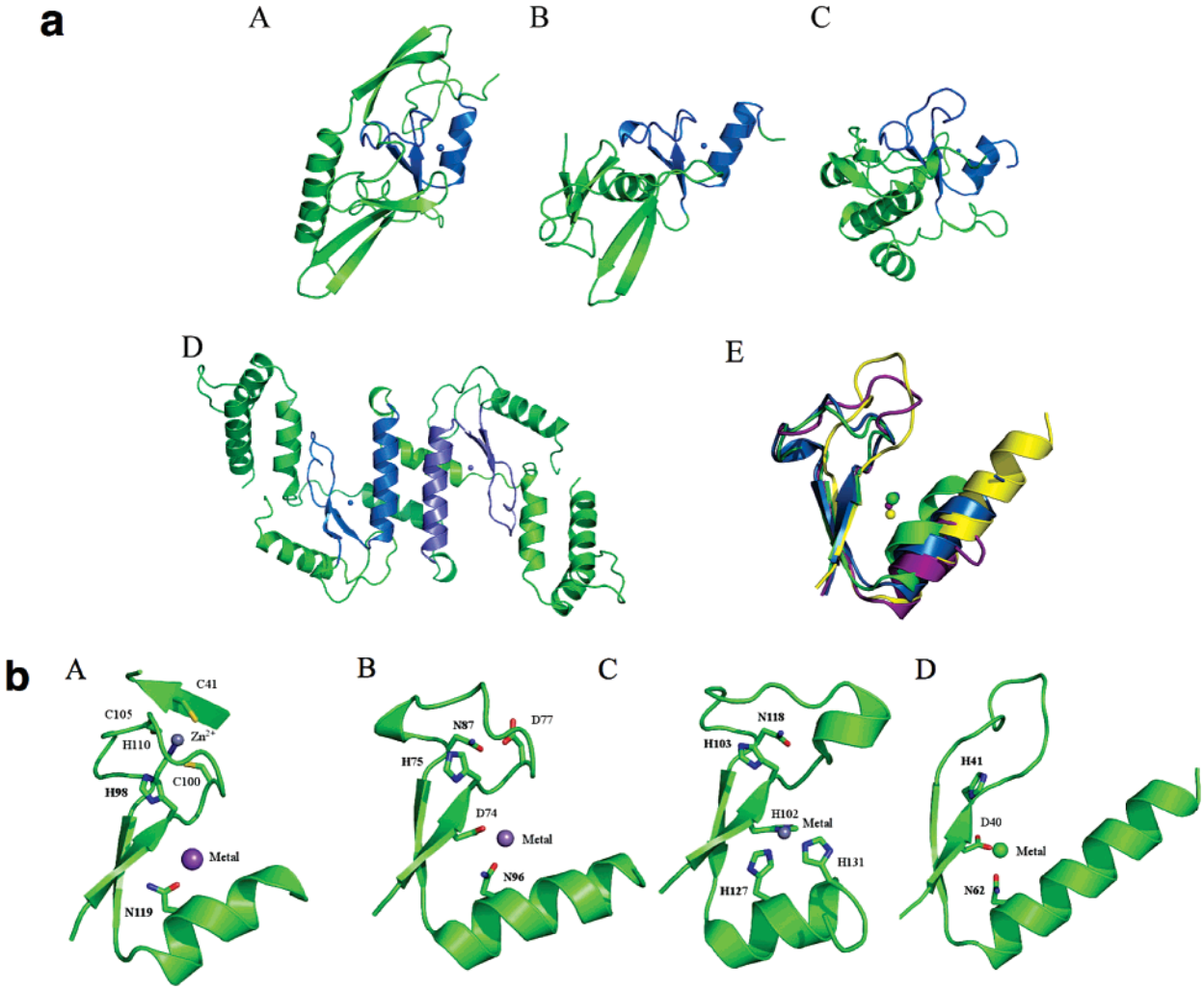

FIGURE 2: Comparison of HNH domain-containing enzymes. (a) The HNH domains of I-PpoI (A), I-HmuI (B), E9 colicin (C), and T4 endonuclease VII (D) are colored blue; the remainder of each enzyme subunit is colored green. Only one monomer is shown, except in the case of T4 endonuclease VII. (E) Superposition of core HNH motifs. (b) Comparison of HNH domain active sites. HNH domains of I-PpoI (A), I-HmuI (B), E9 colicin (C), and T4 endonuclease VII (D) in similar orientations. HNH motif residues are labeled in bold. Note how the initial H and final H/N of the motif are located in identical positions relative to the secondary structure of each protein. PDB entries 1CZ0, 1FSJ, 1U3E, and 1EN7, respectively, were used.

metal-binding residues are sometimes replaced with their amide-containing cousins. As might be expected given this degree of variability, the PD(D/E)XK enzymes do not display a single uniform reaction mechanism: the reaction may require one, two, or even possibly three metal ions to proceed; the position of metal binding sites may be shifted; and a variety of residues and chemistries can be used for proton transfer and transition state stabilization (*7*).

The HNH motif, while generally less well-known than its PD(D/E)XK counterpart, is also found across a broad spectrum of activities, including homing endonucleases, bacterial colicins, restriction endonucleases, transposases, and DNA packaging factors (*12*). The conserved structural core of the HNH domain consists of two antiparallel β-strands, connected by a loop of varying length and flanked by an α-helix with a metal binding site between the two elements. Few significant insertions of additional structural elements have been observed for this domain. The HNH motif (Figure 2) can exist as an independently folded catalytic domain (as observed in colicins) or can be encoded in-frame with additional protein domains that direct the overall DNA binding specificity and cleavage activity (as observed in HNH homing endonucleases).

In contrast to the PD(D/E)XK superfamily, the HNH superfamily displays very strong structural and mechanistic conservation across the active sites of enzymes that are otherwise quite dissimilar (Figure 2b). The first of these is an invariant histidine (**H**NH) that acts as the general base of the DNA cleavage reaction. The position of this catalytic residue at the end of the first β-strand was conserved in all structures that were examined. The second residue of the H**N**H motif stabilizes the position of the two β-strands relative to one another but can be replaced by alternative modes of structural stabilization; the third and final conserved residue is a metal-binding histidine (HN**H**), which can sometimes be substituted with an asparagine (HN**N**).

The proposed catalytic mechanisms for all studied HNH enzymes are virtually identical. The pK_a of the histidine general base is elevated through a hydrogen bond with a nearby carbonyl oxygen from a side chain or backbone carbonyl. The imidazole deprotonates a water molecule that is positioned for an in-line nucleophilic attack on the scissile

phosphate. A single bound metal ion stabilizes the transition state and leaving group but does not participate in activation of the attacking water (*13*).

While representatives of the HNH motif are found across a multitude of biochemical systems, in general this fold family appears to be not as broadly diverged, widespread, or successful as the PD(D/E)XK family, particularly in eukaryotic genomes. It is possible that the difference in the spread and utilization of these folds might reflect different limitations in the "plasticity" of their core folds and corresponding catalytic abilities in such a way that resculpting the HNH motif for new functions through the accumulation of random mutations is more difficult. As a first step in addressing this question, we have systematically mutated the invariant histidine general base in an HNH model enzyme (the I-PpoI homing endonuclease) and examined (i) the ability of individual enzyme mutants to display residual nuclease activity and (ii) the structural mechanism of "rerouting" the role of the histidine general base during phosphodiester hydrolysis for the most active point mutant. The premise of this work is that a mutation that still provides a low level of catalytic activity might provide the initial platform for subsequent reoptimization of activity, either through reversion to a wild-type sequence and structure or through selection and optimization of a novel solution to the constraints of a phosphotransfer reaction.

MATERIALS AND METHODS

Preparation of Materials. Mutants of I-PpoI were prepared with the QuickChange XL site-directed mutagenesis kit (Stratagene, La Jolla, CA) following the commercially provided protocol. Mutagenesis primers were purchased from Operon Biotechnologies, Inc. (Huntsville, AL). All enzyme mutations were verified by direct sequencing of the expression plasmids.

The expression and purification protocols were similar to those previously published (*14*). Protein expression was carried out in LB medium in the presence of 1.5 mM zinc acetate overnight at 16 °C. I-PpoI contains two structural zinc ions in its core fold, which are independent of its active sites but are required for folding. Exogenous zinc in the media improves expression, solubility, and recovery of active enzyme. Cells were lysed in the presence of DNAse in the lysis buffer (added to a concentration of 1 mg/mL). Pure protein fractions were concentrated, mixed with glycerol (final concentration of 50%), and stored at −80 °C. DTT was omitted from the final buffer in those protein preparations destined for calorimetric analyses.

To create the target site plasmid pBEND3-Ppo, DNA oligonucleotides containing the I-PpoI target site (5′-CTAGATGACTCTCTTAAGGTAGCCAA-3′ and its complement 5′-TCGATTGGCTACCTTAAGAGAGTCAT-3′) were purchased from Operon Biotechnologies, Inc., and were annealed by being heated to 95 °C and slow-cooled to room temperature. The substrate pBEND3 (*15*) plasmid was prepared by linearization with the restriction enzymes SalI and XbaI in sequential digests, and ligation of the annealed DNA oligonucleotides into the cut vector yielded the complete target site plasmid pBEND3-Ppo. pBEND3-Ppo was linearized by digestion with ScaI followed by phenol/chloroform extraction.

Cleavage Assays. Cleavage assay reactions were carried out as previously described (*16*, *17*). 10× Ppo cleavage buffers were composed of 100 mM buffer, 500 mM sodium chloride, 100 mM magnesium chloride, and 20 mM DTT. The buffer depended on the pH: sodium citrate for pH 5.0 and 5.5, Bis-Tris for pH 6.0 and 6.5, Tris for pH 7.0−9.0, CHES for pH 9.5 and 10.0, and CAPS for pH 10.5. One hundred six nanograms of linearized substrate plasmid was used per reaction, resulting in a final concentration of 5.0 μM. Reactions also included 0.05 mg/mL BSA. Cleavage by wild-type I-PpoI was assayed at an enzyme concentration of 10 pM, while concentrations of mutant enzymes ranged from 10 pM to 1 μM. Nonenzymatic reaction components were combined first and then prewarmed to 37 °C in a hot block before enzyme was added and the reactions were allowed to proceed at 37 °C for 1 h. Reactions were terminated by the addition of 2× Ppo stop buffer (2% SDS, 100 mM EDTA, 20% glycerol, and 0.2% bromophenol blue). Samples were run on a 0.8% agarose gel, stained with ethidium bromide, and imaged with an Eagle Eye II and EagleSight version 3.22 (both from Stratagene). Quantification was performed with ImageJ 1.32j (*18*).

For longer-term cleavage assays of enzyme mutants that generated no visible product under the conditions described above, reactions were set up similarly, but the mixtures contained 150 ng (7 μM) of linearized plasmid and were incubated in a thermocycler with a heated lid for either 24 or 48 h prior to addition of stop buffer. To assess WT cleavage activity postincubation, 1 nM enzyme was incubated at 37 °C for 1, 24, or 48 h in the presence of 150 ng of nonspecific dsDNA oligonucleotides; 150 ng of linearized pBEND3-Ppo was then added, and the reaction mixture was incubated for 1 h before addition of stop buffer.

Steady State Kinetic Assays. Cleavage reactions were performed essentially as described above, but with a larger reaction volume, a different DNA substrate, and lower enzyme concentrations. Each reaction mixture contained 100 pM radiolabeled double-stranded DNA oligonucleotide and unlabeled oligonucleotide to bring the total dsDNA concentration to either 0.4, 0.7, 1, 2, 10, 50, or 100 nM. Optimal enzyme concentrations were determined empirically to give interpretable data plots over a reasonable time course and were 10 pM for WT and 100 pM for H98Q. Reaction volumes were 100 μL for DNA concentrations from 0.4 to 10 nM and 150 μL for 50 and 100 nM. Before addition of enzyme, reaction mixtures were warmed to 37 °C and a "time-zero" sample was taken. Subsequent aliquots (10 μL) were quenched on a time scale appropriate for the speed of the reaction: for WT every 10−60 s (0.4−10 nM) or 90 s (50 and 100 nM) and for H98Q at 2, 4, 6, 8, 10, 15, and 20 min for 0.4−10 nM and at 5, 10, 15, and 20 min and every 10 min thereafter to 90 min for 50 and 100 nM. All samples were mixed with 10 μL of 2× Ppo stop buffer immediately after being withdrawn from the main reaction tube.

Samples were loaded on a 1.5 mm × 16 cm × 16 cm 10% [40:1 (w/w) acrylamide:bisacrylamide ratio] polyacrylamide gel containing 2% glycerol and run at 175 V in 0.25× TBE for 3 h at 4 °C. The gels were dried on filter paper and exposed to a phospho-imaging screen for 24−48 h. The screens were scanned on a Storm Phosphoimager 840

(Molecular Dynamics), and the bands were quantified using ImageQuant (Molecular Dynamics). Once all data were collected, the quantified values of bands in each lane were used to determine the amounts of DNA cleaved, which were then plotted versus time for each DNA concentration. A linear equation was fit to the first four to seven points of each run to determine the initial velocity at a given DNA concentration, and the reciprocals of these values were then plotted in double-reciprocal plots. The linear-fit equations from these plots were used to calculate the kinetic parameters K_M, k_{cat}, and k_{cat}/K_M.

Isothermal Titration Calorimetry. Two DNA oligonucleotides containing the I-PpoI target site (5′-GGGGGAGTAACTATGACTCTCTTAAGGTAGCCAAAGGGGG-3′ and its complement 5′-CCCCCTTTGGCTACCTTAAGAGAGTCATAGTTACTCCCCC-3′) were purchased from Operon Biotechnologies, Inc., and annealed by being heated to 95 °C and slow-cooled to room temperature to form the DNA target site for binding analysis by isothermal titration calorimetry (ITC). The length of the target sequence was chosen to match that used for earlier gel shift studies (*17*). We added G-C anchors at each end to promote formation of stable heterodimers and to prevent hairpin formation of the pseudopalindromic site.

Prior to ITC experiments, both protein and DNA samples were dialyzed overnight into Ppo ITC buffer [20 mM sodium phosphate (pH 8.0), 50 mM sodium chloride, and 10 mM EDTA] using Tube-O-Dialyzers (G Biosciences, St. Louis, MO). EDTA was included to prevent enzymatic cleavage of DNA during the experiment; the enzyme does not display metal-dependent binding or changes in binding affinity. Following dialysis, samples were quantified by UV spectroscopy using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE). The extinction coefficient (ϵ) of I-PpoI was calculated using ProtParam (*19*) and did not differ between WT and H98Q. The annealed DNA oligonucleotide ϵ value was determined empirically by digesting the dsDNA oligonucleotide to completion with DNAse and using known nucleotide monophosphate ϵ values (*20*) to calculate ϵ for the dsDNA fragment (*21*). Sample concentrations ranged from 1.5 to 2.5 μM I-PpoI and from 30 to 40 μM dsDNA target.

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

X-ray Crystallography of H98Q I-PpoI. Crystals were grown under the same conditions with the same DNA sequence as previously described for wild-type I-PpoI (*14*).

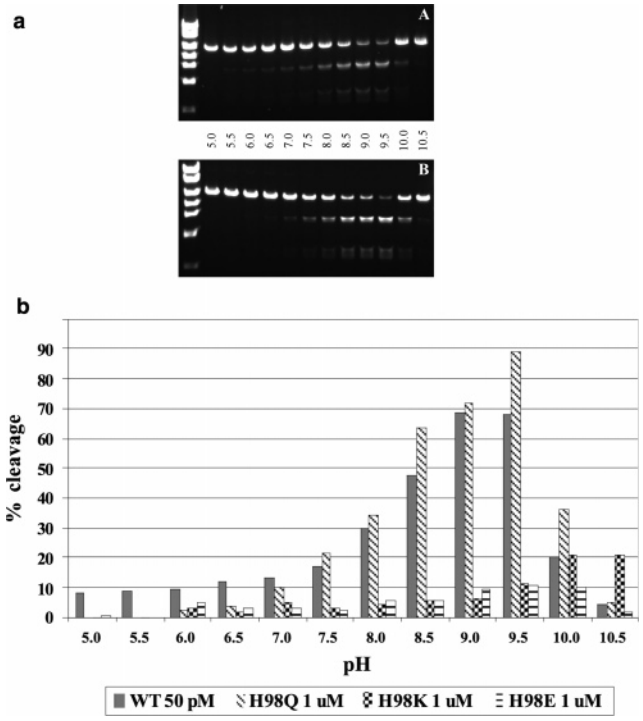

FIGURE 3: Cleavage activity of WT and H98Q I-PpoI. (a) pH profile. Linearized pBEND3-Ppo plasmid was digested with either 50 pM WT or 1 μM H98Q I-PpoI at the specified pH for 1 h as described. Both proteins exhibit maximal cleavage activity at pH 9.0−9.5. (b) Percent cleavage vs pH for all active enzymes. Gels visualizing cleavage of linearized pBEND3-Ppo by WT I-PpoI at 50 pM and H98Q, H98K, and H98E mutant I-PpoI at 1 μM over 1 h at a range of pHs were quantified with ImageJ. Intensities for the two cleavage product bands in each lane were added together and compared to the remaining uncut plasmid in the same lane to determine the percent cleavage.

DNA oligonucleotides were purchased from Oligos, Etc. (Wilsonville, OR). Following brief soaks in a series of cryo solutions composed of artificial mother liquor [0.1 M citrate (pH 5.6), 20 mM NaCl, 2 mM EDTA and 28% PEG 4000] and an increasing level of glycerol (5−20%), crystals were suspended in a fiber loop and flash-frozen in liquid nitrogen. Data were collected on beamline 5.0.1 at the ALS (Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA) using an ADSC CCD area detector. Data were processed and scaled using the DENZO/SCALEPACK program suite (*22*).

The structure was determined by molecular replacement with EPMR (*23*), using the protein and DNA chains from the H98A I-PpoI structure (PDB entry 1CYQ) (*13*) as the search model. Refinement was conducted using the CNS package (*24*) with 10% of reflections set aside for R_{free} (*25*). Visual inspection and manual adjustment of structures were conducted in XtalView (*26*). Solvent molecules in the active site were placed according to manual inspection of residual difference maps and their behavior in refinement; additional waters were selected by CNS and confirmed by manual inspection of density, contact distances, and refinement behavior. The geometric quality of the refined models was evaluated using PROCHECK (*27*). Structure superposition was done using Swiss PDB Viewer (*28*), and additional structural analysis and figure preparation were carried out in PyMOL (*29*). The coordinates of the H98Q structure have been deposited in the Protein Data Bank as entry 2O6M.

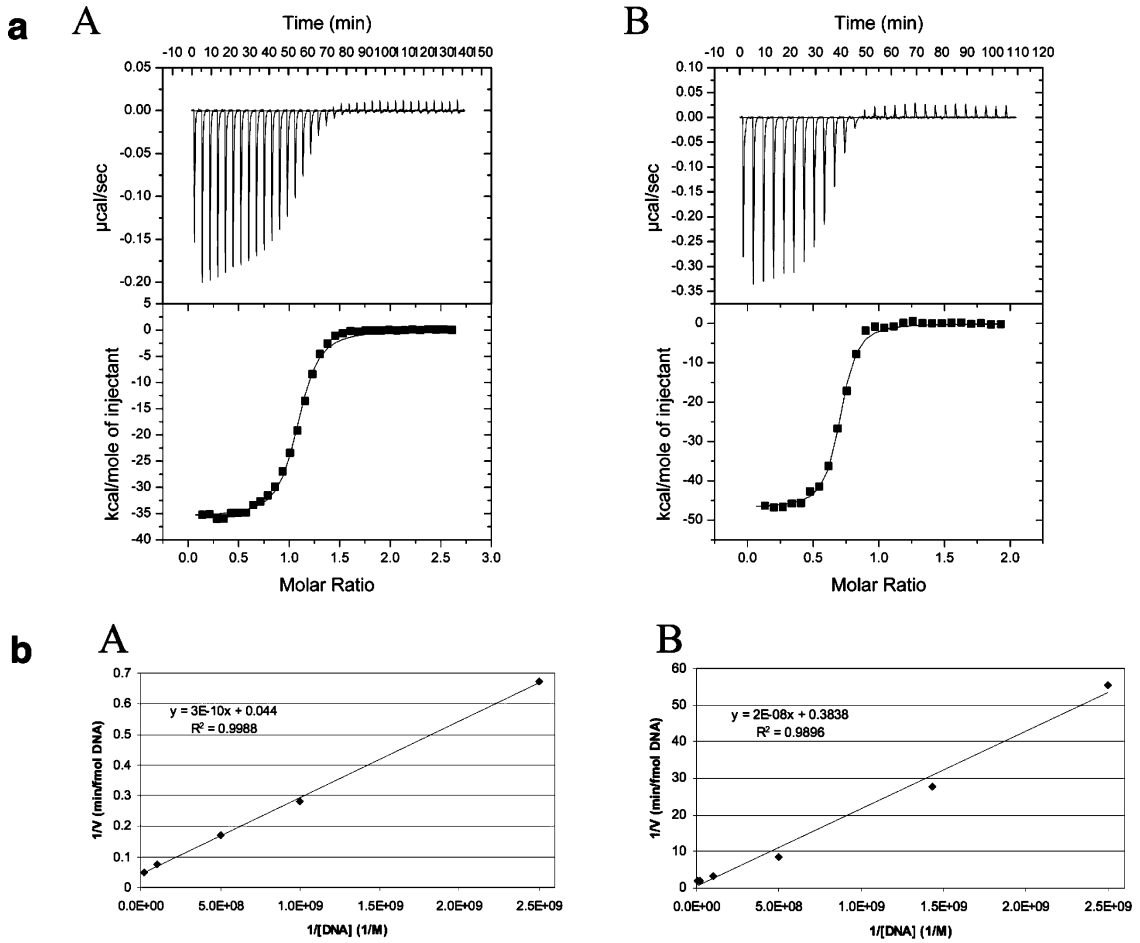

FIGURE 4: Biochemical behavior of WT and H98Q I-PpoI. (a) Binding progression curves measured by isothermal titration calorimetry (ITC). The upper half of each panel displays the raw data trace after baseline subtraction for WT (A) and H98Q (B). The lower half of each panel plots the integrated values of each peak and the curve fit to them. K_D values of 17 ± 1.7 nM for WT and 13 ± 1.9 nM for H98Q were obtained. The differences in molar ratios between these samples are within the error of the experimental system and are related to imprecision in the absolute concentration of protein samples for wild-type and H98Q I-PpoI used in those separate experiments. (b) Steady state kinetics. Double-reciprocal kinetic plots are shown for WT (A) and H98Q (B). The fitted equations were used to calculate reported kinetic parameters. Plots of moles of DNA cleaved vs time were used to determine the initial velocity at each DNA concentration. The reciprocals were then plotted to yield the displayed graphs. Calculated values include a k_{cat} of 2.3×10^{-3} min^{-1} for WT vs a value of 2.6×10^{-5} min^{-1} for H98Q and a k_{cat}/K_M of 3.3×10^5 min^{-1} M^{-1} for WT vs a value of 5.0×10^2 min^{-1} M^{-1} for H98Q.

RESULTS AND DISCUSSION

H98 Mutations and Initial Characterization. The site of the conserved catalytic histidine general base (residue 98) of I-PpoI was mutated as described, and the sequences of all constructs were verified. We had previously generated a catalytically inactive mutation at this position (H98A) (*13*) which served as a negative control for activity comparisons, in addition to the wild-type enzyme.

Initial cleavage experiments that report on residual activity (1 μM enzyme and 5 μM DNA, 1 h digest at 37 °C), conducted at both pH 7.5 and 10.0 [two pHs previously used in I-PpoI cleavage assays (*16*)], revealed measurable, but reduced, cleavage activity only for three mutants (H98Q, H98K, and H98E) and hinted at pH dependences of the cleavage reactions (described in detail below).

The remaining mutants that did not display cleavage activity under these conditions were then assayed in longer-term (24 and 48 h) digests, to confirm the absence of observable catalytic activity (data not shown). These experiments demonstrated that one additional mutant enzyme, H98C, displayed slow, nonspecific digestion of the substrate. No observable cleavage was detected for the remaining mutations at residue 98, including the previously characterized negative control H98A. We also determined that the WT protein lost most of its activity after incubation for 24 h at 37 °C and therefore assume that most observed cleavage by the various enzyme species took place in the first 24 h of the digest.

Table 1: Properties of Wild-Type and H98Q I-PpoI

	WT	H98Q
pH optimum	9.5	9.5
ITC K_D (nM)	17 ± 1.7	13 ± 1.9
K_M (nM)	6.8 ± 0.7	52 ± 0.5
k_{cat} (min^{-1})	$2.3 \times 10^{-3} \pm 0.002$	$2.6 \times 10^{-5} \pm 4 \times 10^{-6}$
k_{cat}/K_M (min^{-1} M^{-1})	3.3×10^5	5.0×10^2

pH Profile Determination. We then determined the relative cleavage progression and pH profile for all of the enzyme mutants with observable specific activity (Figure 3). DNA cleavage was assessed every 0.5 pH unit over the range of 5.0−10.5. At protein concentrations greater than 10 nM, wild type I-PpoI cleaved the DNA substrate to completion at all pHs; at protein concentrations of ≤100 pM, an activity optimum at pH 9.0−9.5 was observed. Cleavage by active mutant enzymes also exhibited strong pH preferences,

although at a much higher protein concentration (1 mM) than WT.

Mutants H98E and H98K displayed low levels of pH-dependent cleavage that was maximal under strongly basic conditions, appearing to plateau at pH 10−10.5 (higher pHs could not be tested due to the limited solubility of the protein and buffer components). In contrast, mutant H98Q displayed both the greatest extent of DNA cleavage of the point mutants and a defined activity profile with a clear optimum at pH 9.5 (Figure 3).

Following these initial experiments, we decided to focus on our most active endonuclease mutant, H98Q, for a more in-depth comparison to wild-type I-PpoI. We determined that the mutant, like the wild-type enzyme, is active in the presence of the divalent metal ions Mg^{2+}, Ca^{2+}, Zn^{2+}, Mn^{2+}, Ni^{2+}, and Co^{2+} (data not shown), agreeing with previously published analyses of I-PpoI activity (*30*−*32*). In the HNH active site, the single bound divalent cation is not involved in activation of the water nucleophile but rather participates only in transition state and leaving group stabilization. This mechanism promotes a promiscuous dependence on metals that can be satisfied by most divalent cation species.

Given that the H98Q mutant displays a pH profile and a metal dependence similar to those of the wild-type enzyme, it seemed possible that the mechanism of H98Q might display features that were similar to those of the wild-type enzyme, but with a much lower efficiency and specific activity. Therefore, a direct comparison of these two catalysts was undertaken, consisting of a detailed thermodynamic binding analysis, steady state kinetic measurements, X-ray structure determinations, and mutagenesis of an additional active site residue in the context of both the wild-type and H98Q backgrounds.

Binding Analysis by ITC. We used isothermal titration calorimetry (ITC) to determine the dissociation constants (K_D) of both WT and H98Q I-PpoI. These experiments yielded high-quality data traces with distinct peaks and easily fitted baseline regions (Figure 4a). Using this method, we obtained average K_D values of 17 ± 1.7 nM for WT and 13 ± 1.9 nM for H98Q. This finding demonstrated that the binding of the H98Q mutant to DNA is unaffected and that differences in the catalytic step of DNA cleavage itself were causing the observed reduction in activity.

ITC measurements of macromolecular binding are prone to fairly large uncertainties in molar ratios of ±20% (approximately what is seen here when WT and H98Q are compared), due to imprecision in the quantitation of protein concentration between separate preparations. However, the resulting calculation of K_d values and free energy changes of binding are more consistent between runs.

Kinetics of WT and H98Q. We performed a standard Michaelis−Menten analysis of DNA cleavage and turnover for WT and H98Q enzymes (Figure 4b and Table 1). The reduced cleavage activity displayed by H98Q corresponds to a 100-fold reduction in k_{cat} (from 2.3×10^{-3} to 2.6×10^{-5} min^{-1}) and a 7-fold increase in K_M (from 7 to 50 nM). This corresponds to a 700-fold reduction in catalytic efficiency (k_{cat}/K_M) and confirms that the significantly lower specific activity of the mutant enzyme is primarily due to a reduced rate of the chemical hydrolysis event upon formation of the ES complex, rather than a significant reduction in DNA binding affinity.

Table 2: X-ray Diffraction Data and Refinement Statistics

Data Collection	
space group	$P3_121$
resolution (Å) (highest shell)	50−2.3 (2.44−2.30)
unit cell dimensions	$a = b = 113.853$ Å, $c = 88.854$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$
wavelength	1
total no. of reflections	313359
no. of unique reflections	29954
completeness	99.6 (96.6)
redundancy	5.6 (3.1)
R_{merge}	6.5 (29.6)
average $I/\sigma I$	16.8 (2.5)
χ^2	0.756
Refinement	
R_{work} (%)[a]	21.1 (0.247)
R_{free} (%)[b]	24.5 (0.287)
rmsd for bond lengths (Å)	0.005
rmsd for bond angles (deg)	1.242
no. of protein residues	324
no. of nucleotides	42
no. of metal ions	6
no. of water molecules	286
average *B*-factors	31.9 (protein), 33.6 (DNA), 56.4 (solvent)

[a] $R_{work} = (\Sigma|F_{obs} - F_{calc}|)/(\Sigma F_{obs})$ for the working data set. [b] $R_{free} = (\Sigma|F_{obs} - F_{calc}|)/(\Sigma F_{obs})$ for a randomly selected subset of 5% of the data.

Structure of H98Q. We next determined the structure of the protein−DNA complex by X-ray crystallography. Our data (Table 2) extended to 2.3 Å resolution and were isomorphous with previous data sets collected on wild-type and H98A enzymes, allowing for direct comparison of structures. The structure was determined by molecular replacement using the protein and DNA chains from the H98A I-PpoI structure as a model (PDB entry 1CYQ) (*13*). Initial maps clearly indicated the position of the bound structural zinc atoms (in the core of the protein fold) and a single bound magnesium ion in each active site, as well as the waters coordinating the magnesium. During model building and initial refinement, density for Q98 was easily identified (Figure 5). Additional water molecules were identified and modeled in the active site region. After refinement, the values of the crystallographic *R*-factors were 21.1 for R_{work} and 24.5 for R_{free}, with 91.8% of residues in the most favorable regions of the Ramachandran plot and no residues in the disallowed regions (Table 2).

Overall, the DNA-bound structures of wild type I-PpoI (PDB entry 1CZ0), H98A (PDB entry 1CYQ) (*13*), and H98Q (reported here; PDB entry 2O6M) are nearly identical, with pairwise C_α rmsds of ∼0.3 Å. The DNA atoms in the endonuclease active sites are in similar configurations, and all structures contain a single bound metal ion in a 6-fold, octahedral coordination in contact with the scissile phosphate (Na^+ for WT and Mg^{2+} for H98A and H98Q). The Q98 side chain occupies a location similar to that of H98 in the wild-type enzyme, with Oϵ1 of glutamine positioned near Nϵ2 of the imidazole ring of the wild-type histidine residue (Figure 5b−d).

The active site water network is also well maintained in H98Q, at least in part as a result of the replacement of histidine with glutamine. The nucleophilic waters themselves superimpose nicely from the three structures and display similar distances and angles from their respective phosphate

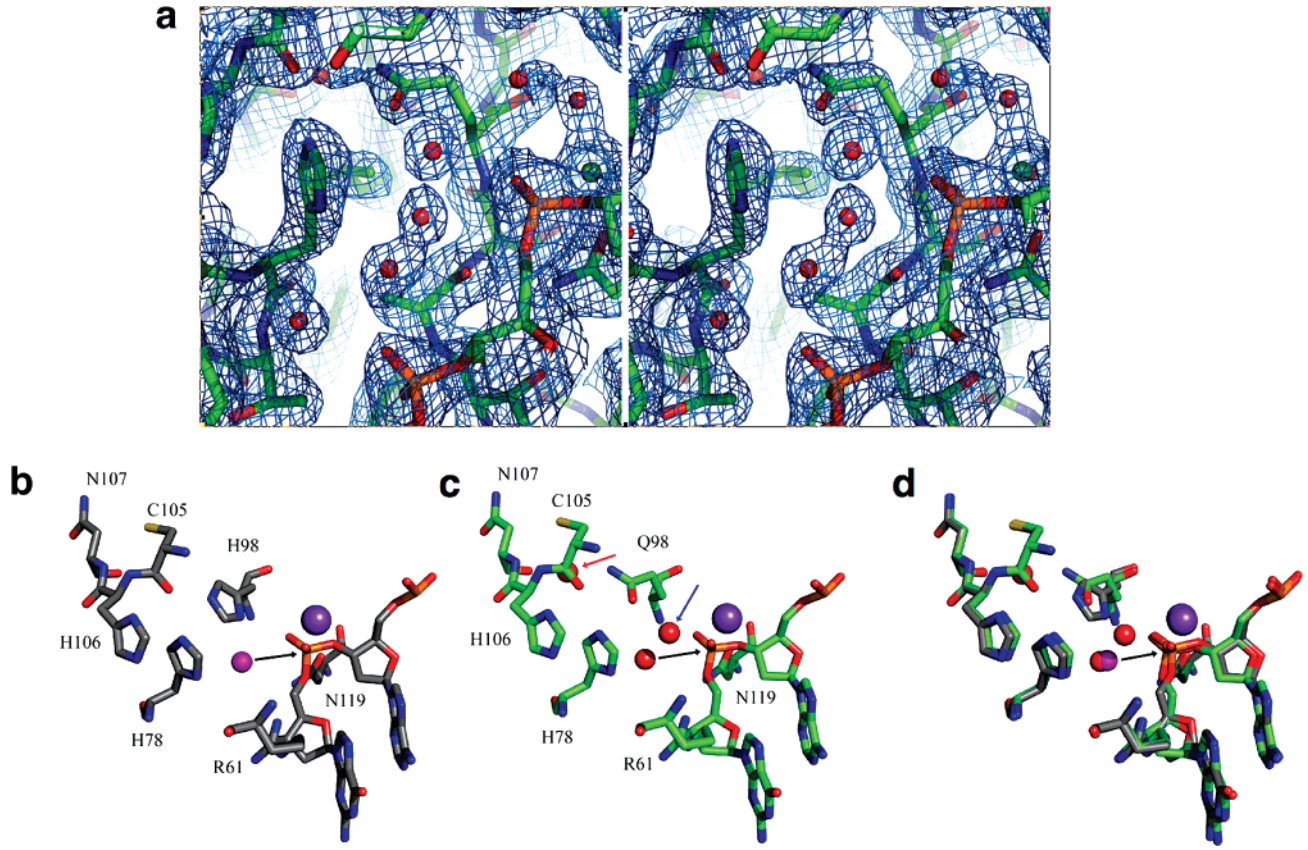

FIGURE 5: Structural analysis of H98Q I-PpoI. (a) Active site electron density. (A) Composite omit map of the active site region of H98Q I-PpoI contoured at 1σ. The map was calculated in CNS prior to the addition of Q98 and active site waters to the model. The nucleophilic water is at the center, with the scissile phosphate to the right, H78 to the left, and Q98 above. (b−d) Active site close-up. Detailed view of the active sites of WT and H98Q I-PpoI, with individual views of WT (b), H98Q (c), and their superposition (d). The bound metal in each structure is shown as a purple sphere. The direction of attack by the nucleophilic water on the scissile phosphate is indicated with a black arrow. Relative to the wild-type enzyme, H98Q displays a unique water molecule bound to Q98 and the backbone carbonyl of N107 (red arrow) as well as a second water that has shifted by approximately 1.5 Å toward the scissile phosphate (blue arrow). PDB entries 1CZ0 and 2O6M are shown.

targets. One ordered water molecule that directly contacts the nucleophile has shifted by ~1.5 Å, and a newly observed additional water molecule has appeared (Figure 5); these ordered solvent molecules are located near the glutamine amide side chain and the scissile phosphate, where they partially fill the space formerly occupied by H98. As discussed below, it appears that the positioning of these waters might represent one of the main contributions of Gln98 to catalysis in the mutant enzyme.

Mechanistic Hypothesis. After examining the structure of the H98Q active site in the context of previous results, we hypothesized that the glutamine at position 98 was not acting as the general base in the reaction. In addition to very unlikely assignment of deprotonation activity to a primary amide group on the glutamine side chain, the wild-type and mutant enzymes display nearly identical pH profiles, a result that would not be expected if the proton transfer mechanism was so drastically altered. After analyzing the active site region further, we identified a histidine residue at position 78 and hypothesized that this residue is taking over as the general base in the H98Q enzyme, but from a much less favorable geometric position (Figure 6). In the hypothetical mechanism, the role of the glutamine residue is more structural than catalytic, serving to maintain an appropriate solvation environment around the scissile phosphate by structurally replacing Nϵ2 of His98 with its own amide nitrogen while maintaining an otherwise isomorphous active site architecture.

His78 is positioned similarly in both the WT and H98Q structures, with Nδ1 of H78 positioned 4.18 Å from the nucleophilic water in the WT structure and 3.80 Å from it in H98Q. The backbone carbonyl of residue 103 forms a hydrogen bond to Nϵ2 of H78, thus elevating the pK_a of Nδ1 and facilitating abstraction of a proton from the nucleophilic water, activating it for attack (Figure 6). In the wild-type enzyme, the Nϵ2 atom of H98 forms a hydrogen bond to the backbone carbonyl of residue 105; similar hydrogen bond contacts are observed to the catalytic histidine of all HNH endonucleases (either from side chains or from backbone oxygens).

An analysis of the geometry of the active site structures of WT and H98Q endonucleases appears to explain the large reduction in catalytic efficiency for the H98Q enzyme. The first observed difference is the distance between the nucleophilic water and Nδ1 of the histidine general base. In the WT enzyme, this distance is 2.68 Å, but in H98Q, it is 3.80 Å. In addition, the angle of proton transfer to Nδ1 of His78 is nonoptimal. In the WT enzyme, the angle of proton transfer is 125.6° and the nucleophilic water is located ~30° above the plane of the histidine ring. In contrast, the equivalent angle for H78 in the mutant enzyme is 95.8°, and the nucleophilic water is approximately 90° off the plane of the

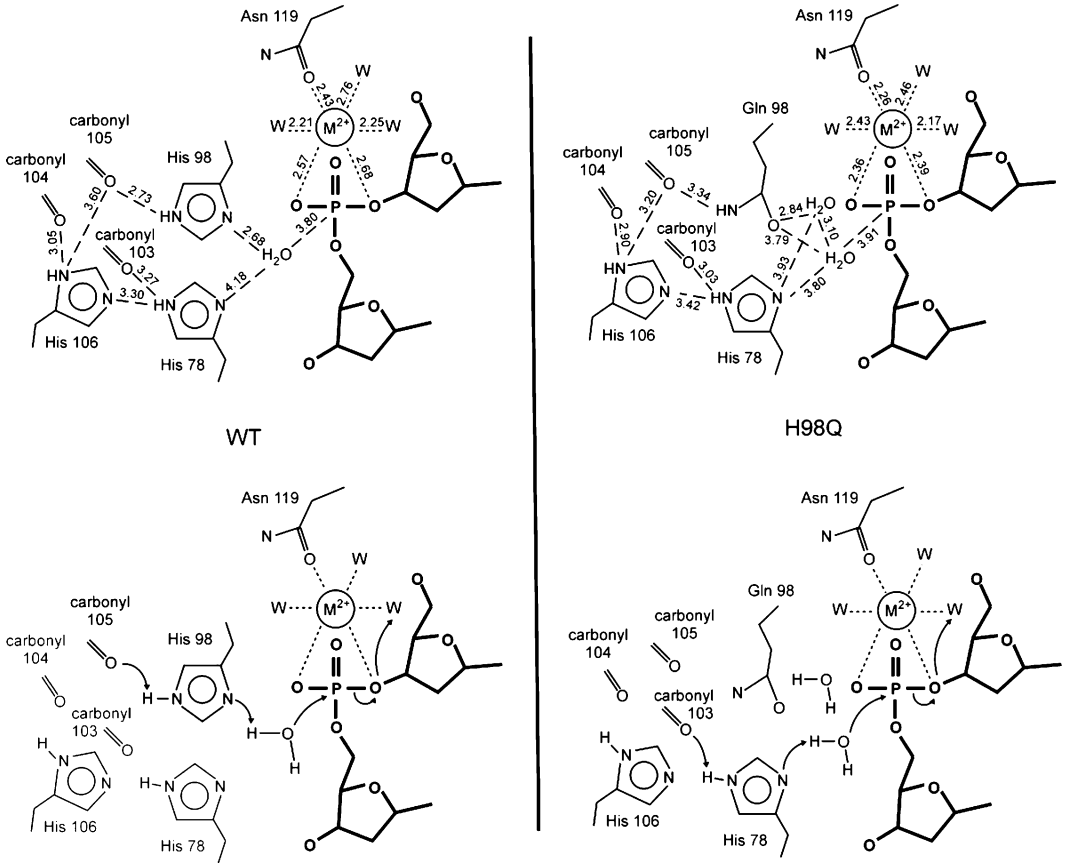

FIGURE 6: Schematic of the active site region and postulated reaction mechanisms of WT (left) and H98Q (right). The upper half of each side displays contact distances between potentially catalytically involved residues. The lower halves diagram the proposed reaction mechanism. Metal contacts are represented by dotted lines, whereas dashed lines mark distances of other contacts. All distances are in angstroms. Arrows indicate movement of electrons during cleavage. The metal position is labeled M^{2+}, and some water molecules are represented by a W. Drawings not to scale. PDB entries 1CZ0 and 2O6M were analyzed.

ring, positioned above the Cγ atom. This geometry is highly unfavorable and most likely the main cause of the reduced rate of reaction in the mutant enzyme.

Validation of the Participation of Histidine 78 in Catalysis by H98Q I-PpoI. The hypothesis that H78 "stands in" for H98 to promote a less efficient DNA hydrolysis reaction can readily be tested by mutagenesis: the proposed mechanism leads to the prediction that mutation of H78 should inactivate H98Q but have little or no effect on the wild-type enzyme. Therefore, we generated an H78A mutation in both WT and H98Q backgrounds. Cleavage assays with the H98Q/H78A double mutant performed over the pH range of 7.0−10.5 showed no detectable activity (Figure 7a), whereas the single H78A (in an otherwise wild-type background) mutant displayed WT levels of activity in an enzyme titration assay (Figure 7b).

Since other mutations at residue 98 either completely eliminate visible activity or result in cleavage at only high pH, it is clear that glutamine 98 is providing a unique contribution that allows His78 to partially restore function. It seems likely from a comparison of the active sites of wild-type, H98A, and H98Q enzymes that glutamine 98 partially complements activity by fortuitously maintaining an architecture of metal-associated solvent molecules that allow for efficient stabilization of the phosphoanion transition state, and possibly the protonation of the 3′-hydroxylate leaving group. In H98A I-PpoI, the position of the nucleophilic water is again virtually unchanged from that observed in the wild-type enzyme. However, the reduction of the side chain to a single methyl carbon also prevents the formation of the

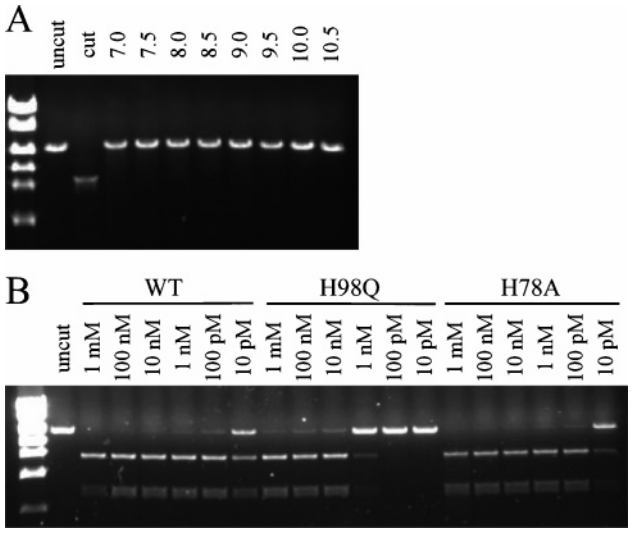

FIGURE 7: Cleavage ability of double and single mutants of H78. (A) Cleavage reactions with 1 μM H98Q/H78A I-PpoI double mutant against linearized pBEND3-Ppo plasmid at the specified pH for 1 h. No detectable cleavage was observed at any pH. (B) Enzyme titrations to determine the relative activity of WT, H98Q, and H78A I-PpoI against linearized pBEND3-Ppo. The activity levels of WT and H78A are comparable, and both are substantially higher than that of H98Q. The label "uncut" indicates linearized pBEND3-Ppo plasmid only and "cut" digestion by 1 μM WT I-PpoI.

ordered solvent structure in the active site that is observed for H98Q, as well as preventing the acquisition of an additional water near the scissile phosphate.

Mutability of the HNH Domain. A histidine side chain at residue 78 can stand in for the primary general base of the HNH active site in the case of I-PpoI. This result is strikingly similar to the observation of second site substitutions across the active sites of PD(D/E)XK nucleases (Figure 1b) and therefore begs the question of why such a mutated construct might not lead to subsequent re-optimization of catalytic efficiency through additional mutations and reorganization of the active site. To phrase the question differently, is the extreme structural conservation of HNH active sites, observed between enzymes such as homing endonucleases and colicins that otherwise are highly diverged, due to constraints that make any alternative active site architectures inherently poor catalysts of phosphoryl transfer?

The compact HNH domain, which is typically fewer than 50 residues in size (Figure 2), is simultaneously involved in protein folding and stability, DNA binding, and catalysis. This strategy allows for an economy of sequence and structure that minimizes the size of endonuclease reading frames, an important feature of both homing endonuclease genes and bacterial colicins, which are often embedded in mobile elements such as introns and plasmids, respectively. In addition, the stability of the core HNH fold allows it to be inserted as a self-contained domain into surrounding protein scaffolds. However, this structural and functional "multitasking" strongly constrains the sequence of the HNH motif so that it cannot easily sample mutational diversity without a significant loss of at least one of its biochemical properties. In contrast, the epitopes and residues of the PD-(D/E)XK motif that are directly involved in catalysis are generally discontinuous surface loops extending from that core fold (Figure 1), allowing them to more freely sample sequence and structural variation during evolution without significantly impacting protein stability or DNA affinity. Furthermore, the planar aromatic structure of an imidazole ring might also impart more tightly constrained geometric requirements for proton transfer than the use of alternate general base chemistries such as the lysine side chains or metal-bound waters that are used by PD(D/E)XK nucleases.

In the example of I-PpoI, histidine 78 is located in the core of the HNH fold, two amino acids from the end of the β-hairpin responsible for directly binding DNA. Any perturbations of the protein backbone around H78 would probably disrupt both DNA binding and packing of the catalytic domain. There does not seem to be general conservation of H78 in other HNH endonucleases. Neither the *Serratia* nuclease nor the E9 colicin has a histidine in the equivalent position, or anywhere else in their active sites. Although I-HmuI has a histidine at a position equivalent to H78 of I-PpoI, it does not make any contacts that could allow it to serve as a general base. Similarly, there is a histidine in the neighborhood of the I-PpoI H78 position in T4 endonuclease VII, but due to the absence of DNA substrate and a noticeably different active site architecture, it is difficult to determine if it could substitute for the catalytic histidine. Therefore, the ability to display the alternative mechanism described above is likely unique to the His-Cys box homing endonuclease branch of the HNH superfamily. For this enzyme, and presumably for its homologues, multiple constraints on residues in the relatively small HNH motif for folding, DNA binding, and cleavage might simply have produced a highly efficient DNA-binding and cleavage module that can be used in many different structural contexts but cannot easily be modified beyond its canonical structure.

ACKNOWLEDGMENT

We thank Django Sussman and Juan-Carlos Pizzaro for assistance with X-ray data collection.

REFERENCES

1. Anantharaman, V., Aravind, L., and Koonin, E. V. (2003) Emergence of diverse biochemical activities in evolutionary conserved structural scaffolds of proteins, *Curr. Opin. Chem. Biol. 7*, 12−20.
2. Abeln, S., and Deane, C. M. (2005) Fold usage on genomes and protein fold evolution, *Proteins: Struct., Funct., Genet. 60*, 690−700.
3. Hegyi, H., Lin, J., Greenbaum, D., and Gerstein, M. (2002) Structure genomics analysis: Characteristics of atypical, common and horizontally transferred folds, *Proteins: Struct., Funct., Genet. 47*, 126−141.
4. Nagano, N., Orengo, C. A., and Thornton, J. M. (2002) One fold with many functions: The evolutionary relationships between TIM barrel families based on their sequences, structures and functions, *J. Mol. Biol. 321*, 741−765.
5. Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D., and Aravind, L. (2006) Evoluationary genomics of the HAD superfamily: Understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes, *J. Mol. Biol. 361*, 1003−1034.
6. Aravind, L., Anantharaman, V., and Koonin, E. V. (2002) Monophyly of class I aminoacyl tRNA synthetase, USPA, ETFP, photolyase, and PP-ATPase nucleotide-binding domains: Implications for protein evolution in the RNA world, *Proteins: Struct., Funct., Genet. 48*, 1−14.
7. Yang, W., Lee, J. Y., and Nowotny, M. (2006) Making and breaking nucleic acids: Two-Mg^{2+}-ion catalysis and substrate specificity, *Mol. Cell 7*, 5−13.
8. Pingoud, A., Fuxreiter, M., Pingoud, V., and Wende, W. (2005) Type II restriction endonucleases: Structure and mechanism, *Cell. Mol. Life Sci. 62*, 685−707.
9. Kosinski, J., Feder, M., and Bujnicki, J. M. (2005) The PD-(D/E)XK superfamily revisited: Identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function, *BMC Bioinf. 6*, 172.
10. Orlowski, J., Boniecki, M., and Bujnicki, J. M. (2007) I-Ssp6803I: The first homing endonuclease from the PD-(D/E)-XK superfamily exhibits an unusual mode of DNA recognition, *Bioinformatics* (in press).
11. Bujnicki, J. M., and Rychlewski, L. (2001) Grouping together highly diverged PD-(D/E)XK nucleases and identification of novel superfamily members using structure-guided alignment of sequence profiles, *J. Mol. Microbiol. Biotechnol. 3*, 69−72.
12. Cymerman, I. A., Obarska, A., Skowronek, K. J., Lubys, A., and Bujnicki, J. M. (2006) Identification of a new subfamily of HNH nucleases and experimental characterization of a representative member, HphI restriction endonuclease, *Proteins 65*, 867−876.
13. 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.
14. Flick, K. E., McHugh, D., Heath, J. D., Stephens, K. M., Monnat, R. J., Jr., and Stoddard, B. L. (1997) Crystallization and preliminary X-ray studies of I-PpoI: A nuclear, intron-encoded homing endonuclease from *Physarum polycephalum*, *Protein Sci. 6*, 2677−2680.
15. Zwieb, C., and Adhya, S. (1994) Improved plasmid vectors for the analysis of protein-induced DNA bending, *Methods Mol. Biol. 30*, 281−294.
16. Mannino, S. J., Jenkins, C. L., and Raines, R. T. (1999) Chemical mechanism of DNA cleavage by the homing endonuclease I-PpoI, *Biochemistry 38*, 16178−16186.

17. Galburt, E. A., Chadsey, M. S., Jurica, M. S., Chevalier, B. S., Erho, 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.
18. Rasband, W. (1997−2006) U.S. National Institutes of Health, Bethesda, MD.
19. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server, in *The Proteomics Protocols Handbook* (Walker, J. M., Ed.) pp 571−607, Humana Press, Totowa, NJ.
20. Cavaluzzi, M. J., and Borer, P. N. (2004) Revised UV extinction coefficients for nucleoside-5′-monophosphates and unpaired DNA and RNA, *Nucleic Acids Res. 32*, e13.
21. Kallansrud, G., and Ward, B. (1996) A comparison of measured and calculated single- and double-stranded oligodeoxynucleotide extinction coefficients, *Anal. Biochem. 236*, 134−138.
22. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, *Methods Enzymol. 276*, 307−326.
23. Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999) Rapid automated molecular replacement by evolutionary search, *Acta Crystallogr. D55* (Part 2), 484−491.
24. 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., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D54* (Part 5), 905−921.
25. Kleywegt, G. J., and Brunger, A. T. (1996) Checking your imagination: Applications of the free R value, *Structure 4*, 897−904.
26. McRee, D. E. (1999) XtalView/Xfit: A versatile program for manipulating atomic coordinates and electron density, *J. Struct. Biol. 125*, 156−165.
27. Laskowski, R. J., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
28. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling, *Electrophoresis 18*, 2714−2723.
29. DeLano, W. L. (2002) *PyMOL*, DeLano Scientific, San Carlos, CA.
30. Lowery, R., Hung, L., Knoche, K., and Bandziulis, R. (1992) Properties of I-PpoI: A rare-cutting intron-encoded endonuclease, *Promega Notes 38*, 8−12.
31. 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.
32. Wittmayer, P. K., and Raines, R. T. (1996) Substrate binding and turnover by the highly specific I-PpoI endonuclease, *Biochemistry 35*, 1076−1083.

BI700418D