Sequence of the *Saccharomyces cerevisiae* *PHR1* gene and homology of the *PHR1* photolyase to *E. coli* photolyase

Gwendolyn B.Sancar

Department of Biochemistry 231H, University of North Carolina, Chapel Hill, NC 27514, USA

Received 12 August 1985; Revised and Accepted 24 October 1985

ABSTRACT

The nucleotide sequence of a 2301 base pair region of *Saccharomyces cerevisiae* DNA containing the *PHR1* gene is reported. Within this region a single open reading frame of 1695 base pairs was found; using the insertional inactivation technique it was shown that part or all of this open reading frame specifies the *PHR1*-encoded photolyase. The amino acid sequence of the 565 amino acid long polypeptide predicted from the *PHR1* nucleotide sequence was compared to the amino acid sequence of *E. coli* photolyase. Overall the sequence homology was 36.5%; however, two short regions near the amino terminus as well as the carboxy-terminal 150 amino acids display significantly greater sequence homology. The presence of these strongly conserved regions suggests that the yeast and *E. coli* photolyase possess common structural and functional domains involved in substrate and/or chromophore binding.

INTRODUCTION

Pyrimidine dimers (cyclobutanedipyrimidines) are the major lethal lesions induced in DNA by far UV radiation (254 nm). A number of different enzymatic pathways for repairing or removing pyrimidine dimers have evolved; among these repair pathways photoreactivation, mediated by DNA photolyase (EC 4.1.99.3), is unique in that it is the only pathway in which pyrimidine dimers are repaired *in situ* rather than being excised from DNA (for recent reviews of enzymatic DNA repair, see ref. 1 and 2). Repair by photolyases is accomplished by a two-step mechanism; first the enzyme binds to dimers in DNA, then upon absorbing near UV or visible light (300-600 nm) the enzyme converts light energy to chemical energy which is used to break the cyclobutane ring thereby restoring the pyrimidines to their original configuration (3,4).

Enzymatic photoreactivation has been found in organisms throughout the phylogenetic tree including bacteria, blue-green algae, protozoa, and marsupials (see ref. 5 for a review). Photolyases from several organisms, including *E. coli* (6-9), *Streptomyces griseus* (10,11), and *Saccharomyces cerevisiae* (12-14), have been isolated and studied *in vitro*. The enzymes

from these three organisms share certain common properties; the bacterial enzymes as well as one of the two photolyases isolated from *S. cerevisiae* are similar in molecular weight (Mr 41,000-54,000) and each of these three enzymes has an intrinsic chromophore which is a flavin derivative (3,6-8,11-13). These similarities, as well as the common substrate specificity of the enzymes, suggest that these photolyases may exhibit common sequences or secondary structures in the regions of the substrate and chromophore binding sites. I have recently demonstrated that the *PHR1* gene of the yeast *S. cerevisiae* encodes a DNA photolyase and that this enzyme is capable of complementing both point mutations and deletions of the *E. coli phr1* gene (15). Based upon the size of the restriction fragment carrying the yeast gene and the ability of *E. coli* to furnish the chromophore of the *PHR1*-encoded photolyase, I suggested that *PHR1* protein (Phr1) corresponds to the flavo-photolyase of *S. cerevisiae* previously characterized by Werbin and coworkers (12,13). In this communication I report the nucleotide sequence of the *PHR1* gene and the predicted amino acid sequence of the Phr1 protein. Phr1 photolyase shows significant sequence homology to *E. coli* photolyase particularly over the carboxy-terminal 150 amino acids of the two proteins. Highly conserved regions of these proteins may reflect common functional domains important in substrate and/or chromophore binding.

METHODS

Construction of Strains and Tn1000 Derivatives of pGBS100

The construction of pGBS100, which is a pBR328 derivative carrying the inducible *tac* promoter and a 3.3 kilobase pair (kbp) *PvuII* fragment containing the *PHR1* gene, has been described previously (15). Transcription of *PHR1* originates at a site within the 3.3 kbp *PvuII* fragment, not from the *tac* promoter (Figure 1A). To obtain a suitable donor strain for the mating experiments described below I constructed strain DHI/*F'**lacI*^Q/pGBS100 as follows. *E. coli* K12 strain DHI (16) was mated with strain CSR603/*F'**lacI*^Q (6,17,18) followed by selection on minimal agar plates (19) containing 10 µg/ml thiamine hydrochloride; only DHI and DHI/*F'**lacI*^Q can grow on this medium. Individual clones were checked for the presence of F factor by testing for sensitivity to phage M13. DHI/*F'**lacI*^Q was then transformed with pGBS100 selecting for resistance to tetracycline. To isolate derivatives of pGBS100 containing Tn1000 insertions, strain DHI/*F'**lacI*^Q carrying pGBS100 was mated with CSR603 (*recA1*, *uvrA6*, *phr1*, *Strep*^R; 17,18). CSR603 transconjugants which had received both the F factor and pGBS100 were identified by

plating a dilution of the mating mixture on luria broth agar containing 50 µg/ml streptomycin, 10 µg/ml tetracycline, and 200 µg/ml ampicillin. Clones containing Tn1000 inserts in *PHR1* were identified as previously described (6). The locations of the Tn1000 insertions in *Phr*⁺ and *Phr*⁻ derivatives of pGBS100 were determined by restriction mapping of the PstI and BamHI sites in the plasmids.

DNA Sequencing

DNA used for sequence analysis was obtained from plasmids propagated in *E. coli* strains. DNA from pGBS100 was isolated by CsCl-ethidium bromide equilibrium gradient centrifugation as previously described (20) with the difference that plasmid DNA was not amplified by chloramphenicol treatment of the cultures prior to DNA extraction. The various restriction fragments carrying portions of *PHR1* were purified by sucrose gradient centrifugation (21). In some cases plasmids YEp13-*PHR1*, YRp7-*PHR1*-C and YRp7-*PHR1*-D (kindly provided by Drs. David Schild and Robert K. Mortimer; 22) were used as sources of DNA which was purified as previously described (20). The majority of the sequence reported here was obtained using the Maxam and Gilbert sequencing technique (23); however, for a few regions the sequence was obtained by the dideoxy method of Sanger *et al.* (24) as modified for M13 by Messing *et al.* (25). All techniques used to label restriction fragments and to obtain and sequence M13 derivatives carrying specific restriction fragments have been reported (26).

Alignment of the *Phr1* and *E. coli* Photolyase Amino Acid Sequences

The amino acid sequences of the *PHR1*-encoded polypeptide and *E. coli* photolyase were compared using the program ALIGN (27) from the Protein Identification Resource of the National Biomedical Research Foundation. The PAM250 matrix (27) was used for the comparison and a value of 6 was used for the matrix bias and for the gap penalty. The significance of the alignment obtained was determined by comparison to alignments obtained with 25 random permutations of the sequences. The sequences in the National Biomedical Research Foundation protein sequence library were examined for homology to *Phr1* and *E. coli* photolyase using the program FASTP developed by Lipman and Pearson (28) with ktup = 2.

RESULTS

Boundaries of *Phr1* and Sequencing Strategy

Plasmid pGBS100 (Figure 1A) contains a 3.3 kbp PvuII fragment from *S. cerevisiae* DNA carrying the *PHR1* gene (15,22). Recently I demonstrated that

Figure 1. A. Restriction map of plasmid pGBS100. The location and direction of transcription of the *tet*, *bla* and *PHR1* genes and the *tac* expression cartridge are indicated by arrows. B. Location of Tn1000 inserts in the 3.3 kbp PvuII fragment from pGBS100 carrying *PHR1* and phenotype of the resultant plasmids in *E. coli* strain CSR603 (*recA1*, *uvrA6*, *phr-1*). Arrows above the line indicate location of Tn1000 insertions resulting in loss of the *Phr⁺* phenotype conferred by pGBS100 while those below the line indicate the locations of insertions which did not inactivate *PHR1*. Scale indicates the position in pGBS100.

pGBS100 also complements mutations in the *E. coli phr1* gene and determined the direction of transcription of *PHR1* (15). To further localize the *PHR1* gene in this plasmid I isolated a number of derivatives of pGBS100 which contain insertions of the transposable element Tn1000 within the 3.3 kbp PvuII fragment and tested the ability of these plasmids to complement the *phr1* mutation in *E. coli*. As can be seen in Figure 1B the locations of the insertion sites define an approximately 1.95 kbp region containing the *phr1* complementing activity. I then determined the nucleotide sequence of a portion of the 3.3 kbp PvuII fragment beginning at the 5' PvuII site (relative to the direction of transcription of *PHR1*) and extending through nucleotide 2301 of the fragment. The sequencing strategy utilized is shown in Figure 2. Approximately 85% of the sequence was determined on both strands.

Nucleotide Sequence of *PHR1*

The sequence of the 2301 bp region of DNA containing the *PHR1* gene is

Figure 2. Strategy used in sequencing the *PHR1* gene. The 2301 bp region, the sequence of which is given in Figure 3, is shown. The PvuII site at left is the site 5' to *PHR1* relative to the direction of transcription. Arrows show the direction and extent of sequence obtained beginning at the indicated restriction sites, and the position of the arrows above or below the line indicate which strand was sequenced. Scale indicates distance from the 5' PvuII site; numbers in parentheses indicate position in pGBS100.

shown in Figure 3. Within this region a single long open reading frame of 1695 bp is found, beginning with an ATG at position 355-357 and ending with TGA at position 2050-2052; the location of this reading frame corresponds well (within the resolution of restriction mapping) with the location of the *phr1* complementing activity as defined by the insertional inactivation data. No other reading frame longer than 450 bp was observed for either strand (Figure 4). Taken together these results clearly indicate that part or all of this 1695 bp reading frame encodes the *PHR1* photolyase.

As is discussed below, the homology of the *PHR1*-encoded protein to *E. coli* photolyase indicates that translation of *PHR1* mRNA begins at or 5' to the AUG corresponding to the ATG at position 586-588 in the DNA sequence. However, within the 1695 bp open reading frame there are two in-frame ATG codons prior to the one at 586-588 (Figure 3). In yeast the consensus sequence at the translational start site of genes expressed at high levels is PuXXATGPuXT (29). Of the three potential start sites for *PHR1*, at positions 355-357, 523-525 and 586-588, the site at 355-357 with sequence ATAATGAAA has the best match to the consensus sequence. However, it should be noted that photolyase is not an abundant protein in *S. cerevisiae* (30,31) so that a good match to the consensus sequence for the translational start site is not necessarily expected.

At present the sequences required for transcription of *PHR1* are unknown,

Figure 3. Nucleotide sequence of the *PHR1* gene and surrounding 5' and 3' regions and predicted amino acid sequence of the 1695 bp open reading frame. Numbering of bases and amino acids is from the first nucleotide in the 5' PvuII site and from the first ATG in the 1695 bp open reading frame respectively. The three potential translational start sites referred to in the text are doubly overlined while the regions 3' to *PHR1* which are similar to sites involved in transcription termination are singly overlined.

nor is it possible on the basis of present data to predict the location of these sites as (i) the analysis of insertion derivatives in *PHR1* has not yet been carried out in yeast and (ii) although there are a number of regions 5' to the ATG at position 586-588 which share partial homology to the conserved consensus sequence TATA^T/_A found 5' to the transcriptional start sites of eukaryotic genes (32), no single region displays an especially good match. Matches 3' to *PHR1* are found to the two different consensus sequences thought

Figure 4. Locations of potential translational initiation and termination sites in the 2301 bp region sequenced. Lines above the horizontal lines indicate initiation codons ATG and GTG while those below the horizontal lines indicate termination codons TAA, TGA and TAG. Horizontal lines ending in rightward pointing arrows indicate reading frames in the same direction as *PHR1* transcription while those ending in leftward pointing arrows indicate reading frames in the opposite direction. The 1695 bp open reading frame is indicated by the thick line.

to be involved in transcription termination in yeast. The region encompassing nucleotides 2106-2134 contains the sequence TGATX₆TAGTX₁₁(77% A+T)TTAT which is similar to the tripartate element TAGX₁₋₁₄TAGT...(A+T rich) TTT identified by Zaret and Sherman (33) as being required for transcription termination and polyadenylation. In addition at nucleotides 2222-2229 the sequence TTTTATA is a perfect match to the consensus sequence identified by Henikoff et al. (34) as necessary but not sufficient for efficient transcription termination.

Langford et al. (35) have proposed that in *S. cerevisiae* the consensus sequence required for splicing of introns from primary transcripts is ↓GTATGT...TACTAACA...PyAG↓ where the splice sites are indicated by the arrows. No sequence matching this consensus sequence is found in the sequence shown in Figure 3 and thus it seems likely that *PHR1* does not contain introns. This is consistent with the expression of *PHR1* in *E. coli* (15).

Codon usage in *PHR1*

Bennetzen and Hall (36) have shown that in yeast genes encoding abundant

Table 1. Codon Usage in *PHR1*

Amino Acid	Codon	Total	Codon Bias Index†	Amino Acid	Codon	Total	Codon Bias Index†		
Phe	UUU	15	0.40	Tyr	UAU	19	0.39		
	UUC*	10			UAC*	12			
Leu	UUA	17	0.27	His	CAU	7	0.46		
	UUG*	14			CAC*	6			
	CUU	5		Gln	CAA*	17	0.81		
	CUC	2			CAG	4			
	CUA	4			Asn	AAU		23	0.36
	CUG	9				AAC*		13	
Ile	AUU*	9	0.26	Lys	AAA	39	0.20		
	AUC*	6			AAG*	10			
	AUA	19		Asp	GAU	20	0.81		
Met	AUG	18	0.62		GAC	9			
	Val	GUU*	11		Glu	GAA*		25	0.56
		GUC*	5			GAG		6	
		GUA	5	Cys		UGU*	5	0.48	
GUG	5	UGC	4						
Ser	UCU*	16	0.46		Trp	UGG	16		0.33
	UCC*	5				Arg	CGU		
	UCA	8		CGC	1				
	UCG	5		CGA	2				
	AGU	9		CGG	2				
	AGC	3		AGA*	14				
Pro	CCU	6	0.59	Gly	AGG	6			
	CCC	2			Gly	GGU*	8	0.52	
	CCA*	13				GGC	5		
	CCG	1				GGA	8		
Thr	ACU*	12	0.58	Gly		GGG	3		
	ACC*	6			Gly	GGU*	8	0.33	
	ACA	9				GGC	5		
	ACG	3				GGA	8		
Ala	GCU*	8	0.52	Gly		GGG	3		
	GCC*	5			Gly	GGU*	8	0.33	
	GCA	10				GGC	5		
	GCG	2				GGA	8		

† Codon Bias Index is as defined in text.

* Codons recognized by most abundant isoacceptor tRNA or used most frequently in abundant yeast proteins. No preference is exhibited for Asp codons among abundant yeast proteins (36).

proteins codon usage is biased in favor of codons recognized by the most abundant tRNA for each amino acid while in less highly expressed genes codon usage appears to be more random. This second pattern of codon usage is found in the *PHR1* gene (Table 1). Overall the codon bias index, defined as the fraction of total codons corresponding to the 22 codons most frequently used in abundant yeast proteins (36), is 0.46 for *PHR1*; this value for the codon bias index is in contrast to values of 0.9-1.0 seen for proteins present at high levels in yeast (36). For *PHR1* the codon bias index remains essentially constant regardless of which of the three potential start codons are utilized and does not vary significantly (index = 0.45-0.47) in the regions between these ATGs. Over the entire *PHR1* coding region, the codons corresponding to the most abundant isoacceptor tRNA species are used less often than other cognate codons for the amino acids phenylalanine, leucine, isoleucine, serine, tyrosine, histidine, asparagine, lysine, arginine and glycine. This pattern of codon usage supports the experimental observation that yeast photolyase is not an abundant protein (30,31).

The *PHR1* photolyase and *E. coli* photolyase

Depending upon which of the three potential translational start sites within the *PHR1* open reading frame is used, the *PHR1* gene encodes a polypeptide of 66189, 59613 or 57168 daltons containing 565, 509 or 488 amino acids respectively. The smallest of these three predicted polypeptides corresponds most closely to the molecular weight of 51-53 kilodaltons reported by Minato and Werbin (12) and Iwatsuki et al. (13) for the yeast flavo-photolyase. However, preliminary data suggests that the first ATG in the *PHR1* open reading frame is the one actually used. This conclusion comes from experiments in which derivatives of pGBS101 (which is identical to pGBS100 except that transcription from the *tac* promoter is in the same direction as *PHR1* transcription (15) were constructed such that the second ATG of *PHR1* is adjacent to the ribosome binding site of *tac*; these constructs display a 10 fold reduction in photoreactivating activity compared to pGBS101, suggesting that they encode a truncated peptide missing a portion of the protein necessary for full photoreactivating activity (data not shown). Therefore, for the purpose of the following discussion I will assume that translation begins at the position corresponding to the first ATG in the *PHR1* open reading frame. The amino acid sequence of the Mr 66189 polypeptide thus encoded is shown below the nucleotide sequence in Figure 4 and the amino acid composition of the polypeptide is shown in Table 2. The polypeptide consists of 38% non-polar, 35% polar, 11% acidic and 16% basic amino acids. With the exception

Table 2. Amino Acid Composition of *PHR1* Photolyase and *E. coli* Photolyase†

Amino Acid	<i>PHR1</i> *		<i>E. coli</i> Photolyase	Amino Acid	<i>PHR1</i> *		<i>E. coli</i> Photolyase
	1st Met	3rd Met			1st Met	3rd Met	
Ala	25	21	50	Lys	49	39	17
Val	26	22	30	Arg	29	23	34
Leu	51	48	43	His	13	11	15
Ile	34	30	19				
Pro	22	19	26	Gly	24	23	28
Phe	25	22	19	Ser	46	37	23
Trp	16	16	15	Thr	30	25	22
Met	18	16	8	Cys	9	9	7
				Tyr	31	27	14
Asp	29	27	20	Asn	36	30	20
Glu	31	26	33	Gln	21	17	28

† Amino acid composition of *E. coli* photolyase is from ref. 26.

* The amino acid composition of *PHR1* photolyase is shown for the polypeptide beginning at the first Met which has the best consensus sequence for a translational start site and the third Met which is near to the site where homology with *E. coli* photolyase begins.

of the amino terminal 68 residues in which 18 basic amino acids and only 3 acidic amino acids are present, no extensive region containing predominantly one class of amino acid is apparent.

We have previously reported the nucleotide sequence of the *E. coli phr* gene and the predicted amino acid sequence of the encoded photolyase (26) which is a flavoprotein (7,8) of Mr 53994 and is composed of 471 amino acids. As can be seen in Table II, although the proportions of acidic, basic, polar and nonpolar residues are quite similar in Phr1 and *E. coli* photolyase, the amino acid compositions of the two proteins are quite dissimilar. The Phr1 composition is also dissimilar to the composition reported by Iwatsuki *et al.* (12) for the yeast flavophotolyase (see discussion). To test the hypothesis that Phr1 and *E. coli* photolyase are structurally and functionally related the extent of similarity between the amino acid sequences of these two proteins was determined by computer analysis using the program ALIGN from the Protein Identification Resource (27). As can be seen in Figure 5, homology is evident over the entire lengths of the proteins with the exception of the amino terminal 75 amino acids of Phr1. To achieve this alignment it was necessary to insert a total of 16 gaps within the Phr1 and *E. coli* photolyase amino acid sequences; however, this number of gaps does not exceed the 4 gaps per 100 amino acids recommended by Doolittle (37) as the maximum number of gaps permissible between authentically related proteins.

Out of a total of 471 possible matches between Phr1 and *E. coli* photo-

Figure 5. Amino acid sequence homology between Phr1 and *E. coli* photolyase. The alignment obtained as described in Methods is shown. Identical amino acids are indicated by a *, while gaps are shown as hyphens. The amino acid sequence of *E. coli* photolyase is from reference 26. Numbering is from the first ATG in the 1695 bp open reading frame for Phr1 and the first amino acid in mature *E. coli* photolyase.

lyase there are 172 identities; thus the overall sequence conservation is 36.5% with an alignment score equal to 25.35 standard deviations from the mean of 25 comparisons of random permutations. The longest and most highly conserved region is located at the carboxy terminal regions of the proteins where from amino acids 421 to 565 (Phr1 numbering) 52% (78/144) of the amino acids are conserved; in addition two short extensively conserved regions are present near the amino terminus at amino acids 80 to 93 (71% identity) and 136 to 158 (52% identity). The region from amino acids 159-420 is less homologous overall with only 24.7% identity and 13 gaps. The fact that only limited regions of Phr1 protein and *E. coli* photolyase are highly conserved

strongly suggests that these regions specify important common structural and functional domains of the enzymes.

The nucleotide sequence homology of the coding regions of the two genes parallels the amino acid homology; overall 40% of nucleotides are conserved (data not shown).

DISCUSSION

The two most important questions which arise from the results presented here concern the relationship of the photolyase encoded by the yeast *PHR1* gene to previously characterized yeast photolyases and the functions of the highly conserved structural domains of Phr1 and *E. coli* photolyase.

Two photolyases have been isolated from commercial preparations of baker's yeast. One of the enzymes is a monomer of Mr 51,000-53,000 which has flavin cofactor tightly bound (12,13) while the second enzyme is a heterodimer composed of two subunits of Mr 82,500 and 54,000 (14). As noted above, Iwatsuki et al. (13) have reported the amino acid composition of the monomeric flavoenzyme as determined by amino acid analysis of purified protein. The amino acid composition of Phr1 photolyase predicted from the nucleotide sequence of *PHR1* reported here does not agree with these authors' results. Nevertheless it is likely that the *PHR1* gene encodes the flavoenzyme of Iwatsuki et al. (13). The molecular weight predicted for the *PHR1* gene product (66,189-57,168 daltons) is similar to the reported molecular weight of the flavoenzyme. Furthermore the results of both DNA sequence analysis and insertion activation experiments reported here indicate that the 3.3. kbp *PvuII* fragment carrying *PHR1* contains a single open reading frame conferring photoreactivation. The fact that *PHR1* is able to complement mutations in the *E. coli phr1* gene, which encodes a monomeric photolyase that in pure form is capable of photoreactivation in the absence of any other proteins (38), indicates that Phr1 has photoreactivating activity alone; this is in contrast to the Mr 54,000 subunit of the heterodimeric yeast photolyase which has no detectable activity in the absence of the Mr 82,500 subunit (14). Finally using a 65% pure preparation of the enzyme Zwetsloot et al. (39) have recently reported that the molecular weight and absorption spectrum of Phr1 protein are consistent with the properties reported by Iwatsuki et al. (13) for the yeast flavoenzyme. Given the difficulty of purifying a protein present in very low quantities in yeast cells (30,31), it is possible that the sample utilized by Iwatsuki et al. (13) contained many minor contaminants which

resulted in an apparent amino acid composition different from that of completely pure enzyme.

The results reported here demonstrate that the photolyases encoded by the *PHR1* gene of *S. cerevisiae* and the *phr1* gene of *E. coli* are structurally as well as functionally related. The properties of these enzymes indicate that both possess a chromophore binding site and a substrate binding site. The chromophore binding sites of the two enzymes are expected to be similar since both have a FAD cofactor (7,8,13); therefore one or more of the highly conserved regions probably constitute the chromophore binding domain. Using the algorithm developed by Lipman and Pearson (28) the amino acid sequences of the yeast and *E. coli* photolyases was compared to those of known FAD-binding proteins in the National Biomedical Research Foundation library; no significant homologies (i.e., optimized alignment scores greater than 70) were detected. This result is not surprising as the secondary and tertiary structures at the flavin binding sites are conserved while the primary sequences of different flavin binding proteins at these sites are quite different (40,41). Wierenga *et al.* (41) have found that the FAD binding domains of *p*-hydroxybenzoate hydroxylase and glutathione reductase consist of a four-stranded parallel β -sheet bordered by two α -helices on one side and by a three-stranded antiparallel β -sheet on the other. While identification of the precise regions constituting the FAD-binding sites in *E. coli* and yeast photolyases awaits solution of the crystal structures of these enzymes, the secondary structure predicted by the Chou-Fasman algorithm (42) for the conserved carboxy-terminal regions of the enzymes suggests the presence of alternating α -helices and β -sheets such as are found in FAD binding domains (not shown). Both *E. coli* photolyase and the Phr1 photolyase are unusually rich in tryptophan; the positions of nine tryptophans are conserved in these enzymes including six conserved tryptophans in the highly conserved carboxy-terminal regions. It is intriguing that aromatic compounds have been shown to stabilize against disproportionation blue neutral flavin radical (43) such as is found in *E. coli* photolyase (8).

Photolyases from yeast (3) and from *E. coli* (9) bind specifically to pyrimidine dimers in DNA; however, the minimum substrate size required by these two enzymes may be quite different. Substrates smaller than oligo dT₁₀ cannot effectively compete with high molecular weight DNA for binding of yeast photolyase (45). In contrast the *E. coli* enzyme can repair dimers in substrates as small as oligo dT₃ (46) and we have recently found that the enzyme makes only 1-2 phosphate contacts when bound to dimer-containing DNA

(unpublished observations). That there is a difference in the substrate binding sites of the two enzymes is also supported by the difference in the dependence of substrate binding on NaCl concentration; the yeast flavoenzyme exhibits a sharp optimum at an ionic strength of 0.175 M (46) while the *E. coli* enzyme has a very broad binding range with a peak at 0.125-0.150 M (9). These results suggest that ionic interactions between the enzyme and the phosphodiester backbone are much more important for binding of the yeast enzyme and thus that the substrate binding site of Phr1 contains more positively charged amino acids than the analogous site of *E. coli* photolyase. Whether the highly basic region of 68 amino acids predicted at the amino terminus of Phr1 plays a role in DNA binding by the enzyme is currently under investigation.

The nucleotide sequences of the *RAD1*, *RAD3* and *RAD6* genes of *S. cerevisiae* have recently been determined (47-49); the proteins encoded by *RAD1* and *RAD3* are involved in nucleotide excision repair (50) while the *RAD6*-encoded protein is required for postreplication repair of DNA damage (51). It has been noted that both Rad3 and Rad6 proteins have extremely acidic carboxy-termini (48,49) and it has been proposed that these regions may be involved in interactions with nucleosomes or other DNA repair proteins (48,49). No such acidic region is present in Phr1.

ACKNOWLEDGEMENTS

I thank Kathy A. Franklin and J. David Taylor for technical assistance and Dr. Clyde Hutchison for performing the sequence alignments and homology searches. This work was supported by grant numbers GM31082 and GM35123 from NIH, American Cancer Society Institutional Grant number IN-15-Y, and a grant from the University Research Council, University of North Carolina at Chapel Hill.

REFERENCES

1. Lindahl, T. (1982) *Ann. Rev. Biochem.* 51, 61-87.
2. Sutherland, B.M. (1981) in *The Enzymes*, Boyer, P.D., Ed., Vol. XIV Part A, pp. 481-518, Academic Press, New York.
3. Rupert, C.S. (1962) *J. Gen. Physiol.* 45, 725-741.
4. Setlow, R.B., Carrier, W.L. and Bollum, F.J. (1965) *Proc. Natl. Acad. Sci. USA* 53, 1111-1118.
5. Harm, H. (1976) in *Photochemistry and Photobiology of Nucleic Acids*, Wang, S.Y., Ed., Vol. II, pp. 219-263, Academic Press, New York.
6. Sancar, G.B., Smith, F.W. and Sancar, A. (1983) *Nucleic Acids Res.* 11, 6667-6678.
7. Sancar, A. and Sancar, G.B. (1984) *J. Mol. Biol.* 172, 223-227.
8. Jorns, M.S., Sancar, G.B. and Sancar, A. (1984) *Biochemistry* 23, 2673-2679.
9. Sancar, G.B., Smith, F.W. and Sancar, A. (1985) *Biochemistry* 24, 1849-1855.

10. Eker, A.P.M. and Fichtinger-Schepman, A.M.J. (1975) *Biochim. Biophys. Acta* 378, 54-63.
11. Eker, A.P.M., Dekker, R.H. and Berends, W. (1981) *Photochem. Photobiol.* 33, 65-72.
12. Minato, S. and Werbin, H. (1971) *Biochemistry* 10, 4503-4508.
13. Iwatsuki, N., Joe, C.O. and Werbin, H. (1980) *Biochemistry* 19, 1172-1176.
14. Boatwright, D.T., Madden, J.J., Denson, J. and Werbin, H. (1975) *Biochemistry* 14, 5418-5421.
15. Sancar, G.B. (1985) *J. Bacteriol.* 161, 769-771.
16. Hanahan, D. (1983) *J. Mol. Biol.* 166, 557-580.
17. Sancar, A. and Rupert, C.S. (1978) *Mutat. Res.* 51, 133-137.
18. Sancar, A., Hack, A.M. and Rupp, W.D. (1979) *J. Bacteriol.* 137, 692-693.
19. Miller, J.H. (1972) *Experiments in Molecular Genetics*, pp. 432, Cold Spring Harbor Laboratory, Cold Spring Harbor.
20. Sancar, G.B., Sancar, A., Little, J.W. and Rupp, W.D. (1982) *Cell* 28, 523-530.
21. Hughes, C. and Meynell, G.G. (1977) *Mol. Gen. Genet.* 151, 175-179.
22. Schild, D., Johnston, J., Chang, C. and Mortimer, R.K. (1984) *Molec. Cell. Biol.* 4, 1864-1870.
23. Maxam, A.M. and Gilbert, W. (1980) in *Methods in Enzymology*, Moldave, K. and Grossman, L., Eds., Vol. 65, 499-560, Academic Press, New York.
24. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
25. Messing, J., Crea, R. and Seeburg, P.H. (1981) *Nucleic Acids Res.* 9, 309-321.
26. Sancar, G.B., Smith, F.W., Lorence, M.C., Rupert, C.S. and Sancar, A. (1984) *J. Biol. Chem.* 259, 6033-6038.
27. Dayhoff, M.O. (1979) *Atlas of Protein Sequence and Structure*, Vol. 5, Supplement 3, pp. 1-8 and 353-358, National Biomedical Research Foundation, Washington.
28. Lipman, D.J. and Pearson, W.R. (1985) *Science* 227, 1435-1441.
29. Dobson, M.J., Tuite, M.F., Roberts, N.A., Kingsman, A.J. and Kingsman, S.M. (1982) *Nucleic Acids Res.* 10, 2625-2637.
30. Harm, H. and Rupert, C.S. (1968) *Mutation Res.* 6, 355-370.
31. Yasui, A. and Laskowski, W. (1975) *Int. J. Radiat. Biol.* 28, 511-518.
32. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* 50, 349-383.
33. Zaret, K.S. and Sherman, F. (1982) *Cell* 28, 563-573.
34. Henikoff, S., Kelly, J.D. and Cohen, E.H. (1983) *Cell* 33, 607-614.
35. Langford, C.J., Klinz, F.-J., Donath, C. and Gallwitz, D. (1984) *Cell* 36, 645-653.
36. Bennetzen, J.L. and Hall, B.D. (1982) *J. Biol. Chem.* 257, 3026-3031.
37. Doolittle, R.F. (1981) *Science* 214, 149-159.
38. Sancar, A., Smith, F.W. and Sancar, G.B. (1984) *J. Biol. Chem.* 259, 6028-6032.
39. Zwetsloot, J.C.M., Vermeulen, W., Hoeijmakers, J.H.J., Yasui, A., Eker, A.P.M. and Bootsma, D. (1985) *Mutation Res.* 146, 71-77.
40. Rossman, M.G., Liljas, A., Brändén, C.-I. and Banaszak, L.J. (1975) in *The Enzymes*, Boyer, P.D., Ed., Vol. XI, Part A, pp. 61-102, Academic Press, New York.
41. Wierenga, R.K., Drenth, J. and Schulz, G.E. (1983) *J. Mol. Biol.* 167, 725-739.
42. Chou, P.Y. and Fasman, G.D. (1978) *Ann. Rev. Biochem.* 47, 251-276.
43. Müller, F., Brüstlein, M., Hemmerich, P., Massey, V. and Walker, W.H. (1972) *Eur. J. Biochem.* 25, 573-580.
44. Setlow, J.K. and Bollum, F.J. (1968) *Biochim. Biophys. Acta* 157, 233-237.
45. Jorns, M.S., Sancar, G.B. and Sancar, A. (1985) *Biochemistry* 24, 1856-1861

46. Madden, J.J. and Werbin, H. (1974) *Biochemistry* 13, 2149-2154.
47. Yang, E. and Friedberg, E.C. (1984) *Molec. Cellular Biol.* 4, 2161-2169.
48. Reynolds, P., Higgins, D.R., Prakash, L. and Prakash, S. (1985) *Nucleic Acids Res.* 13, 2357-2372.
49. Reynolds, P., Weber, S. and Prakash, L. (1985) *Proc. Natl. Acad. Sci. USA*, 82, 168-172.
50. Wilcox, D.R. and Prakash, L. (1981) *J. Bacteriol.* 148, 618-623.
51. Prakash, L. (1981) *Mol. Gen. Genet.* 184, 471-478.

NOTE ADDED IN PROOF:

While this manuscript was under review, Yasui and Langeveld (*Gene* 36, 349-355, 1985) also published a nucleotide sequence of the *Saccharomyces cerevisiae* *PHR1* gene; those authors reached similar conclusions regarding the homology between the yeast and *E. coli* photolyases although there are some minor differences between the amino acid alignments presented. However, there are a number of differences between both the nucleotide and amino acid sequences reported by Yasui and Langeveld and the sequences reported here. These differences are as follows (numbering refers to the sequences presented in this work): (i) the A at position 29 was absent, resulting in loss of the XbaI site near the 5' PvuII site; (ii) the T at position 213 was missing; (iii) the sequence GTCATG at position 583-588 was reported as GCAATG; (iv) the C at position 848 is reported as a G; (v) the T at position 859 was reported as an A; (vi) the sequence GAT at position 952-954 was reported as AGT; (vii) the C at position 1407 was reported as a G; (viii) the G at position 1448 was reported as an A; (ix) the G at position 1771 was reported as an A; (x) the sequence GGACC at position 2238-2242, which constitutes an *AvaII*/*Sau96I* recognition site, was reported as GAACC; (xi) the sequence CGCGGTTA at positions 2263-2271 were reported as CGGTTA. These discrepancies result in differences in the amino acid assignments at the following sites (the assignments of Yasui and Langeveld are in parentheses): (i) Ser 169 (Thr); (ii) Asp 200 (Ser); (iii) Glu 365 (Gly); (iv) Glu 473 (Lys). In addition, although there are no differences in the nucleotide sequences at positions 1915-1917 and 1933-1935, these codons were incorrectly reported by Yasui and Langeveld as encoding Glu rather than Gln. It should be noted that for the sequence presented here the sequence at each of the regions of discrepancy was determined at least twice and in all cases except one the sequence was determined for both strands. In addition the sequence discrepancies involving restriction sites have been confirmed by restriction mapping and by sequencing from these sites.