Cloning the *Drosophila* homolog of the xeroderma pigmentosum complementation group C gene reveals homology between the predicted human and *Drosophila* polypeptides and that encoded by the yeast *RAD4* gene

Karla A. Henning, Carolyn Peterson¹, Randy Legerski¹ and Errol C. Friedberg*

Laboratory of Molecular Pathology, Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX 75235 and ¹Department of Molecular Genetics, University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA

Received December 17, 1993; Accepted December 21, 1993

EMBL accession no. Z28622

ABSTRACT

A human xeroderma pigmentosum group C (XPC) cDNA has been previously isolated by functional complementation (Legerski and Peterson, *Nature*, 359, 70–73, 1992). Sequence analysis did not reveal protein motifs which might suggest a possible biochemical function for the putative XPC protein. In order to identify functional domains in the translated XPC sequence the homologous gene from *Drosophila melanogaster*, designated *XPC^{DM}*, was cloned by DNA hybridization. Sequence analysis of an apparently full-length cDNA revealed an open reading frame which can encode a predicted polypeptide of 1293 amino acids. Significant homology of the C-terminal 346 amino acids with both the human XPC and *Saccharomyces cerevisiae* Rad4 protein sequences is observed, suggesting that these proteins are functional homologs.

INTRODUCTION

Xeroderma pigmentosum (XP) is an autosomal recessive disorder which is characterized by sensitivity to sunlight and a dramatically increased risk of cancer of sun-exposed areas of the body, including the skin, eyes and tongue (1). Cells from XP patients are hypersensitive to ultraviolet (UV) light and UV-mimetic chemicals, and are defective in nucleotide excision repair following exposure to these agents (1). Cell fusion studies have identified at least eight genetic complementation groups in XP, designated A–G for the classical form of the disease and V for the variant form (2).

Biochemical studies have shown that non-dividing cells from XP group C patients can efficiently repair UV-induced damage in actively transcribed DNA, but appear to be deficient in the repair of lesions in transcriptionally inactive regions of the genome, and on the non-transcribed strand of active genes (3, 4). An XPC cDNA was cloned by complementation of XP-C

cells in culture using an Epstein–Barr virus-based vector system (5). This cDNA clone is 3.8 kb in length and can encode a predicted protein of 823 amino acids (5, 6). The putative XPC protein is hydrophilic, with 34.1% charged residues, and is also quite basic (5, 6). Northern analysis failed to identify a transcript in mRNA isolated from many XP-C patients (5). Recently, another XPC cDNA has been independently cloned and sequenced. This cDNA is 3.6 kb in length and adds an additional 117 amino acids to the N-terminus of the predicted protein sequence [Dr Chikahide Masutani, The Institute of Physical and Chemical Research (RIKEN), Japan, personal communication].

Several genes from the yeast *S. cerevisiae* have been shown to be homologs of human genes which are defective in patients with the hereditary disease XP. Identification of the yeast homolog of XPC might therefore be expected to provide insights into the functional role of the XPC protein and the molecular pathology of XP. Limited homology between a region of the predicted XPC protein sequence and that of the *S. cerevisiae* Rad4 sequence was noted previously, but the significance of this homology was uncertain (5).

In order to further evaluate the significance of the tentative homology between the human XPC and yeast Rad4 proteins we elected to isolate the *Drosophila* homolog of the XPC gene so that the polypeptide region of interest could be evaluated by comparing all three sequences. In addition, the availability of the cloned *Drosophila* gene would provide an opportunity to determine whether a yeast gene other than *RAD4* might be the true functional homolog of XPC. *Drosophila* homologs of human excision repair genes have been isolated previously. The *D. melanogaster* homolog of the human XPA protein shares 45% amino acid identity (64% similarity) with the human sequence (7). Additionally, the *D. melanogaster* homolog of the human *XPB* gene has been identified as the *haywire* gene and defects in this gene have been shown to confer UV radiation sensitivity and central nervous system defects in some cases (8). The predicted *haywire* polypeptide shares 66% amino acid identity with the

*To whom correspondence should be addressed

human protein and 50% identity with its *S. cerevisiae* counterpart, Ssl2 protein (8). One might therefore expect that the human *XPC* gene is also conserved between humans, *Drosophila* and yeast. We have cloned the *D. melanogaster* homolog of the human *XPC* gene, designated *XPC^{DM}*, and show that the C-terminal 346 amino acids of the predicted polypeptide share significant homology with similar regions of both the translated human *XPC* and yeast *RAD4* genes. No other yeast gene has been identified as a more conserved homolog of the human *XPC* and *Drosophila XPC^{DM}* genes.

MATERIALS AND METHODS

Southern hybridization

Restriction digestion of the plasmid pXPC3 (5) with *Sfi*I releases a 3.5 kb fragment of the human *XPC* cDNA sequence. This fragment was purified on a 0.7% Sea Plaque gel (FMC) and labeled directly (9) with [α -³²P]-dATP and [α -³²P]-dCTP by the random primer method (10). Genomic DNA was isolated from the SV40-transformed human cell line VA13, the mouse cell line A9, *Schizosaccharomyces pombe* strain 972h- and *S. cerevisiae* strain SX46. Chicken and *Drosophila* genomic DNA were from Clontech. Southern transfers were made to positively charged nylon filters and hybridized at 45°C in Hybrisol I (Oncor). Heterologous hybridizations were washed to a final stringency of 2×SSPE, 0.1% SDS at 45°C, and identical sequences were washed to a final stringency of 2×SSPE, 1.0% SDS at 68°C.

Genomic library screening

A *D. melanogaster* 6–16 h embryo (Canton S wild-type) genomic library in 1 vector Charon 4A (11) was obtained from the American Type Culture Collection (ATCC 37332). Plaque lifts were made with Sureblot discs (Oncor) and processed by established procedures (12). Filters were prehybridized for approximately 3 h at 65°C in 5×SSPE, 5×Denhardt's, 0.5 % SDS. Then 10⁶ cpm/ml of the *XPC* probe was boiled with 25–50 µg/ml single-stranded herring sperm DNA, added to the prehybridization mix and the filters were hybridized overnight at 65°C in a shaking water bath. The filters were washed twice at room temperature and once at 45°C in 2×SSPE, 0.1% SDS,

and autoradiographed overnight at –80°C with an intensifying screen.

cDNA library screening

A *D. melanogaster* 0–4 h embryo (strain *dp cn bw*) cDNA library in vector pNB40 (13) was kindly provided by Dr Steve Wasserman, Department of Biochemistry, UT Southwestern Medical Center. Plaque lifts were made using Hybond N+ (Amersham) and processed as suggested by the manufacturer. Filters were hybridized with the *XPC^{DM}*sub68 probe as described above for genomic library screening, and washed to a final stringency of 2×SSPE, 1% SDS at 68°C.

Sequence analysis

The insert contained in pcD1.3 was completely sequenced on both strands using Sequenase version 2.0 (United States Biochemical). The 5' and 3' ends of the cDNA were sequenced using primers to vector sequences. Subclones were similarly sequenced using pUC forward and reverse primers. The remaining gaps were filled using sequence-derived primers.

RESULTS

Cloning and nucleotide sequence of the *Drosophila XPC^{DM}* cDNA

The human *XPC* cDNA was used to probe Southern blots of human, mouse, chicken, *Drosophila*, *Schizosaccharomyces pombe* and *S. cerevisiae* genomic DNA. Hybridizing fragments were observed with human, mouse, chicken, and *Drosophila* DNA (Fig. 1). A clone was isolated from a *Drosophila* genomic library by hybridization with the human *XPC* cDNA. Southern analysis of this genomic clone revealed a 4.0 kb *Eco*RI fragment which hybridized to the human *XPC* cDNA. This fragment size was in complete agreement with that originally observed on Southern blots of *Drosophila* genomic DNA (Fig. 1). This 4.0 kb fragment was subcloned into pBGS19 (14) and was designated *XPC^{DM}*sub68. Subcloning and further analysis localized the hybridizing region to an 800 bp *Hind*III fragment designated 68sub1 (Fig. 2). Sequence analysis confirmed strong nucleotide sequence conservation between 68sub1 and the human *XPC* cDNA over the region of the human *XPC* sequence previously

Figure 1. Evolutionary conservation of the *XPC* gene. DNA was restricted with *Eco*RI enzyme and subjected to conventional Southern analysis during which genomic DNA (10 µg per lane) from various species was hybridized with a human *XPC* cDNA probe. Lane 1, human genomic DNA; lane 2, mouse genomic DNA; lane 3; *D. melanogaster* genomic DNA; lane 4, chicken genomic DNA. Lanes 1 and 2 were exposed for 16 h at room temperature; lanes 3 and 4 were exposed for 24 h at –80°C with an intensifying screen.

Figure 2. Restriction map of the *Drosophila* genomic clone *XPC^{DM}*sub68, and the *XPC^{DM}* cDNA clone, pcD1.3. The genomic subclone 68sub1, which hybridized to the human *XPC* cDNA, is indicated by a dark box. Dashed lines identify restriction sites conserved between the genomic and cDNA clones (H, *Hind*III; R, *Eco*RI).

Figure 4. Homology of the XPC^{DM} protein to the human XPC and yeast Rad4 protein sequences. (A) Diagrammatic representation of the yeast Rad4, *D. melanogaster* XPC^{DM} and human XPC protein sequences. Striped boxes identify the conserved portion of each peptide. (B) Alignment of the protein sequences over the evolutionarily conserved region. Identical amino acids are shaded. XPC^{DM} protein, amino acids 947–1293; human XPC protein, amino acids 517–867; *S. cerevisiae* Rad4 protein, amino acids 302–668.

is 4.3 kb long with an open reading frame of 3879 bp corresponding to a predicted polypeptide of 1293 residues (Fig. 3). The sequence AATAAA, corresponding to a consensus polyadenylation signal, is present beginning at nucleotide 4162 and the polyA tail begins at nucleotide 4199.

Comparison of the human XPC, *Drosophila* XPC^{DM} and *S. cerevisiae* RAD4 sequences

The predicted XPC^{DM} protein is very basic with an isoelectric point of 10.05 (as determined by the isoelectric program of GCG). The basic nature of XPC^{DM} is similar to that of the human XPC protein, which has an isoelectric point of 9.54. The yeast Rad4 protein is also basic, with an isoelectric point of 7.93. All three of the protein sequences contain acidic stretches, which are characteristic of nuclear proteins which associate with chromatin, including nucleolin, High Mobility Group (HMG) proteins and the *S. cerevisiae* Rad6 protein (15–18).

XPC^{DM} shows significant conservation at both the nucleotide and protein sequence level with the human XPC cDNA in the 3' region of the coding sequence, corresponding to the region of homology tentatively identified between the human XPC and the *S. cerevisiae* Rad4 proteins (Fig. 4). The C-terminal 346 amino acids of XPC^{DM} are 50% identical (68% similar) to the human XPC protein and 27% identical (48% similar) to the

Table 1. Potential bipartite nuclear targeting signals

Protein	Residue	Bipartite motif
XPC ^{DM}	922	KK apvlpsvqn IRKdR
	1195	KK tvhlrlpglm RicKK
	1275	KK likglirer IKKKy
XPC	44	KK slskvsqgk RKRgc
Rad4	31	RR vrrknqplpd aKKKf
	377	RR yaqwmnskr KRRit
	388	RK rritkddfg KwfRK

S. cerevisiae Rad4 protein (Fig. 4). The three proteins share 18.8% identity (41.6% similarity) overall in this region (Fig. 4). Pair-wise comparisons of the conserved region of the protein sequences were done with the Bestfit program from GCG. A total of 10 randomizations of the second sequence were also analyzed to determine an average score for two unrelated sequences of this length. The alignment of the yeast Rad4 and *Drosophila* XPC^{DM} protein sequences scored 18 standard deviations above the mean; the yeast Rad4 and human XPC proteins scored 16 standard deviations above the mean; and the human XPC and *Drosophila* XPC^{DM} proteins scored 64 standard deviations above the mean. These scores are statistically significant, supporting the contention that these three proteins are evolutionarily related. The sequences of the N-terminal portions of the proteins are not conserved and show considerable variation in length (Fig. 4A). Nonetheless, all three proteins contain large numbers of charged residues which are divided into acidic and basic regions, suggesting that they may be functionally important.

The human XPC and *Drosophila* XPC^{DM} proteins both contain large numbers of serine residues in the amino-terminal portion of their protein sequences, which may indicate that they are phosphorylated *in vivo*. Phosphorylation has been suggested as a possible mechanism for regulating the transport of proteins into the nucleus (19). In the XPC^{DM} sequence, consensus phosphorylation sites are present for cAMP- and cGMP-dependent protein kinase, casein kinase II, protein kinase C and tyrosine kinase. Notably, it has been previously shown that casein kinase II can partially complement the UV radiation sensitivity of XP-C and XP-D cells in culture, suggesting a role for casein kinase II-dependent protein phosphorylation as a modulator of cellular responses to DNA damage (20).

The XPC^{DM} protein sequence also contains three potential bipartite nuclear targeting sequences (Table 1). This motif consists of two basic residues followed by a spacer of any 10 residues, with three out of the next five residues being basic (19, 21), and has been identified in a number of nuclear proteins, including p53 protein, DNA polymerase α , several topoisomerases and several proteins involved in nucleotide excision repair. The revised human XPC protein sequence also contains a bipartite signal, and the yeast Rad4 protein contains three of these motifs.

DISCUSSION

The isolation and sequence analysis of the *Drosophila* homolog of the human XPC gene has added convincing evidence to the contention that the *S. cerevisiae* homolog of XPC is the RAD4 gene. The majority of the cloned yeast nucleotide excision repair genes have been shown to have human homologs. Indeed the only exceptions likely constitute cases in which a concerted search for homologous genes has not yet been initiated. Hence, it is clear

that the genes, and by inference the biochemistry, of this DNA repair mode are evolutionarily conserved in eukaryotes. Among the cloned yeast nucleotide excision repair genes those that are most highly conserved are the essential genes *RAD3* and *SSL2*, in which amino acid identity with their human homologs (*XPD* and *XPB*, respectively) extends along the entire length of the polypeptide. For non-essential genes such as *RAD2*, *RAD10* and *RAD14* the general rule is that limited domains of the predicted amino acid sequence share significant homology with domains in the human homologs, as is the case with the *RAD4* and *XPC* genes examined in the present study. Such conservation notwithstanding, Koken *et al.* (22) were unable to detect correction of repair-defective phenotypes by microinjection or DNA transfection of the *Drosophila haywire* cDNA into XP-B fibroblasts or into homologous rodent cell lines. Similarly, we were unable to show correction of the UV radiation sensitivity of XP-C fibroblasts by transfection of an episomal plasmid expressing the *XPC^{DM}* cDNA (data not shown). This lack of complementation presumably reflects the inability of the *Drosophila* proteins to form functional complexes with one or more mammalian repair proteins *in vivo*.

The observation that the C-terminal portion of the *Drosophila* protein is highly conserved in humans and yeast suggests that it contains important functional domains. In this regard it is of interest that several spontaneous and induced mutations in the yeast *RAD4* gene that result in defective nucleotide excision repair, have been predicted to result in truncated polypeptides (23) which delete the C-terminal third of the Rad4 protein. The identification of functional domains within the XPC protein will be important in the evaluation of mutations associated with the disease XP, and in the eventual clarification of a biochemical function for the XPC protein. The highly conserved amino acid residues identified by the comparison of the human, *Drosophila*, and yeast sequences suggests that biochemical characterization of the *Drosophila* and yeast proteins might be very informative in this regard.

ACKNOWLEDGEMENTS

These studies were supported by research grants CA44247 (ECF) and CA52461 (RL) from the USPHS. We thank our laboratory colleagues for helpful discussions and review of the manuscript and Dr Chikahide Masutani for communicating unpublished results.

REFERENCES

- Cleaver, J.E. and Kraemer, K. (1989) In Scriver, C.R., Beaudet, A.L., Sly, W.S. and Valle, D. (eds), *The Metabolic Basis of Inherited Disease*. McGraw-Hill, New York, Vol. 2, pp. 2949–2971.
- Hoeijmakers, J.H.J. and Bootsma, D. (1990) *Cancer Cells*, **2**, 311–320.
- Kantor, G.J., Barsalou, L.S. and Hanawalt, P.C. (1990) *Mutat. Res.*, **235**, 171–180.
- Venema, J., van Hoffen, A., Karcagi, V., Natarajan, A.T., van Zeeland, A.A. and Mullenders, L.H.F. (1991) *Mol. Cell. Biol.*, **11**, 4128–4134.
- Legerski, R. and Peterson, C. (1992) *Nature*, **359**, 70–73.
- Legerski, R. and Peterson, C. (1992) *Nature*, **360**, 610.
- Shimamoto, T., Kohno, K., Tanaka, K. and Okada, Y. (1991) *Biochem. Biophys. Res. Commun.*, **181**, 1231–1237.
- Mounkes, L.C., Jones, R.S., Liang, B.-C., Gelbart, W. and Fuller, M.T. (1992) *Cell*, **71**, 925–937.
- Daum, H.A., III, White, H.W., Seidell, C.M. and Johnson, P.A. (1991) *BioTechniques*, **11**, 784–790.
- Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.*, **132**, 6–13.
- Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) *Cell*, **15**, 687–701.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Brown, N.H. and Kafatos, F.C. (1988) *J. Mol. Biol.*, **203**, 425–437.
- Spratt, B.G., Hedge, P.J., te Heesen, S., Edelman, A. and Broome-Smith, J.K. (1986) *Gene*, **41**, 337–342.
- Ptashne, M. (1988) *Nature*, **335**, 683–689.
- Lapeyre, B., Bourbon, H. and Amalric, F. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 1472–1476.
- Wen, L., Huang, J., Johnson, B. and Reeck, G.R. (1989) *Nucleic Acids Res.*, **17**, 1197–1214.
- Sung, P., Prakash, S. and Prakash, L. (1988) *Genes Dev.*, **2**, 1476–1485.
- Dingwall, C. and Laskey, R.A. (1991) *Trends Biochem. Sci.*, **16**, 478–481.
- Teitz, T., Eli, D., Penner, M., Bakhanashvili, M., Naiman, T., Timme, T., Wood, C. M., Moses, R.E. and Canaani, D. (1990) *Mutat. Res.*, **236**, 85–97.
- Friedberg, E.C. (1992) *Trends Biochem. Sci.*, **17**, 347.
- Koken, H.M.M., Vreeken, C., Bol, S.A.M., Cheng, C., Jaspers-Dekker, I., Hoeijmakers, J.H.J., Eeken, J.C.J., Weeda, G. and Pastink, A. (1992) *Nucleic Acids Res.*, **20**, 5541–5548.
- Couto, L.B. and Friedberg, E.C. (1989) *J. Bacteriol.*, **171**, 1862–1869.