3084–3089 *Nucleic Acids Research, 1998, Vol. 26, No. 13* *© 1998 Oxford University Press*

The *XRCC2* DNA repair gene from human and mouse encodes a novel member of the *recA/RAD51* family

Richard Cartwright, Cathryn E. Tambini, Paul J. Simpson and John Thacker*

Medical Research Council, Radiation and Genome Stability Unit, Harwell, Oxfordshire OX11 0RD, UK

Received April 15, 1998; Revised and Accepted May 8, 1998

DDBJ/EMBL/GenBank accession nos Y08837, Y17033, Y17040

ABSTRACT

We recently identified a positional candidate for the *XRCC2* DNA repair gene at human chromosome 7q36.1. We have now cloned the cDNA for this gene from both human and mouse and show that it is a highly conserved novel member of the *recA/RAD51* recombination repair gene family. The cDNA is able to complement significantly the phenotype of a unique cell line, *irs1*, which shows extreme sensitivity to DNA cross-linking agents and genetic instability. This phenotype is consistent with a role for the *XRCC2* gene in recombination repair in somatic cells, suggesting that in addition to *RAD51*, other members of this gene family have an important function in high fidelity repair processes in mammals. Despite this function, the *XRCC2* gene transcript is expressed at a very low level in somatic tissue, but is elevated in mouse testis, suggesting an additional role in meiosis.

INTRODUCTION

Severe damage to DNA, such as double-strand breaks or interstrand cross-links, creates considerable problems for the maintenance of genomic integrity. Studies with microorganisms, such as bacteria and yeast, have revealed that repair of this type of damage is largely controlled by a specific set of genes that are also involved in genetic recombination (1–4). Thus, defects in these genes give rise to cellular sensitivity to agents such as ionizing radiation and bifunctional chemicals and to loss of homologous recombination capacity. Although less is known of this type of DNA repair in mammals, it is clear that there is structural conservation of some of the gene products involved (5–9).

The *recA/RAD51* family of genes has been identified in the archaea, prokaryotes and eukaryotes as highly conserved components of genetic recombination (10–12). These genes are involved in early steps of homologous recombination, to promote homologous DNA pairing and strand exchange. Mammalian homologues of this family of genes have recently been identified through their conservation of protein sequence and the human Rad51 protein has been shown to retain a number of functional similarities to bacterial and yeast orthologues (13,14). The potential importance of mammalian Rad51 functions has been emphasized recently by the finding of interactions with the tumour suppressor protein p53 and with the breast cancer susceptibility proteins Brca1 and Brca2 (15–20).

We have been working towards the cloning of the gene defective in a unique cell line derived in our laboratory (*irs1*), which is sensitive to a range of DNA-damaging agents. The phenotype of the *irs1* cell line, showing in particular extreme sensitivity to DNA cross-linking agents and genetic instability (21,22), is consistent with a role for the defective gene in recombination repair (23). We have mapped this gene, provisionally named *XRCC2*, to human chromosome 7q36.1 (23) and recently identified a candidate gene from the ability of genomic DNA from that region to complement the defect in *irs1* (24). We have now cloned the full-length human *XRCC2* cDNA and similarly identified the mouse gene to show that its structure is highly conserved in mammals. The gene is a novel member of the *recA/RAD51* family, demonstrating that, in addition to *RAD51*, other members of this gene family have an important role in mammalian repair processes.

MATERIALS AND METHODS

Cell culture, transfection and complementation testing

Cell lines were cultured and exposed to the DNA-damaging agent mitomycin C as described previously (23). Aliquots of 10 µg cDNA in an expression vector carrying a hygromycin resistance selection gene (25) were transfected into 10^6 *irs1* cells by electroporation (400 V, 500 µF; BioRad gene pulser). Clonal lines of successfully transfected cells were identified by their growth in medium containing 400 µg/ml hygromycin. Following selection, the ability of the cDNA to complement the defect in *irs1* was tested by respreading the transfected lines and parental lines in different concentrations of mitomycin C. A pool of ~100 transfected lines, as well as several individual lines, were tested for resistance to mitomycin C. Similarly, cycling cultures of parental and cDNA-transfected lines were tested for the frequency of chromosomal aberrations, by scoring Giemsa stained metaphase spreads according to standard criteria (26).

Isolation of the full-length *XRCC2* genes

A human cDNA library derived from HeLa cells (kindly supplied by Dr R.Legerski, University of Texas, Houston, TX) (25) was screened with probes made by PCR, using primers designed to the putative *XRCC2* cDNA fragment (GenBank accession no Y08837) identified previously (24). Other cDNA libraries were made

*To whom correspondence should be addressed. Tel: +44 1235 834393; Fax: +44 1235 834776; Email j.thacker@har.mrc.ac.uk

in-house (R.Cartwright, unpublished results) or kindly supplied by Dr I.D.Hickson (Institute of Molecular Medicine, Oxford, UK). A 140 kb PAC carrying the *XRCC2* gene (from the RPC1 library, constructed by P.de Jong; 27) was identified previously (24). This was restricted and subcloned into pBluescript II (Stratagene); fragments carrying the 5′-end of *XRCC2* were identified by probing Southern blots with synthetic oligonucleotides derived from cDNA sequences. These fragments were sequenced by primer walking using an ABI377 automated sequencer with dye termination chemistry. Sequences were manipulated and translated using programs included in the Wisconsin Sequence Package v.8 (Genetics Computer Group, Madison, WI), supplied by the UK HGMP Resource Centre.

Rapid amplification of cDNA ends (RACE)

Double-stranded cDNA containing RACE adaptors was synthesized from 1 µg poly(A)$^+$ RNA from various sources (CP lymphoblastoid cell line, primary HF12 human fibroblasts, mouse L929 cell line, mouse liver and mouse testis), using the Marathon cDNA amplification kit (Clontech) and following the supplier's protocol. Nested PCR was performed using a touchdown protocol (28) from 65 to 55°C using adaptor primers and gene-specific primers. A mixture of Thermoprime plus DNA polymerase (Advanced Biotechnologies) and Pfu DNA polymerase (Stratagene) was used to obtain high fidelity products.

RNA analysis

Total RNA was isolated using Trizol reagent (Gibco) from whole tissues of 67-day-old male CBAH/1768.2 fos mice. Aliquots of 75 µg were subsequently used for poly(A)$^+$ RNA isolation using oligo(dT)$_{25}$ Dynabeads (Dynal) following the manufacturer's protocol. mRNA was reverse transcribed with the thermostable rTth reverse transcriptase (Perkin Elmer) using the manufacturer's conditions, varying the primer annealing temperature. Reactions were also done in the absence of Mn^{2+} to test for the presence of contaminating DNA and none was found. To prepare northern blots, poly(A)$^+$ RNA was isolated from exponentially growing cell lines using the mRNA Direct kit (Dynal) following the manufacturer's protocol. An aliquot of 20 µg poly(A)$^+$ RNA was separated on a formaldehyde–agarose gel and transferred onto Hybond N membrane (Amersham). Hybridization was to an 840 bp cDNA probe corresponding to the open reading frame of *XRCC2*.

RESULTS

Cloning human and mouse *XRCC2* full-length cDNAs

Using a previously identified cDNA fragment, mapping specifically to genomic clones that complemented the phenotype of the *irs1* cell line (24), we screened a number of cDNA libraries by PCR for the presence of the gene. We found that commonly there was either no corresponding cDNA or that a truncated version of the gene (i.e. without a suitable start codon) was present, probably accounting for earlier difficulties in attempting to clone the gene by transfection of cDNA libraries into the defective *irs1* cell line (23). Ultimately, multiple screening from one library (25) yielded a clone that included a potential translation start codon (see below). RACE procedures were also used to identify 5′-sequences, but despite screening large numbers of products derived from both primary and immortalized human cell lines, we succeeded in adding only ~40 bases to the 5′-end. Several RACE products showed a longer additional sequence, but were found to be spurious, with either incorrectly spliced versions of *XRCC2* or *XRCC2* spliced onto other identifiable cDNAs. The total length of human *XRCC2* cDNA sequence obtained was 1580 bp, with a potential start codon at position 83. This start codon conforms with the consensus (29) at the important –3 position but not at the +4 position (<u>ATG</u>T instead of <u>ATG</u>G) and predicts an open reading frame of 840 bp. The 3′-UTR of the gene contains a sequence closely matching the consensus of the *Alu* repeat (30) and has one of the less common variants of the poly(A) signal sequence (ATTAAA) (31).

Sequences from the mouse *XRCC2* gene were cloned by degenerate PCR, using primers designed to the central region of the human *XRCC2* gene. This was followed by 5′- and 3′-RACE to complete the sequence to a similar relative position as the human gene. An identical putative start codon is found in the mouse sequence, giving an ORF 6 bp shorter than the human gene. The nucleotide sequences of the mouse and human ORFs are highly conserved, with 79% base pair identity (not shown). The mouse gene has no *Alu*-like repeat at the 3′-end, but has an identical poly(A) signal site to the human *XRCC2* gene.

Human and mouse predicted protein sequences are compared in Figure 1; these are 78% identical. Within the predicted protein sequence, it is seen that the characteristic nucleotide binding domains of the RecA/Rad51 family are present. The first domain has the conserved GxxxxGKT structure (32), while the second domain has the characteristic hydrophobic β-sheet.

Comparison is also made in Figure 1 between the predicted protein sequences of *XRCC2* and those of the human *RAD51* (5,6) and *DMC1* (33) genes; these genes were chosen for comparison because they represent orthologues of known family members in other organisms, such as yeast. In yeast, Dmc1 is a meiosis-specific protein, while Rad51 functions in both meiosis and mitosis (2,3). This comparison shows that the Xrcc2 protein is structurally distinct but retains a number of highly conserved residues outside the nucleotide binding domains. It is also evident from this comparison that the predicted size of the Xrcc2 protein is smaller than that of other members of this family. For this reason and because the putative start codon does not entirely conform with the established consensus (see above and Fig. 2A), we sequenced genomic DNA further upstream of the identified 5′-region, to establish that no further coding regions were present.

Genomic DNA sequencing

We had previously established that restriction enzyme sites characteristic of a CpG island (34) were present on a 110 kb human genomic fragment containing the *XRCC2* gene, located at chromosome 7q36.1 (24). Sequencing a region of human genomic DNA upstream of the 5′-end of *XRCC2* revealed that this CpG island was in close proximity to the gene (Fig. 2A). 5′ CpG islands mark the start of housekeeping genes and are typically ~1 kb long, but may include more than one 5′ exon (34,35). Therefore, we sequenced a further 1700 bp of this upstream region. Additionally, an anonymous mouse BAC genomic sequence (accession no. AC004093), found to contain the mouse *XRCC2* gene, revealed a CpG island located at the same point as in the human sequence, relative to the 5′-end of the *XRCC2* gene (Fig. 2B). However, comparison of the human and mouse genomic sequences failed to reveal further upstream sequences which were

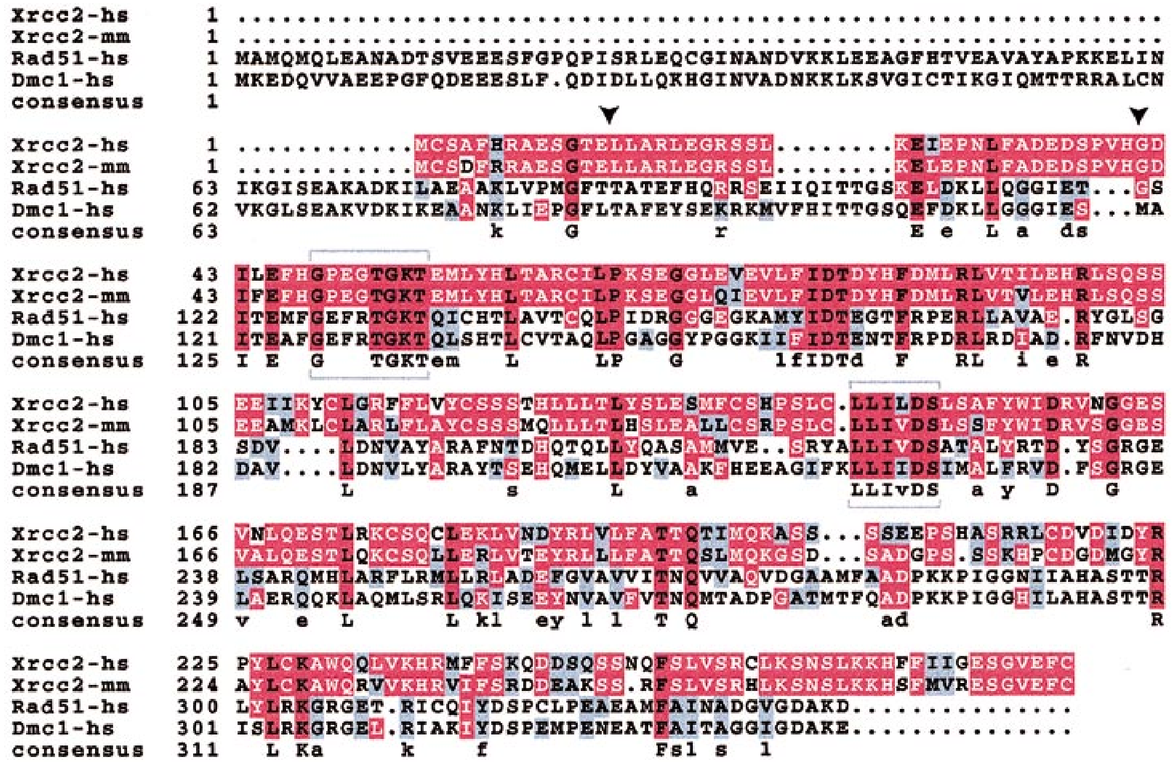

Figure 1. Comparison of the human and mouse Xrcc2 predicted protein sequences with each other and with other human members of the RecA/Rad51 family (GCG Pileup program with default values). White lettering on red indicates identity of the Xrcc2 sequences to each other and to either Rad51 or Dmc1; black on red indicates identity of all four sequences; black on blue indicates amino acid similarities. Intron–exon boundaries for the three exons of the *XRCC2* genes are shown by arrowheads. The two nucleotide binding domains are boxed.

sufficiently homologous to be considered as parts of the gene. Taken with our cDNA library and RACE work (see above), this evidence suggests that it is unlikely that further transcribed sequences are present upstream of the identified 5′-end of the *XRCC2* gene.

Sequencing human genomic DNA also revealed that the *XRCC2* gene is divided into three exons, with intron–exon boundaries positioned as shown in Figure 1. Boundary sequences conformed to the consensus (36), with AG at the 3′ splice site and GT at the 5′ splice site (data not shown).

Functional complementation

To check that the identified cDNA is able to complement the defect in the *irs1* cell line, a human cDNA in a mammalian expression vector was transfected into *irs1*. This cDNA contained the complete ORF as defined above. As seen in Figure 3, the cDNA gave a highly significant increase in the surviving fraction of *irs1*, by a factor of ~10, in response to challenge with the DNA cross-linking agent mitomycin C. This cDNA did not correct the sensitivity of two other mitomycin C sensitive cell lines, *irs3* and *irs1*SF (37), showing that it is specific for the *irs1* defect (data not shown). Similarly, the full-length gene was able to give a significant reduction in the level of chromosomal instability shown by the *irs1* line, in either individual transfected cell lines or a pooled set of many lines (Table 1).

Table 1. Frequency of chromosomal aberrations in wild-type (V79 line) cells, *XRCC2*-defective cells (*irs1* line) and cloned lines of *irs1* transfected with the human *XRCC2* cDNA

Cell line	Mitotic cells scored	Chromosomal aberrations Gaps	Breaks	Exchanges	Aberrations per 100 cells
Wild-type	100	0	0	0	2
irs1	127	12	11	10	26
irs1 + cDNA (pooled lines)	100	0	0	0	0
irs1 + cDNA (line 1)	100	2	0	0	2
irs1 + cDNA (line 2)	107	3	3	3	8

Expression of the *XRCC2* gene

It was likely, from the lack of representation of the *XRCC2* gene in cDNA libraries and that no ESTs existed for this gene in the public databases, that the transcript expression level would be low. Despite using large amounts of mRNA from human primary and immortalized cell lines, we had difficulty in obtaining a satisfactory signal on northern blots using the full-length human cDNA as probe. In contrast, we could readily detect transcripts from other genes mapping close to *XRCC2* at 7q36.1 (data not

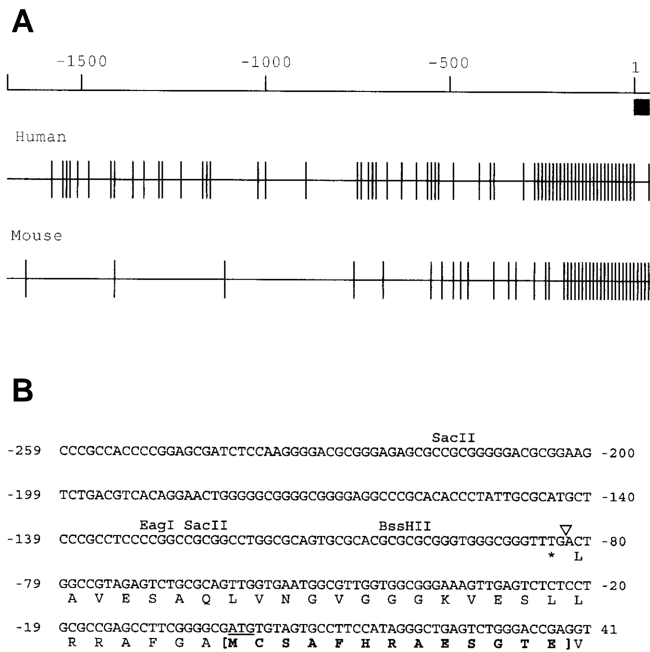

Figure 2. Genomic DNA sequence at the 5′-end of *XRCC2*. (**A**) Human genomic sequence, including exon 1, with the translation start site underlined. The maximum extent of RACE products is shown by an arrowhead. The conceptual protein sequence is shown as far as an upstream stop codon, with exon 1 shown in brackets. Restriction sites characteristic of CpG islands (34) are shown above the sequence. (**B**) Distribution of CpG sites (vertical bars) in a 1730 bp region of the human and mouse genome upstream of *XRCC2* exon 1 (black box). A scale in base pairs, with numbering as for (A), is shown above the sequences. The mouse genomic sequence is from accession no. AC004093.

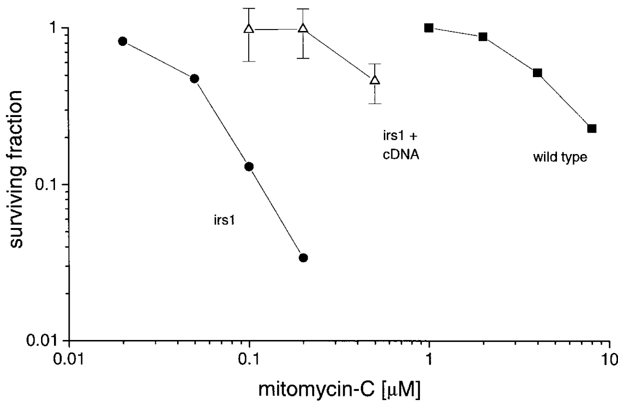

Figure 3. Survival of the *XRCC2*-defective cell line (*irs1*) and wild-type cells (V79 hamster), relative to several lines transfected with the human *XRCC2* gene, following treatment with mitomycin C. The mean and SE of data found for five independent transfected lines and a pooled set of lines are shown.

shown). However, we did detect a signal in 20 μg mRNA from an immortalized mouse cell line, L929, showing a 1.8 kb transcript (Fig. 4A).

Commercially available multiple tissue northern blots also failed to show any signals with mouse or human *XRCC2* probes. In view of this overall low abundance, we used RT-PCR on

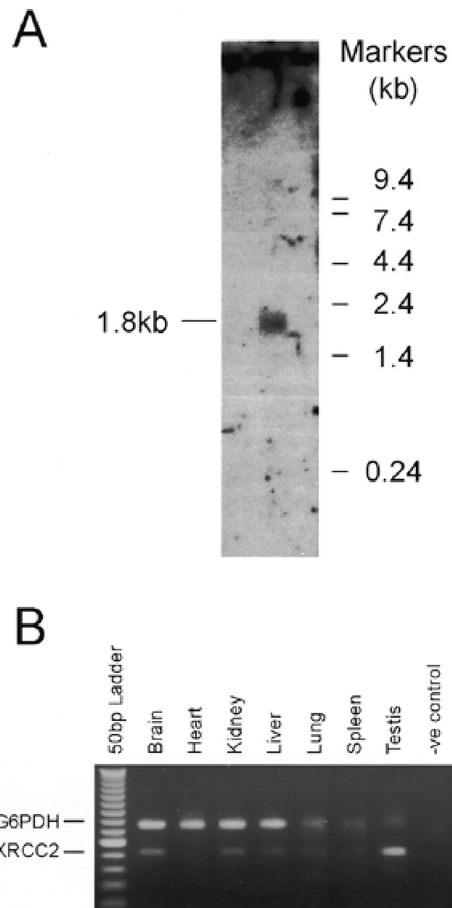

Figure 4. Expression of the mouse *XRCC2* gene. (A) Northern analysis of 20 μg mRNA from L929 mouse cells (centre lane), probed with the full-length *XRCC2* gene. The transcript size is indicated to the left. (B) RT-PCR of mouse tissue RNA, with *GAPDH* as a control; amounts of RNA used were 3 ng (*GAPDH*) and 4 μg (*XRCC2*).

mRNA from mouse tissues to look at tissue-specific expression of *XRCC2*. As an approximate control for the amount of mRNA present, primers specific for the housekeeping gene *GAPDH* were used on the same samples. Figure 4B shows that amplified products were present in all tissues, but that there was a reproducibly elevated level of *XRCC2* transcript in testis.

DISCUSSION

The homologies identified from the predicted protein structure of the *XRCC2* gene suggest that it forms part of a homologous recombination pathway involved in the repair of severe damage to DNA. This predicted role for the gene, as suggested earlier (23), is consistent with the phenotype of the *irs1* cell line and constitutes the first example of a gene from this pathway identified by its complementing activity.

The *irs1* cell line is moderately sensitive to several different types of DNA-damaging agents, including ionizing radiation (3-fold), UV light (3-fold) and ethyl methanesulphonate (10-fold), although it is especially sensitive to mitomycin C (60-fold) (21). This extreme sensitivity of *irs1* to DNA cross-linking agents such as mitomycin C is likely to arise because this

type of damage can be handled only by homologous recombination, while damage such as DNA double-strand breaks caused by ionizing radiation can be repaired by other pathways. For example, it has been shown that double-strand breaks in mammalian cells are repaired predominantly by non-homologous end-joining processes and, in contrast to *irs1*, cell lines defective in end-joining are extremely sensitive to ionizing radiation but not to other agents (38). The sensitivity of *irs1* to agents that cause mostly single-stranded damage, such as UV light and alkylating agents, is likely to arise as a result of a defect in post-replication repair (see also below). If single-stranded damage is not repaired prior to DNA replication (S phase), then a gap will commonly form in the newly synthesized strand and the resulting double-stranded damage will require homologous recombination for its resolution (3,4).

Homologous recombination takes advantage of the sequence similarity between duplicated regions of DNA, for example sister chromatids, to repair the damaged sequence (3). The repair will be carried out with high fidelity providing the duplicated region is identical in sequence to the damaged region. Conversely, loss of this recombination repair process should result in a loss of repair fidelity. It is of considerable interest therefore that *irs1* cells show genetic instability, with increased formation of spontaneous and radiation-induced gene mutations and chromosomal aberrations, relative to wild-type cells (15,39,40; Table 1 and unpublished data).

The introduction of the cloned human *XRCC2* cDNA into the *irs1* hamster cell line gives a significant but not complete reversal of the sensitivity and instability phenotype (Fig. 3 and Table 1). The inability to completely revert this phenotype is unlikely to be due to a difference between the human and hamster *XRCC2* genes, because we have already shown that human genomic DNA carrying the *XRCC2* gene can completely revert the *irs1* phenotype (24). However, we have also found that the gene transcript is normally expressed at a very low level (see Results). We consider, therefore, that it is possible that expression of *XRCC2* under a strong promoter, while complementing the phenotype in part, may lead to some adverse effects on the transfected cell lines.

The very low level of transcript expression, for a gene that is clearly very important for survival from damage by DNA cross-linking agents, indicates that the gene is tightly regulated. It is possible that this regulation relates to cell cycle position; there is evidence for regulated expression of recombination proteins through the cell cycle in both yeast and mammalian cells. *RAD51* gene transcript levels increase sharply in S phase of *Saccharomyces cerevisiae* (12) and levels of Rad51 and Rad52 proteins increase during S phase and peak in G_2 in human and rodent cells (41). Also, the human Rad51 protein is found to concentrate in multiple foci within the nucleus of mitotic cells at S phase (42). These findings are again consistent with a major role for homologous recombination processes in post-replication repair of damage. In this context, it is interesting to note that the *irs1* cell line shows loss of the S phase-dependent resistance to X-rays found in wild-type cells (43). In addition, the increased expression of *XRCC2* in mouse testis, relative to other tissues tested, is suggestive of a role for the gene product in meiosis. In this respect *XRCC2* may again have similarities to *RAD51*; the *RAD51* transcript is highly expressed in mouse testis (5) and Rad51 protein is associated with chromosome synapsis in human and mouse meiotic cells (44).

Do studies of homologous recombination in other organisms give any precedents for the possible functional role of *XRCC2?* In the bakers yeast *S.cerevisiae*, for example, there are four different members of the *recA*/*RAD51* family: *ScRAD51*, *ScDMC1*, *ScRAD55* and *ScRAD57* (2). Each of these genes has a non-redundant function in recombination, although it is known that the ScRad51, ScRad55 and ScRad57 proteins work in a complex (45,46). As noted above, close structural (and probably functional) homologues of *ScRAD51* and *ScDMC1* exist in mammals, but as yet no close homologues of *ScRAD55* and *ScRAD57* have been found in other organisms. While it is possible that *XRCC2* is a functional homologue of one of these latter two genes, it is difficult on present information to be sure of this. Comparison of the predicted protein structures shows that *XRCC2* has a similar level of homology (20–25% identity, 45% similarity) to each of the four *S.cerevisiae* proteins. Similarly, *XRCC2* is structurally distinct from other human members of this gene family (Fig. 1), including recently discovered genes (47). Gene duplication and structural divergence is a relatively common feature in the evolution of higher organisms (48), but it is also possible that a high level of structural homology is not conserved in the evolution of some members of the *recA*/*RAD51* family. In accordance with this, we have recently identified a member of this family from the yeast *Schizosaccharomyces pombe* that also does not closely resemble any specific member of the *S.cerevisiae* group (unpublished data).

Biochemical studies of the *XRCC2* gene product, now in progress, should resolve these questions of functional similarities. Given the presence of an elevated level of transcript in germinal tissue, it will be particularly important to assess the function of this gene in both mitotic and meiotic tissue.

ACKNOWLEDGEMENTS

The work of the authors was supported in part by European Communities contract F13P-CT92-0007. We are grateful for the support of Steven Scherer and Lap-Chee Tsui (Hospital for Sick Children, Toronto, Canada) in the early stages of this work.

REFERENCES

1 Kowalczykowski,S.C., Dixon,D.A., Eggleston,A.K., Lauder,S.D. and Rehrauer,W.M. (1994) *Microbiol. Rev.*, **58**, 401–465.
2 Game,J.C. (1993) *Semin. Cancer Biol.*, **4**, 73–83.
3 Shinohara,A. and Ogawa,T. (1995) *Trends Biochem. Sci.*, **20**, 387–391.
4 Cox,M.M. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 11764–11766.
5 Shinohara,A., Ogawa,H., Matsuda,Y., Ushio,N., Ikeo,K. and Ogawa,T. (1993) *Nature Genet.*, **4**, 239–243.
6 Yoshimura,Y., Morita,T., Yamamoto,A. and Matsushiro,A. (1993) *Nucleic Acids Res.*, **21**, 1665.
7 Muris,D.F.R., Bezzubova,O., Buerstedde,J.M., Vreeken,K., Balajee,A.S., Osgood,C.J., Troelstra,C., Hoeijmakers,J.H.J., Ostermann,K., Schmidt,H. *et al*. (1994) *Mutat. Res.*, **315**, 295–305.
8 Shen,Z.Y., Denison,K., Lobb,R., Gatewood,J.M. and Chen,D.J. (1995) *Genomics*, **25**, 199–206.
9 Kanaar,R., Troelstra,C., Swagemakers,S.M.A., Essers,J., Smit,B., Franssen,J.-h., Pastink,A., Bezzubova,O.Y., Buerstedde,J.-M., Clever,B. *et al*. (1996) *Curr. Biol.*, **6**, 828–838.
10 Sandler,S.J., Satin,L.H., Samra,H.S. and Clark,A.J. (1996) *Nucleic Acids Res.*, **24**, 2125–2132.
11 Roca,A.I. and Cox,M.M. (1997) *Prog. Nucleic Acid Res. Mol. Biol.*, **56**, 129–223.
12 Basile,G., Aker,M. and Mortimer,R.K. (1992) *Mol. Cell. Biol.*, **12**, 3235–3246.
13 Benson,F.E., Stasiak,A. and West,S.C. (1994) *EMBO J.*, **13**, 5764–5771.
14 Gupta,R.C., Bazemore,L.R., Golub,E.I. and Radding,C.M. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 463–468.
15 Lim,D.-S. and Hasty,P. (1996) *Mol. Cell. Biol.*, **16**, 7133–7143.

16 Sturzbecher,H.-W., Donzelmann,B., Henning,W., Knippschild,U. and Buchhop,S. (1996) *EMBO J*., **15**, 1992–2002.
17 Buchhop,S., Gibson,M.K., Wang,S.W., Wagner,P., Sturzbecher,H.-W. and Harris,C.C. (1997) *Nucleic Acids Res*., **25**, 3868–3874.
18 Scully,R., Chen,J., Plu,A., Xiao,Y., Weaver,D., Feunteun,J., Ashley,T. and Livingston,D.M. (1997) *Cell*, **88**, 265–275.
19 Mizuta,R., LaSalle,J.M., Cheng,H.-L., Shinohara,A., Ogawa,H., Copeland,N., Jenkins,N.A., Lalande,M. and Alt,F.W. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 6927–6932.
20 Sharan,S.K., Morimatsu,M., Albrecht,U., Lim,D.-S., Regel,E., Dinh,C., Sands,A., Eichele,G., Hasty,P. and Bradley,A. (1997) *Nature*, **386**, 804–810.
21 Jones,N.J., Cox,R. and Thacker,J. (1987) *Mutat. Res*., **183**, 279–286.
22 Thacker,J., Ganesh,A., Stretch,A., Benjamin,D.M., Zahalsky,A.J. and Hendrickson,E.A. (1994) *Mutagenesis*, **9**, 163–168.
23 Thacker,J., Tambini,C.E., Simpson,P.J., Tsui,L.-C. and Scherer,S.W. (1995) *Hum. Mol. Genet*., **4**, 113–120.
24 Tambini,C.E., George,A.M., Rommens,J.M., Tsui,L.-C., Scherer,S.W. and Thacker,J. (1997) *Genomics*, **41**, 84–92.
25 Peterson,C. and Legerski,R. (1991) *Gene*, **107**, 279–284.
26 Savage,J.R.K. (1976) *J. Med. Genet*., **13**, 103–122.
27 Ioannou,P.A. and de Jong,P.J. (1996) In Dracopoli,N.C. *et al*. (eds), *Current Protocols in Human Genetics*. John Wiley & Sons, New York, NY, Unit 5.15.
28 Don,R.H., Cox,P.T., Wainwright,B.J., Baker,K. and Mattick,J.S. (1991) *Nucleic Acids Res*., **19**, 4008.
29 Kozak,M. (1991) *J. Biol. Chem*., **266**, 19867–19870.
30 Kariya,Y., Kato,K., Hayashizaki,Y., Himeno,S., Tarui,S. and Matsubara,K. (1987) *Gene*, **53**, 1–10.
31 Birnstiel,M.L., Busslinger,M. and Strub,K. (1985) *Cell*, **41**, 349–359.
32 Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) *EMBO J*., **8**, 945–951.
33 Habu,T., Taki,T., West,A., Nishimune,Y. and Morita,T. (1996) *Nucleic Acids Res*., **24**, 470–477.
34 Bickmore,W.A. and Bird,A.P. (1992) *Methods Enzymol*., **216**, 224–244.
35 Gardiner-Garden,M. and Frommer,M.J. (1987) *J. Mol. Biol*., **196**, 261–282.
36 Mount,S.M. (1982) *Nucleic Acids Res*., **10**, 459–472.
37 Collins,A.R. (1993) *Mutat. Res*., **293**, 99–118.
38 Jackson,S.P. and Jeggo,P.A. (1995) *Trends Biochem. Sci*., **20**, 412–415.
39 Tucker,J.D., Jones,N.J., Allen,N.A., Minkler,J.L., Thompson,L.H. and Carrano,A.V. (1991) *Mutat. Res*., **254**, 143–152.
40 Okayasu,R., Pantelias,G.E. and Iliakis,G. (1993) *Mutat. Res*., **294**, 199–206.
41 Chen,F., Nastasi,A., Shen,Z., Brenneman,M., Crissman,H. and Chen,D.J. (1997) *Mutat. Res*., **384**, 205–211.
42 Tashiro,S., Kotomura,N., Shinohara,A., Tanaka,K., Ueda,K. and Kamada,N. (1996) *Oncogene*, **12**, 2165–2170.
43 Cheong,N., Wang,X., Wang,Y. and Iliakis,G. (1994) *Mutat. Res*., **314**, 77–85.
44 Barlow,A.L., Benson,F.E., West,S.C. and Hulten,M.A. (1998) *EMBO J*., **17**, 5207–5215.
45 Hays,S.L., Firmenich,A.A. and Berg,P. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 6925–6929.
46 Sung,P. (1997) *Genes Dev*., **11**, 1111–1121.
47 Cartwright,R., Dunn,A.M., Simpson,P.J., Tambini,C.E. and Thacker,J. (1998) *Nucleic Acids. Res*., **26**, 1653–1659.
48 Ohno,S. (1970) *Evolution by Gene Duplication*. George Allen and Unwin, London, UK.