J Mol Evol (1997) 45:535–548

JOURNAL OF MOLECULAR EVOLUTION

Molecular Evolution of the Photolyase–Blue-Light Photoreceptor Family

Satoru Kanai,[1] Reiko Kikuno,[1] Hiroyuki Toh,[1] Haruko Ryo,[2] Takeshi Todo[3]

[1] Department of Bioinfomatics, Biomolecular Engineering Research Institute, 6-2-3, Furuedai, Suita, Osaka, 565 Japan
[2] Department of Radiation Biology, Faculty of Medicine, Osaka University, 2-2 Yamada-oka, Suita, Osaka, 565 Japan
[3] Radiation Biology Center, Kyoto University, Yoshida-konoecho, Sakyo-ku, Kyoto 606-01 Japan

Received: 23 October 1996 / Accepted: 1 April 1997

Abstract. The photolyase–blue-light photoreceptor family is composed of cyclobutane pyrimidine dimer (CPD) photolyases, (6-4) photolyases, and blue-light photoreceptors. CPD photolyase and (6-4) photolyase are involved in photoreactivation for CPD and (6-4) photoproducts, respectively. CPD photolyase is classified into two subclasses, class I and II, based on amino acid sequence similarity. Blue-light photoreceptors are essential light detectors for the early development of plants. The amino acid sequence of the receptor is similar to those of the photolyases, although the receptor does not show the activity of photoreactivation. To investigate the functional divergence of the family, the amino acid sequences of the proteins were aligned. The alignment suggested that the recognition mechanisms of the cofactors and the substrate of class I CPD photolyases (class I photolyases) are different from those of class II CPD photolyases (class II photolyases). We reconstructed the phylogenetic trees based on the alignment by the NJ method and the ML method. The phylogenetic analysis suggested that the ancestral gene of the family had encoded CPD photolyase and that the gene duplication of the ancestral proteins had occurred at least eight times before the divergence between eubacteria and eukaryotes.

Key words: Evolution — Protein — DNA photolyase — Blue-light photoreceptor — Phylogenetic analysis — Photoreactivation

Introduction

UV light induces DNA damage such as CPDs and (6-4) photoproducts. Corresponding to the damage, there are two types of DNA photolyases. One of them specifically exerts the activity on CPD whereas another repairs only (6-4) photoproducts (reviewed by Sancar 1996). The former is called CPD photolyase and the latter (6-4) photolyase. The amino acid sequences of the two enzymes are similar to each other (Todo et al. 1996), although the chemical structures of their substrates are quite different (Brash 1988; Mitchell and Nairn 1989; Taylor 1995).

CPD photolyases are divided into two subclasses, class I and II, based on the sequence similarity, which is here referred to as class I photolyases and class II photolyases, respectively. CPD photolyases contain two types of cofactors. One of them is two-electron–reduced FAD, which acts as the active-site cofactor. Another cofactor is MTHF, or 8-HDF, which acts as the photoantenna. Based on the kind of photoantenna, class I photolyases are further divided into two subgroups, which are here denoted as MTHF-type photolyase and 8-HDF–type photolyase, respectively (Yasui et al. 1994). The photoreactivation mechanism of MTHF-type or 8-HDF–type photolyases has been investigated in great detail: (1)

Abbreviations: UV, ultraviolet; CPD, cyclobutane pyrimidine dimer; (6-4) photoproducts, pyrimidine (6-4) pyrimidone photoproducts; FAD, flavin-adenine dinucleotide; MTHF, 5,10-methenyltetrahydrofolate; 8-HDF, 8-hydroxy-5-deazaflavin; NJ, neighbor-joining; ML, maximum likelihood; AIC, Akaike information criterion

Correspondence to: S. Kanai; e-mail skanai@beri.co.jp

the photoantenna (MTHF or 8-HDF) absorbs a blue-light photon; (2) the excitation energy is transferred from the photoantenna to active-site cofactor (FAD) by dipole–dipole interaction; (3) the excited FADH radical donates an electron to a CPD, splitting the cyclobutane ring; and (4) finally, the electron is transferred back to FADH, accompanied by the generation of the two canonical bases (Hearst 1995; Kim et al. 1991, 1992). The efficiency of energy transfer from 8-HDF to FAD is 98% in eubacterium *Anacystis nidulans* 8-HDF–type photolyase, while that from MTHF to FAD is 63% in an *Escherichia coli* MTHF-type photolyase (Kim et al. 1992). Therefore, it was suggested that 8-HDF–type photolyases are able to eliminate CPD more rapidly than MTHF-type photolyases (Malhotra et al. 1992). The crystal structure of MTHF-type photolyase from *E. coli* was determined (Park et al. 1995). The structure consists of two domains, an N-terminal α/β domain and a C-terminal helical domain, connected by a loop of 72 residues. MTHF binds in a cleft between the two domains, whereas FAD is included in the helical domain and is accessible through a hole in the surface of the domain. CPD binding sites have not been identified, but the hole in the helical domain has shape and polarity suitable for CPD binding. Therefore, the hole is considered to correspond with putative CPD binding site (Park et al. 1995). In contrast, investigation of the photoreactivation mechanism of class II photolyases has not advanced. However, the cofactors of a few class II enzymes have been elucidated. The class II photolyase from *Drosophila melanogaster* has FAD and MTHF (Kim et al. 1996a), while FAD and 8-HDF are included in the enzyme from an archaebacterium, *Methanobacterium thermoautotrophicum* (Kiener et al. 1989). Yasui et al. (1994) reported that the enzyme from a eukaryote, *Potorous tridactylis*, includes FAD, although the second cofactor has not been detected in the enzyme.

(6-4) photolyases have been identified recently (Todo et al. 1993), but the presence of cofactors has not been confirmed and the precise photoreactivation mechanism has not been well elucidated. The fluorescence and action spectra of *Xenopus laevis* (6-4) photolyase suggested that the cofactor composition of the enzyme is different from those of CPD photolyases (Kim et al. 1996b). On the other hand, it was suggested that the enzyme includes FAD as the cofactor of the *X. laevis* enzyme (Todo et al. 1997). Further studies are required for the identification of the cofactors for (6-4) photolyases. Two (6-4) photolyase homologues have been found in *Homo sapiens* (Todo et al. 1996; Hsu et al. 1996). However, the homologues, which contain FAD and MTHF as the cofactors, show neither (6-4) nor CPD photolyase activity (Hsu et al. 1996). The function of the homologues has been unknown, although Hsu et al. (1996) suggested that the homologues may act as blue-light photoreceptors.

Blue-light photoreceptors are essential light detectors for the early development of plants (Ahmad and Cashmore 1993). In addition, they mediate phototropism, hypocotyl elongation, stomatal opening, and expression of specific genes (Kaufman 1993; Short and Birggs 1994; Liscum and Hangarter 1994). The nucleotide sequences of the cDNAs for the receptors were recently determined (Ahmad and Cashmore 1993; Batschauer 1993; Malhotra et al. 1995). The deduced amino acid sequences were unexpectedly similar to those of photolyases (Ahmad and Cashmore 1993; Batschauer 1993; Malhotra et al. 1995). The receptors from the eukaryotes *Sinapis alba* and *Arabidopsis thaliana* overexpressed in *E. coli* also contain FAD and MTHF as the cofactors, although the receptors do not show photoreactivation activity (Malhotra et al. 1995).

Table 1 summarizes the classification of the family. The functions carried by the members of the protein family are highly divergent. However, the evolutionary process of the functional divergence of the family has not been sufficiently studied. Yasui et al. (1994) reported the molecular phylogeny of CPD photolyases by the NJ method (Saitou and Nei 1987), although they used only 13 sequences, which did not include (6-4) photolyases or blue-light photoreceptors. Now, 22 sets of sequence data of this family are available. We compared the amino acid sequences to reveal the evolutionary relationship among the members of the photolyase–blue-light photoreceptor family. The evolutionary history of the family will be discussed based on the reconstructed phylogenetic trees.

Materials and Methods

The 22 amino acid sequences used in this study are listed in Table 1. A multiple sequence alignment was constructed with the program Clustal W 1.4 (Higgins et al. 1991; Thompson et al. 1994), which was then modified by visual inspection with an alignment editor, BIORESEARCH/AE 3.0 (Fujitsu Ltd. 1995). To examine the conservation of aligned sites, amino acid residues were classified into six physicochemically similar groups based on the criteria of Schwartz and Dayhoff (1978): (1) positively charged group: Lys, Arg, His; (2) negatively charged group: Asp, Asn, Glu, Gln; (3) small hydrophilic group: Gly, Pro, Ser, Thr; (4) hydrophobic group: Leu, Ile, Met, Val; (5) aromatic group: Phe, Tyr, Trp; (6) Cys. The last category consists of only Cys residue.

All the sites containing gaps were excluded from the alignment for the following phylogenetic analyses. At first, we performed the NJ inference to obtain an overview of the evolutionary relationship of the photolyase–blue-light photoreceptor family. The genetic distance for each aligned pair was calculated with the program PROTDIST in PHYLIP 3.5c (Felsenstein 1993, 1996), where the amino acid substitution model, PAM001 (Dayhoff et al. 1978), was used. Then, an unrooted NJ tree was constructed with NEIGHBOR in PHYLIP 3.5c. The statistical significance of each cluster in the tree was evaluated with 1,000 iterations of bootstrap resamplings and tree reconstructions (Felsenstein 1985) using PROTDIST, NEIGHBOR, SEQBOOT, and CONSENSE in PHYLIP 3.5c.

To further examine the phylogenetic relationship obtained by the NJ method, we employed the ML method (Felsenstein 1981; Kishino et

Table 1. List of proteins used in the current analyses[a]

Type of protein (abbreviation)	Species	Kingdom	Abbreviated name	Accession No. (DB)/References	Characterized cofactors
MTHF type class I CPD photolyase (C1MPHR)	*Bacillus firmus*	Eubacterium	PHR_BACF	Q04449 (sp)	FAD, MTHF (Malhotra et al. 1994)
	Escherichia coli	Eubacterium	PHR_ECOL	P00914 (sp)	FAD, MTHF (Johnson et al. 1988)
	Salmonella typhimurium	Eubacterium	PHR_SALT	P25078 (sp)	FAD, MTHF (Li and Sancar 1991)
	Saccharomyses cerevisiae	Eukaryote	PHR_SCER	P05066 (sp)	FAD, MTHF (Sancar et al. 1987; Johnson et al. 1988)
	Neurospora crassa	Eukaryote	PHR_NEUC	P27526 (sp)	FAD, MTHF (Eker et al. 1994)
8-HDF type class I CPD photolyase (C18PHR)	*Anacystis nidulans*	Eubacterium	PHR_ANAN	P05327 (sp)	FAD, 8-HDF (Eker et al. 1990)
	Synechocystis sp.	Eubacterium	PHR_SYNE	U51943 (gb)	
	Streptomyces griseus	Eubacterium	PHR_STRG	P12768 (sp)	FAD, 8-HDF (Eker et al. 1981)
	Halobacterium halobium	Archaebacterium	PHR_HALH	P20377 (sp)	FAD, 8-HDF (Iwasa et al. 1988)
(6-4) photolyase (64PHR)	*Drosophila melanogaster*	Eukaryote	64_DROM	D83701 (gb)	
(6-4) photolyase homologue (64PHR)	*Homo sapiens*	Eukaryote	64_HUMA1	D83702 (gb)	FAD, MTHF (Hsu et al. 1996)
	Homo sapiens	Eukaryote	64_HUMA2	Hsu et al. 1996	FAD, MTHF (Hsu et al. 1996)
Blue-light photoreceptor (BLR)	*Chlamydomonas reinhardtii*	Eukaryote	BLR_CHRE	S57795 (pir)	
	Sinapis alba	Eukaryote	BLR_SIAL	P40115 (sp)	FAD, MTHF (Malhotra et al. 1995)
	Arabidopsis thaliana	Eukaryote	BLR_ARTH	S66907 (gb)	FAD, MTHF (Malhotra et al. 1995)
Class II CPD photolyase (C2PHR)	*Myxococcus xanthus*	Eubacterium	PHR_MXU	U44437 (gb)	
	Methanobacterium thermoautotrophicum	Archaebacterium	PHR_METH	P12769 (sp)	FAD, 8-HDF (Kiener et al. 1989)
	Drosophila melanogaster	Eukaryote	PHR_DM	S52047 (pir)	FAD, MTHF (Kim et al. 1996a)
	Carassius auratus	Eukaryote	PHR_CA	A45098 (pir)	
	Oryzias latipes	Eukaryote	PHR_OLAP	D26022 (gb)	
	Monodelphis domestica	Eukaryote	PHR_OPPO	D31902 (gb)	
	Potorous tridactylis	Eukaryote	PHR_PTRI	D26020 (gb)	FAD (Yasui et al. 1994)

[a] ''DB'' denotes the symbol of database : sp, SWISS-PROT release 33.0; gb, GenBank release 95.0; pir, PIR protein sequence database release 49.0

al. 1990). However, it was difficult to apply the ML analysis to all the 22 sequences, since a large number of possible trees are generated. We, therefore, selected the proteins derived from nodes with less than 50% bootstrap probabilities in the NJ analysis, as well as the representative proteins of each subgroup. Then, nine amino acid sequences were selected, which were subjected to the ML analysis with PROTML in MOLPHY 2.3b3 (Adachi and Hasegawa 1996). In the analysis, AIC of each possible tree was calculated based on the four different amino acid substitution models—the JTT (Jones et al. 1992), Dayhoff (Dayhoff et al. 1978), JTT-F, and Dayhoff-F models (Adachi and Hasegawa 1995). AIC is defined as $-2 \times$ (log-likelihood) $+ 2 \times$ (number of free parameters) (Akaike 1974). The tree with minimal AIC was considered to be the most appropriate tree. The statistical significance of each cluster in the ML tree was evaluated by the bootstrap analysis with 1,000 iterations.

The obtained phylogenetic trees were drawn with TREETOOL 2.0.2 (Maciukenas and McCaughey 1994).

Results and Discussion

Functional Implications Derived From Multiple Sequence Alignment

Figure 1 shows a multiple alignment of the photolyase–blue-light photoreceptor family. Since the N- and C-terminal regions of the members were highly diverged in residue and length, they were excluded from the alignment. The crystal structure of class I photolyase from *E. coli* has been determined recently (Park et al. 1995), which revealed MTHF and FAD binding sites of the enzyme. In addition, the CPD binding sites were putatively assigned to a hole of the crystal structure. These sites play important roles for the function of the enzyme. Therefore, it is expected that the comparison of the residues at these sites would provide much information about the functional divergence of the protein family. Table 2 shows the summary of the comparative study, although the alignment in Fig. 1 also includes the same information. As shown in Table 2, the protein family was divided into five groups based on the functional difference as follows: group A, MTHF-type photolyases; group B, 8-HDF–type photolyases; group C, (6-4) photolyases; group D, blue-light photoreceptors; and group E, class II photolyases. For simplicity, (6-4) photolyase homologues were included in group C, although their functions have remained unknown. The functional classification roughly corresponded with the phylogenetic clustering as described below. In the current approach, class I photolyases were classified into two groups, based on the available second cofactors (groups A and B). In contrast, group E, or the group of class II CPD photolyases, was not further classified, because identification of the second cofactors of the photolyase has not advanced (see Table 1).

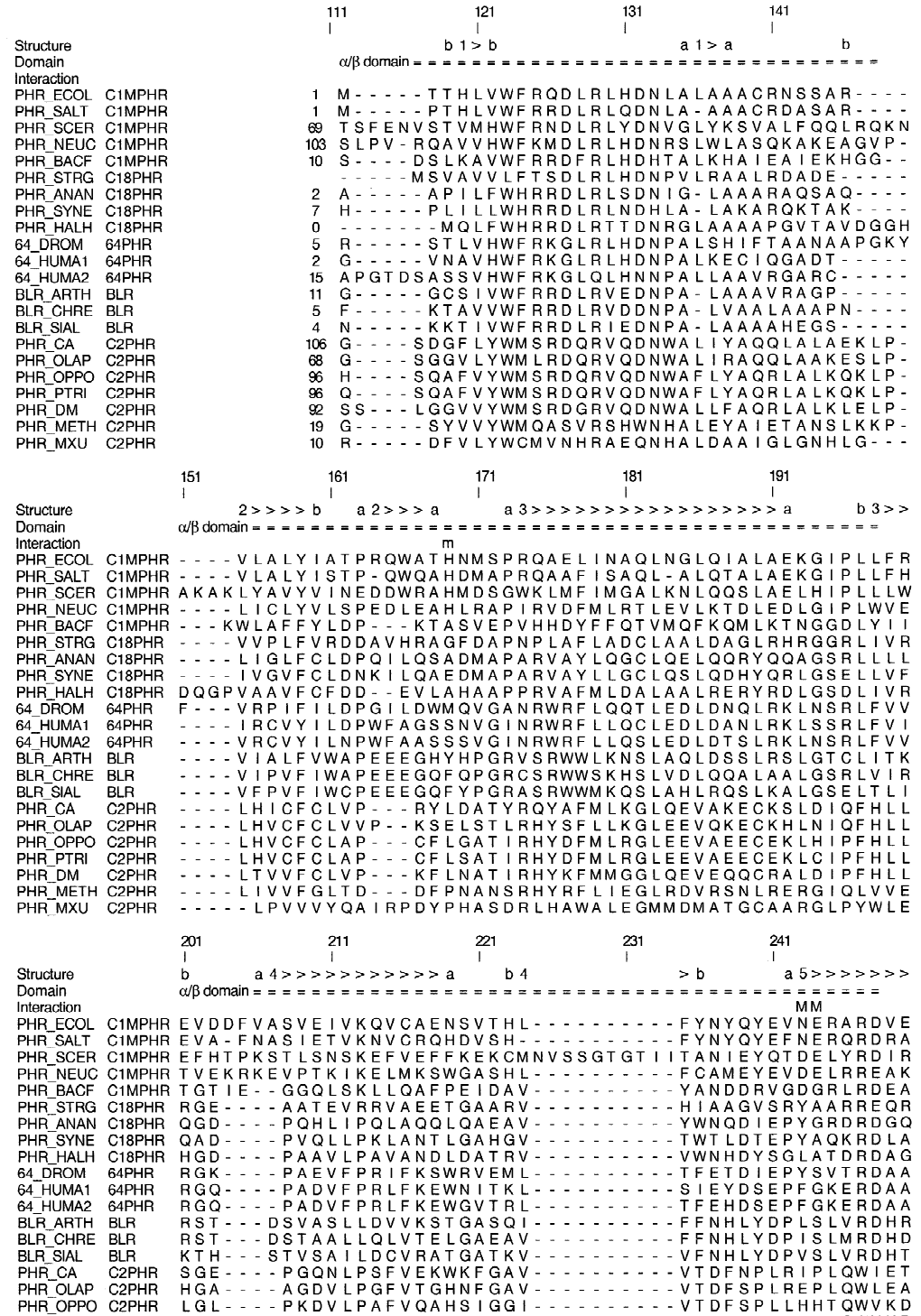

Fig. 1. Alignment of the amino acid sequences of DNA photolyases and blue-light photoreceptors. The *residue number* of beginning amino acid of each sequence is shown at the left side of the sequence. The information about the secondary structures, domains, interactions with cofactors, binding with CPD, and electron transfer derived from the crystal structure of DNA photolyase from *E. coli* (Park et al. 1995) is also shown in the alignment. The line ''*Structure*'' shows the secondary structures (a: α helix, b: β strands, 3: 3_{10} helix). For example, in the case of ''an>>>>a,'' where ''an'' means the beginning of the *n*-th α helix, ''>'' indicates the continuation of the structure. The last ''*a*'' denotes the end of the structure. The line ''*Domain*'' indicates the regions of the two domains and an interdomain loop in the DNA photolyase. The line ''*Interaction*'' denotes the residue sites which interact with the cofactors. The characters ''*f*'' and ''*F*'' indicate residues which interact with FAD, while ''*m*'' and ''*M*'' indicate residues which interact with MTHF. The *capital letter* means the direct H bond to the cofactors. The *lowercase letter* indicates the water-mediated H bond to the cofactors. ''*C*'' shows putative CPD binding site. Abbreviated name of each sequence is shown in Table 1.

```
                          251                 261                 271                 281                 291
                          |                   |                   |                   |                   |
Structure                 > > > a                   b 5 > > b                                              a 6 > > > > > >
Domain                    α/β domain = = = = = = = = = = = = = = = = > Inter domain loop = = = = = = = = = = = = = = = = = = = =
Interaction                                                                                                  C
PHR_ECOL    C1MPHR        V E R A L R - - - - - - N V V C E G F D D S V I L P P G - A V M T G N - H E M Y K V F T P F K N A W
PHR_SALT    C1MPHR        V E K L P - - - - - - - S V I C E G F D D S V I L A P G - - A M T G N - H E M Y K V F T P F K N A W
PHR_SCER    C1MPHR        L L E N E D - - - - - H R L Q L K Y Y H D S C I V A P G - L I T T D R - G T N Y S V F T P W Y K K W
PHR_NEUC    C1MPHR        L V K L L A E G E K G E K M A A D V V H D T C V V M P G - A L Q S G S - G G Q Y A V Y S P W F R A W
PHR_BACF    C1MPHR        A E H F L A K - - - - Q S I P F Y T F E D A Y L T E P D - Q V L K K D - G T P Y K V F T P Y Y K A W
PHR_STRG    C18PHR        I R E A L A D S - - - G R E L H V H D A V V T A L A P G - R V V P T G G K D H F A V F T P Y F R R W
PHR_ANAN    C18PHR        V A A A L K T - - - - A G I R A V Q L W D Q L L H S P D - Q I L S G S - G N P Y S V Y G P F W K N W
PHR_SYNE    C18PHR        V A Q A L R E - - - - R G L A I A T E W D Q L M H H P G - E V L T Q A - G S P Y T V Y T P F W K N W
PHR_HALH    C18PHR        V R D A L D A - - - - A G V A H A Q F H D A V H H R P G - E I R T N A - G D P Y S V Y T Y F W R K W
64_DROM     64PHR         V Q K L A K A - - - - E G V R V E T H C S H T I Y N P E - L V K A K N L G K A P I T Y Q K F L G I V
64_HUMA1    64PHR         I K K L A T E - - - - A G V E V I V R I S H T L Y D L D - K I I E L N G G Q P P L T Y K R F Q T L I
64_HUMA2    64PHR         I M K M T K E - - - - A G V E V V T E N S H T L Y D L D - R I I E L N G Q K P P L T Y K R F Q A I I
BLR_ARTH    BLR           A K D V L T A - - - - Q G I A V R S F N A D L L Y E P W - E V T D E L - G R P F S M F A A F W E R C
BLR_CHRE    BLR           C K R G L T A - - - - A G V A H R T F N G D M L Y E P W - D V L D P N - K Q P Y S T F D D F W N S V
BLR_SIAL    BLR           V K E K L V E - - - - R G I S V Q S Y N G D L C M S P G - R Y T V K R - A N L L L V L I L T G K K C
PHR_CA      C2PHR         V K K H L P - - - - - A D V P F I Q V D A H N V V - P C W E A S G K L E Y G A R T I R G K I T K L L
PHR_OLAP    C2PHR         V K K G L P - - - - - E D I P F I Q V D A H N I V - P C W V A S P K L E Y S A R T I R G K I T N L L
PHR_OPPO    C2PHR         V Q D G L P - - - - - K Q V P F V Q V D A H N I V - P C W I A S D K Q E Y G A R T I R H K I H D R L
PHR_PTRI    C2PHR         V Q D A L P - - - - - R Q V P F V Q V D A H N I V - P C W V A S D K Q E Y G A R T I R H K I H D R L
PHR_DM      C2PHR         V G K A L P - - - - - K S V P L V Q V D A H N V V - P L W V A S D K Q E Y A A R T I R N K I N S K L
PHR_METH    C2PHR         A A G A L H - - - - - - - I P L T Q V E S N V I V - P V E T A S D K E E Y S A G T F K P K I K R H L
PHR_MXU     C2PHR         A A K A L D - - - - - - - V P L F A V D A S C V V - P M Q R I A T R Q - I G A Y T L R P K L K K L W
```

```
                          301                 311                 321                 331                 341
                          |                   |                   |                   |                   |
Structure                 > > >      > a
Domain                    Inter domain loop = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = = =
Interaction
PHR_ECOL    C1MPHR        L K R - - L R E G - M P E C V A A P K V R S S G S - - - - I E P S P S - - - - I - T L N Y P R Q S -
PHR_SALT    C1MPHR        L R - - - L K E D - I P P C V P A P K I R V S G L - - - - S T P L T P - - - - V - S L N Y P Q Q E -
PHR_SCER    C1MPHR        V L Y V N N Y K K S T S E I C H L H I I E P L K Y N E T F E L K P F Q - - - - - - Y S L P D E F L
PHR_NEUC    C1MPHR        I K H - - I E E N - - P E C L E I Y E K P G P N P P G T K E K H E N L F A C S I P E A P E G K R L R
PHR_BACF    C1MPHR        A K - - - - E R K R T P A V I K R D V L L G S V H K G T A P - - - - - - - - - - - D R E A E T L F N
PHR_STRG    C18PHR        E A E G V R G T Q T A P R T V R V P D G V A S D P L P - - - - - - - - - - - - - D R D C V E N L S
PHR_ANAN    C18PHR        Q A - - - Q P K P - T P V A T P T E L V D L S P E Q L T A I A P L L L - - - - S - E L P T L K Q L G
PHR_SYNE    C18PHR        S Q - - - L P K T - S P V P T P K D L Q G L T P A E K E K L A P L E P - - - - L - A I P Q L A D L G
PHR_HALH    C18PHR        Q - - - - D R E K - N P P A P E P E P - - A D L A A D T A L A D T S P - - - - - - - L P S V Q E L G
64_DROM     64PHR         - - - - - - E Q L K V P K V L G V P E K L K K M P T P P K D E V E Q K D S A A Y - D C P T I K Q L V
64_HUMA1    64PHR         S K - - - M E P L E I P V E T I T S E V I E K C T T P L S D D H D E K - - - - Y - G V P S L E E L G
64_HUMA2    64PHR         S R - - - M E L P K K P V G L V T S R Q M E S C R A E I Q E N H D E T - - - - Y - G V P S L E E L G
BLR_ARTH    BLR           L S M - - P Y D P E S P L L P P K K I I S G - - - - D V S K C V A D P - - - - L - V F E D D S E K G
BLR_CHRE    BLR           R A M - - P V P P P F P V S A P A S M P A V P A A V P S M T V A E V D - - - - W - F F T P E Q E - A
BLR_SIAL    BLR           L D M - - S V E S - V V L P P W R L M P L S A A E T V W A C S V E E - - - - L - G L E N E A E K P
PHR_CA      C2PHR         P - - - - E F L T E I P L V D T H P H - S A S R A A E P V D W E E V L - - - - S - S L E V E R S V G
PHR_OLAP    C2PHR         S - - - - E F L T D F P L V D K H P F - S A T K T A K A V D W D K T L - - - - A - S L K V D R T V G
PHR_OPPO    C2PHR         P - - - - H F L T E F P P V I C H P Y - P S N I Q A E P V D W N A C R - - - - A - G L Q V D R S V K
PHR_PTRI    C2PHR         P - - - - H F L T E F P P V I C H P Y - T S N V Q A E P V D W N G C R - - - - A - G L Q V D R S V K
PHR_DM      C2PHR         G - - - - E Y L S E F P P V V R H P H G T G C K N V N T V D W S A A Y - - - - A - S L Q C D M E V D
PHR_METH    C2PHR         K R - - - - - - F M V P L R M R T L K M D S L D L E P G P E F E D A V R - - - - - D F R A P E D L E
PHR_MXU     C2PHR         P - - - E Y L D R A V P N R A V K A A A A G R K L E P D F A T S D A R - - - - - - E S R E S L D A F
```

```
                          351                 361                 371                 381                 391
                          |                   |                   |                   |                   |
Structure                                            a 7 > > > > > > > > > > > a   a     8 > > > > > a
Domain                    Inter domain loop = = = = = = = = > Helical domain = = = = = = = = = = = = = = = = = = = = = = = =
Interaction                                                                                   F       C               F F
PHR_ECOL    C1MPHR        - - - - F D T - A H F P V E E K A A I A Q L R Q F C Q N G - - A G E Y E Q Q R D F P A - - V E G T S
PHR_SALT    C1MPHR        - - - - F D - - A L F P V E E N A V I A Q L R Q F C A Q - - - A D E Y A L R R D F P A - - V D G T S
PHR_SCER    C1MPHR        Q Y I P K S K W C L P D V S E E A A L S R L K D F L G T K - - S S K Y N N E K D M L Y - - L G G T S
PHR_NEUC    C1MPHR        D D E K A R Y H S L W P A G E H E A L K R L E K F C D E A - - I G K Y A E R R N I P A - - M Q G T S
PHR_BACF    C1MPHR        N L I K K C S Y D W S A I G E E H A I K R L Q M F T K K R - - L S G Y K A N R D F P S - - I T G T S
PHR_STRG    C18PHR        P - - - - - - - G L A R G G E E A G R K L V T S W L N G P - - M A D Y E D G H D D L A G - - D A T S
PHR_ANAN    C18PHR        F - - D W D G G F P V E P G E T A A I A R L Q E F C D R A - - I A D Y D P Q R N F P A - - E A G T S
PHR_SYNE    C18PHR        F - - I W D Q P L P L T P G E E A A E Q R L D W F V A H G - - L E E Y Q Q N R N F P A - - L D G T S
PHR_HALH    C18PHR        - - F A E P E A A V P D A G T A A A R S L L D A F R E S G D - I Y R Y E D R R D Y P H - - E E P T S
64_DROM     64PHR         K R P E E L G P N K F P G G E T E A L R R M E E S L K D E I W V A R F E K P N T A P N S L E P S T T
64_HUMA1    64PHR         F D T D G L S S A V W P G G E T E A L T R L E R H L E R K A W V A N F E R P R M N A N S L L A S P T
64_HUMA2    64PHR         F P T E G L G P A V W Q G G E T E A L A R L D K H L E R K A W V A N Y E R P R M N A N S L L A S P T
BLR_ARTH    BLR           S - - N A L L A R A W S P G W S N G D K A L T T F I N G P - - L L E Y S K N R R K A D - - S A T T S
BLR_CHRE    BLR           S - - S D Q L K F K W K P G V G G A I S E L E H F L A E R - - L T E F E H D R A K V D - - R D S T S
BLR_SIAL    BLR           S - - N A L L T R A W S P G W S N A D K I L N E F I E K Q - - L I D Y A K N S K K V V - - G N S T S
PHR_CA      C2PHR         - - - - E V - - D W A Q P G T S G M N M L E S F I D Q R - - L R L F A T H R N N P N - - Y D A L S
PHR_OLAP    C2PHR         - - - - E P - - K L A K P G T E A G L A M L E S F I D V R - - L K L F G T Q R N D P N - - A A A L S
PHR_OPPO    C2PHR         - - - - E V - - S W A K P G T A S G L T M L Q S F I S Q R - - L P Y F G S D R N N P N - - K D A L S
PHR_PTRI    C2PHR         - - - - E V - - S W A K P G T A S G L T M L Q S F I A E R - - L P Y F G S D R N N P N - - K D A L S
PHR_DM      C2PHR         - - - - E V - - Q W A K P G Y K A A C Q Q L Y E F C S R R - - L R H F N D K R N D P T - - A D A L S
PHR_METH    C2PHR         P - - - - - - - S V F R G G T S T A L S I F S E F L R E K - - L E C F E R Y R N D P V - - K N C L S
PHR_MXU     C2PHR         D F D H S V A P I Q E R G G R K A G L D A L Q A F V H Q R - - L E G Y D E G R N D P G - - L A R Q S
```

Fig. 1. Continued.

Table 2(a) shows the comparison of the MTHF binding residues. X-ray crystallographic analysis revealed that seven residues of MTHF-type photolyase from *E. coli* (PHR_ECOL in Table 2) are involved in the MTHF binding function of the enzyme. Surprisingly, five out of the seven residues were not conserved among the members of group A. Only Glu residue at alignment site 536 was invariant, and alignment site 243 was occupied by the residues of the negatively charged group. However, the remaining five sites included physicochemically different residues. Payne et al. (1990) reported that MTHF-type photolyase from *E. coli* shows photolyase activity,

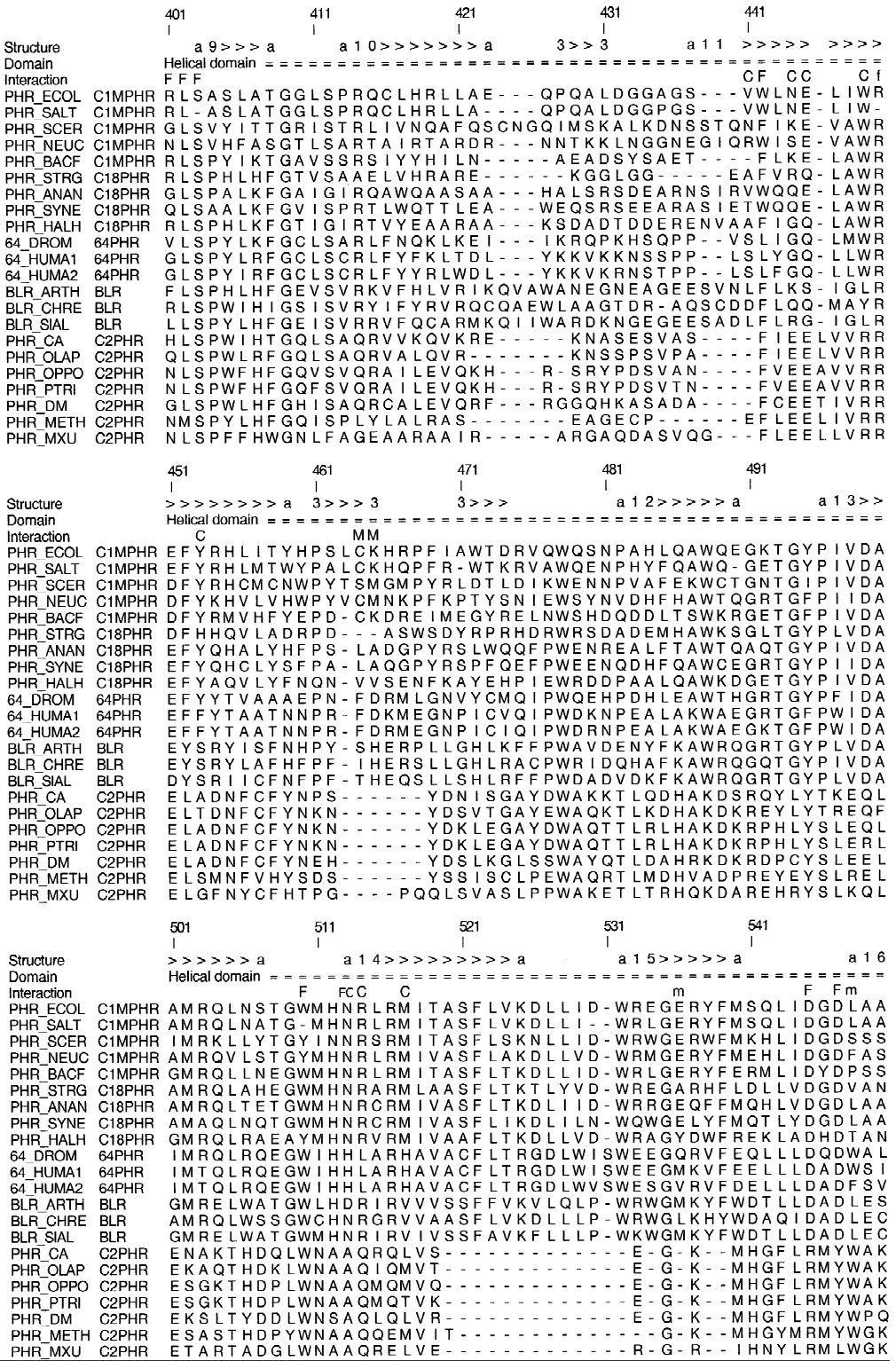

Fig. 1. Continued.

even without MTHF. The observation suggested that the second cofactor or photoantenna is not essential for the photolyase activity, which could explain the weak conservation of the MTHF binding sites among the members of group A. Blue-light photoreceptors utilize MTHF as the second cofactor as well. However, these sites of the members of group D were occupied by physicochemically different residues from those of PHR_ECOL, even at the alignment sites 243 and 536. The high variability of MTHF binding sites in the members of group and D could be explained by the weak functional constraint described above. Another possible interpretation of the

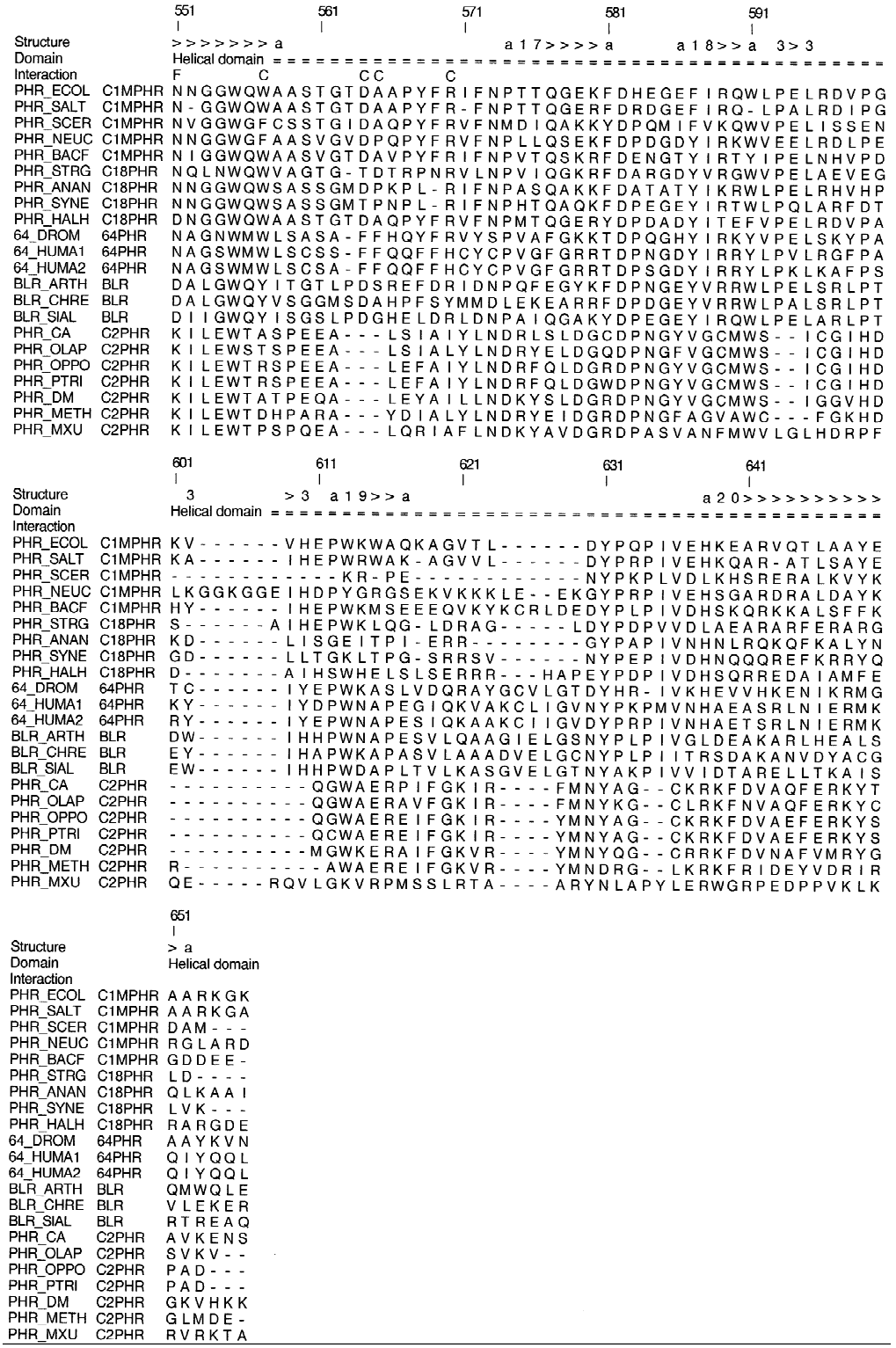

Fig. 1. Continued.

observation is that the members of group D may have a different MTHF binding mechanism than those of group A. It is difficult to definitively describe the problem at this stage, and further study is required. There were no data available to identify the 8-HDF binding sites of the members of group B. Therefore, we just examined the residues of the members of group B, which correspond to the seven MTHF binding sites. As expected, these sites were not conserved among the members of group B. In addition, the physicochemical characters of the amino acid residues at the five alignment sites, 169, 243, 244, 464 and 465, of the members of group B were quite

Table 2. Comparisons of residues of (a) MTHF, (b) FAD, and (c) CPD binding sites[a]

(a) MTHF binding sites

		Site	169	243	244	464	465	536	548
		Interaction	m	M	M	M	M	m	m
Group A	PHR_ECOL	C1MPHR	H	N	E	C	K	E	L
	PHR_SALT	C1MPHR	H	N	E	C	K	E	L
	PHR_SCER	C1MPHR	H	D	E	S	M	E	S
	PHR_NEUC	C1MPHR	H	D	E	C	M	E	F
	PHR_BACF	C1MPHR	A	D	G	C	K	E	P
Group B	PHR_STRG	C18PHR	G	Y	A	–	–	A	V
	PHR_ANAN	C18PHR	A	Y	G	L	A	E	L
	PHR_SYNE	C18PHR	E	Y	A	L	A	E	L
	PHR_HALH	C18PHR	A	L	A	V	V	Y	T
Group C	64_DROM	64PHR	M	Y	S	F	D	Q	W
	64_HUMA1	64PHR	S	F	G	F	D	M	W
	64_HUMA2	64PHR	S	F	G	F	D	V	F
Group D	BLR_ARTH	BLR	Y	L	S	S	H	M	L
	BLR_CHRE	BLR	F	I	S	I	H	L	L
	BLR_SIAL	BLR	F	V	S	T	H	M	L
Group E	PHR_CA	C2PHR	L	I	P	–	–	–	W
	PHR_OLAP	C2PHR	E	E	P	–	–	–	W
	PHR_OPPO	C2PHR	L	H	H	–	–	–	W
	PHR_PTRI	C2PHR	L	H	H	–	–	–	W
	PHR_DM	C2PHR	L	L	P	–	–	–	W
	PHR_MXU	C2PHR	P	I	I	–	–	–	W
	PHR_METH	C2PHR	P	I	Q	–	–	–	W

(b) FAD binding sites

		Site	385	399	400	401	402	403	442	450	510	513	545	547	550	551
		Interaction	F	F	F	F	F	F	F	f	F	F	F	F	f	F
Group A	PHR_ECOL	C1MPHR	Y	T	S	R	L	S	W	R	W	N	D	D	A	N
	PHR_SALT	C1MPHR	Y	T	S	R	L	–	W	–	–	N	D	D	A	N
	PHR_SCER	C1MPHR	Y	T	S	G	L	S	F	R	Y	N	D	D	S	N
	PHR_NEUC	C1MPHR	Y	T	S	N	L	S	W	R	Y	N	D	D	S	N
	PHR_BACF	C1MPHR	Y	T	S	R	L	S	F	R	W	N	D	D	S	N
Group B	PHR_STRG	C18PHR	Y	T	S	R	L	S	F	R	W	N	D	D	N	N
	PHR_ANAN	C18PHR	Y	T	S	G	L	S	W	R	W	N	D	D	A	N
	PHR_SYNE	C18PHR	Y	T	S	Q	L	S	W	R	W	N	D	D	A	N
	PHR_HALH	C18PHR	Y	T	S	R	L	S	F	R	Y	N	D	D	N	D
Group C	64_DROM	64PHR	F	T	T	V	L	S	L	R	W	H	D	D	L	N
	64_HUMA1	64PHR	F	P	T	G	L	S	L	R	W	H	D	D	I	N
	64_HUMA2	64PHR	Y	P	T	G	L	S	L	R	W	H	D	D	V	N
Group D	BLR_ARTH	BLR	Y	T	S	F	L	S	F	R	W	D	D	D	S	D
	BLR_CHRE	BLR	F	T	S	R	L	S	F	R	W	N	D	D	C	D
	BLR_SIAL	BLR	Y	T	S	L	L	S	F	R	W	N	D	D	C	D
Group E	PHR_CA	C2PHR	F	L	S	H	L	S	F	R	W	A	R	Y	K	K
	PHR_OLAP	C2PHR	F	L	S	Q	L	S	F	R	W	A	R	Y	K	K
	PHR_OPPO	C2PHR	F	L	S	N	L	S	F	R	W	A	R	Y	K	K
	PHR_PTRI	C2PHR	F	L	S	N	L	S	F	R	W	A	R	Y	K	K
	PHR_DM	C2PHR	F	L	S	G	L	S	F	R	W	A	R	Y	Q	K
	PHR_MXU	C2PHR	Y	Q	S	N	L	S	F	R	W	A	R	L	K	K
	PHR_METH	C2PHR	F	L	S	N	M	S	F	R	W	A	R	Y	K	K

different from those of group A, although the several residues at remaining two sites, 536 and 548, were similar or identical between the members of groups A and B. The residues of the seven sites of the members of group B seemed to be rather similar to those of group D, although the functional meaning of the similarity was not clear. Here, we refrain from discussion of the second cofactor binding sites of groups C and E, because the

Table 2. Continued

		(c) Putative CPD binding sites														
			Hydrophobic							Polor						
		Site	293	441	449	442	517	557	565	389	444	445	513	514	564	570
Group A	PHR_ECOL	C1MPHR	F	V	W	Y	M	W	A	R	N	E	N	R	D	R
	PHR_SALT	C1MPHR	F	V	W	Y	M	W	A	R	N	E	N	R	D	R
	PHR_SCER	C1MPHR	F	N	W	Y	M	F	A	K	K	E	N	R	D	R
	PHR_NEUC	C1MPHR	Y	R	W	Y	M	F	P	R	S	E	N	R	D	R
	PHR_BACF	C1MPHR	F	–	W	Y	M	W	A	R	K	E	N	R	D	R
Group B	PHR_STRG	C18PHR	F	A	W	H	M	W	D	H	R	Q	N	R	T	R
	PHR_ANAN	C18PHR	Y	V	W	Y	M	W	P	R	Q	E	N	R	D	R
	PHR_SYNE	C18PHR	Y	T	W	Y	M	W	P	R	Q	E	N	R	T	R
	PHR_HALH	C18PHR	Y	A	W	Y	M	W	A	R	G	Q	N	R	D	R
Group C	64_DROM	64PHR	Y	S	W	Y	H	W	F	N	G	Q	H	L	F	R
	64_HUMA1	64PHR	Y	S	W	F	H	W	F	R	G	Q	H	L	F	H
	64_HUMA2	64PHR	Y	S	W	F	H	W	F	R	G	Q	H	L	F	H
Group D	BLR_ARTH	BLR	F	L	L	S	V	Y	S	R	K	S	D	R	D	R
	BLR_CHRE	BLR	F	D	Y	S	V	Y	A	R	Q	Q	N	R	D	Y
	BLR_SIAL	BLR	L	L	L	S	V	Y	G	S	R	G	N	R	D	R
Group E	PHR_CA	C2PHR	R	–	R	A	L	A	–	R	E	E	A	Q	–	I
	PHR_OLAP	C2PHR	R	–	R	T	M	T	–	R	E	E	A	Q	–	L
	PHR_OPPO	C2PHR	R	–	R	A	M	R	–	R	E	E	A	Q	–	I
	PHR_PTRI	C2PHR	R	–	R	A	T	R	–	R	E	E	A	Q	–	I
	PHR_DM	C2PHR	R	–	R	A	L	A	–	R	E	E	A	Q	–	I
	PHR_MXU	C2PHR	R	–	R	G	L	P	–	R	E	E	A	Q	–	A
	PHR_METH	C2PHR	K	E	R	S	M	D	–	R	E	E	A	Q	–	L

[a] The information is derived from the alignment in Fig. 1. See legend of Fig. 1 for notation of the characters in the line ''*Interaction*''

second cofactors of (6-4) photolyases and the most of class II enzymes have not been identified.

FAD is considered to be included in all the members of the protein family, and 14 FAD binding amino acid residues have been found in the crystal structure of PHR_ECOL. Alignment sites 385, 400, and 402 were occupied by physicochemically similar residues, respectively. Alignment sites 403 and 450 were invariant except for the deletion in MTHF-type photolyase from *Salmonella typhimurium* (PHR_SALT). Alignment site 510 was occupied by aromatic residues, which was also deleted in PHR_SALT. Alignment site 442 was also occupied by aromatic residues, except for three members of group C. Thus, FAD binding sites seemed to be highly conserved, comparing to the case of MTHF binding sites. However, alignment sites 399, 513, 545, 547, and 551 were occupied by the residues with different physicochemical characters between group E and the other groups. As discussed below, group E was distantly related to the other members of the family. In addition, the members of group E included a long deletion in the C-terminal helical domain (alignment sites 520 to 539). Such a deletion was not found in the other members of the family. The different conservation pattern between class II photolyase and the others may reflect the change of FAD binding mechanism caused by the deletion in class II photolyase or the insertion in the others.

CPD binding sites were putatively assigned in the crystal structure of PHR_ECOL. The putative sites constitute a hole in the structure, to which CPD is supposed to be bound. One face of the hole consists of seven hydrophobic residues, while seven polar residues form another surface of the hole. Table 2(c) summarizes the comparison of the residues at the putative CPD binding sites. CPD is the substrate for the members of both group A and B. Alignment sites 449, 513, 514, 517, and 570 were invariant among the members of groups A and B. In addition, four sites, 293, 389, 445, and 557, were occupied by physicochemically similar residues, and two sites, 442 and 564, were nearly invariant. Site 565 was occupied by physicochemically similar residues, except for 8-HDF–type photolyase from *Streptomyces griseus* (PHR_STRG). The observation suggested that these sites are important for CPD binding activity of the members of groups A and B. The members of group D follow a conservation pattern similar to those of groups A and B, although the members of group D do not have photolyase activity. The members of group C share residues similar or identical to those of groups A, B, and D at the alignment sites 293, 389, 513, and 570. However, the amino acid residues at the alignment sites 444, 513, 514, 517, and 565 were conserved among the members of group C. These residues were physicochemically different from corresponding residues of the members of groups A, B, and D. (6-4) photolyase from *D. melanogaster* (64_DROM) uses (6-4) photoproduct as substrate instead

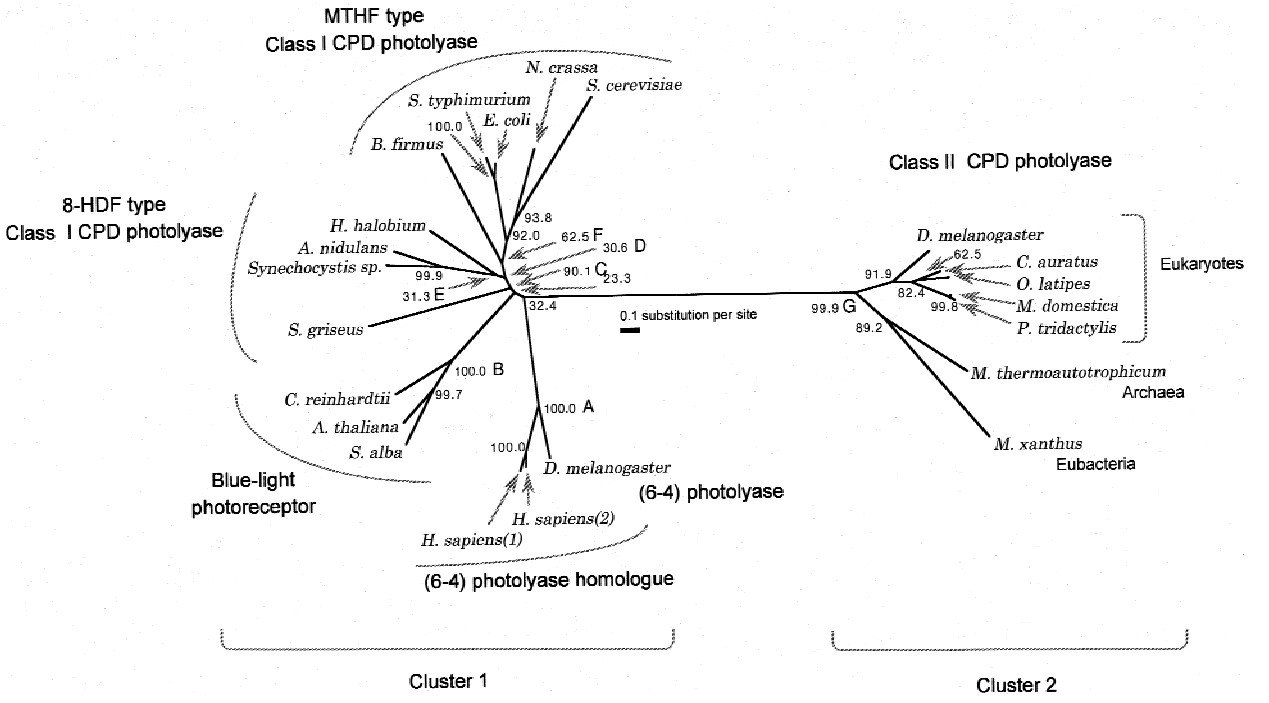

Fig. 2. An unrooted phylogenetic tree obtained by the neighbor-joining method. The *numbers* at the nodes indicate the bootstrap probabilities.

of CPD, and the difference in the conservation pattern may be correlated with the substrate specificity of the enzymes, although (6-4) photolyase homologues from *H. sapiens* (64_HUMA1 and 2) show neither CPD nor (6-4) photolyase activity (Hsu et al. 1996). On the other hand, the members of group E also utilize CPD as their substrate. However, they showed a conservation pattern similar to those of groups A and B at only the three sites 389, 445, and 517, and the remaining 11 sites diverged highly from those of groups A and B. As described above, group E was distantly related to the other members of the protein family, and the C-terminal helical domain of class II photolyase included a long deletion. The deletion was observed close to and within the putative CPD binding sites. The observation suggested that the CPD binding mechanism of class II photolyase is different from those of MTHF-type photolyase and 8-HDF–type photolyase, although they share CPD as their substrate.

Phylogeny of the Photolyase–Blue-Light Photoreceptor Family

Figure 2 shows an unrooted phylogenetic tree by the NJ method, which was divided into two clusters. The clusters are here referred to as clusters 1 and 2. Cluster 1 was composed of (6-4) photolyases, blue-light photoreceptors, and class I photolyases, while cluster 2 consisted of only class II photolyases. The phylogenetic clustering roughly corresponded to a functional classification of the family in the previous section, except for the 8-HDF–type photolyase.

In cluster 1, MTHF-type photolyases, (6-4) photolyases, and blue-light photoreceptors constituted three distinctive subclusters, which corresponded to groups A, B, and D in the previous section. On the other hand, 8-HDF–type photolyases did not form a single subcluster. That is, the functional classification of group B had no evolutionary meaning. As shown in Fig. 2, the roots for the subclusters (6-4) photolyase, blue-light photoreceptor, and MTHF-type photolyase were located at nodes A, B, and F, respectively. The tree topology suggested that present (6-4) photolyases have derived from an ancestral enzyme at node A, which carried (6-4) photolyase activity. Similarly, an ancestral protein corresponding to node B was considered to carry blue-light photoreceptor activity, from which current photoreceptors have evolved. The present MTHF-type photolyases originated from node F, which corresponded to an ancestral CPD photolyase with MTHF. This view is supported by the high bootstrap probabilities for nodes A, B, and F, which were 100.0%, 100.0%, and 62.5%, respectively. On the other hand, the tree topology suggested that there are two independent lineages to 8-HDF–type photolyase. One of them branched off at node C, including only one enzyme from *S. griseus*, whereas another lineage consisted of two enzymes from eubacteria: *Synecocystis* sp.; *A. nidulans;* and an enzyme from an archaebacterium, *Halobacterium halobium.* The branching point for the latter lineage was node D. Thus, the topology of the two lineages was not consistent with the pattern of ordinary species divergence and suggested two independent lineages to 8-HDF–type photolyase. The lineage consisting of the enzyme from *S. griseus* was statistically significant,

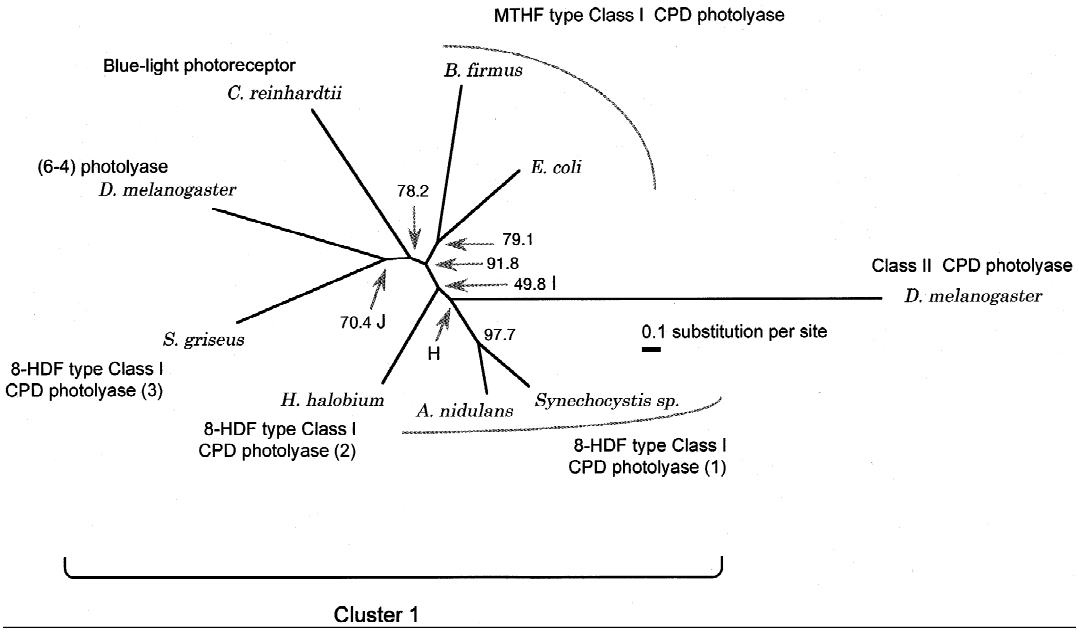

Fig. 3. An unrooted phylogenetic tree obtained by the maximum likelihood method. The *numbers* at the nodes indicate the bootstrap probabilities.

since the bootstrap probability for nodes C is high (90.1%). On the other hand, the bootstrap probability for nodes D is quite low (30.6%). In addition, the bootstrap probability for node E, which corresponds with the divergence point between eubacteria and archaebacterium in the former lineage, is only 31.3%. Thus, the topology about the divergence of 8-HDF–type photolyases from *Synecocystis* sp., *A. nidulans,* and *Halobacterium halobium* was not statistically significant.

In cluster 2, class II photolyase was divided into two subclusters at the node G. One of them includes the enzymes from eukaryotes, whereas another consists of the enzymes from an archaebacterium and a eubacterium. Therefore, node G is considered to correspond to the gene duplication of the enzymes before the divergence between eubacteria and archaebacteria. The tree topology in cluster 2 is considered to be significant because high bootstrap probabilities are evaluated for most of the nodes in the cluster.

As described above, the NJ tree included several nodes with low bootstrap probabilities. However, some of these nodes were related to positions in the tree crucial to a description of the evolutionary history of the family. To check the tree topologies at the nodes, ML analysis was applied to the aligned sequences. However, it requires enormous computational time to examine all of the 22 sequences. Therefore, we selected the following nine sequences, which were related to the crucial nodes or representatives of each subcluster: four 8-HDF–type photolyases from *A. nidulans, Synechocystis* sp., *S. griseus,* and *H. halobium;* MTHF-type photolyase from *E. coli* and *Bacillus firmus;* class II photolyase from *D. melanogaster;* (6-4) photolyase from *D. melanogaster;* and blue-light photoreceptor from *Chlamydomonas reinhardtii.* Only one enzyme from cluster 2 was included in the selection, since the topology of cluster 2 was shown to be statistically significant by NJ analysis. The enzyme was used as an outgroup of the remaining eight sequences from cluster 1.

We examined four amino acid substitutions models for the ML analysis, among which the JTT-F model produced a tree with minimal AIC. Figure 3 shows the unrooted tree. The differences between the minimal AIC and the AICs of the other possible trees were greater than 1.0, which suggested that the tree topology shown in Fig. 3 is statistically significant.

In the ML tree, eight sequences from cluster 1 of the NJ tree formed a cluster against class II photolyase from *D. melanogaster.* Hereafter, the cluster was also referred to as cluster 1. As in the case of the NJ analysis, (6-4) photolyase, blue-light photoreceptor, and MTHF-type photolyase also constituted distinctive subclusters in cluster 1. However, the relative position of each subcluster in the tree was slightly different from that of the NJ tree. Contrary to the result by NJ analysis, most of the nodes in cluster 1 of the ML tree contained high bootstrap probabilities, greater than 70%. We, therefore, considered that the topology of cluster 1 in the ML tree was more reliable than that in the NJ tree. The evolutionary divergence of the three subclusters will be discussed in the next section.

The NJ tree suggested two independent lineages to 8-HDF–type photolyase. In contrast, the ML analysis suggested that there are three independent lineages to 8-HDF–type photolyase. In addition, the relative positions of the enzymes were different from those in the NJ tree. The enzymes from *A. nidulans* and *Synechocystis* sp. branched at node H, while the enzymes from *H. ha-*

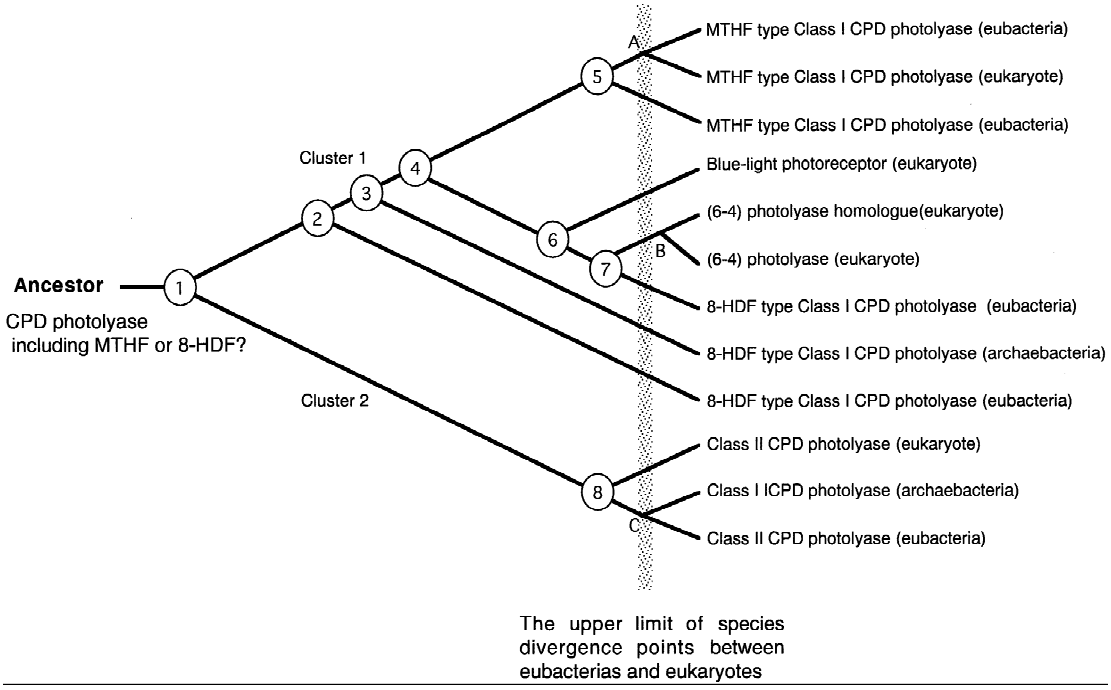

Fig. 4. Scheme of evolutionary process for the photolyase-photoreceptor family. The *circles with figures* indicate the nodes corresponding to gene duplication before the divergence between eubacteria and eukaryotes.

lobium diverged at node I. The locations of the nodes suggested early and independent divergence of the enzymes in cluster 1, while the enzyme from *S. griseus* diverged from the lineage to the subcluster to (6-4) photolyase at node J. The bootstrap probability for node J was high, 70.4%, while the bootstrap probability for node I was only 49.8%. The probability for node H is not shown in the figure, because it was identical to that for the node I. Thus, the ML analysis, as well as the NJ analysis, suggested several independent lineages to 8-HDF–type photolyases, although the branching pattern of the lineages in the ML tree was different from that in the NJ tree. Therefore, the tree topology among the lineages could not be uniquely determined in the current approach.

Evolutionary Scheme of Photolyase–Blue-Light Photoreceptor Family

As described above, the NJ tree is composed of two clusters. The two clusters are connected by the longest branch in the tree. In addition, both clusters include the enzymes from three primary domains of organisms, archaebacteria, eubacteria, and eukaryotes. The observations suggest that the root of the family may be placed on the longest branch. Introducing the putative root, the schematic phylogenetic tree is redrawn (Fig. 4) where the topologies of cluster 1 and 2 were reconstructed based on those in the ML and NJ tree, respectively. The node 1 is the putative root of the tree, where the first gene duplication occurs. Both clusters include CPD photolyases, which are derived from eukaryotes, eubacteria, and archaebacteria. The observation suggests that the ancestral protein at node 1 was a CPD photolyase. However, we could not identify the second cofactor of the ancestral enzyme, because both MTHF and 8-HDF are used as the second cofactors for CPD photolyases belonging to clusters 1 and 2.

We searched for such nodes as correspond to the divergence between proteins from eubacteria and those from eukaryotes or archaebacteria in the schematic phylogenetic tree. Then, nodes A and C were selected as putative species divergence points between eubacteria and eukaryotes (see ‘‘The upper limit of species divergence points between eubacteria and eukaryotes’’ in Fig. 4). Here, we assumed that archaebacteria are evolutionarily closer to eukaryotes than eubacteria (Iwabe et al. 1989). Nodes 1 to 8 (Fig. 4) in the upstream to the putative divergence points were, then, considered to correspond to the gene duplications, which had occurred before the species divergence between eubacteria and eukaryotes. As described above, node 1 corresponded to the putative root of the tree, where the ancestral genes for class I photolyases and class II photolyases diverged. The gene duplication of the ancestral enzyme is considered to have occurred at least eight times, and the prototypes of the current genes formed before the divergence between eubacteria and eukaryotes. The genomic redundancy of the photolyase genes in the ancestral organisms may be an adaptation against the high UV radiation on the ancient Earth.

It is interesting and important to investigate which

was the second cofactor of the ancestral CPD photolyase at node 1, MTHF or 8-HDF. In addition, the problem is related to the change in the second cofactor during the evolution of the protein family. As described above, 8-HDF–type photolyase derived from several different lineages. Figure 4 shows a model with three lineages based on the ML analysis. However, the interpretation of the model differs depending on the second cofactor of the ancestral enzyme. MTHF is widely distributed over the current living organisms, while 8-HDF is rare. Therefore, it seems likely that MTHF was used as the second cofactor in the ancestral enzyme. If so, 8-HDF–type photolyase had independently appeared three times from the enzymes in cluster 1 (nodes 2, 3, and 7 in Fig. 4). In this case, MTHF-type photolyase was present on the line connecting nodes 4 to 7 via node 6. Blue-light photoreceptor had functionally diverged from the CPD photolyase at node 6. The gene duplication at node 7 was followed by the functional divergence of each copy. One of them evolved to be (6-4) photolyase, while another remained CPD photolyase, although the second cofactor changed from MTHF to 8-HDF. In cluster 2, 8-HDF–type enzyme appeared at node C. On the other hand, there is a report that 8-HDF is a more efficient photoantenna than MTHF, whose ratio of electron transfer is one and a half times higher than MTHF, including enzyme (Kim et al. 1992). Thus, 8-HDF seemed more suitable to adapt to the high UV environment on the ancient Earth, and the ancestral enzyme may have used 8-HDF as the photoantenna. Then, 8-HDF gradually replaced MTHF as UV radiation on Earth decreased, because MTHF was more abundant than 8-HDF. This idea also seemed as likely as the former one. In this case, the ancestral proteins on the line, connecting nodes 1 to 4 through nodes 2 and 3, were considered to have been 8-HDF–type photolyase. Similarly, 8-HDF–type photolyase were present on the line connecting nodes 4 and 7 via node 6. The change of the second cofactor from 8-HDF to MTHF occurred on three lines in cluster 1. One of them connected nodes 4 and 5, from which MTHF-type photolyase appeared. The second line connected node 6 to blue-light photoreceptor. The third one was a line connecting nodes 7 and B. In cluster 2, the change of the second cofactor from 8-HDF to MTHF occurred at node 8. As described above, it was difficult to identify the second cofactor of the ancestral enzyme in the current approach. Future study of this problem will reveal the adaptation strategy of ancestral organisms on ancient Earth.

References

Adachi J, Hasegawa M (1995) Improved dating of the human/chimpanzee separation in the mitochondrial DNA tree: heterogeneity among amino acid sites. J Mol Evol 40:622–628

Adachi J, Hasegawa M (1996) MOLPHY (programs for molecular phylogenetics) 2.3b3. Institute of Statistical Mathematics, Tokyo

Ahmad M, Cashmore AR (1993) HY4 gene of *A. thaliana* encodes a protein with characteristics of a blue-light photoreceptor. Nature 366:162–166

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Contr 19:716–723

Batschauer A (1993) A plant gene for photolyase: an enzyme catalyzing the repair of UV-light-induced DNA damage. Plant J 4:705–709

Brash DE (1988) UV mutagenic photoproducts in *Escherichia coli* and human cells: a molecular genetics perspective on human skin cancer. Photochem Photobiol 48:59–66

Dayhoff MO, Schwartz RM, Orcutt BC (1978) A model of evolutionary change in proteins. In: Dayhoff MO (ed) Atlas of protein sequence structure. National Biomedical Research Foundation, Washington, DC, p 345

Eker APM, Dekker RH, Berends W (1981) Photoreactivating enzyme from *Streptomyces griseus*—IV. On the nature of the chromophoric cofactor in *Streptomyces griseus* photoreactivating enzyme. Photochem Photobiol 33:65–72

Eker APM, Kooiman P, Hessels JK, Yasui A (1990) DNA photoreactivating enzyme from the cyanobacterium *Anacystis nidulans.* J Biol Chem 265:8009–8015

Eker APM, Yajima H, Yasui A (1994) DNA photolyase from the fungus *Neurospora crassa.* Purification, characterization and comparison with other photolyases. Photochem Photobiol 60:125–133

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Felsenstein J (1993) PHYLIP (phylogeny inference package) version 3.5c. Department of Genetics, University of Washington, Seattle

Felsenstein J (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. In: Doolittle RF (ed) Methods in enzymology, vol 266. Academic Press, San Diego, p 418

Hearst JE (1995) The structure of photolyase: using photon energy for DNA repair. Science 268:1858–1859

Higgins DG, Bleasby AJ, Fuchs R (1991) CLUSTAL V: improved software for multiple sequence alignment. Comput Appl Biosci 8:189–191

Hsu DS, Zhao X, Zhao S, Kazantsev A, Wang RP, Todo T, Wei YF, Sancar A (1996) Putative human blue-light photoreceptors hCRY1 and hCRY2 are flavoproteins. Biochemistry 35:13871–13877

Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T (1989) Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc Natl Acad Sci USA 86:9355–9359

Iwasa T, Tokutomi S, Tokunaga F (1988) Photoreactivation of *Halobacterium halobium:* action spectrum and role of pigmentation. Photochem Photobiol 47:267–270

Johnson JL, Hamm-Alvarez S, Payne G, Sancar GB, Rajagopalan KV, Sancar A (1988) Identification of the second chromophore of *Escherichia coli* and yeast DNA photolyases as 5,10-methenyltetrahydrofolate. Proc Natl Acad Sci USA 85:2046–2050

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282

Kaufman LS (1993) Transduction of blue-light signals. Plant Physiol 102:333–337

Kiener A, Husain I, Sancar A, Walsh C (1989) Purification and properties of Methanobacterium thermoautotrophicum DNA photolyase. J Biol Chem 264:13880–13887

Kim S-T, Heelis PF, Okumura T, Hirata Y, Mataga N, Sancar A (1991) Determination of rates and yields of interchromophore (folate → flavin) energy transfer and intermolecular (flavin → DNA) electron transfer in *Escherichia coli* photolyase by time-resolved fluorescence and absorption spectroscopy. Biochemistry 30:11262–11270

Kim S-T, Heelis K, Sancar A (1992) Energy transfer (Deazaflavin →

$FADH_2$) and electron transfer ($FADH_2 \rightarrow T\langle\rangle T$) kinetics in *Anacystis nidulans* photolyase. Biochemistry 31:11244–11248

Kim S-T, Malhotra K, Ryo H, Sancar A, Todo T (1996a) Purification and characterization of *Drosophila melanogaster* photolyase. Mutat Res 363:97–104

Kim S-T, Malhotra K, Taylor JS, Sancar A (1996b) Purification and partial characterization of (6-4) photoproduct DNA photolyase from *Xenopus laevis.* Photochem Photobiol 63:292–295

Kishino H, Miyata H, Hasegawa M (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J Mol Evol 31:151–160

Li YF, Sancar A (1991) Cloning, sequencing, expression and characterization of DNA photolyase from *Salmonella typhimurium*. Nucleic Acids Res 19:4885–4890

Liscum E, Hangarter RP (1994) Mutational analysis of blue-light sensing in *Arabidopsis.* Plant Cell Environ 17:639–648

Maciukenas M, McCaughey M (1994) TREETOOL 2.0.2. Ribosomal RNA Database Project, University of Illinois

Malhotra K, Kim S-T, Walsh C, Sancar A (1992) Roles of FAD and 8-hydroxy-5-deazaflavin chromophores in photoreactivation by Anacystis nidulans DNA photolyase. J Biol Chem 267:15406–15411

Malhotra K, Kim S-T, Sancar A (1994) Characterization of a medium wavelength type DNA photolyase: purification and properties of photolyase from *Bacillus firmus.* Biochemistry 33:8712–8718

Malhotra K, Kim S-T, Batschauer A, Dawut L, Sancar A (1995) Putative blue-light photoreceptors from *Arabidopsis thaliana* and *Sinapsis alba* with a high degree of sequence homology to DNA photolyase contain the two photolyase cofactors but lack DNA repair activity. Biochemistry 34:6892–6899

Mitchell DL, Nairn RS (1989) The biology of the (6-4) photoproduct. Photochem Photobiol 49:805–819

Park H-W, Kim S-T, Sancar A, Deisenhofer J (1995) Crystal structure of DNA photolyase from *Escherichia coli.* Science 268:1868–1872

Payne G, Wills M, Walsh C, Sancar A (1990) Reconstruction of *Escherichia coli* photolyase with flavins and flavin analogues. Biochemistry 29:5706–5711

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sancar A (1996) No ''end of history'' for photolyases. Science 272: 48–49

Sancar GB, Smith FW, Heelis PF (1987) Purification of the yeast PHR1 photolyase from an *Escherichia coli* overproducing strain and characterization of the intrinsic chromophores of the enzyme. J Biol Chem 262:15457–15465

Schwartz RM, Dayhoff MO (1978) Matrices for detecting distant relationships. In: Dayhoff MO (ed) Atlas of protein sequence and structure. National Biomedical Research Foundation, Washington, DC, p353

Short TW, Birggs WR (1994) The transduction of blue light signals in higher plants. Annu Rev Plant Physiol Plant Mol Biol 45:143–171

Taylor JS (1995) DNA, sunlight and skin cancer. Pure Appl Chem 67:183–190

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Todo T, Takemori H, Ryo H, Ihara M, Matsunaga T, Nikaido O, Sato K, Nomura T (1993) A new photoreactivating enzyme that specifically repairs ultraviolet light-induced (6-4) photoproducts. Nature 272:109–112

Todo T, Ryo H, Yamamoto K, Toh H, Inui T, Ayaki H, Nomura T, Ikenaga M (1996) Similarity among the *Drosophila* (6-4) photolyase, a human photolyase homologue, and the DNA photolyase-blue-light photoreceptor family. Science 272:109–112

Todo T, Kim S-T, Hitomi K, Otoshi E, Inui T, Morioka H, Kobayashi H, Ohtsuka E, Toh H, Ikenaga M (1997) Flavin adenine dinucleotide as a chromophore of the *Xenopus* (6-4) photolyase. Nucleic Acids Res 25:764–768

Yasui A, Eker APM, Yasuhira S, Yajima H, Kobayashi T, Takao M, Oikawa A (1994) A new class of DNA photolyases present in various organisms including aplacental mammals. EMBO J 13: 6143–6151