Prepr. Am. Chem. Soc. Div. Polym. Chem. **35** (no. 2), 807 (1994).
46. K. Mislow and P. Bickart, *Isr. J. Chem.* **15**, 1 (1976–1977); H. Wynberg *et al.*, *J. Am. Chem. Soc.* **87**, 2635 (1965).
47. For leading references, see E. J. Enemark and T. D. P. Stack, *Agnew. Chem. Int. Ed. Engl.* **34**, 996 (1995); M. Fujiki, *J. Am. Chem. Soc.* **116**, 6017 (1994); S. Hanessian, A. Gomtsyan, M. Simard, S. Roelens, *ibid.*, p. 4495; J. M. Schnur *et al.*, *Science* **264**, 945 (1994); M. Mascal, C. J. Moody, A. I. Morrell, A. M. Z. Slawin, D. J. Williams, *J. Am. Chem. Soc.* **115**, 813 (1993); J. M. Lehn, *Makromol. Chem. Macromol. Symp.* **69**, 1 (1993); Y. Ito, E. Ihara, M. Murakami, *Angew. Chem. Int. Ed. Engl.* **31**, 1509 (1992); O. Vogl, G. D. Jaycox, C. Kratky, W. J. Simonsick Jr., K. Hatada, *Acc. Chem. Res.* **25**, 408 (1992); P. Magnus, W. Danikiewicz, T. Katch, J. C. Huffman, K. Folting, *J. Am. Chem. Soc.* **112**, 2465 (1990).
48. The work performed at Polytechnic University was supported by the Chemistry and Materials Divisions of the National Science Foundation and by the Petroleum Research Fund administered by the American Chemical Society. These studies were supported at Osaka University by the Ministry of Education, Science, and Culture (grant number 050-44 096). The effort at Lawrence Livermore National Laboratory was supported by the U.S. Department of Energy under contract W-7405-ENG-48, and financial support in Israel was made available by the Weizmann Institute. The Brooklyn-Osaka collaboration was aided by a National Science Foundation Japan–United States Fellowship to M.M.G. We are grateful to our students and co-workers who made this work possible and whose names are in the references and to K. Mislow and H. Morawetz for advice during the course of this work.

RESEARCH ARTICLE

Crystal Structure of DNA Photolyase from *Escherichia coli*

Hee-Won Park, Sang-Tae Kim,* Aziz Sancar, Johann Deisenhofer

Photolyase repairs ultraviolet (UV) damage to DNA by splitting the cyclobutane ring of the major UV photoproduct, the *cis,syn*-cyclobutane pyrimidine dimer (Pyr<>Pyr). The reaction is initiated by blue light and proceeds through long-range energy transfer, single electron transfer, and enzyme catalysis by a radical mechanism. The three-dimensional crystallographic structure of DNA photolyase from *Escherichia coli* is presented and the atomic model was refined to an *R* value of 0.172 at 2.3 Å resolution. The polypeptide chain of 471 amino acids is folded into an amino-terminal α/β domain resembling dinucleotide binding domains and a carboxyl-terminal helical domain; a loop of 72 residues connects the domains. The light-harvesting cofactor 5,10-methenyltetrahydrofolylpolyglutamate (MTHF) binds in a cleft between the two domains. Energy transfer from MTHF to the catalytic cofactor flavin adenine dinucleotide (FAD) occurs over a distance of 16.8 Å. The FAD adopts a U-shaped conformation between two helix clusters in the center of the helical domain and is accessible through a hole in the surface of this domain. Dimensions and polarity of the hole match those of a Pyr<>Pyr dinucleotide, suggesting that the Pyr<>Pyr "flips out" of the helix to fit into this hole, and that electron transfer between the flavin and the Pyr<>Pyr occurs over van der Waals contact distance.

The ultraviolet component of sunlight leads to the formation of cyclobutadipyrimidines (pyrimidine dimers, Pyr<>Pyr) in DNA. These lesions block replication and transcription and thus have cytotoxic and mutagenic effects (*1*). With the gradual depletion of the stratospheric ozone layer, a higher flux of UV is reaching the surface of the Earth and hence more Pyr<>Pyr are being formed in the biosphere with potentially serious consequences (*2*). Cells survive the harmful effects of Pyr<>Pyr by excision repair or by photoreactivation. In photoreactivation, the enzyme DNA photolyase (*3*) binds to Pyr<>Pyr and, upon excitation by blue light, splits the cyclobutane ring and restores the intact bases (*4*).

$$E + Pyr<>Pyr \xrightarrow{\text{Light (400 nm)}} E + Pyr\text{-}Pyr$$

Photolyase has been found in bacteria, archaebacteria, and eukaryotes including goldfish, rattlesnake, and marsupials (*1*). The enzyme consists of a single polypeptide chain of 454 to 614 amino acids (*5*) and two noncovalently attached prosthetic groups. One of these is the catalytic cofactor $FADH^-$ (*6*); the other one is a light-harvesting cofactor. Photolyases are classified into two groups according to the type of light-harvesting cofactors: The folate class enzymes bind 5,10-methenyltetrahydrofolylpolyglutamate (MTHF), and the deazaflavin class enzymes contain 7,8-didemethyl-8-hydroxy-5-deazariboflavin.

A blue-light photoreceptor with strong sequence similarity to microbial photolyases but with no photolyase activity has been discovered in plants (*7*). Conversely, animal photolyases have only weak sequence similarity to microbial photolyases (*5*). Both the blue-light photoreceptor and the animal photolyases have FAD and MTHF as cofactors (*8*) suggesting functional similarities with the *Escherichia coli* photolyase described below.

As apparent from the reaction scheme, binding of the enzyme and catalysis are distinct both temporally and mechanistically. Binding is independent of light, catalysis is light-initiated. Photolyase is one of the rare specific DNA binding proteins that bind to its target in both double- and single-stranded DNA with comparable affinities (*9*). The binding constants are $K_D = 10^{-9}$ M and $K_D = 10^{-8}$ M, respectively for double- and single-stranded DNA containing a thymine dimer (T<>T) (*4*, *9*). Photolyase also binds to T<>T in the dinucleotide form, albeit with considerably lower affinity [$K_D = 10^{-4}$ M (*10*)]. Thus, both the dimer and the backbone deformation caused by the dimer contribute to binding.

Catalysis is initiated by absorption of a blue-light photon and leads to splitting of two C-C bonds. Five steps can be distinguished in the overall reaction. First, the light-harvesting cofactor absorbs a photon; second, the excitation energy is transferred to the catalytic cofactor; third, an electron is transferred to the Pyr<>Pyr in the substrate; fourth, the C5-C5 and C6-C6 sigma bonds of the cyclobutane ring are broken; fifth, the electron is transferred back to the flavin and the now intact DNA dissociates from the enzyme (*4*).

In addition to energy and electron transfer during the repair reaction, two other photoinduced electron transfer processes have been observed in DNA photolyase from *E. coli*. (i) During purification, the catalytic cofactor of the enzyme is oxidized to the catalytically inert neutral radical form, $FADH^\bullet$ (*11*). This form can be reactivated by irradiation with white light, which causes electron transfer from Trp^{306}

H.-W. Park is in the Department of Biochemistry and J. Deisenhofer is in the Department of Biochemistry and the Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX 75235. S.-T. Kim and A. Sancar are in the Department of Biochemistry and Biophysics, University of North Carolina School of Medicine, Chapel Hill, NC 27599.

*Present address: Procter and Gamble Far East, Inc., Japan Headquarters and Technical Center, 17, Koyocho Naka 1-chome, Higashinada-ku, Kobe 658, Japan.

to the excited quartet state of $FADH^{\bullet}$ (*12*). (*ii*) The $FADH^{\bullet}$ form of enzyme does split Pyr<>Pyr by means of 280-nm light by electron transfer from Trp^{277}, which is at the DNA binding site (*13*).

Thus, photolyase is a biological system that performs resonant energy transfer and three distinct electron transfer reactions. Similar processes occur during the photosynthetic light reactions, and structural information on photosynthetic reaction centers (*14*) was of critical importance for understanding the data contributed by spectroscopic and biochemical studies of these reactions. In order to better understand the structure-function relations in photolyase, we have performed an x-ray structure analysis.

Crystals of DNA photolyase from *E. coli*, a member of the folate class (*15*), and from *Anacystis nidulans*, a member of the deazaflavin class (*16*) have been reported. Here we describe the three-dimensional structure of DNA photolyase from *E. coli*, which we determined by x-ray crystallography at 2.3 Å resolution. The structure reveals the conformation, distance, and relative orientation of the cofactors $FADH^{\bullet}$ and MTHF, and their interaction with the protein. It also indicates the binding site and an unexpected binding mode of substrate.

Structure determination. We overexpressed, purified, and crystallized DNA photolyase from *E. coli* as described (*15*). The enzyme purified from overproducing *E. coli* contains substoichiometric MTHF with three to six glutamate residues (*17*). To obtain a homogeneous enzyme preparation, we removed the intrinsic folate and replaced it with a stoichiometric amount of folate monoglutamate. Thus, the enzyme used in our studies had the flavin in the neutral radical form ($FADH^{\bullet}$) and the folate in the monoglutamate form. With this enzyme, we obtained two crystal forms, one of which (form II) was suitable for structure analysis.

Crystals of form II have the symmetry of space group *P1* and the unit cell dimensions a = 62.6 Å, b = 72.2 Å, c = 58.5 Å, α = 99.1°, β = 101.5°, and γ = 72.0° (*15*). Two molecules in the unit cell are related by almost perfect twofold symmetry. We determined the crystal structure with a combination of multiple isomorphous replacement, solvent flattening, noncrystallographic symmetry averaging, and crystallographic refinement (Table 1). The refined atomic model described here has an R value of 0.172 (R_{free} = 0.246) for 32619 reflections with $F > 2\sigma(F)$ between 10 Å and 2.3 Å resolution. It includes residues 1 to 469 and the two cofactors; residues 470 and 471 are disordered in the crystal. The quality of the model is excellent as indicated by a low deviation of the model from ideal geometry (Table 2), by a correlation of 0.896 between final model and electron density (*18*), and by more than 90 percent of the residues in the most favored region of the Ramachandran diagram (*19*). Only residue Arg^{8} remained in a disallowed region, despite fre-

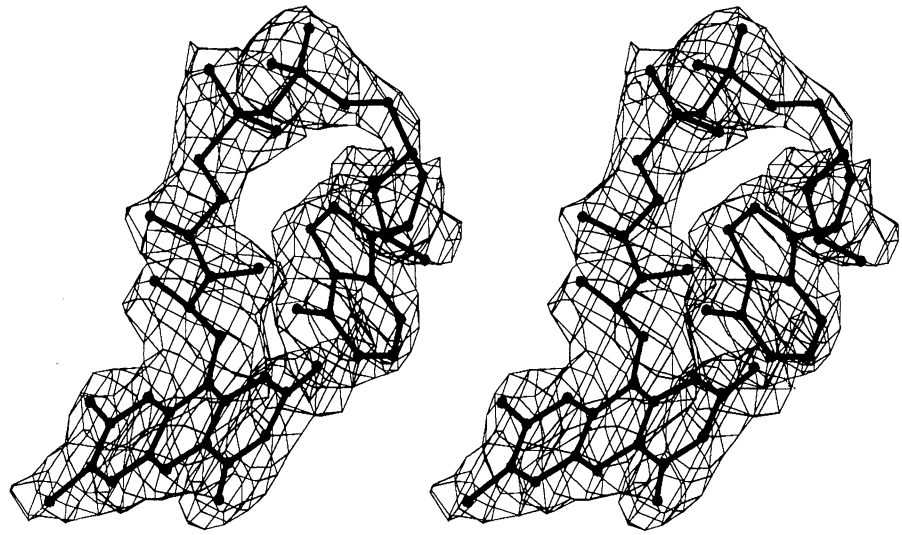

Fig. 1. Stereo pair: FAD cofactor (thick lines) with electron density (thin lines), contoured at 1σ.

Table 1. Crystal structure determination. Complete diffraction data were measured on dual multiwire area detectors of Xuong-Hamlin design with graphite monochromated CuKα x-rays generated by a Rigaku RU-200 rotating anode generator. The generator was operated in fine focus mode at 50 kV, 100 mA. A Huber three-circle goniostat with stepper motors was used to position and rotate crystals for data collection. Data were processed with the program XDS (*47*) and then merged and scaled with the program PROTEIN (*48*). Extensive screening to determine an adequate mother liquor for stabilizing photolyase crystals was unsuccessful. Therefore, highly concentrated solutions of heavy atoms were made in water and added directly to the protein droplet containing crystals. A difference-Patterson map of the thimerosal-1 derivative was calculated with PROTEIN and showed three heavy-atom binding sites. These data were used to calculate single isomorphous replacement (SIR) phases at 5 Å resolution with the MIRPH option in PROTEIN. The SIR phases were then improved by a solvent-flattening method (*49*) with the program MOLBOUND (M. Schneider). The combined phases from SIR and solvent-flattening were used to find the right enantiomorph. Additional sites of thimerosal-1 and sites of thimerosal-2, K_2OsCl_6, and K_2PtCl_4 were found by difference Fourier synthesis with SIR phases based on thimerosal-1 derivative. Multiple isomorphous replacement (MIR) phases with all the derivative data were calculated with the program MLPHARE (*50*) at 3 Å resolution. The noncrystallographic transformation that relates the two molecules in the asymmetric unit was determined by real space search with PROTEIN and further refined with the program CCMAX in the DEMON package (written and provided by F. M. D. Vellieux). The MIR phases were improved by density averaging in DEMON. An atomic model was built with the graphics program O (*51*). Density averaging and model building were iterated until the complete polypeptide chain and cofactors were fitted. The fitted model was refined by conventional positional refinement protocol in XPLOR (*52*).

	Native	Thimerosal-1	Thimerosal-2	K_2OsCl_6	K_2PtCl_4
Resolution (Å)	2.3	3.0	3.0	3.0	3.0
Reflections					
Measured (N)	76311	37497	47090	52337	33848
Unique (N)	33086	15473	17201	16404	13786
R_{merge}* (%)	4.31	4.98	4.89	4.83	4.60
Completeness [$\|F\| > 2\sigma$] (%)	78.3	81.1	90.0	85.7	72.0
Soaking					
Conc. (mM)		1.0	0.2	(2%)	0.5
Time (hours)		72	26	55.5	11
$\|\Delta F\|/\|F\|$†		0.16	0.12	0.09	0.14
Binding sites (PhP)‡		5	2	2	2
		1.6	1.1	0.6	0.6
Figure of merit	0.49				

*$R_{merge} = \Sigma_{hkl}[\Sigma_i|I_{hkl,i} - \langle I_{hkl}\rangle|]/\Sigma_{hkl,i}\langle I_{hkl}\rangle$, where $I_{hkl,i}$ is the intensity of an individual measurement of the reflection with Miller indices h, k, and l, and $\langle I_{hkl}\rangle$ is the mean intensity of that reflection. †$|\Delta F|/|F|$ represents the relative isomorphous difference between the native and the derivative structure factor amplitudes and is calculated using the expression $\Sigma_{hkl}(||F_{hkl,P}| - |F_{hkl,PH}||)/\Sigma_{hkl}|F_{hkl,P}|$, where $|F_{hkl,P}|$ and $|F_{hkl,PH}|$ refer to the measured structure factor amplitudes of the native and the derivative. ‡PhP represents the phasing power of a derivative and is the ratio of the root mean square (rms) amplitude of the heavy-atom structure factor to the rms lack of closure.

quent attempts to rebuild the polypeptide chain. As a representative example of the crystallographic data, Fig. 1 shows the refined FAD model together with its electron density.

Overall structure. The polypeptide chain of DNA photolyase is folded into five β strands, 20 α helices, five short 3_{10} helices, and connecting segments. Overall views of the model and the positions of secondary structure elements in the amino acid sequence are shown in Fig. 2. These structure elements combine to form two major domains, an α/β domain (residues 1 to 131) and a helical domain (residues 204 to 469). A long interdomain loop (residues 132 to 203) wraps around the α/β domain and connects it to the helical domain.

The α/β domain adopts a fold typical for dinucleotide binding domains (*20*): A β sheet with five parallel strands (β1 to β5) is covered by two α helices on each side (α1, α3, and α4, α5, respectively) (Fig. 3A); the short helix α2 continues from strand β2. Residues Asn^{108} and Glu^{109} in helix α5 interact with the MTHF cofactor.

The helical domain includes helices α7 to α20 and resembles a slab with the approximate dimensions 60 Å by 40 Å by 25 Å (Fig. 3B). Its most prominent feature is a flat face with a hole in the center leading to the FAD binding site. The helices of the domain can be grouped into two clusters with the FAD binding site in between. Cluster I comprises α7 to α11; α6 from the interdomain loop also adds to this cluster and contributes to the flat domain face. Cluster I contains all residues that interact with the phosphate oxygens of the FAD.

Cluster II consists of α12 to α20; α13 to α16 form a four-helix bundle within this cluster. Residues within this bundle have polar interactions with the riboflavin part of the FAD. The apoenzyme does not bind substrate (*21*). This suggests that binding of FAD stabilizes the structure by keeping the two helix clusters together; the clusters probably have to move apart for binding of FAD by the apoenzyme. The polypeptide chain connecting clusters I and II contains two 3_{10} helices and contributes part of the MTHF binding pocket.

The contact between the α/β domain and the helical domain in photolyase buries 1121 $Å^2$ of accessible surface area of the α/β domain, and 1051 $Å^2$ of accessible surface area of the helical domain. Carbon atoms contribute 55 percent, and polar atoms (nitrogen and oxygen) contribute 45 percent of the contact surface; these fractions are typical for surfaces of water-soluble proteins (*22*). Of the polar atoms, 16 percent in α/β domain and 19 percent in helical domain belong to charged residues. Bound water molecules fill gaps between the domains. Thus, the interaction between the two domains of photolyase is predominantly polar with 16 H bonds or charge interactions. In contrast, only four H bonds or charge inter-

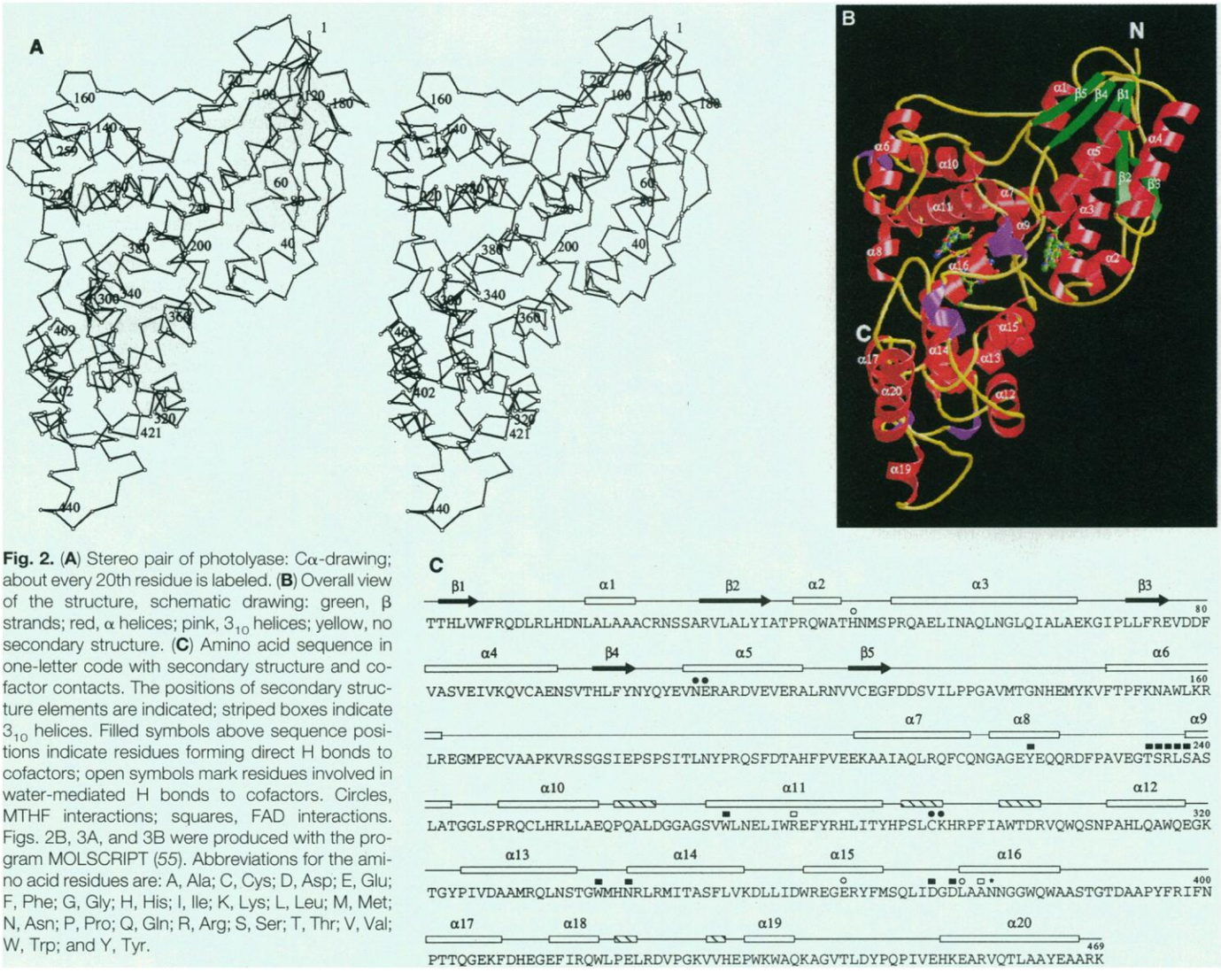

Fig. 2. (**A**) Stereo pair of photolyase: Cα-drawing; about every 20th residue is labeled. (**B**) Overall view of the structure, schematic drawing: green, β strands; red, α helices; pink, 3_{10} helices; yellow, no secondary structure. (**C**) Amino acid sequence in one-letter code with secondary structure and cofactor contacts. The positions of secondary structure elements are indicated; striped boxes indicate 3_{10} helices. Filled symbols above sequence positions indicate residues forming direct H bonds to cofactors; open symbols mark residues involved in water-mediated H bonds to cofactors. Circles, MTHF interactions; squares, FAD interactions. Figs. 2B, 3A, and 3B were produced with the program MOLSCRIPT (*55*). Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

actions between protein residues link the two helix clusters in the helical domain.

We searched for structural similarities between photolyase and other proteins using the program WHATIF (*23*) and the protein structure database distributed with it. We found the best matches between the α/β domain and domains with dinucleotide binding fold in other proteins—for example, glycinamide ribonucleotide transformylase [Protein Data Bank (PDB) code 1grc], apo liver alcohol dehydrogenase (PDB code 8adh), and aspartate carbamoyltransferase (PDB code 8atc) (*24*). In view of the wide distribution of the dinucleotide binding fold (*20*), this is not surprising. The other region of photolyase for which several matches appeared consists of residues 341 to 385 with α14, α15, and α16 within cluster II of the helical domain. Among the proteins containing segments with similar folds are arthropodan hemocyanin (PDB code 1hc6), engrailed homeodomain (PDB code 1hdd), citrate synthase (PDB code 2cts), cytochrome c551 (PDB code 451c), and biotin operon repressor biotin holoenzyme synthetase complex (PDB code 1bib) (*24*). In view of the short length of the matched chain segments and of the wide variety of functions of the matched proteins, we do not believe that these structural similarities indicate evolutionary relationships.

Cofactors. The cofactor MTHF binds to the enzyme in a shallow groove formed by residues 44, 108, 109, 292, 293, and 375 (Fig. 4A). The side chain of Glu^{109} is perfectly positioned to form two H bonds with the nitrogens N1 and N3 of the pteridine ring. The hydroxyl group of the pteridine is in H bonding distance both with the side chain oxygen and the peptide nitrogen of Asn^{108}. The nitrogen N8 of the pteridine is in H-bond distance to a water molecule. The nitrogen N10 of MTHF and the peptide carbonyl oxygen of Cys^{292} approach each other to 3.1 Å. This interaction is indicated in Fig. 4A and could be a charge interaction of the carbonyl with the positive charge on the five-membered ring of MTHF. The side chain of Lys^{293} forms a salt bridge with the α carboxylate of the Glu moiety of MTHF. The side chain of this Glu has only weak electron density because of the disorder in the crystal. The position of MTHF at the periphery of the enzyme and its solvent accessibility explain why this cofactor readily dissociates from the enzyme during purification and why an enzyme totally lacking MTHF retains its native conformation and normal substrate binding affinity (*25*). Of the residues that form the MTHF binding site, Glu^{109} is conserved in all folate class photolyases, except for the enzyme from *Bacillus firmus* in which MTHF has its absorption maximum at 410 nm instead of ~380 nm (*26*).

The catalytic cofactor FAD is bound to photolyase in a U-shaped conformation with the isoalloxazine and adenine rings in close proximity. This conformation is different from the FAD conformation in all 13 FAD binding proteins deposited in the Protein Data Bank (*27*). The FAD connects helix clusters I and II in the helical domain (Fig. 4B). Residues from helix cluster I bind the diphosphate moiety of FAD. Six of eight polar interactions between the phosphate oxygens and the protein are formed by residues 234 to 238 (Thr-Ser-Arg-Leu-Ser) near the NH_2-terminal end of α9; the other two polar interactions involve the side chains of Tyr^{222} in α8 and of Trp^{271} in α11. In addition, Trp^{338} from helix cluster II also interacts with one of the phosphates. The remaining interactions between residues from helix cluster II and FAD involve the riboflavin moiety. The side chain oxygens of Asn^{341} are in H-bonding distance to the 2′ hydroxyl group, and the side chain oxygens of Asn^{378} are in H-bonding distance to the nitrogen N5 of the isoalloxazine ring. The peptide nitrogen of Asp^{374} forms an H bond to one of the carbonyl groups of the isoalloxazine ring. Parallel to the isoalloxazine ring and in van der Waals contact to it, residues Arg^{344} and Asp^{372} form a salt bridge. Several water molecules participate in the H-bonding network around the FAD (Fig. 4B).

In addition to these interactions of the FAD with its surroundings in the protein, there are also polar interactions within the FAD. The 3′ hydroxyl group of the adenine moiety forms an H bond with one of the phosphate oxygens, and the 3′ hydroxyl group of the riboflavin moiety is in H-bonding distance to the nitrogen N1 of the isoalloxazine ring. The latter interaction may contribute to the stabilization of the active $FADH^-$ form of the enzyme. The edges of the isoalloxazine ring and of the adenine ring are accessible to solvent through an opening in the surface of the helical domain.

Most of the residues whose side chains interact with FAD are conserved in eight known sequences of photolyases from microorganisms [for a sequence alignment, see (*26*, *28*)]. These are Tyr^{222}, Thr^{234}, Ser^{235}, Asn^{341}, Arg^{344}, Asp^{372}, and Asn^{378} (Asp in *Halobacterium halobium* photolyase). They represent only a small subset of the conserved residues in the FAD binding domain.

Functional implications. Several functional aspects of photolyase have implications for a wide range of biological phenomena: The enzyme binds to DNA containing Pyr<>Pyr with high specificity and affinity in an essentially sequence-independent manner (*29*), it absorbs light, transfers energy between two chromophores, and carries out catalysis by a radical mechanism. We comment on these various features in relation to the crystal structure.

1) The first feature deals with energy transfer. Ordinarily, although photolyase

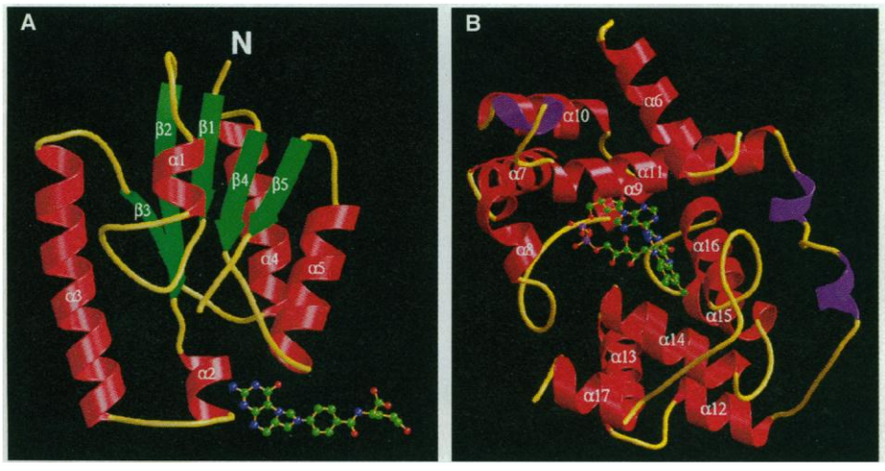

Fig. 3. Structural domains of DNA photolyase; colors as in Fig. 2B. (**A**) α/β domain with MTHF; (**B**) helical domain with FAD.

Table 2. Result of refinement.

Resolution (Å)	10–2.3
R_{work}*	0.172
R_{free}†	0.246
Number of reflections [$\|F\| > 2\sigma(\|F\|)$]	32619
Number of atoms used in refinement	8083
Number of atoms included in non-crystallographic symmetry restraints	7302
rms coordinate error from SIGMAA (*53*) (Å)	0.27
rms deviation of bond lengths (Å)	0.012
rms deviation of bond angles (degrees)	1.85

*$R_{work} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$, where $|F_{obs}|$ and $|F_{calc}|$ are observed and calculated structure factor amplitudes, respectively; summation over the reflections used in refinement. †R_{free} is equivalent to R_{work} except that the summation is over reflections (5 percent of the total) set aside for an unbiased test of the progress of refinement (*54*).

depleted of MTHF can carry out repair by direct photoexcitation of $FADH^-$, about 80 percent of the photons that initiate catalysis in the holoenzyme are absorbed by MTHF because of its higher extinction coefficient compared to $FADH^-$ (*21*). The excitation energy is transferred from MTHF to $FADH^-$ in *E. coli* photolyase with a time constant of about 200 ps and 62 percent efficiency (*30*) by the radiationless resonant mechanism described by the Förster theory (*31*). On the basis of the time constant and assuming random orientation of the chromophores, a distance of 21 Å was calculated between MTHF and $FADH^-$ (*30*) and, within experimental error, the same distance was calculated between MTHF and $FADH^\bullet$ (*30*). The distance actually observed in the atomic model described here is considerably shorter—that is, 16.8 Å between the centers of mass of the atoms connected by conjugated double bonds of MTHF and of the isoalloxazine ring. The discrepancy exists because the transition dipole moments of MTHF and the isoalloxazine ring of FAD are perpendicular to each other, and the transition dipole moment of FAD is also perpendicular to the vector between the cofactors; this brings the orientation factor κ^2 in the Förster formula (*31*) close to zero rather than the value 2/3 which was used in estimating the distance from the spectroscopic data.

We conclude that energy transfer between the chromophores in *E. coli* photolyase is not a rate-limiting step for DNA repair, and that there is no evolutionary pressure to optimize it. Energy transfer in photolyases of the deazaflavin class occurs with a time constant of 50 ps and 100 percent efficiency (*32*); FAD and deazaflavin in these enzymes must have relative orientations more favorable for energy transfer.

2) The second characteristic feature of photolyase is specific binding to UV damaged DNA. Photolyase binds to its substrate in any sequence context with comparable affinities (*29*). Because the enzyme specifically recognizes Pyr<>Pyr both in the dinucleotide form and with about 10^5-times higher affinity in a polydeoxynucleotide (*10*), both Pyr<>Pyr and the surrounding DNA backbone must contribute to binding. The structure of substrate-free DNA photolyase that we report here, in combination with the available functional data, supports the following model for substrate binding. DNA containing a Pyr<>Pyr binds to the flat surface of the helical domain with the phosphate backbone of one strand following a trace of positive electrostatic potential that runs across this surface (Fig. 5A). The trace includes the arginine residues 226, 342, and 397 (Fig. 5B). Photolyase contacts four to five phosphates in the damaged strand around the dimer (*33*) and mutations in each of these arginine residues makes the affinity 10^2- to 10^3-times lower without affecting the quantum yield of repair (*34*). The arginines are located near the rim of the hole in the flat surface which leads to the FAD binding site. This hole has the right dimension to allow a thymine dimer to approach the FAD to van der Waals contact distance. The residues lining the hole are hydrophobic on one side and polar on the other (Fig. 5B); this asymmetry fits well with the asymmetric polarity of a thymine dimer, in which the cyclobutane ring is hydrophobic and the opposite edges of the thymine bases have nitrogens and oxygens capable of forming H bonds. If bound in the hole, a thymine dimer would have polar interactions with the adenine ring of FAD, and with residues Asn^{273}, Glu^{274}, and Asn^{341}; it would be in van der Waals contact with the isoalloxazine ring of FAD, and with residues Trp^{277}, Tyr^{281}, and Trp^{384}. This mode of binding would allow electron transfer between the FAD and the substrate over a short distance and with high efficiency.

The only way the Pyr<>Pyr can bind in the fashion described is for the Pyr<>Pyr

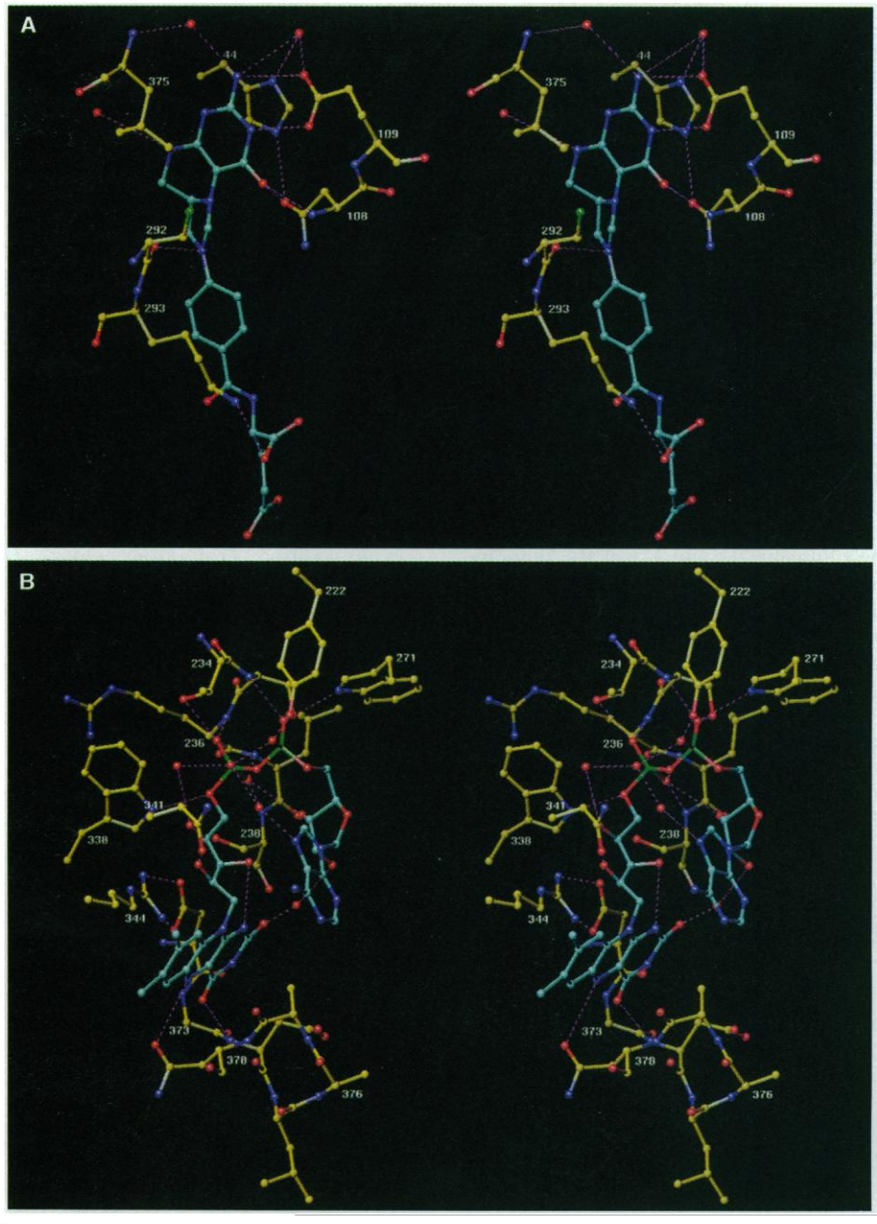

Fig. 4. Cofactor binding in DNA photolyase. (**A**) MTHF binding; (**B**) FAD binding. Possible H bonds are indicated in purple. Atom colors are yellow, carbons in protein; light blue, carbons in cofactors; blue, nitrogens; red, oxygens; green, phosphorus (A) or sulfur (B). Selected residue numbers are indicated in white.

to flip out from the helix into the hole. Although some controversy exists about the conformation of Pyr<>Pyr containing DNA (*35*), there is a consensus that the hydrogen bonds of the two bases that make up the dimer are substantially weakened and that the helix structure is perturbed on the 3′ side of the dimer (*36*). Weakening of these H bonds could facilitate the flipping out of the dimer on binding to DNA photolyase. In this proposed binding mode the intradimer phosphate remains outside of the hole, consistent with the finding that prior ethylation of this phosphate does not interfere with binding (*33*). Flipping out of a base from a DNA duplex was first observed in the crystal structure of the complex of Hha I DNA cytosine-5-methyltransferase with a covalent reaction intermediate of its substrate (*37*). The structure of uracil-DNA glycosylase also strongly suggested flipping of the target base out of the DNA duplex as an important part of substrate binding by the enzyme (*38*).

Yeast photolyase binds to the 1,2-d(GpG) intrastrand crosslinks of cisplatin with an affinity 50 times lower than that for Pyr<>Pyr and there is indirect evidence that this photolyase-cisplatin DNA complex may have a different structure from that of the photolyase-Pyr<>Pyr DNA complex (*39*). We docked the crystal structure of cis-Pt(NH3)$_2$[d(GpG)] (*40*) to the atomic model of photolyase and found that the hole cannot accommodate the adduct; this observation is in agreement both with the lower binding affinity and with a different binding mode.

Pyr<>Pyr DNA glycosylases are the only other class of enzymes that bind to Pyr<>Pyr with high affinity and specificity. T4 endonuclease V (T4 endo V) is an example of these enzymes and its crystal structure has been determined (*41*). This small (16 kD) protein binds to Pyr<>Pyr in DNA and cleaves the glycosylic bond of the 5′ residue and then the intradimer phosphodiester bond. Enzyme-substrate crystals for this enzyme are not available, but site-specific mutagenesis studies (*42*) have identified a potential binding site involving a Trp-Tyr-Lys-Tyr-Tyr sequence. Similar to photolyase, T4 endo V has a groove-like depression on its surface which contains a cluster of basic amino acids (Fig. 5A) (*41*); the backbone of the DNA helix could potentially bind in this groove. In both enzymes, aromatic residues may serve to position the substrate correctly vis-à-vis the catalytic center by interacting with Pyr<>Pyr (*43*). However, the spatial arrangement of these residues is rather different in the two enzymes and there is no evidence for a requirement for base flipping during the binding of T4 endo V. Thus, it appears that the two enzymes bind to Pyr<>Pyr by different mechanisms. Indeed, while photolyase binds to Pyr<>Pyr in single- and double-stranded DNAs with comparable affinities, T4 endo V binds to Pyr<>Pyr in double-stranded DNA with an affinity that is at least 1000-times higher (*44*).

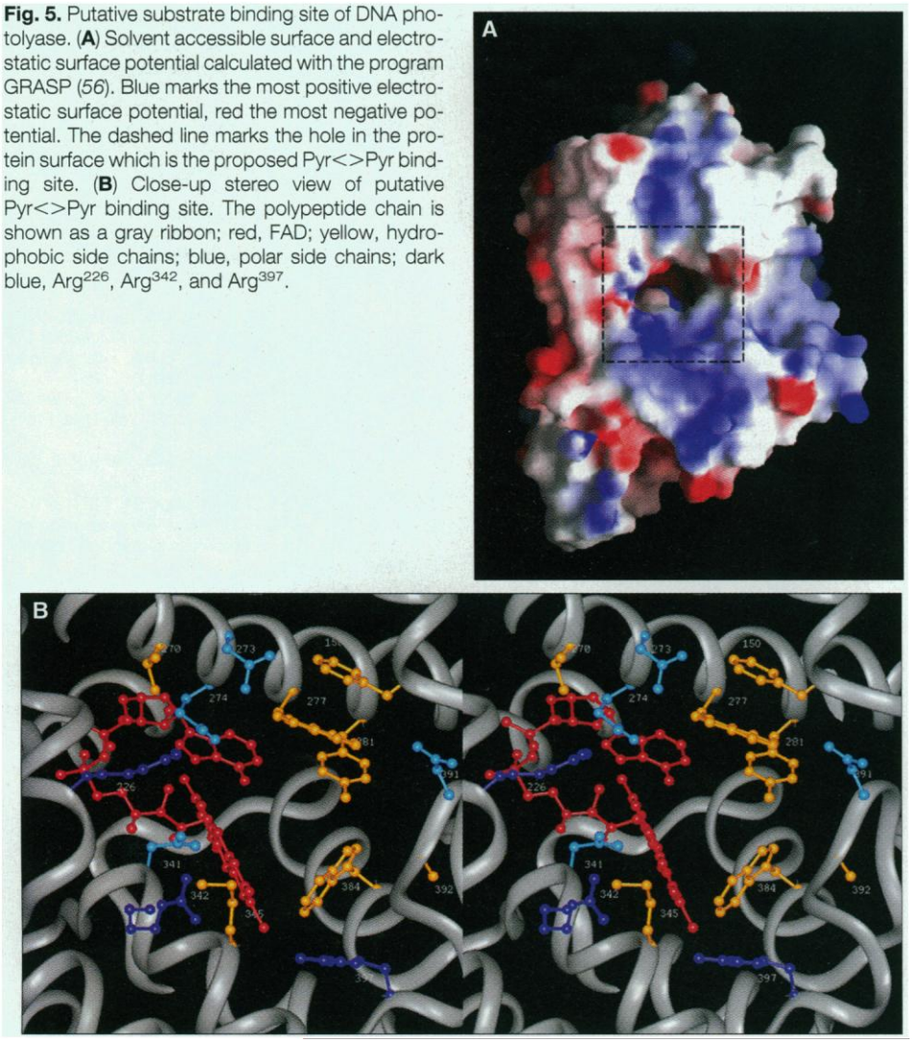

Fig. 5. Putative substrate binding site of DNA photolyase. (**A**) Solvent accessible surface and electrostatic surface potential calculated with the program GRASP (*56*). Blue marks the most positive electrostatic surface potential, red the most negative potential. The dashed line marks the hole in the protein surface which is the proposed Pyr<>Pyr binding site. (**B**) Close-up stereo view of putative Pyr<>Pyr binding site. The polypeptide chain is shown as a gray ribbon; red, FAD; yellow, hydrophobic side chains; blue, polar side chains; dark blue, Arg^{226}, Arg^{342}, and Arg^{397}.

3) Photolyase is one of the few enzymes that catalyzes reactions by radical mechanism. Photolyase initiates splitting of Pyr<>Pyr by donating an electron to generate a Pyr<>Pyr radical anion (*30*). The proposed mode of binding would allow electron transfer over a short distance. After splitting of Pyr<>Pyr, the electron returns back to the FADH$^{\bullet}$ to regenerate the catalytically active FADH^{-}. The tight fit of Pyr<>Pyr in the hole would prevent leakage of charge to the solvent. Thus, both electron transfer and recovery of the cofactor are expected to occur with high efficiency consistent with the experimental data indicating a quantum yield of about 1.0 for repair by FADH^{-} in the excited singlet state (*10*). Upon breakage of the cyclobutane ring, the distances between the two pyrimidines would increase from a carbon-carbon bond length to van der Waals distance; this could force the repaired substrate out of the hole and thus lead to rapid dissociation from the enzyme.

The other two electron transfer reactions occurring with photolyase involve two tryptophan residues. The excited singlet state of Trp^{277} splits Pyr<>Pyr by electron transfer (*13*) with high efficiency. The proximity of Trp^{277} to the proposed substrate binding site is, therefore, another satisfying feature of the model. Finally, photoreduction of excited FADH$^{\bullet}$ requires intact Trp^{306}, implying electron transfer from this residue (*12*). Trp^{306} is more than 13 Å away from FAD and is by no means the tryptophan closest to FAD; it is near the protein surface and thus accessible to reducing agents to restore its neutral state after electron transfer. Several possible pathways for electrons from Trp^{306} to FAD can be proposed on the basis of the structure. One pathway would

involve the tryptophans 359 and 382, which lie between donor and acceptor; this pathway would include three gaps. However, site-directed mutagenesis studies do not support this model (*45*). Another possible pathway with only two gaps would involve the polypeptide chain between residues 358 and 366, and the side chain of Phe^{366}.

The most important questions yet to be answered concern details of substrate binding and of the catalytic process. We hope to address these with co-crystals of enzyme and substrate, which will open the way to study both the stationary and the dynamic structure of the complex. The co-crystals may be an ideal system to follow the structural changes during the time course of the enzymatic reaction by time-resolved crystallography (*46*) with the Laue technique and short exposure times. The advantage of photolyase over other systems where time-resolved crystallography has been conducted is that light is a necessary ingredient of catalysis by photolyase, so that the reaction and the accompanying conformational change can be induced throughout the crystal by a short (picosecond to nanosecond) light flash.

REFERENCES AND NOTES

1. W. Harm, *Biological effects of ultraviolet radiation* (Cambridge Univ. Press, Cambridge, 1980).
2. R. Stolarski *et al.*, *Science* **256**, 342 (1992).
3. C. S. Rupert, S. H. Goodgal, R. M. Herriott, *J. Gen. Physiol.* **41**, 451 (1958).
4. S.-T. Kim and A. Sancar, *Photochem. Photobiol.* **57**, 895 (1993); A. Sancar, *Biochemistry* **33**, 2 (1994).
5. S. Yasuhira and A. Yasui, *J. Biol. Chem.* **267**, 25644 (1992); T. Todo *et al.*, *Mutat. Res. DNA Repair* **315**, 213 (1994); T. Kato Jr. *et al.*, *Nucleic Acids Res.* **22**, 4119 (1994).
6. The active form of FAD in photolyase is the two-electron reduced form, $FADH^-$ (*4*). During purification it is oxidized to the catalytically inactive blue neutral radical form, $FADH^•$. In referring to this cofactor, these designations are used only when the redox status of the cofactor is relevant to the point being made. Otherwise, we use FAD, which encompasses all three redox states of flavin.
7. M. Ahmad and A. R. Cashmore, *Nature* **366**, 162 (1993).
8. K. Malhotra, S.-T. Kim, A. Batschauer, L. Dawut, A. Sancar, *Biochemistry* **34**, 6892 (1995); S.-T. Kim, K. Malhotra, A. Sancar, unpublished observation.
9. G. B. Sancar, F. W. Smith, A. Sancar, *Biochemistry* **24**, 1849 (1985); I. Husain and A. Sancar, *Nucleic Acids Res.* **15**, 1109 (1987).
10. S.-T. Kim and A. Sancar, *Biochemistry* **30**, 8623 (1991).
11. M. S. Jorns, G. B. Sancar, A. Sancar, *ibid.* **24**, 1856 (1985); G. Payne, P. F. Heelis, B. Rohrs, A. Sancar, *ibid.* **26**, 7121 (1987); M. S. Jorns, B. Wang, S. P. Jordan, L. P. Chanderkar, *ibid.* **29**, 552 (1990).
12. P. F. Heelis, T. Okamura, A. Sancar, *ibid.* **29**, 5694 (1990); Y. F. Li, P. F. Heelis, A. Sancar, *ibid.* **30**, 6322 (1991); S.-T. Kim, A. Sancar, C. Essenmacher, G. T. Babcock, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8023 (1993).
13. Y. F. Li and A. Sancar, *Biochemistry* **29**, 5698 (1990); S.-T. Kim, Y. F. Li, A. Sancar, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 900 (1992).
14. J. Deisenhofer and H. Michel, *Annu. Rev. Cell Biol.* **7**, 1 (1991).
15. H.-W. Park, A. Sancar, J. Deisenhofer, *J. Mol. Biol.* **231**, 1122 (1993).
16. K. Miki *et al.*, *ibid.* **233**, 167 (1993).
17. J. L. Johnson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2046 (1988); S. Hamm-Alvarez, A. Sancar, K. V. Rajagopalan, *J. Biol. Chem.* **264**, 9649 (1989).
18. Calculated with the program OOPS, G. J. Kleywegt and T. A. Jones, unpublished. Available through anonymous FTP from Uppsala, Sweden.
19. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993).
20. G. E. Schulz, *Curr. Opin. Struct. Biol.* **2**, 61 (1992).
21. G. Payne, M. Wills, C. Walsh, A. Sancar, *Biochemistry* **29**, 5706 (1990); G. Payne and A. Sancar, *ibid.*, p. 7715.
22. J. Janin and C. Chothia, *J. Biol. Chem.* **265**, 16027 (1990).
23. G. Vriend and C. Sander, *Proteins* **11**, 52 (1991).
24. The references to the proteins are made by the Protein Data Bank tracking code numbers. The authors are (i) Glycinamide ribonucleotide transformylase [PDB code 1grc, root mean square deviation of α carbons (rmsd) 3.1 Å, 89 residues], P. Chen *et al.*, *J. Mol. Biol.* **227**, 283 (1992). (ii) Apo liver alcohol dehydrogenase (PDB code 8adh, rmsd 3.7 Å, 99 residues), F. Colonna-cesari *et al.*, *J. Biol. Chem.* **261**, 15273 (1986). (iii) Aspartate carbamoyltransferase (PDB code 8atc, rmsd 3.6 Å, 92 residues), H. Ke, W. N. Lipscomb, Y. Cho, R. B. Honzatko, *J. Mol. Biol.* **204**, 725 (1988). (iv) Arthropodan hemocyanin (PDB code 1hc6, rmsd 3.8 Å, 45 residues), A. Volbeda and W. G. J. Hol, *J. Mol. Biol.* **209**, 249 (1989). (v) Engrailed homeodomain (PDB cde 1hdd, rmsd 3.2 Å, 45 residues), C. R. Kissinger, B. Liu, E. Martin-blanco, T. B. Kornberg, C. O. Pabo, *Cell* **63**, 579 (1990). (vi) Citrate synthase (PDB code 2cts, rmsd 3.4 Å, 41 residues), S. Remington, G. Wiegand, R. Huber, *J. Mol. Biol.* **158**, 111 (1982). (vii) Cytochrome c551 (PDB code 451c, rmsd 3.6 Å, 43 residues), Y. Matsuura, T. Takano, R. E. Dickerson, *J. Mol. Biol.* **156**, 389 (1982). (viii) Biotin operon repressor biotin holoenzyme synthetase complex (PDB code 1bib, rmsd 3.6 Å, 41 residues), K. P. Wilson, I. M. Shewchuk, R. G. Brennan, A. J. Otsuka, B. W. Matthews, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9257 (1992).
25. S. Hamm-Alverez, A. Sancar, K. V. Rajagopalan, *J. Biol. Chem.* **265**, 18656 (1990).
26. K. Malhotra, S.-T. Kim, A. Sancar, *Biochemistry* **33**, 8712 (1994).
27. F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977).
28. G. B. Sancar, *Mutat. Res. DNA Repair* **236**, 147 (1990); H. Yajima, H. Inoue, A. Oikawa, A. Yasui, *Nucleic Acids Res.* **19**, 5359 (1991).
29. D. L. Svoboda, C. A. Smith, J.-S. A. Taylor, A. Sancar, *J. Biol. Chem.* **268**, 10694 (1993).
30. S.-T. Kim *et al.*, *Biochemistry* **30**, 11262 (1991); T. Okamura *et al.*, *J. Am. Chem. Soc.* **113**, 3143 (1991); S.-T. Kim *et al.*, *ibid.* **116**, 3115 (1994).
31. Th. Förster, *Disc. Farad. Soc.* **27**, 7 (1959).
32. K. Malhotra, S.-T. Kim, C. Walsh, A. Sancar, *J. Biol. Chem.* **267**, 15406 (1992); S.-T. Kim, P. F. Heelis, A. Sancar, *Biochemistry* **31**, 11244 (1992).
33. I. Husain, G. B. Sancar, S. R. Holbrook, A. Sancar, *J. Biol. Chem.* **262**, 13188 (1987).
34. M. E. Baer and G. B. Sancar, *ibid.* **268**, 16717 (1993); K. Malhotra, S.-T. Kim, Y. F. Li, A. Sancar, unpublished data.
35. S. N. Rao, J. W. Keepers, P. Kollman, *Nucleic Acids Res.* **12**, 4789 (1984); D. A. Pearlman, S. R. Holbrook, D. H. Pirkle, S.-H. Kim, *Science* **227**, 1304 (1985); I. Husain, J. Griffith, A. Sancar, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2558 (1988); C.-I. Wang and J.-S. Taylor, *ibid.* **88**, 9072 (1991).
36. F. N. Hayes, D. L. Williams, R. L. Ratlift, A. J. Varghese, C. S. Rupert, *J. Am. Chem. Soc.* **93**, 4940 (1971); J. Kemmink, R. Boelens, T. Koning, G. A. Van der Marel, J. H. Van Boom, *Nucleic Acids Res.* **15**, 4645 (1987); J. S. Taylor, D. S. Garrett, I. R. Brockie, D. L. Svoboda, J. Telser, *Biochemistry* **29**, 8858 (1990).
37. S. Klimasauskas, S. Kumar, R. J. Roberts, X. Cheng, *Cell* **76**, 357 (1994).
38. R. Savva, K. McAuley-Hecht, T. Brown, L. Pearl, *Nature* **373**, 487 (1995).
39. M. E. Fox, B. J. Feldman, G. Chu, *Mol. Cell. Biol.* **14**, 8071 (1994); J. T. Reardon and A. Sancar, unpublished results.
40. S. E. Sherman, D. Gibson, A. H.-J. Wang, S. J. Lippard, *Science* **230**, 412 (1985).
41. K. Morikawa *et al.*, *ibid.* **256**, 523 (1992).
42. A. Recinos and R. S. Lloyd, *Biochemistry* **27**, 1832 (1988); D. G. Stump and R. S. Lloyd, *ibid.*, p. 1839; M. Ishida, Y. Kanamori, N. Hori, T. Inaoka, E. Ohtsuka, *ibid.* **29**, 3817 (1990).
43. B. J. Lee, H. Sakashita, T. Ohkubo, M. Ikehara, T. Doi *et al.*, *ibid.* **33**, 57 (1994).
44. P. C. Seawell, T. J. Simon, A. K. Ganesan, *ibid.* **19**, 1685 (1980); T. Inaoka, M. Ishida, E. Ohtsuka, *J. Biol. Chem.* **264**, 2609 (1989).
45. K. Malhotra and A. Sancar, unpublished observation.
46. K. Moffat, *Annu. Rev. Biophys. Biophys. Chem.* **18**, 309 (1989).
47. W. Kabsch, *J. Appl. Crystallogr.* **21**, 916 (1988).
48. W. Steigemann, *PROTEIN: A Program System for the Crystal Structure Analysis of Proteins* (Max-Planck-Institut für Biochemie, Martinsried bei Munich, Germany, 1992).
49. B.-C. Wang, *Methods Enzymol.* **115**, 90 (1985).
50. Z. Otwinowski, in *Isomorphous Replacement and Anomalous Scattering*, W. Wolf, P. R. Evans, A. G. W. Leslie, Eds. (Science and Engineering Research Council, Warrington, UK, 1991), p. 80.
51. T. A. Jones, M. Bergdoll, M. Kjeldgaard, in *Crystallographic and Modeling Methods in Molecular Design*, C. E. Bugg and S. E. Ealick, Eds. (Springer Verlag, New York, 1990), p. 189.
52. A. T. Brünger, *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, 1992).
53. R. J. Read, *Acta Crystallogr.* **A42**, 140 (1986).
54. A. T. Brünger, *Nature* **355**, 472 (1992).
55. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).
56. A. Nicholls, *GRASP: Graphical Representation and Analysis of Surface Properties* (Columbia University, New York, 1992).

57. We thank R. G. Kurumbail, J. E. Hunt, G. B. Sancar, K. Malhotra, Y. F. Li, and L. Dawut for helpful discussions and advice, D. Diggs and K. Watkins for technical assistance, and D. B. Staber for help with the manuscript. Supported by NIH grant GM31082 (A.S.), and Welch Foundation grant I-1185 (J.D.). Coordinates have been deposited with the Brookhaven Protein Data Bank.

27 March 1995; accepted 22 May 1995