The C-Terminal Domain of the Adenine-DNA Glycosylase MutY Confers Specificity for 8-Oxoguanine·Adenine Mispairs and May Have Evolved from MutT, an 8-Oxo-dGTPase†

David M. Noll, Arhonda Gogos, Joshua A. Granek, and Neil D. Clarke*

Department of Biophysics and Biophysical Chemistry, Johns Hopkins School of Medicine, Baltimore, Maryland 21205

Received February 11, 1999; Revised Manuscript Received April 7, 1999

ABSTRACT: MutY is an adenine-DNA glycosylase with specificity for mismatches involving 8-oxoguanine (^{o}G·A) or guanine (G·A). In addition to a 25 kDa catalytic domain common to all members of its DNA glycosylase superfamily, MutY has a 14 kDa C-terminal domain. Sequence analyses suggest that this C-terminal domain is distantly related to MutT, a pyrophosphohydrolase specific for 2′-deoxy-8-oxoguanosine triphosphate (d^{o}GTP). Here we present biochemical evidence that the MutT-like domain of MutY is the principal determinant of ^{o}G specificity. First, MutY dissociates approximately 1500-fold more slowly from ^{o}G-containing product DNA than from G-containing product, but a truncated protein lacking the C-terminal domain dissociates as rapidly from ^{o}G-DNA as the full-length protein dissociates from G-DNA. Second, MutY removes adenine from ^{o}G·A mismatches almost 30-fold faster than from G·A mismatches in a pre-steady-state assay, but deletion of the C-terminal domain reduces this specificity for ^{o}G·A to less than 4-fold. The kinetic data are consistent with a model in which binding of ^{o}G to the C-terminal domain of MutY accelerates the pre-steady-state glycosylase reaction by facilitating adenine base flipping. The observation that ^{o}G specificity derives almost exclusively from the C-terminal domain of MutY adds credence to the sequence analyses and suggests that specificity for ^{o}G·A mismatches was acquired by fusion of a MutT-like protein onto the core catalytic domain of an adenine-DNA glycosylase.

The oxidative DNA damage product 8-oxoguanine (^{o}G)[1] is a particularly mutagenic lesion because of its propensity to mispair with adenine (*1*). ^{o}G·A mispairs can arise via two pathways: (i) oxidation of guanine in the nucleotide pool followed by misincorporation of d^{o}GTP opposite adenine during replication and (ii) oxidation of guanine in DNA followed by misincorporation of adenine (*2*, *3*). In *Escherichia coli* and in humans, there is one enzyme that prevents the misincorporation of d^{o}GTP opposite adenine (MutT) and another that corrects the misincorporation of adenine opposite ^{o}G (MutY). MutT prevents incorporation of d^{o}GTP by hydrolyzing it to d^{o}GMP (*2*). MutY excises adenine that has been misincorporated opposite ^{o}G by cleaving the glycosylic bond connecting the adenine base to the DNA backbone (*4*, *5*). MutY will also remove adenine that is mispaired with several other bases, including guanine and cytosine (*4*, *6*−*10*).

MutY is a member of a highly divergent family of DNA glycosylases that have very different substrate specificities (*11*). Similarity among diverse members of this family was established by the structure determinations of endonuclease III and AlkA, which showed that the proteins share most of a two-domain fragment that includes the active site cleft (*12*−*14*). This region of structural similarity defines a catalytic domain for the superfamily. AlkA has an additional N-terminal domain of unknown function. MutY and its closest homologues also have a domain not found in other members of the family, but in this case, it is C-terminal to the catalytic domain.

The structure of the catalytic domain of MutY (cd-MutY; ends with Lys225) was recently determined with adenine bound in the active site (*15*). As expected, cd-MutY is very similar in structure to endonuclease III and to the catalytic domain of AlkA. The burial of adenine in the active site of cd-MutY suggests that the mismatched base must flip out of the DNA double helix in order to be bound by the enzyme. The protein−DNA complexes of two other DNA glycosylases from different structural families (T4-endo V and uracil-DNA glycosylase) suggest that base flipping is a general feature of DNA glycosylases (*16*, *17*).

Unlike most DNA glycosylases, which remove damaged bases or bases not normally found in DNA, MutY removes adenine, a normal base, when mispaired with ^{o}G or G. The crystal structure of the cd-MutY−adenine complex suggests some candidate residues that might play a role in the recognition of the base opposite the adenine, but the crystal structure does not include the C-terminal domain. The experiments described here show that the most critical determinants for recognition of ^{o}G do, in fact, lie within this domain.

† Support for this research was provided by the American Cancer Society.

* To whom correspondence should be addressed: Department of Biophysics and Biophysical Chemistry, Johns Hopkins School of Medicine, 725 N. Wolfe St., WBSB 713, Baltimore, MD 21205. Telephone: (410) 614-0338. Fax: (410) 502-6910. E-mail: nclarke@bs.jhmi.edu.

[1] Abbreviations: ^{o}G, 7,8-dihydro-8-oxo-2′-deoxyguanosine or 8-oxoguanine; cd-MutY, catalytic domain of MutY; SAX, strong anion exchange; MBP, maltose binding protein; ab, abasic (i.e., apurinic or apyrimidinic).

10.1021/bi990335x CCC: $18.00
Published on Web 04/28/1999

There has been significant confusion in the literature concerning the relative specificity of MutY for ^{o}G·A and G·A mispairs (*18*−*20*), the presence or absence of an associated lyase activity (*19*−*21*), and the role, if any, of the C-terminal domain (*21*, *22*). Work by David and colleagues has recently shown that much of this confusion stems from the fact that MutY turns over extremely slowly, especially with ^{o}G·A substrates (*5*, *23*, *24*). As a result, assays have generally been performed under conditions (i.e., excess protein) where the simplifying assumptions of Michaelis−Menten kinetics are invalid and can lead to incorrect inferences regarding steady-state rates and K_m values. In addition, because the very low turnover rate may indicate a slow rate of dissociation from product DNA, reported binding constants may not reflect true equilibrium values. To avoid these problems, we have used single-turnover adenine glycosylase kinetics and a direct measurement of product dissociation rates to determine the substrate specificities of full-length MutY and of the catalytic domain of MutY (cd-MutY). These results reveal a critical role for the C-terminal domain of MutY in the specific recognition of ^{o}G in ^{o}G·A mispairs. In addition to the experimental evidence that the C-terminal domain recognizes ^{o}G, we present sequence analyses that suggest that this domain is evolutionarily related to MutT, a hydrolase specific for d^{o}GTP. In combination, these results suggest that specificity for ^{o}G·A mismatches was acquired by fusion of a MutT-like protein onto the catalytic domain of an adenine-DNA glycosylase.

MATERIALS AND METHODS

Materials. All DNA synthesis reagents, including the ^{o}G phosphoramidite, were purchased from Glen Research. T4 polynucleotide kinase and Factor Xa protease were purchased from New England Biolabs. Uracil-DNA glycosylase was purchased from Trevigen.

Oligonucleotide Substrate Preparation. All substrate oligonucleotides were synthesized using standard phosphoramidite chemistry and commercially available reagents. Two variants each of two complementary strands were synthesized: d-TGC-CTG-CAG-AAG-GC**X**-ACT-GCA-GAT-CGT-CAC, where **X** is either deoxyuridine (U) or adenine, and d-GTG-ACG-ATC-TGC-AGT-**Y**GC-CTT-CTG-CAG-GCA, where **Y** is either guanine (G) or 8-oxoguanine (^{o}G). Oligonucleotide substrates were cleaved from the support and deprotected by treatment with concentrated NH_4OH at 55 °C for 24 h. For deprotection of oligomers containing ^{o}G, the solution was supplemented with 0.25 M β-mercaptoethanol. Oligomers were purified using anion exchange chromatography on a Dynamax II SAX column. Oligonucleotides containing A or U at the mispaired position were 5′ end-labeled using polynucleotide kinase and [γ-^{32}P]ATP and purified with Pharmacia G-25 microspin columns. Labeled duplexes were prepared by mixing the two complementary strands in a 1:1.2 ratio of the A or U strand to the G or ^{o}G strand in the presence of trace amounts of the labeled strand. The duplexes were then annealed in the glycosylase reaction buffer [20 mM Tris (pH 7.6), 80 mM NaCl, 1 mM DTT, 2 mM EDTA, and 2.9% glycerol] by heating to 85 °C for 5 min and then cooling to room temperature.

Construction, Overexpression, and Purification of MutY and cd-MutY. MutY and the catalytic domain of MutY (cd-MutY; ends with Q226) were expressed in *E. coli* strain TB1 as maltose binding protein (MBP) fusions, using the pMal-c2 vector (New England Biolabs). The MutY portions of the fusion genes were confirmed by DNA sequencing. Cultures were grown in LB medium to an OD_{600} of 0.6 and then induced with 0.3 mM IPTG for 3 h. The cell pellet was collected, resuspended in buffer A [0.5 M NaCl, 20 mM Tris (pH 7.4), 1 mM EDTA, and 1 mM DTT] in the presence of protease inhibitors (Pefabloc SC, leupeptin, and pepstatin), and frozen at −20 °C overnight. The cells were then thawed and lysed by sonication. The MBP fusion proteins were purified by affinity chromatography according to a New England Biolabs procedure. Fusion proteins were cleaved with Factor Xa protease (30 μg/mg MBP fusion protein, 36 h at room temperature); cleavage of the fusion protein leaves an extra Ile-Ser dipeptide N-terminal to the start of MutY. The digestion mixture was diluted to 0.1 M NaCl and passed over a heparin column, eluted with buffer A, and then passed over an amylose column to remove uncleaved fusion protein. MutY and cd-MutY were further purified to apparent homogeneity by gel filtration chromatography (Sephadex 75). The purified proteins were stored at −20 °C in buffer A and 10% glycerol.

Sequence Analysis. The relationship of the MutY C-terminal domain to other protein sequences was investigated using the database search program PSI-BLAST, which iteratively updates a position-specific scoring matrix as new homologues are identified (*25*). Using residues 217−390 of *E. coli* MutY as the probe sequence, PSI-BLAST identified only a few homologues from closely related bacterial species before converging after the first round. The top scoring sequence that failed to meet the default threshold for addition to the profile was the human homologue of MutY. Since the biochemical activity of this homologue has been confirmed (*26*, *27*), addition of this sequence to the profile was deemed justified. Addition of this sequence allowed many more distantly related sequences to be identified in subsequent rounds. The results of each round of the PSI-BLAST search can be found at http://biophysics.med.jhmi.edu/clarke/MutYMutT.html. The relationship between MutT and the C-terminal domain of MutY from several organisms was also revealed using a search of hidden Markov models for proteins of known structure at http://www.cse.ucsc.edu/research/compbio (*28*). The PRSS program (*29*, *30*) (http://fasta.bioch.virginia.edu/fasta/prss.html) was used to show that there is a significant relationship, by the sequence jumbling criterion, between human MutY and both *E. coli* MutY ($P < 0.00001$) and *E. coli* MutT ($P < 0.002$).

Product Dissociation Kinetics. The rate of product DNA release was determined using an electrophoretic mobility shift assay. MutY (300 nM final concentration) or cd-MutY (150 nM final concentration) was added to a 30 nM solution of labeled ^{o}G·A or G·A duplex in glycosylase reaction buffer (final volume of 100 μL) and incubated at 37 °C for 15 min to generate an abasic site opposite either ^{o}G or G. At these protein concentrations, all of the product DNA appears to be bound (data not shown). Formation of abasic product was confirmed by base-catalyzed cleavage of an aliquot (data not shown). Following removal of a 10 μL aliquot as a zero time point for the dissociation kinetics, 10 μL of a 300 μM solution of unlabeled DNA duplex containing ^{o}G opposite an abasic site (^{o}G·ab) was added. This results in a 1000-

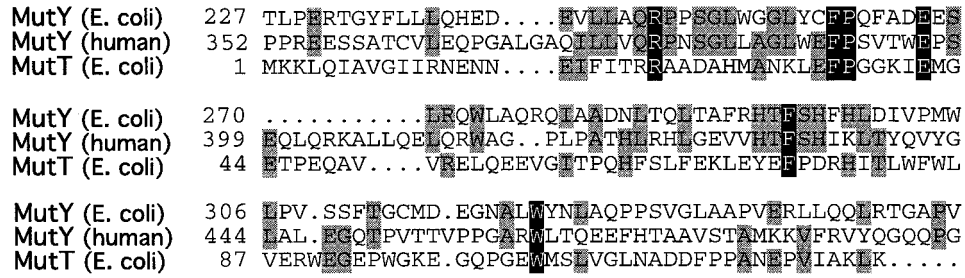

FIGURE 1: Sequence alignment of the *E. coli* MutY C-terminal domain, the human MutY C-terminal domain, and *E. coli* MutT. The alignment was obtained from the PSI-BLAST output, with the addition of several residues at the N-terminus of MutT. The human MutY sequence is included because it serves as a statistically significant bridge between the *E. coli* MutY and MutT sequences as assessed by the PRSS jumbling test (see Materials and Methods) (*29*, *30*).

fold excess of competitor DNA over bound DNA and a 100−200-fold excess of competitor DNA over protein. Aliquots were removed at various times (5 min to 68 h) and electrophoresed on a $^1/_2\times$ TBE 8% nondenaturing polyacrylamide (19:1) gel at 200 V while submerged in 1× TBE at room temperature to prevent warming of the gel. ^{o}G·ab competitor DNA was prepared by annealing equimolar amounts of the ^{o}G-containing strand with the U-containing strand, followed by treatment with 2 units of uracil-DNA glycosylase for 1 h in glycosylase reaction buffer. Controls using radiolabeled U strand confirmed that this protocol generates an abasic site opposite ^{o}G (data not shown). Dissociation assays were carried out in triplicate, and quantified by phosphorimaging. The time courses for the dissociation from product DNA and for conversion of the substrate to product (described below) were fitted to a single-exponential decay curve using either KaleidaGraph or SigmaPlot. For the fast dissociating complexes (i.e., >80% dissociation at 5 min), it was assumed that variation in the gel loading time was the greatest source of error, so errors were estimated by calculating dissociation rates based on time points of 5.0 ± 1.5 min. The reported value for cd-MutY dissociation from G·A DNA is a lower limit because dissociation was greater than 95% complete by the first time point.

Single-Turnover Kinetics. Enzymatic assays were carried out in the glycosylase reaction buffer. Single-turnover glycosylase rates were determined with protein in excess of substrate using a Kin-Tek model RQF-3 pulsed quench-flow apparatus. In some cases, the rate was sufficiently slow that manual mixing was used. Various concentrations of MutY or cd-MutY were mixed with an equal volume of mispaired substrate at a lower concentration and allowed to react for 0.5−150 s, and then the reactions were quenched by the addition of 2.1 volumes of 0.1 M NaOH. The protein and DNA concentrations that were used are listed in the figure legends. At protein concentrations of >150 nM, MBP fusion proteins were used because of the availability of more concentrated protein stocks; control experiments at lower concentrations showed no significant effect of the MBP on the rate of glycosylase activity for either MutY or cd-MutY (data not shown). The quenched alkaline solutions were heated to 90 °C for 15 min to cleave the abasic site. The solutions were cooled to room temperature and neutralized with acetic acid. The DNA was precipitated, dried, dissolved in 90% formamide loading buffer, and electrophoresed on a 1× TBE 20% denaturing (8 M urea) polyacrylamide (19:1) gel submerged in a 45 °C water bath. Glycosylase assays were quantitated by phosphorimaging, and the fraction cleaved was calculated from the intensity of the substrate and product bands, correcting for a small fraction of substrate DNA (<1%) that comigrated with the product in the absence of enzyme and a small fraction of substrate DNA (<8%) that appeared to be nonrepairable (uncleaved after a 30 min incubation with a 5000-fold excess of full-length MutY). Rates for glycosylase activity at saturating protein concentrations were determined by fitting the observed rate measured at different protein concentrations to a hyperbola.

RESULTS

Sequence Analysis. Proteolytic studies and sequence comparisons have shown that MutY consists of a catalytic domain that is similar in all members of the endoIII/MutY/AlkA superfamily, and a C-terminal domain that is unique to the MutY subfamily (*21*, *22*, *31*). To determine whether the extra domain is related to any other proteins, a PSI-BLAST search was carried out using the *E. coli* MutY C-terminal domain sequence (see Materials and Methods). After several iterations, all of the annotated homologues that were identified were either MutY homologues or MutT homologues. MutT is a d^{o}GTP pyrophosphohydrolase and a member of a large family of phosphohydrolases with differing specificities but with conserved catalytic function (*32*). Subsequent iterations of PSI-BLAST were required to identify other members of this family. This suggests that the relationship between MutY and MutT is closer than that between MutY and other phosphohydrolases which recognize substrates other than d^{o}GTP. Figure 1 shows an alignment of the C-terminal domains of *E. coli* and human MutY with the complete sequence of *E. coli* MutT.

The PSI-BLAST results suggest that MutT and the C-terminal domain of MutY might share an evolutionary relationship. Since MutT and MutY each recognize a substrate that includes 8-oxoguanine (d^{o}GTP in the case of MutT and ^{o}G·A DNA mispairs in the case of MutY), it seemed plausible that the MutT-like domain of MutY confers specificity for ^{o}G. To investigate the role of the C-terminal domain, kinetic experiments were conducted using full-length MutY and a version of the protein in which the C-terminal domain has been deleted (cd-MutY).

Kinetics of Dissociation from Abasic Product DNA. The steady-state turnover of MutY with ^{o}G·A mispair-containing DNA is much slower than glycosylic bond cleavage itself, suggesting that product dissociation may be rate-limiting (*5*). To test this idea directly and to see what role the C-terminal

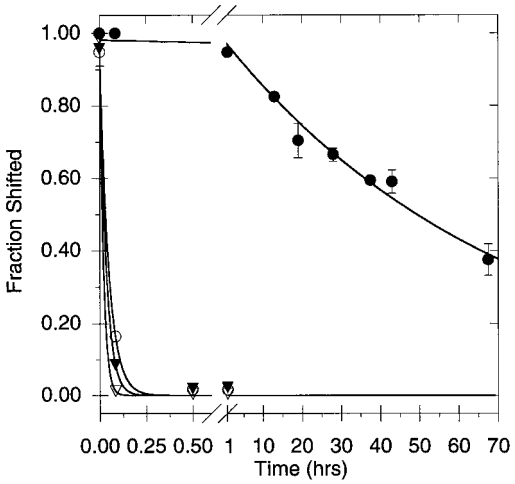

FIGURE 2: Determination of the rates of dissociation of MutY (circles) and cd-MutY (triangles) from °G·ab (black) or G·ab (white) mispair-containing DNA. The fraction shifted was calculated from triplicate measurements using an electrophoretic mobility shift assay as described in Materials and Methods. Error bars have been omitted in instances where they are smaller than the symbol. The solid lines represent single-exponential decays fit to the data (note the break in the *x*-axis and the change in scale). Dissociation rates are listed in Table 1.

Table 1: Glycosylase and Dissociation Rates[a]

	glycosylase rate ($\times 10^3$ s^{-1})		dissociation rate (h^{-1})	
	°G·A	G·A	°G·ab	G·ab
MutY	226 ± 5	7.8 ± 1.1	0.014 ± 0.001	21 ± 5
cd-MutY	3.4 ± 0.7	0.9 ± 0.1	29 ± 6	>52

[a] Derivation of rates and errors is described in Materials and Methods.

domain plays in this slow turnover, the rate of product dissociation was measured using different combinations of proteins and substrates. Dissociation rates were determined using a gel shift competition assay (Materials and Methods) in which saturating concentrations of MutY or cd-MutY were allowed to react with labeled DNA containing °G·A or G·A mispairs. Under these conditions, essentially all of the substrate was converted to the respective abasic product (°G·ab or G·ab). A large excess of unlabeled °G·ab DNA was then added to compete with the labeled product DNA for the rebinding of any protein that might dissociate. At various times following the addition of competitor, an aliquot was removed and analyzed by gel electrophoresis to determine the amount of radiolabeled DNA still bound by protein.

Figure 2 shows the fraction of product DNA that remains bound to protein as a function of time. The differences in dissociation kinetics are dramatic. While the half-life for dissociation of full-length MutY from °G·ab DNA is approximately 48 h, the same protein dissociates from G·A DNA with a half-life of less than 3 min. Furthermore, tight binding to °G·ab DNA appears to be attributable to the C-terminal domain since cd-MutY, which lacks the C-terminal domain, dissociates at a similar rate from both °G·ab and G·ab DNA, and these rates of dissociation are similar to that of full-length MutY bound to G·ab DNA (Figure 2 and Table 1). Although the estimated errors for the fast-dissociating complexes are fairly large (see Materials and Methods), these differences are significantly less dramatic than the roughly 1500-fold slower dissociation observed for MutY on °G·ab DNA.

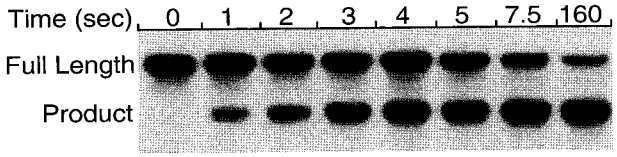

FIGURE 3: Sample autoradiograph of a representative glycosylase time course at an °G·A concentration of 15 nM and a MutY concentration of 30 nM. The time course determination was conducted using an RQF-3 pulsed quench-flow apparatus with a 0.1 N NaOH quench solution. Glycosylase activity was scored by cleavage of the abasic site followed by separation using PAGE as described in Materials and Methods.

Glycosylase Kinetics. The very slow dissociation of MutY from G·ab DNA confirms the suggestion that product dissociation is rate-limiting under steady-state conditions (*5*). Thus, to determine the rate of glycosylic bond cleavage, it is necessary to measure the pre-steady-state rate. Pre-steady-state rates were measured under conditions in which MutY is in molar excess over substrate. This ensures that each DNA molecule is repaired by a separate protein molecule, making the observed rate of abasic DNA formation independent of product dissociation.

Figure 3 is an autoradiogram of a representative time course using a rapid-quench apparatus to follow glycosylase kinetics for a 30 nM solution of °G·A mixed with a 60 nM solution of MutY. In the absence of glycosylase activity (time 0), the radiolabeled A-containing strand is insensitive to treatment with base and only full-length substrate is observed. However, incubation with MutY rapidly produces a base sensitive species that corresponds to the expected adenine-excised product.

Similar experiments were conducted using both MutY and cd-MutY with both °G·A and G·A substrates. Figure 4A shows one example using the same protein and DNA concentrations used for Figure 3. As is clear from this figure, the rate of adenine excision from °G·A by full-length MutY (●) is much faster than the rate of adenine excision from G·A mispairs. This difference has been previously noted (*5*), but the use of a rapid-quench device in these experiments allows the magnitude of the difference to be accurately determined. Of greater note is the fact that the preference of MutY for °G·A mispairs (compare circles) is nearly abolished by deletion of the C-terminal domain (compare triangles), but deletion of the domain has only a small effect on G·A glycosylase activity (compare white symbols).

Assays similar to those shown in Figure 4A were performed at a substrate concentration of 1 nM and protein concentrations ranging from 5 to 1250 nM. The rates obtained from these kinetics experiments are plotted in Figure 4B. The rate of °G·A repair by MutY (●) is dependent on protein concentration at relatively low concentrations, but becomes much less dependent at higher concentrations, asymptotically approaching a maximum rate. For the other combinations of protein and substrate, errors in the data are large enough that it is uncertain whether there is any concentration dependence at low protein concentrations. However, it is clear that these protein−substrate pairs have also reached their maximum intrinsic rate given that the rates are essentially indistinguishable over a 250-fold range of protein concentrations. Thus, the differences in glycosylase activity at the highest protein concentrations are not attributable to differences in protein−substrate binding affinity, but

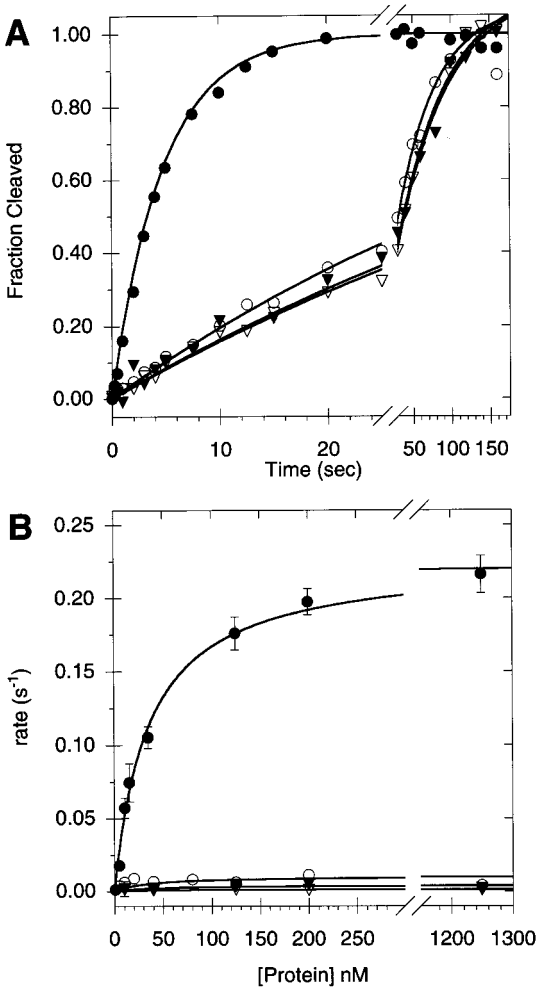

FIGURE 4: Determination of maximal glycosylase rates of MutY (circles) and cd-MutY (triangles) with ^{o}G·A (black) or G·A (white) mispair-containing DNA. Panel A is a plot of the fraction cleaved vs time for the protein and DNA concentrations used in the autoradiogram shown in Figure 3. Despite slower glycosylase rates for full-length MutY with G·A and for cd-MutY with both ^{o}G·A and G·A, the rapid-quench apparatus was used to facilitate direct comparison. In all four cases, the data can be fit (solid lines) to a single-exponential decay (note the break in the *x*-axis and the change in scale). Panel B is a plot of rate (as determined by single-exponential fits to individual time courses as described above) vs protein concentration. Individual time courses were determined with a DNA concentration of 1 nM and protein concentrations ranging from 5 to 1250 nM. The protein concentration dependence of the glycosylase rate has been fit to a hyperbola (solid line). Glycosylase rates are listed in Table 1.

rather reflect differences in the rate-determining step. The maximal rates of reaction obtained from fits to these data are summarized in Table 1.

The most striking results are that MutY cleaves adenine almost 30-fold more rapidly from mispairs with ^{o}G than with G, and that this difference is due in large part to the C-terminal domain. Thus, using two completely different criteria, the C-terminal domain of MutY seems to be the principal determinant of ^{o}G specificity. First, the C-terminal domain confers tight ^{o}G-specific binding as measured by dissociation from abasic product DNA, and second, it confers an increase in the intrinsic adenine glycosylase rate when the basepair partner of adenine is ^{o}G.

How does the C-terminal domain increase the intrinsic adenine glycosylase rate? We note first that structural data and comparisons to other DNA glycosylases suggest that the adenine must be flipped out of the DNA helix in order to be excised, and that flipping of the adenine could be rate-limiting. If base flipping is, in fact, rate-limiting, then binding events that promote base flipping would be expected to increase the rate of repair (*33*). We suggest that this occurs in MutY by binding of the C-terminal domain to ^{o}G. In our model, this binding disrupts the ^{o}G·A base pairing interactions and allows adenine to flip out of the DNA more readily. We favor a model in which recognition of ^{o}G by the C-terminal domain is due to direct binding because of the apparent evolutionary relationship with MutT, a protein known to bind to d^{o}GTP.

DISCUSSION

Recent kinetic and biochemical analyses suggest that the assays routinely used to characterize MutY can be misleading under some conditions or if interpreted incorrectly. Potential sources of confusion that have been addressed include the inappropriate application of Michaelis−Menten assumptions in kinetic experiments (*5*; D. M. Noll and N. D. Clarke, unpublished results), the untested assumption of thermodynamic equilibrium in binding experiments (this work), the use of coupled glycosylase−lyase assays in measuring glycosylase activity (*24*, *34*), and the uncertain significance of Schiff base cross-linking as a surrogate measure of enzyme activity or binding affinity (*24*, *34*). These and perhaps other sources of confusion have led to debate over the role of the MutY C-terminal domain (*21*, *22*). In this paper, we have avoided these problems by measuring directly the rate of dissociation from product DNA and the pre-steady-state rate of adenine glycosylase activity.

The conclusion of these experiments, consistent with our earlier conclusion based on less direct evidence (*22*), is that the C-terminal domain of MutY confers nearly all of the specificity for ^{o}G·A versus G·A mispairs. This specificity manifests itself in a dramatically longer time for dissociation from ^{o}G·ab DNA than from G·ab DNA, and in a sizable enhancement in the rate of adenine excision when A is mispaired with ^{o}G. For reasons discussed above, we believe both effects can be attributed to the direct binding of ^{o}G to the C-terminal domain. Inspection of the structure of the MutY catalytic domain and of a model for how this domain might bind to DNA (*15*) suggests that the C-terminus of the protein is near the DNA and that the C-terminal domain could very well bind to the base opposite the flipped out adenine (not shown). Although the bound ^{o}G need not be in a flipped configuration itself, such a configuration would obviously satisfy the requirement of our kinetic model that ^{o}G binding destabilize the ^{o}G·A base pair and promote base flipping of the adenine.

There is good evidence that ^{o}G·A and not G·A is the biologically relevant substrate for MutY. Overexpression of MutM, an ^{o}G-DNA glycosylase, completely suppresses the mutagenic effects of MutY inactivation, suggesting that G·A pairs, which are not a substrate for MutM, are not a significant source of mutations as far as MutY correction is concerned (*35*). It may seem paradoxical, then, that the C-terminal domain confers on MutY both a faster pre-steady-state rate of adenine release for ^{o}G·A mispairs and a much slower steady-state rate for the same substrate due to tight

product binding. In vivo, though, it is possible that other factors catalyze the dissociation of MutY from product DNA. The most likely candidates are proteins involved in subsequent steps of repair (*36*, *37*). In fact, as has been noted by others, it may well be advantageous for MutY to remain tightly bound to ^{o}G·ab DNA until the rest of the repair machinery is available (*4*). Otherwise, ^{o}G might be excised by MutM, creating an ab·ab site that retains no genetic information and is susceptible to double-strand breakage.

The sequence identity between *E. coli* MutY and *E. coli* MutT is extremely low (10% with two gaps). Conservation of key catalytic residues, often a useful way of substantiating weak evolutionary relationships, is not applicable in this case because the evolution of the C-terminal domain would not have depended on retention of MutT catalytic activity. Nevertheless, the sequences do appear to be significantly related by several criteria. First, there is the PSI-BLAST search itself. Second, in a search of hidden Markov models for proteins of known structure, a significant degree of similarity was detected between *E. coli* MutT and the C-terminal domains of MutY from several species, especially human (see Materials and Methods; data not shown). Third, a sequence jumbling test also suggested a significant degree of similarity between human MutY and *E. coli* MutT (Materials and Methods).

Most importantly, the experiments described here strongly suggest that the C-terminal domain binds ^{o}G, and thus provide a biochemical link to the substrate specificity of MutT. This biochemical connection lends substantial credence to the sequence analyses, and suggests that MutT and the C-terminal domain of MutY have a common evolutionary origin. Since MutY homologues from bacteria to humans contain the extra C-terminal domain relative to other DNA glycosylases with different specificities, it appears that the acquisition of this domain was an early event in the evolution of adenine-DNA glycosylases. We suggest that MutY arose as a gene fusion between a core DNA glycosylase with specificity for adenine mispairs and a MutT-like protein. The modification of specificity at a single base pair by the acquisition of an entire domain is a remarkable example of the modularity of protein evolution.

ACKNOWLEDGMENT

We thank Dr. Roger McMacken for the use of the RQF-3 pulsed quench-flow apparatus and Dr. Paul Miller for the use of an ABI 392 DNA synthesizer and ancillary oligonucleotide synthesis and purification facilities.

REFERENCES

1. Grollman, A. P., and Moriya, M. (1993) *Trends Genet. 9*, 246.
2. Maki, H., and Sekiguchi, M. (1992) *Nature 355*, 273.
3. Shibutani, S., Takeshita, M., and Grollman, A. P. (1991) *Nature 349*, 431.
4. Michaels, M. L., Tchou, J., Grollman, A. P., and Miller, J. H. (1992) *Biochemistry 31*, 10964.
5. Poello, S. L., Leyes, A. E., and David, S. S. (1998) *Biochemistry 37*, 14756.
6. Radicella, J. P., Clark, E. A., and Fox, M. S. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 9674.
7. Au, K. G., Cabrera, M., Miller, J. H., and Modrich, P. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 9163.
8. Tsai, W. J., Liu, H. F., and Lu, A. L. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 8779.
9. Lu, A. L., and Chang, D. Y. (1988) *Genetics 118*, 593.
10. Lu, A. L., and Chang, D. Y. (1988) *Cell 54*, 805.
11. Nash, H. M., Bruner, S. D., Scharer, O. D., Kawate, T., Addona, T. A., Spooner, E., Lane, W. S., and Verdine, G. L. (1996) *Curr. Biol. 6*, 968.
12. Kuo, C. F., McRee, D. E., Fisher, C. L., O'Handley, S. F., Cunningham, R. P., and Tainer, J. A. (1992) *Science 258*, 434.
13. Labahn, J., Scharer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, G. L., and Ellenberger, T. E. (1996) *Cell 86*, 321.
14. Yamagata, Y., Kato, M., Odawara, K., Tokuno, Y., Nakashima, Y., Matsushima, N., Yasumura, K., Tomita, K., Ihara, K., Fujii, Y., Nakabeppu, Y., Sekiguchi, M., and Fujii, S. (1996) *Cell 86*, 311.
15. Guan, Y., Manuel, R. C., Arvai, A. S., Parikh, S. S., Mol, C. D., Miller, J. H., Lloyd, R. S., and Tainer, J. A. (1998) *Nat. Struct. Biol. 5*, 1058.
16. Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., and Tainer, J. A. (1996) *Nature 384*, 87.
17. Vassylyev, D. G., Kashiwagi, T., Mikami, Y., Ariyoshi, M., Iwai, S., Ohtsuka, E., and Morikawa, K. (1995) *Cell 83*, 773.
18. Lu, A. L., Tsai, W. J., and Cillo, J. (1995) *J. Biol. Chem. 270*, 23582.
19. Lu, A. L., Yuen, D. S., and Cillo, J. (1996) *J. Biol. Chem. 271*, 24138.
20. Bulychev, N. V., Varaprasad, C. V., Dorman, G., Miller, J. H., Eisenberg, M., Grollman, A. P., and Johnson, F. (1996) *Biochemistry 35*, 13147.
21. Manuel, R. C., and Lloyd, R. S. (1997) *Biochemistry 36*, 11140.
22. Gogos, A., Cillo, J., Clarke, N. D., and Lu, A. L. (1996) *Biochemistry 35*, 16665.
23. David, S. S., and Williams, S. D. (1998) *Chem. Rev. 98*, 1221.
24. Williams, S. D., and David, S. S. (1998) *Nucleic Acids Res 26*, 5123.
25. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res. 25*, 3389.
26. McGoldrick, J. P., Yeh, Y.-C., Solomon, M., Essigmann, J. M., and Lu, A.-L. (1995) *Mol. Cell. Biol. 15*, 989.
27. Slupska, M. M., Baikalov, C., Luther, W. M., Chiang, J.-H., Wei, Y.-F., and Miller, J. H. (1996) *J. Bacteriol. 178*, 3885.
28. Karplus, K., Barrett, C., and Hughey, R. (1999) *Bioinformatics* (in press).
29. Smith, T. F., and Waterman, M. S. (1981) *J. Mol. Biol. 147*, 195.
30. Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2444.
31. Manuel, R. C., Czerwinski, E. W., and Lloyd, R. S. (1996) *J. Biol. Chem. 271*, 16218.
32. Bessman, M. J., Frick, D. N., and O'Handley, S. F. (1996) *J. Biol. Chem. 271*, 25059.
33. Stivers, J. T., Pankiewicz, K. W., and Watanabe, K. A. (1999) *Biochemistry 38*, 952.
34. Zharkov, D. O., and Grollman, A. P. (1998) *Biochemistry 37*, 12384.
35. Michaels, M. L., Cruz, C., Grollman, A. P., and Miller, J. H. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7022.
36. Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E., and Tainer, J. A. (1998) *EMBO J. 17*, 5214.
37. Waters, T. R., Gallinari, P., Jiricny, J., and Swann, P. F. (1999) *J. Biol. Chem. 274*, 67.

BI990335X