resent an excess of C→U/T deamination events on the coding strand resulting from transcription, paralleling recent experiments in yeast showing increased mutation rates with higher levels of transcription (*21*). However, the exposure of short single-stranded regions during RNA synthesis is probably too transient to account for the level of asymmetry observed between the strands (*18*).

In contrast, the alternative process, transcription-coupled repair, is highly strand-specific and also predicts the observed differences between C→T and G→A substitution frequencies. In *E. coli*, this repair is known to act on ultraviolet-induced pyrimidine dimers and is targeted to the transcribed (that is, noncoding) strand (*22*). Because C→T transitions are the primary mutations induced by pyrimidine dimers (*23*), transcription-coupled repair will result in a deficit of C→T changes on the transcribed strand, which translates into an excess of C→T over G→A changes on the coding strand (*24*), as detected in the genes that we analyzed. If this process is responsible for the observed transitional asymmetry, the fraction of C→T changes at dipyrimidine sites would exceed 70%, the expectation based on the trinucleotide composition of the *E. coli* genome (*25*). Of the loci in Fig. 2, this fraction is above 80% in *putP* and *gutB*, supporting transcription-coupled repair at dipyrimidine sites as the cause of asymmetry; but at *mdh*, *phoA*, and *crr*, this fraction is below 65%. Because transcription-mediated repair systems have been hypothesized to operate on other types of DNA damage, including deamination of C, any C→U/T change on the transcribed strand could be preferentially corrected (*18*, *26*).

If the C→T versus G→A asymmetry is introduced during transcription, we would expect that cryptic genes, which are expressed only occasionally on an evolutionary timescale, would not display such bias. We did not detect any difference between complementary transition rates for the cryptic gene *celC* in *E. coli* (*27*), although the sample size is admittedly small (five C→T changes compared with four G→A changes; $P > 0.5$). A role for transcription-coupled repair in the evolution of enterobacterial genes has two implications regarding the process of mutation: (i) The rates of certain mutations will decline with increasing levels of gene expression, as recently suggested (*28*, *29*), because frequent transcription increases the opportunity for transcription-coupled repair; and (ii) DNA damage, rather than spontaneous replication errors, causes a substantial fraction of naturally occurring mutations. The excess of C→T over G→A substitutions represents nearly 20% of all changes in the genes analyzed, making this a minimum estimate of all naturally occurring mutations attributable to unrepaired DNA damage.

REFERENCES AND NOTES

1. J. D. Roberts, S. Izuta, D. C. Thomas, T. A. Kunkel, *J. Biol. Chem.* **269**, 1711 (1994).
2. S. Izuta, J. D. Roberts, T. A. Kunkel, *ibid.* **270**, 2595 (1995).
3. T. Q. Trinh and R. R. Sinden, *Nature* **352**, 544 (1991).
4. X. Veaute and R. P. P. Fuchs, *Science* **261**, 598 (1993).
5. A. Kornberg and T. Baker, *DNA Replication* (Freeman, New York, ed. 2, 1992).
6. C.-I. Wu and N. Maeda, *Nature* **327**, 169 (1987).
7. A. R. Fersht and J. W. Knill-Jones, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4251 (1981).
8. M. Bulmer, *J. Mol. Evol.* **33**, 305 (1991).
9. C.-I. Wu, *Nature* **352**, 114 (1991).
10. E. F. Boyd, K. Nelson, F.-S. Wang, T. S. Whittam, R. K. Selander, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1280 (1994).
11. K. Nelson and R. K. Selander, *J. Bacteriol.* **174**, 6886 (1992).
12. K. Nelson, T. S. Whittam, R. K. Selander, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6667 (1991).
13. K. Nelson and R. K. Selander, *ibid.* **91**, 10227 (1994).
14. W. P. Maddison and D. R. Maddison, *MacClade v3.0. Analysis of Phylogeny and Character Evolution* (Sinauer, Sunderland, MA, 1992).
15. D. L. Swofford, *Phylogenetic Analysis Using Parsimony (PAUP), Version 3.0s* (Illinois Museum of Natural History, Champaign, IL, 1991).
16. P. M. Sharp and W.-H. Li, *Mol. Biol. Evol.* **4**, 222 (1987).
17. T. Ikemura, *ibid.* **2**, 13 (1985).
18. A. Skandalis, B. N. Ford, B. W. Glickman, *Mutat. Res.* **314**, 21 (1994).
19. T. Lindahl and B. Nyberg, *Biochemistry* **13**, 3405 (1974).
20. L. A. Frederico, T. A. Kunkel, B. R. Shaw, *ibid.* **32**, 6523 (1993).
21. A. Datta and S. Jinks-Robertson, *Science* **268**, 1616 (1995).
22. I. Mellon and P. C. Hanawalt, *Nature* **342**, 95 (1989).
23. C. Coulondre and J. H. Miller, *J. Mol. Biol.* **117**, 577 (1977).
24. A. R. Oller, I. J. Fijalkowski, R. L. Dunn, R. M. Schaaper, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11036 (1992).
25. G. J. Phillips, J. Arnold, R. Ivarie, *Nucleic Acids Res.* **15**, 2611 (1987).
26. C. S. Downes, A. J. Ryan, R. T. Johnson, *Bioessays* **15**, 209 (1993).
27. B. G. Hall and P. M. Sharp, *Mol. Biol. Evol.* **9**, 654 (1992).
28. A. Eyre-Walker and M. Bulmer, *Genetics* **140**, 1407 (1995).
29. O. Berg and M. Martelius, *J. Mol. Evol.* **41**, 449 (1995).
30. R. K. Selander, J. Li, E. F. Boyd, F.-S. Wang, K. Nelson, in *Bacterial Systematics and Diversity*, F. G. Preist, A. Ramos-Cormenzana, R. Tindall, Eds. (Plenum, New York, 1994), pp. 17–49.
31. R. F. DuBose, D. E. Dykhuizen, D. L. Hartl, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7036 (1991).
32. Sequences for *gutB* yielded several equally parsimonious trees with different numbers of C→T substitutions: The frequency of C→T used in Fig. 2 is based on the lowest value, which would serve to underestimate the actual degree of asymmetry.

33. We thank members of our laboratories for comments on the manuscript. Supported by grants from the NIH to L.C., M.A.R., and H.O.

8 December 1995; accepted 26 February 1996

Similarity Among the *Drosophila* (6-4)Photolyase, a Human Photolyase Homolog, and the DNA Photolyase–Blue-Light Photoreceptor Family

Takeshi Todo,* Haruko Ryo, Kazuo Yamamoto, Hiroyuki Toh, Taiichiro Inui, Hitoshi Ayaki, Taisei Nomura, Mituo Ikenaga

Ultraviolet light (UV)–induced DNA damage can be repaired by DNA photolyase in a light-dependent manner. Two types of photolyase are known, one specific for cyclobutane pyrimidine dimers (CPD photolyase) and another specific for pyrimidine (6-4) pyrimidone photoproducts [(6-4)photolyase]. In contrast to the CPD photolyase, which has been detected in a wide variety of organisms, the (6-4)photolyase has been found only in *Drosophila melanogaster*. In the present study a gene encoding the *Drosophila* (6-4)photolyase was cloned, and the deduced amino acid sequence of the product was found to be similar to the CPD photolyase and to the blue-light photoreceptor of plants. A homolog of the *Drosophila* (6-4)photolyase gene was also cloned from human cells.

Cyclobutane pyrimidine dimers (CPDs) and pyrimidine (6-4) pyrimidone photoproducts [(6-4)photoproducts] are the two major classes of cytotoxic, mutagenic, and carcinogenic photoproducts produced in DNA when cells are irradiated with UV light (*1*–*3*). The phenomenon of photoreactivation—the reduction of lethal and mutagenic effects of UV radiation by simultaneous or subsequent irradiation with near UV or visible light—has been identified in a variety of organisms (*1*, *4*). The enzyme responsible, called DNA photoreactivating enzyme (CPD photolyase), repairs CPDs by reverting them to normal bases using light energy (*4*, *5*). We discovered another type of photolyase in *Drosophila melanogaster* (*6*, *7*) that catalyzes the light-dependent repair of (6-4)photoproducts rather than CPDs. To investigate the mechanism by which the (6-4)photoproducts are repaired, we cloned the *Drosophila* (6-4)photolyase gene.

Escherichia coli normally do not photoreactivate (6-4)photoproducts (*7*) and thus would be expected to show an increased resistance to UV light when engineered to

express a (6-4)photolyase in the presence of photoreactivating light. This functional complementation assay enabled us to clone the (6-4)photolyase complementary DNA (cDNA) from a *Drosophila* expression library. *Escherichia coli* cells carrying the plasmid containing the *E. coli phr*$^+$ gene (encoding CPD photolyase) were used as the host cells for screening the library. In those cells, CPDs were repaired efficiently by CPD photolyase in the presence of photoreactivating light; thus, the amount of (6-4)photoproduct alone determined their sensitivity to UV light. After repeated UV irradiation of *E. coli* transformed with the *Drosophila* cDNA library followed by visible-light illumination, we isolated a cDNA clone (pDm64PR) (*8*) that conferred light-dependent UV resistance on *recA*$^-$ *uvrA*$^-$ *phr*$^+$ *E. coli* (Fig. 1).

The cDNA was 1960 base pairs (bp) long and contained a single long open reading frame encoding 540 amino acids with a calculated molecular mass of 62.9 kD. Northern (RNA) blot analysis showed that a 2-kb transcript corresponding to the cDNA was expressed at high levels only in adult ovary (*9*). This expression pattern is consistent with our previous work showing that the activity levels of *Drosophila* (6-4)photolyase are highest in ovary and embryo (*10*). The cDNA hybridized to band 39 B-C of *Drosophila* polytene chromosome (*9*).

To verify that the cDNA coded for (6-4)photolyase, we measured the enzymatic activity of the recombinant protein expressed in *E. coli*. The recombinant protein purified from *E. coli* crude extracts by affinity chromatography with UV-irradiated DNA attached to beads (*11*) had a molecular mass of 62 kD (Fig. 2A). As seen in Fig. 2B, the recombinant protein bound to the UV-irradiated DNA probe TC-3 and to probes from which CPDs were removed with *E. coli* photolyase, but not to probes from which (6-4)photoproducts were removed with *Drosophila* (6-4)photolyase, thus showing that the binding activity of the recombinant protein is specific for (6-4)photoproduct.

To investigate whether the recombinant protein could photorepair the (6-4)photoproduct, we added the UV-irradiated DNA probe to the recombinant protein and exposed the mixture to fluorescent light. Neither purified *Drosophila* (6-4)photolyase nor the recombinant protein bound to the DNA probe (Fig. 2C). This absence of binding was a result of the loss of (6-4)photoproduct from the UV-irradiated DNA probe, because the binding of the monoclonal antibody specific for (6-4)photoproduct (64M-2) (*12*) was also lost. The (6-4)photoproduct photoreactivating activity of the recombinant protein was quantified by means of an enzyme-linked immunosorbent assay (ELISA) (Fig. 2D). Binding of the 64M-2 antibody to UV-irradiated DNA decreased with increasing amounts of the recombinant protein, whereas binding of the TDM-2 antibody, which specifically recognizes CPDs (*12*), did not change. Together, these results show that the cDNA cloned in pDm64PR encodes the (6-4)photolyase.

There was a striking and unexpected amino acid sequence similarity between the *Drosophila* (6-4)photolyase, class I CPD photolyases, and the blue-light photoreceptors (Fig. 3). The CPD photolyase gene has been isolated from 14 organisms (*4*, *13*), and on the basis of deduced amino acid sequence similarities, the genes have been grouped in two classes (*13*). The photolyase genes from microorganisms belong to class I and those from higher eukaryotes belong to class II (Fig. 3). One exception to this rule is the gene from the archaebacterium *Methanobacterium thermoautotrophicum*, which belongs to class II. The blue-light photoreceptor is the essential light detector for early plant development (*14*). Although this protein

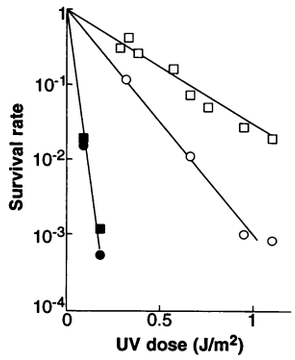

Fig. 1. Effects of photoreactivation on the survival of UV-irradiated *E. coli* SY32 (pRT2) cells (*8*) carrying an empty vector pSPORTS (circles) or pDm64PR (squares). After UV irradiation, the *E. coli* cells were kept in the dark (closed symbols) or illuminated (open symbols).

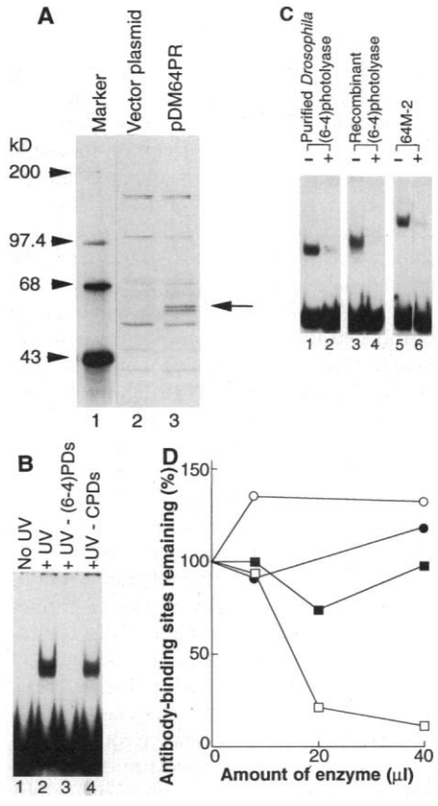

Fig. 2. Photoreactivating activity of cloned (6-4)photolyase expressed in *E. coli*. (**A**) Protein gel showing purified recombinant *Drosophila* (6-4)photolyase from *E. coli* (*11*). Cell extracts were prepared from *E. coli* carrying an empty vector (lane 2) or the recombinant plasmid pDm64PR (lane 3), and recombinant protein was purified by affinity chromatography with UV-irradiated DNA bound to beads. A 6-μl sample of each preparation was resolved by SDS–polyacrylamide gel electrophoresis and the proteins detected by silver staining. Lane 1, molecular size markers. The arrow indicates the 62-kD protein band (*19*). (**B**) Gel shift analysis showing (6-4)photoproduct–specific binding activity of recombinant (6-4)photolyase. UV-irradiated TC-3 DNA probe (*20*) was mixed with purified *Drosophila* (6-4)photolyase (*21*) (lane 3) or *E. coli* photolyase (lane 4) and exposed to fluorescent light for 30 min. The reaction mixture was extracted with phenol, and probe DNA was recovered by ethanol precipitation. The DNA was assayed for protein binding by gel shift with recombinant (6-4)photolyase. Lane 1, unirradiated control. Lane 2, UV-irradiated but no photoreactivating treatment. (6-4)PDs, (6-4)photoproducts). (**C**) Gel shift analysis showing (6-4)photoproduct–specific photoreactivating activity of recombinant (6-4)photolyase. UV-irradiated TC-3 DNA probe (*20*) was exposed to fluorescent light for 30 min with (+) (lanes 2, 4, and 6) or without (−) (lanes 1, 3, and 5) recombinant *Drosophila* (6-4)photolyase. The probe DNA was deproteinized and assayed for protein binding by gel shift with purified *Drosophila* (6-4)photolyase (lanes 1 and 2), recombinant *Drosophila* (6-4)photolyase (lanes 3 and 4), and the antibody directed against (6-4)photoproduct (64M-2, lanes 5 and 6). (**D**) Disappearance of the binding site for the (6-4)photoproduct–specific antibody in UV-irradiated DNA. Repair of UV damage in the photoreactivated DNA was quantified by ELISA by using the antibody specific for (6-4)photoproduct (64M-2, open symbols) or for CPDs (TDM-2, closed symbols). UV-irradiated salmon sperm DNA was mixed with recombinant *Drosophila* (6-4)photolyase (square) or vector plasmid extract (circle) (*22*).

T. Todo, T. Inui, H. Ayaki, M. Ikenaga, Radiation Biology Center, Kyoto University, Yoshida-konoecho, Sakyo-ku, Kyoto 606-01, Japan.
H. Ryo and T. Nomura, Department of Radiation Biology, Faculty of Medicine, Osaka University, 2-2 Yamada-oka, Suita, Osaka 565, Japan.
K. Yamamoto, Department of Biology, Faculty of Science, Tohoku University, Aramaki, Aoba-ku, Sendai 980-77, Japan.
H. Toh, Department of Biochemical Engineering and Science, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Iizuka, Fukuoka 820, Japan.
*To whom correspondence should be addressed.

does not show CPD photoreactivating activity, it has strong sequence similarity to the class I CPD photolyases (*14*, *15*); thus, these proteins may be considered as composing a CPD photolyase–blue-light photoreceptor family. The *Drosophila* (6-4)photolyase has 20 to 22% sequence identity with the class I CPD photolyase and 22 to 24% identity with the blue-light photoreceptor over the entire protein. The *Drosophila* (6-4)photolyase also has less but significant similarity with the class II CPD photolyase, with sequence identity as high as 13% (Fig. 3).

Amino acid sequence comparison reveals that the *Drosophila* (6-4)photolyase, the class I CPD photolyases, and the blue-light photoreceptors are closely related. The homology with the CPD photolyases is unexpected because the structures of the (6-4)photoproducts and CPDs are very different (*3*), and thus the mechanism of photoreactivation was thought to be different (*16*). Both CPD photolyases and the blue-light photoreceptors use flavin-adenine dinucleotide ($FADH^-$) as the catalytic cofactor (*5*, *15*, *17*). The crystal structure of the CPD photolyase from *E. coli* has shown that the enzyme consists of two major domains, an α/β domain and a helical domain, the latter of which contains the FAD binding site (*18*). This domain corresponds to the highly conserved region between the *Drosophila* (6-4)photolyase and the CPD photolyase–blue-light photoreceptor (regions c and d underlined in Fig. 3). Furthermore, the amino acids predicted to interact with FAD in the *E. coli* enzyme are conserved in the *Drosophila* (6-4)photolyase (marked with an asterisk in Fig. 3). These results suggest that the *Drosophila* (6-4)photolyase also uses $FADH^-$ as the catalytic cofactor, and that the basic mechanism for receiving and converting light energy may be the same as that of the CPD photolyase and blue-light photoreceptor.

A homology search of the National Center for Biotechnology Information–expressed sequence tag (EST) human cDNA database using the BLAST program revealed a partial sequence, R19031, that was highly related to the *Drosophila* (6-4)photolyase. Polymerase chain reaction primers were designed from the EST and used to generate a probe for screening a human brain cDNA library. A 2-kb cDNA was cloned, and after extension of the cDNA by the rapid amplification of 5′ cDNA end (5′-RACE method), a 3-kb cDNA was isolated and sequenced. The predicted amino acid sequence of the human protein had 48% identity with the *Drosophila* (6-4)photolyase over the entire protein (Fig. 3); therefore, we refer to it as human (6-4)photolyase homologous protein. Northern blot analysis indicated that the transcript is expressed in multiple tissues (9).

Living organisms use light energy in various ways. Proteins classified within the DNA photolyase–blue-light photoreceptor family are responsible not only for protecting organisms from the damaging effects of UV light, but also for providing an early developmental signal in plants. Although proteins in this family are functionally divergent, they have a common mechanism for receiving and converting light energy in a variety of phenomena by using blue light (wavelength of 350 to 450 nm). The biological function of the human (6-4)photolyase homolog is an interesting puzzle that remains to be solved.

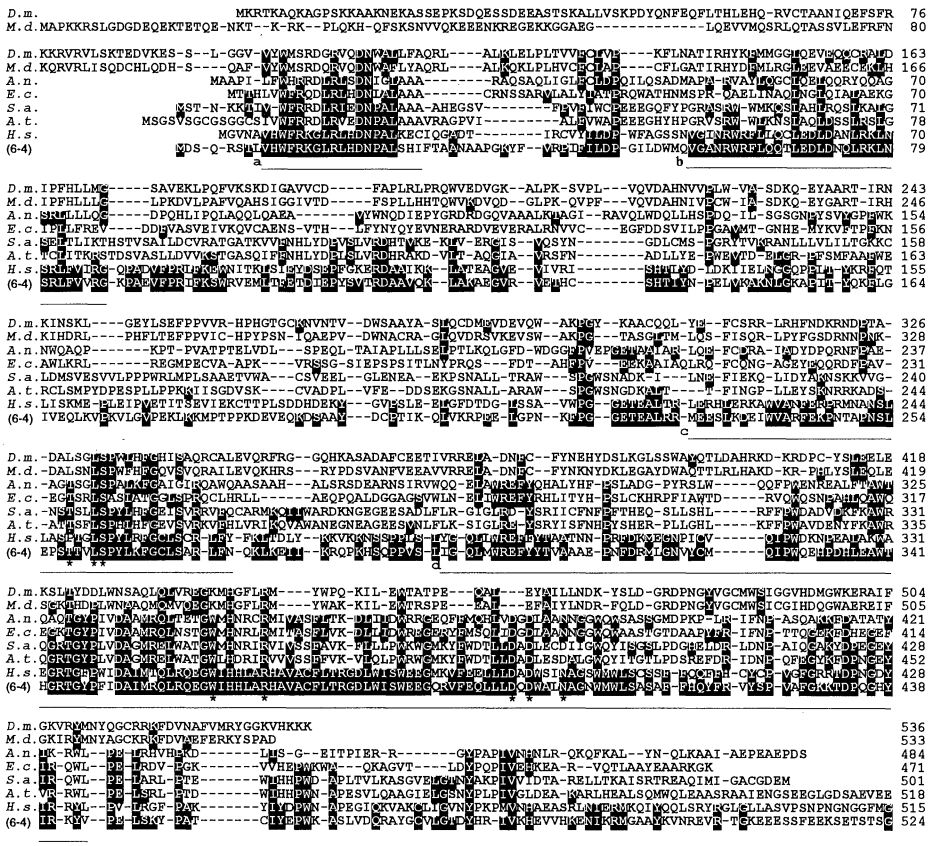

Fig. 3. Comparison of the amino acid sequence of *Drosophila* (6-4)photolyase and its human homolog with the CPD photolyase–blue-light photoreceptor family (*4*, *13*, *14*, *15*, *23*). The amino acid sequences of the *Drosophila* (6-4)photolyase [listed as (6-4)] and the human homolog (*H.s.*, *Homo sapiens*) were aligned with class II CPD photolyases (*D.m.*, *Drosophila melanogaster*; *M.d.*, *Monodelphis domestica*), class I CPD photolyases (*A.n.*, *Anacystis nidulans*; *E.c.*, *E. coli*), and with the blue-light photoreceptors of plants (*S.a.*, *Sinapis alba*; *A.t.*, *Arabidopsis thaliana*). Amino acid identities with *Drosophila* (6-4)photolyase are shown as white letters on a black background. The photolyases of *Saccharomyces cerevisiae*, *Halobacterium halobium*, *Streptomyces griseus*, *Potorous tridactylis*, *Oryzias latipes*, *Carassius auratus*, *Bacillus firmus* (*24*), *Neurospora crassa*, *Salmonella typhimurium*, and *Methanobacterium thermoautotrophicum* and the blue-light photoreceptor of *Chlamydomonas reinhardtii* (*25*) were also analyzed for the sequence similarity. The four conserved regions of the alignment were underlined (a, b, c, and d). Amino acids predicted to interact with FAD are marked with an asterisk below the alignment. The amino acid sequences were aligned by a program with a tree-based algorithm (*26*) and Dayhoff's score table (MDM78) (*27*). The obtained alignment was modified by visual inspection to increase sequence similarity. A dash indicates a gap for insertion or deletion of amino acid residues. The nucleotide sequences for the *Drosophila* (6-4)photolyase and the human homolog have been deposited in GenBank (accession numbers D83701 and D83702, respectively). Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

REFERENCES AND NOTES

1. W. Harm, *Biological Effects of Ultraviolet Radiation* (Cambridge Univ. Press, Cambridge, 1980); E. C. Freidberg, G. C. Walker, W. Siede, *DNA Repair and Mutagenesis* (ASM Press, Washington, DC, 1995).
2. D. E. Brash, *Photochem. Photobiol.* **48**, 59 (1988).
3. D. L. Mitchell and R. S. Nairn, *ibid.* **49**, 805 (1989).
4. G. B. Sancar, *Mutat. Res.* **236**, 147 (1990).
5. A. Sancar, *Biochemistry* **33**, 2 (1994).
6. T. Todo and H. Ryo, *Mutat. Res.* **273**, 85 (1992).
7. T. Todo *et al.*, *Nature* **361**, 371 (1993).
8. The cDNA library was constructed in pSPORTS1 (BRL) by standard procedures, with ovary cDNA from a *Drosophila phr*$^-$ strain as standard material. The *E. coli* strain SY32 (JM107 *phr::Cm*^{r} *uvrA::Kan*^{r}

recA) carrying plasmid pRT2 was used as the host cell for screening. Plasmid pRT2 was constructed by inserting a 3.3-kb Hinc II fragment containing the *E. coli phr*$^+$ gene into the Pvu II site of pACYC184 (Bio Labs). The cDNA library was introduced into *E. coli* strain SY32 (pRT2) and incubated with 1 mM isopropylthio-β-galactoside (IPTG) for 1 hour. After plating on LB agar plates, transformed cells were exposed to 0.1 to 0.6 J/m^2 of UV light from a germicidal lamp and then illuminated for 15 min with visible light at a distance of 24 cm from a FL15D lamp (National) with a 7-mm thick soft glass filter (PR treatment). Surviving cells were collected from the plates the next day and were subjected to a second round of selection with UV plus PR treatment. After 13 rounds of selection, four single colonies were selected from the illuminated plate, and their UV sensitivity with or without PR treatment was determined. Plasmid DNA was isolated from one of the light-dependent UV-resistant colonies and designated pDm64PR.

9. T. Todo *et al.*, data not shown.
10. T. Todo *et al.*, *Mutat. Res.* **315**, 213 (1994).
11. *Drosophila* (6-4)photolyase was expressed in *E. coli* as a lacZ–α peptide fusion protein. To prepare recombinant *Drosophila* (6-4)photolyase, we incubated *E. coli* SY2 (*phr::Cm*^{r} *uvrA::Kan*^{r} *recA::Tet*^{r}) carrying pDm64PR in 150-ml cultures with 1 mM IPTG for 12 hours and then obtained lysates by sonication of bacterial pellets in sonic buffer [50 mM tris-HCl (pH 7.4), 1 mM EDTA, and 10% glycerol]. The lysates were cleared by centrifugation at 12,000*g* for 20 min. To purify protein that bound UV-irradiated DNA, we applied the cleared lysate to UV-treated DNA attached to beads. UV-DNA affinity beads were prepared as follows. Salmon sperm DNA [500 μg, previously digested with Hind III and irradiated with UV light (25 kJ/m^2)] was labeled with 500 mg of photobiotin (Vector Laboratory) by irradiation in an ice bath 10 cm below a sunlamp (Toshiba, FL20SE) with a UV-34 filter (Hoya, Tokyo). Biotin-labeled DNA was bound to streptoavidin paramagnetic beads (Promega). Approximately 400 μg of UV-irradiated DNA was bound to beads. The cleared lysates were mixed with the UV-DNA affinity beads in 1 ml of binding buffer [salmon sperm DNA (1 mg/ml), 200 mM NaCl, 50 mM tris-HCl (pH 7.4), 1 mM EDTA, and 10% glycerol]. After 15 min to allow for binding, the beads were washed twice with 1 ml of washing buffer [350 mM NaCl, 50 mM tris-HCl (pH 7.4), 1 mM EDTA, and 10% glycerol], and protein bound to the beads was eluted with 100 μl of elution buffer [1 M NaCl, 50 mM tris-HCl (pH 7.4), 1 mM EDTA, and 10% glycerol]. After repeating the elution process three times, each 100 μl of eluted solution was combined, and the resultant 300 μl of solution was used as the recombinant *Drosophila* (6-4)photolyase or control plasmid extract.
12. T. Matsunaga *et al.*, *Photochem. Photobiol.* **57**, 934 (1993).
13. A. Yasui *et al.*, *EMBO J.* **13**, 6143 (1994).
14. M. Ahmad and A. R. Cashmore, *Nature* **366**, 162 (1993).
15. K. Malhotra *et al.*, *Biochemistry* **34**, 6892 (1995).
16. S.-T. Kim *et al.*, *J. Biol. Chem.* **269**, 8535 (1994).
17. C. Lin *et al.*, *Science* **269**, 968 (1995).
18. H.-W. Park, S.-T. Kim, A. Sancar, J. Deisenhofer, *ibid.* **268**, 1866 (1995).
19. The recombinant protein migrated as a doublet of ~62 kD. The lower band may represent a degradation product.
20. A 73-bp DNA fragment enriched in the sequence TC (TC-3) was irradiated with 15 kJ/m^2 of UV light and then used in the gel shift assay.
21. The *Drosophila* (6-4)photolyase was purified from the embryo of *Drosophila phr*$^-$ strain, as in (*7*).
22. Salmon sperm DNA irradiated with UV light (8 J/m^2) and purified recombinant (6-4)photolyase were mixed in 400 μl of reaction buffer and exposed to fluorescent light for 30 min. The mixture was phenol extracted, ethanol precipitated, and used for the ELISA assay (*7*, *10*).
23. T. Kato Jr. *et al.*, *Nucleic Acids Res.* **22**, 4119 (1994).
24. K. Malhotra, S.-T. Kim, A. Sancar, *Biochemistry* **33**, 8712 (1994).
25. G. D. Small, B. Min, P. A. Lefebrre, *Plant Mol. Biol.* **28**, 443 (1995).
26. G. J. Barton and M. J. E. Sternberg, *J. Mol. Biol.* **198**, 327 (1987).
27. R. M. Schwartz and M. O. Dayhoff, in *Atlas of Protein Sequence and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Washington, DC, 1978), pp. 353–359.
28. We thank H. Iwasaki for critical reading of the manuscript. Supported by the Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science, and Culture of Japan (numbers 05270101 and 07263240). We also thank the Integrated Molecular Analysis of Genome Expression (IMAGE) consortium for the use of their resources.

25 January 1996; accepted 5 March 1996

Silk Properties Determined by Gland-Specific Expression of a Spider Fibroin Gene Family

Paul A. Guerette, David G. Ginzinger, Bernhard H. F. Weber,* John M. Gosline†

Spiders produce a variety of silks that range from Lycra-like elastic fibers to Kevlar-like superfibers. A gene family from the spider *Araneus diadematus* was found to encode silk-forming proteins (fibroins) with different proportions of amorphous glycine-rich domains and crystal domains built from poly(alanine) and poly(glycine-alanine) repeat motifs. Spiders produce silks of different composition by gland-specific expression of this gene family, which allows for a range of mechanical properties according to the crystal-forming potential of the constituent fibroins. These principles of fiber property control may be important in the development of genetically engineered structural proteins.

Individual spiders generate up to seven mechanically distinct silk fibers by drawing liquid-crystalline proteins (*1*) from separate gland-spinneret complexes (*2*), but it is not known how spiders modulate the mechanical properties of silks. Silks are macromolecular composites of amorphous protein domains that are cross-linked and reinforced by β-sheet microcrystals; the degree of crystalline cross-linking and reinforcement largely determines functionally important mechanical properties (*3*, *4*). For example, spider dragline silk, which forms the spider's safety line and the frame of its web, contains 20 to 30% crystal by volume (*4*, *5*), forming a fiber that is stiff (initial Young's modulus, 10 GPa), strong (tensile strength, 1.5 GPa), and tough (energy to break, 150 MJ m^{-3}) (*4*). Spider viscid silk, which forms the spiral, glue-coated, capture portion of the web, contains ≤5% crystal by volume (*6*) and is mechanically similar to a lightly cross-linked rubber, with low stiffness (initial Young's modulus, 3 MPa) and high extensibility (*4*).

Factors that influence the formation and size of the crystals include the primary and secondary structure of the proteins, control of the genes that encode these proteins, and the chemical and mechanical processing of the proteins during spinning. X-ray diffraction studies on silks and synthetic polypeptides indicate that the amino acid sequence motifs that form β-sheet crystal domains are either poly(alanine) repeats (for example, spider silk) (*4*, *5*, *7*) or repeat motifs where glycine alternates with either alanine or serine (for example, the GAGAGS repeat in *Bombyx mori* cocoon silk) (*8*, *9*). In addition, a large fraction of the protein in silk appears noncrystalline or "amorphous" by x-ray diffraction (*8*), and therefore other sequence motifs may direct the organization of amorphous domains. In the fibroins of the spider *Nephila clavipes*, the crystal-forming poly(alanine) blocks are separated by glycine-rich domains that are similar to the β turn–forming motifs in elastomeric proteins (*10*). We will refer to these as amorphous domains, although they may contain a degree of local order.

To investigate the control of crystallinity in silks, we examined the amino acid sequences of the silk proteins from the orb-weaving spider *Araneus diadematus*. Complementary DNA (cDNA) libraries were constructed from the major ampullate gland (MA; dragline silk), the flagelliform gland (FL; viscid silk), and the cylindrical gland (CY; cocoon silk) (*11*). These libraries were screened with probes that were based on the previously cloned spidroin-1 and spidroin-2 from the spider *N. clavipes* (*12*), hereafter called NCF-1 and NCF-2. Four distinct, partial silk cDNAs were isolated, each of which contained silklike protein repeat motifs, 78 to 101 residues of a nonsilk COOH-terminal protein domain, and a unique 3′ untranslated region (UTR). The silklike protein repeats of the four cDNAs are shown in Fig. 1 and are named ADF-1 through ADF-4 (ADF, *A. diadematus* fibroin).

Of the silklike protein encoded by the

P. A. Guerette and J. M. Gosline, Department of Zoology, University of British Columbia, 6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada.
D. G. Ginzinger and B. H. F. Weber, Department of Medical Genetics, University of British Columbia, 2125 East Mall, Vancouver, British Columbia V6T 1Z4, Canada.

*Present address: Molekulare Humangenetik, University of Würzburg, D-97074 Würzburg, Germany.
†To whom correspondence should be addressed.