Isolation of novel human and mouse genes of the *recA/RAD51* recombination-repair gene family

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

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

ABSTRACT

Genes from the *recA/RAD51* family play essential roles in homologous recombination in all organisms. Using sequence homologies from eukaryotic members of this family we have identified fragments of two additional mammalian genes with homology to *RAD51*. Cloning the full-length cDNAs for both human and mouse genes showed that the sequences are highly conserved, and that the predicted proteins have characteristic features of this gene family. One of the novel genes (*R51H2*) occurs in two forms in human cDNA, differing extensively at the 3′ end, probably due to an unusual form of alternative splicing. The new genes (*R51H2* and *R51H3*) were mapped to human chromosomes 14q23-24 and 17q1.2, respectively. Expression studies showed that *R51H2* is expressed at lower levels than *R51H3*, but that expression of both genes occurs at elevated levels in the testis compared with other tissues. The combination of gene structure conservation and the transcript expression patterns suggests that these new members of the *recA/RAD51* family may also function in homologous recombination-repair pathways.

INTRODUCTION

Genetic recombination is a vital process in all cells, not only to create genetic diversity but also to repair damage to DNA. In bacteria, the RecA protein has a central role in recombination, mediating the search for homologous regions of DNA molecules and promoting DNA strand exchange (1,2). The Rad51 protein of the yeast *Saccharomyces cerevisiae* (ScRad51) has structural homology to RecA (3) as well as functional similarities, including the formation of underwound nucleoprotein filaments on duplex DNA and the promotion of homologous pairing and strand exchange reactions *in vitro* (4,5). In the last few years, functional homologues (orthologues; 6) of the *ScRAD51* gene have been isolated from human and mouse tissue (*HsRAD51* and *MmRAD51*, respectively) (7–9); recent studies with purified HsRad51 show that both nucleoprotein filament formation and pairing, and strand exchange are conserved (10,11). Intriguingly, although the *RAD51* gene of *S.cerevisiae* is not essential for survival, deletion of the orthologue in mice gives embryonic lethality, leading to the suggestion of a potentially important role

for the mammalian *RAD51* gene in other aspects of DNA metabolism (12,13). The potential importance of mammalian Rad51 interactions has been highlighted recently by the finding of physical interactions with the tumour suppressor protein p53, which has a central role in the control of the cell cycle and apoptosis (14,15), as well as with the breast cancer-susceptibility gene products Brca1 and Brca2 (16,17).

Saccharomyces cerevisiae has two other recombination-repair proteins with structural homology to RecA, ScRad55 and ScRad57 (18); these gene products display a temperature-conditional mutant phenotype and appear to work in a complex with ScRad51 (19,20). Most recently, ScRad55 and ScRad57 have been shown to exist as a heterodimer which interacts with ScRad51 in the presence of replication protein A to stimulate DNA strand exchange (21). These observations suggest that the three RecA homologues in *S.cerevisiae* do not have redundant functions, but rather have distinct roles to play in the complex process of recombination. No mammalian orthologues of the ScRad55 and ScRad57 genes have yet been identified.

In a recent study we cloned a human gene, provisionally named *XRCC2*, that complements a mutant cell line (*irs1*) sensitive to agents causing severe types of DNA damage. We found that the XRCC2 predicted protein has structural homology to the RecA/Rad51 family, and more specifically to HsRad51 (22). This finding suggested that, in addition to the possibility of finding orthologues of *ScRAD55* and *ScRAD57*, further structural homologues of *RAD51* are present in humans. Therefore, we have used the *XRCC2* sequence, and others from the *recA/RAD51* family, to attempt to find novel members of this family. We have searched the growing collection of expressed-sequence tags (ESTs) available in the public nucleotide databases, using programs such as TBLASTN (23), which can be queried with conserved regions of gene families. These searches may reveal significant matches to give a partial gene sequence. An EST identified in this way can then be used to facilitate the isolation of a full-length gene copy, so that structural and functional studies can be performed.

MATERIALS AND METHODS

Identification of structural homologues of *HsRAD51*

Human or mouse ESTs were identified using sub-routines of the BLAST program (23) with the DDBJ/EMBL/GenBank or EST databases. The queries were based on conserved regions of the

---

predicted protein sequences of members of the *RAD51* gene family, such as *ScRAD51*, *HsRAD51*, *ScRAD55* and *HsXRCC2* (DDBJ/EMBL/GenBank accession nos M88470, D14134, U01144 and Y08837, respectively). Providing the conserved regions chosen for searching (such as the two nucleotide-binding domains; 24,25) were not too small, these searches found sequences which on re-searching protein databases (e.g. SwissProt) had significant similarity to the *RAD51* family ($P < 10^{-5}$ for at least one member of this family), without detecting related gene families (such as the functionally diverse AAA family of ATPases; 26). For example, the first of these (accession no. R50193) had 62% identity over 90 amino acids of the HsRad51 sequence, while others had lower levels of homology (e.g. accession no. T92120 had 34% identity over 50 amino acids of HsRad51). These two human ESTs showed no sequence overlap and aligned to different parts of the HsRad51 protein, but sequencing showed that they were parts of the same gene. In subsequent searches, ESTs derived from mouse tissue (accession nos AA118958 and AA124439) were found to have a high level of nucleotide homology to R50193 (77% identity over 300 bp), suggesting that there is a mouse orthologue of this gene. Human and mouse sequences from a different structural homologue of *RAD51* were similarly identified. In this case, overlapping mouse ESTs were identified first (accession nos AA049653 and AA260430 with 39% identity over 50 amino acids of HsRad51) and subsequently a human EST was found (accession no. N51784 with 76% identity over 131 bp of the mouse sequence). In matching this human EST to the mouse orthologue, it was apparent that a region of ~340 bp, including the first nucleotide-binding domain, was missing in the human clone. This was verified by finding another EST (accession no. D59413) which overlapped N51784 but also had part of the missing sequence (i.e. was contiguous with the mouse sequence). Note that in the text we are using the term 'structural homologue' to indicate that these new genes are structurally most closely related to the *RAD51* gene (as seen in database searches; see Discussion), but that their functional relationships are as yet unknown.

Cloning and sequencing full-length genes

ESTs, available as I.M.A.G.E. Consortium cDNA clones (27), were obtained from the Resource Centre of the UK Human Genome Mapping Project, Cambridge (I.M.A.G.E. Clone IDs: 576685, 573969, 478233, 746795, 692441, 259579). Additionally, a human cDNA library derived from HeLa cells (kindly supplied by Dr R.Legerski, University of Texas, Houston) (28) was screened with probes made by the polymerase chain reaction (PCR), using primers designed to sections of sequenced ESTs. Probes were cloned and verified by sequencing. Bacteria containing the library or a sub-library enriched for the gene by PCR-analysis of fractions, were screened on Hybond N filters (Amersham) at 2×10^5 bacteria/filter, with ^{32}P-labelled probes. Bacterial colonies hybridizing to the probes were respread onto fresh filters for at least one further round of screening, then purified and mini-preparations of their DNA were checked for the presence of the gene by PCR with gene-specific primers. Where an EST or isolated cDNA clone lacked a specific segment of the gene (see above), this was isolated using PCR with specific primers on a cDNA library, followed by cloning into the pCRII vector (Invitrogen). Other cDNA libraries used to screen for specific PCR products (see Results) were either made in-house

(R.Cartwright, unpublished) or made by Dr D.Simmons and kindly supplied by Dr I.D.Hickson (Institute of Molecular Medicine, Oxford).

Sequencing of the genes was performed either manually or using ABI373A (by Alta Bioscience, Birmingham) or ABI377 automated sequencers with dye-termination chemistry, using both vector and gene-specific primers. Sequences were manipulated and translated using programs included in the Wisconsin Sequence Package, Version 8, Genetics Computer Group, Madison, through the UK HGMP Resource Centre.

Chromosomal mapping

Filters with human genomic PAC clones (library RPCI1), constructed by P.de Jong, Buffalo (29) were screened using radiolabelled cDNA probes as described above. Filters and positive clones were provided by the UK HGMP Resource Centre. Mini-preparations of PAC DNA were nick-translated with biotin–dUTP (Gibco Life Technologies) and mixed with hybridization buffer in 50% formamide. Primary human fibroblasts were fixed, dropped onto slides and G-banded with trypsin (30); mitotic cell images were captured using a colour CCD camera (Optivision). The cells were then fixed, denatured in 70% formamide/2× SSC at 70°C for 2 min and the probe mixture hybridized overnight at 37°C. Slides were washed with 2× SSC in 50% formamide at 37°C. The PAC sequences were indirectly labelled with the fluorochrome FITC through an avidin reaction and amplified once using biotinylated anti-avidin. Chromosomes were counterstained with propidium iodide and analysed with a UV fluorescent microscope; images were recaptured and combined with their G-band image using Adobe Photoshop.

In addition, for one of the novel genes (*R51H3*), DNA samples derived from a panel of monochromosomal rodent–human hybrids (31) were screened by PCR with primers designed to the 3′ end of the gene. Primer sequences were: forward 5′-AATCTTCCCGA-CAGCCAACAGGTT-3′; reverse 5′-AGTGCCAGGTGGCAGT-AAACAGCA-3′; these showed no product formation with mouse or hamster genomic DNA.

Northern blot and RT–PCR analysis

Multiple-tissue northern blots were purchased from Clontech (human #7759-1) and hybridized using Rapidhyb (Clontech) using the supplier's protocol for cDNA probes. cDNA probes were made using PCR with primers designed to the sequenced genes, as follows: *R51H2*, forward primer 5′-TGTGGATCCCT-CACAGAGATTACA-3′; reverse primer 5′-TGTGCATCAAA-ACCCCTTCTG-3′; *R51H3*, forward primer 5′-GGACCTGGT-TTCTGCAGACCTGGA-3′; reverse primer 5′-GCAGCACAT-CCAGCATCTGGAAGA-3′.

Total RNA was isolated using Trizol reagent (Gibco Life Technologies) from whole tissues of 67 day old male CBAH/1768.2 fos mice. The RNA concentration was measured spectrophotometrically and 75 µg was used for poly(A)$^+$-RNA isolation. Poly(A)$^+$-RNA was isolated using oligo(dT$_{25}$) Dynabeads (Dynal) following the supplier's protocol. mRNA was reverse transcribed in the following reaction (20 µl): 1× reaction buffer (10 mM Tris–HCl pH 8.3, 90 mM KCl), 1 mM MnCl$_2$, 200 µM each dNTP, 0.75 µM gene-specific reverse primer (*R51H2*, 5′-GAGAATCTGCCTTCTCTCTGAATC-3′; GAPDH, 5′-GA-GAAACCTGCCAAGTATGATGAC-3′) and 5 U r*Tth* reverse transcriptase (Perkin Elmer). The mixture was heated at 68°C for

A

```
   1 CGGACGCGTGGGCGCGGGGAAACTGTGTAAAGGGTGGGGAAACTTGAAAGTTGGATGCTGCAGACCCGGC   70
                        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1 ..................ctgtgtaaagggtggggaaacttgaaagttggatgctgcagacccggc   48

  71 ATGGGGTAGCAAGAAACTAAAACGAGTGGGTTTATCACAAGAGCTGTGTGACCGTCTGAGTAGACATCAGA  140
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  49 atggggtagcaagaaactaaaacgagtgggtttatcacaagagctgtgtgaccgtctgagtagacatcaga  118
     M  G  S  K  K  L  K  R  V  G  L  S  Q  E  L  C  D  R  L  S  R  H  Q

 141 TCCTTACCTGTCAGGACTTTTTATGTCTTTCCCCACTGGAGCTTATGAAGGTGACTGGTCTGAGTTATCG  210
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 119 tccttacctgtcaggacttttttatgtctttccccactggagcttatgaaggtgactggtctgagttatcg  188
     I  L  T  C  Q  D  F  L  C  L  S  P  L  E  L  M  K  V  T  G  L  S  Y  R

 211 AGGTGTCCATGAACTTCTATGTATGGTCAGCAGGGCCTGTGCCCCAAAGATGCAAACGGCTTATGGGATA  280
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 189 aggtgtccatgaacttctatgtatggtcagcagggcctgtgccccaaagatgcaaacggcttatgggata  258
      G  V  H  E  L  L  C  M  V  S  R  A  C  A  P  K  M  Q  T  A  Y  G  I

 281 AAAGCACAAAGGTCTGCTGATTTCTCACCAGCATTCTTATCTACTACCCTTTCTGCTTTGGACGAAGCCC  350
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 259 aaagcacaaaggtctgctgatttctcaccagcattcttatctactaccctttctgctttggacgaagccc  328
      K  A  Q  R  S  A  D  F  S  P  A  F  L  S  T  T  L  S  A  L  D  E  A

 351 TGCATGGTGGTGTGGCTTGTGGATCCCTCACAGAGATTACAGGTCCACCAGGTTGTGGAAAAACTCAGTT  420
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 329 tgcatggtggtgtggcttgtggatccctcacagagattacaggtccaccaggttgtggaaaaactcagtt  398
      L  H  G  G  V  A  C  G  S  L  T  E  I  T  G  P  P  G  C  G  K  T  Q  F

 421 TTGTATAATGATGAGCATTTTGGCTACATTACCCACCAACATGGGAGGATTAGAAGGAGCTGTGGTGTAC  490
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 399 ttgtataatgatgagcattttggctacattacccaccaacatgggaggattagaaggagctgtggtgtac  468
      C  I  M  M  S  I  L  A  T  L  P  T  N  M  G  G  L  E  G  A  V  V  Y

 491 ATTGACACAGAGTCTGCATTTAGTGCTGAAAGACTGGTTGAAATAGCAGAATCCCGTTTTCCCAGATATT  560
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 469 attgacacagagtctgcatttagtgctgaaagactggttgaaatagcagaatcccgtttttcccagatatt  538
      I  D  T  E  S  A  F  S  A  E  R  L  V  E  I  A  E  S  R  F  P  R  Y

 561 TTAAACACTGAAGAAAAGTTACTTTTCACAAGTAGTAAAGTTCATCTTTATCGGGAACTCACCGTGATGA  630
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 539 ttaaacactgaagaaaagttacttttcacaagtagtaaagttcatctttatcgggaactcaccgtgatga  608
      F  N  T  E  E  K  L  L  L  T  S  S  K  V  H  L  Y  R  E  L  T  C  D  E

 631 AGTTCTACAAAGGATTGAATCTTTGGAAGAAGAAATTATCTCAAAAGGAATTAAACTTGTGATTCTTGAC  700
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 609 agttctacaaaggattgaatctttggaagaagaaattatctcaaaaggaattaaacttgtgattcttgac  678
      V  L  Q  R  I  E  S  L  E  E  E  I  I  S  K  G  I  K  L  V  I  L  D

 701 TCTGTTGCTTCTGTGGTCAGAAAGGAGTTTGATGCACAACTTCAAGGCAATCTCAAAGAAAGAAACAAGT  770
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 679 tctgttgcttctgtggtcagaaaggagtttgatgcacaacttcaaggcaatctcaaagaaagaaacaagt  748
      S  V  A  S  V  V  R  K  E  F  D  A  Q  L  Q  G  N  L  K  E  R  N  K

 771 TCTTGGCAAGAGAGGCATCCTCCTTGAAGTATTTGGCTGAGGAGTTTTCAATCCCAGTTATCTTGACGAA  840
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 749 tcttggcaagagaggcatcctccttgaagtatttggctgaggagttttcaatcccagttatcttgacgaa  818
      F  L  A  R  E  A  S  S  L  K  Y  L  A  E  E  F  S  I  P  V  I  L  T  N

 841 TCAGATTACAACCCATCTGAGTGGAGCCCTGGCTTTCAGGCAGACCTGGTGTCTCCCAGCTGATGATTTG  910
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 819 tcagattacaacccatctgagtggagccctggcttctcaggcagacctggtgtctcccagctgatgatttg  888
      Q  I  T  T  H  L  S  G  A  L  A  S  Q  A  D  L  V  S  P  A  D  D  L

 911 TCCCTGTCTGAAGGCACTTCTGGATCCAGCTGTGTGATAGCCGCACTAGGAAATACCTGGAGTCACAGTG  980
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 889 tccctgtctgaaggcacttctggatccagctgtgtgatagccgcactaggaaatacctggagtcacagtg  958
      S  L  S  E  G  T  S  G  S  S  C  V  I  A  A  L  G  N  T  W  S  H  S

 981 TGAATACCCGGCTGATCCTCCAGTACCTTGATTCAGAGAGAAGACAGATTCTTATTGCCAAGTCCCCTCT 1050
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 959 tgaatacccggctgatcctccagtacctttgattcagagagaagacagattcttattgccaagtcccctct 1028
      V  N  T  R  L  I  L  Q  Y  L  D  S  E  R  R  Q  I  L  I  A  K  S  ·P  L

1051 GGCTCCCTTCACCTCATTTGTCTACACCATCAAGGAGGAAGGCCTGGTTCTTCAAGCCTATGGAAATTCC 1120
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1029 ggctcccttcacctcatttgtctacaccatcaaggaggaaggcctggttcttcaagcctatggaaattcc 1098
      A  P  F  T  S  F  V  Y  T  I  K  E  E  G  L  V  L  Q  A  Y  G  N  S
                                                             g  g  e  k  p

1121 TAGAGACAGATAAATGTGCAAACCTGTTCATCTTGCCAAGAAAAATCCGCTTTTCTGCCACGAAACAAA  1190
     ||||  ||||                    |||||||||||||||||||||||||
1099 tagggata..............ctgtgacctttgtctagagttgatgggggtgtgatttgtgaaataaa  1168
```

B

```
1051 tacaccatcaaggaggaaggcctggttcttcaaggccaagagaagccataggggatactgtgacctttgtc 1120
     ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
1018 TACACCATCAAGGGGGAAGGCCTGGTTCTTCAAGGCCACGAAAGACCATAGGGGATACTGTGACCTTTGTC 1087
      y  t  i  k  e  e  g  l  v  l  q  g  g  e  k  p  *
      Y  T  I  K  G  E  G  L  V  L  Q  G  H  E  R  P  *

1121 tagagttgatgggggtgtgatttgtgaaataaaacaggaccgtactgcttggaagaaggaaacggaagct 1190
     ||||    |||||||| |||||||  |||  ||||||||||| ||||  |||||||||||||||||||||
1088 TAGTGCTGAT.TGCATGTGACTCATGAAATGAAACAGGACTGCGCCTGCTTGGAAAAAGGAAACGGAAGCC 1156

1191 gacataatggggattaattagttgattgctgttgagatggtaacagatttgctcctaaaccattgagcta 1260
     |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
1157 AACATAATGAGGATTAATTGGTTGGTTGCTGTTGAGGTGGTAAC.........................A 1201

1261 gcgatttcagacctagcagggaaggtgaagatgaagaagcctttgttcaggtctctagatgtctagggct 1330
     |  ||||| |||||||    ||||||||||||||||||||| ||||||   |||||||||||||||||
1201 GTGATTTCAGACC.....CGGAAGGTGAAGATGAAGAAGCCTTTATCCA.GTCTCTGGATG.CAGAGGCT 1264

1331 gagggctttgccgccatgggatgtcaacagccataataataatttgcacttata.tagcacctttcaacc 1399
     |||||||||||||||| |||||||||||||||||||||||||||||||||||||  || |||||||||||
1265 AGGGGCTCCACCACCGTGGGATGTCAGCGGCCATCGTAATAATTTGCACTTACACAAGCACCTTTCAGCC 1334

1400 aggcacctcaaagcggtttaaccacattaattaattaaagcccacaatcctcctggggagaggaggagga 1470
     |||||| ||||||  ||||||||||||||||||||||||||||||||| ||||||||||||||||| |||
1335 ATGCCCCTCAAAGTGGTTCAGCCACATTAATTAATTAAAGCCCACAATCCCCCCTAGGGAGCAGGAGGG 1404

1471 tgactaacaagatttgtaattacaggagggaacatttccgaataaagtattgttccgccaaaaaaaaaaa 1540
     ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
1405 GGACTAACAAGATTTGTAATTACAGAAGGGAAAATTTCCGAATAAAGTATTGTTCCGCCAAAAAAAAAAAA 1474
```

10 min and then at 55°C for 30 min. The reaction was placed on ice and the following added: chelating buffer (Perkin Elmer) to 0.8×, $MgCl_2$ to 1.5 mM, gene-specific forward primer to 0.15 µM (*R51H2*, 5′-TAAGCTTGTGATTGTTGACTCCAT-3′; GAPDH, 5′ TGATGGTATTCAAGAGAGTAGGGAG-3′) and water to 100 µl. This mixture was further incubated at 94°C for 2 min and then cycled at 94°C for 30 s, 57°C for 30 s and 72°C for 30 s, followed by an extension at 72°C for 7 min. Reactions were also done in the absence of Mn^{2+} to test for the presence of contaminating DNA, and none was found.

RESULTS

Identification of novel human members of the *recA/RAD51* family

Searches of the databases, using queries based on conserved regions of predicted amino acid sequences of the *recA*/*RAD51* gene family, detected several human and mouse ESTs with significant similarities to specific regions. In these searches, we were able to define clearly members of the RecA/Rad51 gene family, separately from members of related families (see Materials and Methods). Two novel genes were detected, which we have provisionally designated as *RAD51 homologue 2* (*R51H2*) and *homologue 3* (*R51H3*), since these genes have no obvious closer structural homologues among the known yeast genes in the RecA/Rad51 family. This numbering reflects the fact that the *XRCC2* gene was the first human structural homologue of *HsRAD51* to be identified (22), and that other members of this family which have a close structural homologue in yeast such as the recently identified human *DMC1* gene (32) are not included. While this report was in preparation, *R51H2* was also identified by others using degenerate PCR based on the sequence of a gene (*REC2*) from the smut fungus, *Ustilago maydis*, and was provisionally named *hREC2* (33) (but see below and Discussion).

Cloning and sequencing of full-length cDNAs

The ESTs identified for human homologue 2 (*HsR51H2*) did not contain the 5′ end of the gene. A full-length *HsR51H2* sequence was identified by probing a human cDNA library; this was sequenced and found to have an open-reading frame of 1050 bp, with a start codon (GGCATGG) closely conforming to the consensus identified by Kozak (34). However, when we compared this sequence with that derived from the ESTs in the database and the sequence recently published as *hREC2* by Rice *et al.* (33), we found that our sequence differed substantially at the 3′ end (Fig. 1A). This difference encompassed the final 14 bp (five amino acids) of the open-reading frame, although the stop codon is in the same relative position and has the same sequence (TAG), and the whole 3′ untranslated region (445 bp). There is a characteristic poly(A) site close to the end of our *HsR51H2* sequence (AATAAA; Fig. 1A), while this sequence was not

---

Figure 1. (**A**) Comparison of DNA and protein sequences of human *hREC2* (top lines, upper case) and *HsR51H2* (bottom lines, lower case). The start (ATG) and stop (TAG) codons are underlined, as well as the putative poly(A) site (AATAAA) in *R51H2*. Note the disparity at the 3′ end of the nucleotide sequence; the protein sequence is identical apart from five amino acids at the 3′ end. (**B**) Comparison of the 3′ ends of the human (top line, lower case) and mouse (bottom line, upper case) *R51H2* sequences. Overall the nucleotide sequences have 85% identity. The stop (TAG) codon and the putative poly(A) site (AATAAA) are underlined.
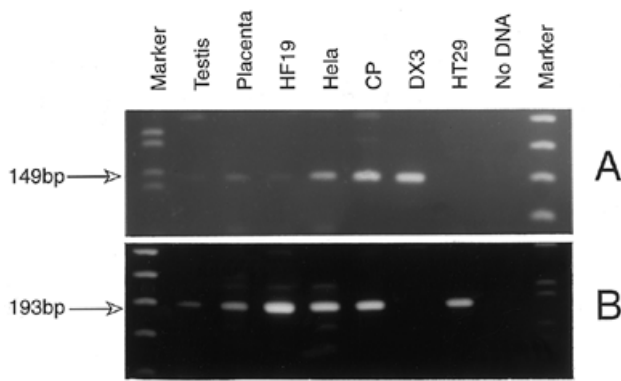
Figure 2. Presence of PCR products from the two forms of the *R51H2* gene (see text) in human cDNA libraries from different cell/tissue sources. PCR product sizes are shown to the left. Libraries are as follows: testis (Clontech HL1142q), placenta [cDNA from Clontech poly(A)⁺ RNA], HeLa (ex. R.Legerski), HF19 (primary normal human fibroblast), CP (normal human lymphoblastoid line), DX3 (melanoma) and HT29 (colon carcinoma). (**A**) Primers designed to 3′ end of *R51H2*; (**B**) Primers designed to 3′ end of *hREC2*. Markers were the 1 kb ladder and the 50 bp ladder (Gibco-BRL).

present in *hREC2* (33). We also compared *HsR51H2* with its mouse counterpart (Fig. 1B), and found in contrast that the overall identity (85.1%) was maintained into the 3′ UTR (84.7% identity over the region that differed between the two forms of the human sequence; Fig. 1B). The published human sequence (*hREC2*) is only 37.4% identical to the mouse counterpart in the 3′ UTR. The small part of the coding frame which is dissimilar in the human sequences and the mouse sequence is compared in Figure 1B.

To test for the prevalence of the two forms of the human *R51H2* sequence, we designed PCR primers to amplify these independently from cDNA libraries derived from different sources. Interestingly, as shown in Figure 2, both forms were detected in the library we used to derive the full-length human *HsR51H2* sequence (lanes marked HeLa). Similarly, both forms were found in libraries from some other sources (testis, placenta, HF19, CP; Fig. 2), although in some, the *R51H2* product level was somewhat lower than the *hREC2* product level, while further libraries contained one form or the other (*R51H2* in DX3; *hREC2* in HT29).

The other novel gene, *R51H3*, encoded human and mouse sequences with 73.9% identity at the nucleotide level (including 5′ and 3′ UTRs; not shown) and 82.3% in predicted protein sequences (Fig. 3A). The predicted open-reading frame for the human nucleotide sequence is 984 bp, while the mouse sequence has an extra 3 bp. The translation initiation site (human AAC**ATG**G, mouse ACT**ATG**G) is in good agreement with the consensus (34). In common with *R51H2*, the 5′ UTR is relatively GC-rich. A putative short nuclear localization sequence, as well as the nucleotide-binding domains characteristic of the RecA/Rad51 family are marked in Figure 3A. The first domain has the conserved [GxxxxGKT] structure, with additional glycines at alternate sites as commonly found in other adenine nucleotide-binding proteins (24), while the second domain has the characteristic hydrophobic β-sheet preceded by glycine at three residues upstream. One of the less common variants (35) of the poly(A) signal sequence (AATATA) is present in both the human and mouse gene (not shown).
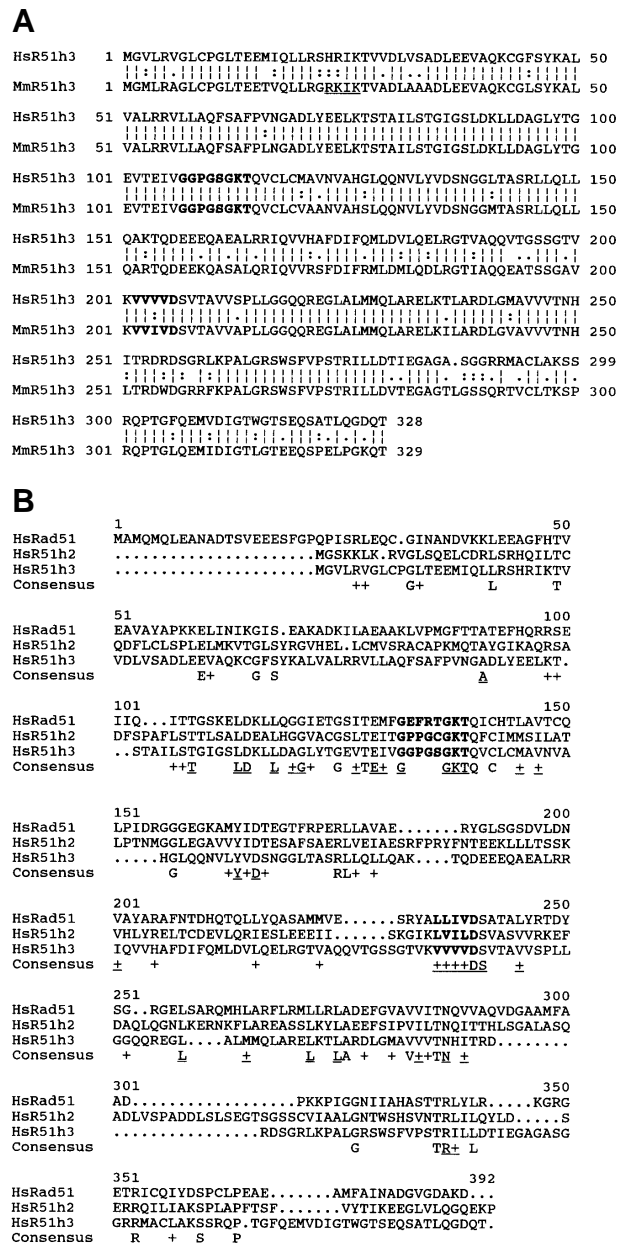

Figure 3. (**A**) Comparison of predicted protein sequences of human and mouse R51h3 using GCG 'GAP' (similarity, 90.8%; identity, 82.3%). A possible nuclear localization signal (underlined), and the two nucleotide-binding domains (bold) are indicated. (**B**) Comparison of predicted protein sequences of human Rad51 and the new genes (HsR51h2, HsR51h3) using the GCG 'Pileup' program, with a consensus based on identities (letters) and conservative changes (+) in sequence (where I=L=V=M, D=E, S=T, A=G, F=W=Y, N=Q and H=R=K). Where the consensus is underlined, the sequence is conserved also between these human sequences and the conserved residues in 23 RecA orthologues from different bacterial species (44). The nucleotide-binding domains are shown in bold.

In Figure 3B the regions of protein conservation of the human R51h2 and R51h3 are marked in comparison with HsRad51. Further areas of sequence conservation are seen, especially in the region of ~75 amino acids surrounding the first nucleotide-binding domain; these conserved sequences are found in RecA-like proteins in the archaea, prokaryotes and eukaryotes (36). A
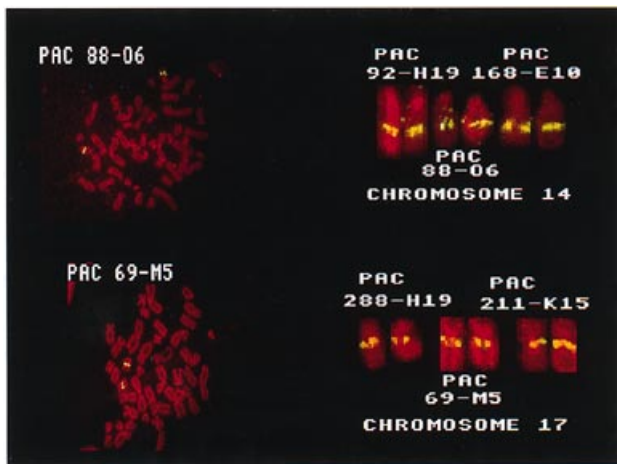
Figure 4. Human chromosome localization of *R51H2* (top) and *R51H3* (bottom) by FISH, using labelled PACs as probes (PAC numbers from the RPCI1 library).

second region of significant protein sequence conservation includes the second nucleotide-binding domain and ~50 amino acids downstream (Fig. 3B); this region includes the sequence that in the RecA protein appears to be responsible for homologous DNA pairing (37,38).

Chromosomal localization

Genomic DNA segments, isolated from a human PAC library, were used to localize the new genes by FISH on normal human metaphases. Since there may to be two different forms of the homologue 2 gene (see above), three PACs identified with the *HsR51H2* probe were labelled and tested for chromosomal position. All three PACs mapped to the same chromosomal region, 14q23-24, in agreement with the recently published data for *hREC2* (Fig. 4, top). The PACs isolated by probing with *HsR51H3* cDNA mapped to human chromosome 17 (Fig. 4, bottom); screening a monochromosomal DNA library with primers specific for the 3′ end of *HsR51H3* gave a product with only the chromosome 17 DNA, verifying this mapping (data not shown). Combined G-banded and fluorescent images for *HsR51H3* localized the gene to 17q1.2. The chromosomal positions of these new members of the *recA/RAD51* family differ from both *HsRAD51* (chromosome 15q) (8) and *XRCC2* (chromosome 7q36.1) (39,40).

Expression of mRNA

The tissue specificity of human gene expression was investigated by northern blot analysis with cDNA probes, as shown in Figure 5A. Probes specific to *HsR51H2* hybridized to mRNA of ~1.8 kb but this was expressed at a very low level in all tissues, with perhaps an increased level in testis. In view of this overall low expression, RT–PCR was carried out on mRNA from mouse tissues using gene-specific primers. To control for the amount of mRNA present, primers specific to the housekeeping gene GAPDH were used on the same samples. Figure 5B shows that amplified products were present in all tissues with the exception of lung tissue. Although RT–PCR can at best be only semi-quantitative, after correction for the amounts of mRNA present, we found that
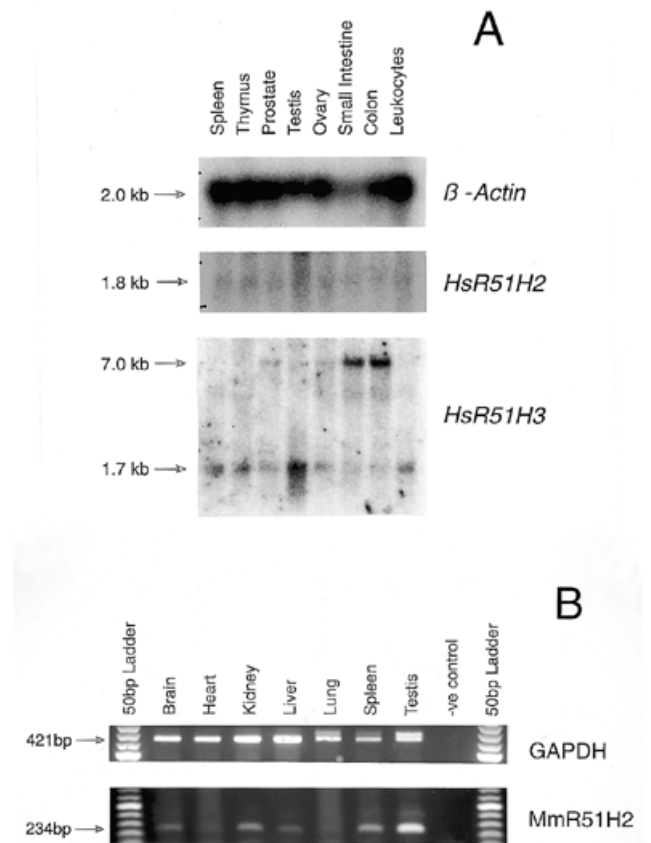

Figure 5. Tissue expression of *R51H2* and *R51H3*. (**A**) Northern analysis of human tissues using ~2 μg mRNA (Clontech MTN blot II). Arrows to the left indicate transcript size; β-actin probe shows loading differences. (**B**) RT–PCR of mouse tissues, with *GAPDH* as loading control; approximate amounts of mRNA used were 20 and 0.5 ng for *R51H2* and *GAPDH*, respectively. PCR product sizes are indicated to the left.

there were reproducibly elevated levels of *R51H2* product in the spleen and especially in the testis.

Probes specific to *HsR51H3* hybridized to an mRNA of ~1.7 kb (Fig. 5A) and again was present in all tissues; in the prostate, small intestine and colon, the probe hybridized to an additional transcript of ~7 kb presumably representing pre-mRNA. The 1.7 kb transcript is overexpressed 5–10-fold in the testis when compared with the other tissues.

DISCUSSION

We have identified two novel human members of the *recA/RAD51* recombination-repair gene family, to add to the *XRCC2* gene we recently identified (22), and found that these structural homologues are highly conserved in humans and rodents. The sequence of one of these structural homologues, *HsR51H2*, as well as its mouse counterpart, have recently been described independently by others under the names *hREC2* and *mREC2*, respectively (33). However, we find that there are discrepancies between our human gene data (*HsR51H2*) and that of *hREC2*. Firstly, the gene we isolated differs in its sequence from a few codons before the stop codon through to the end of the 3′ UTR. In assessing the significance of this finding, we compared our human sequence with the mouse sequence, and found that these

were highly homologous (Fig. 1B), while this was not true for *hREC2* and *mREC2*. Since only one form of the mouse gene has been found, and this was present in libraries derived from normal 4 week old mice, this raised the possibility that our human gene sequence (*R51H2*) may be the 'correct' form of this gene. However, the *hREC2* form was sequenced from an I.M.A.G.E. clone derived from breast tissue, and cDNAs representing parts of this sequence derived from at least one other library (foetal liver spleen) are also present in the EST database, suggesting that this form is not an artefact. The use of PCR with primers spanning the junction where the sequence differs, showed that different cell lines/tissues have either one of the gene forms, or both (Fig. 2). These findings suggest that both forms of the *R51H2* gene are transcribed—perhaps to different extents in different tissues— and that the isoforms represent alternative splicing events at the 3′ end of the gene. Precedents for this intriguing finding have been found rarely; for example, genes having more than one 3′-terminal exon, situated in tandem, which can act as alternative sites of transcription termination in different tissues (41). We are at present examining genomic DNA from this region to see if this type of gene structure accounts for our findings with *R51H2*.

A further difference is in the expression pattern of *R51H2*, relative to that reported for *hREC2*. We found that the level of expression of *R51H2* was too low in northern analysis to assess clearly whether a human tissue-specific pattern occurs, while Rice *et al.* (33) suggested that several human tissues had higher levels of *hREC2* expression (placenta, lung, liver, pancreas, spleen, etc., but not testis). They contrasted this pattern with that found for *RAD51* in the mouse, which is more abundant in testis, ovary, spleen and thymus (8). In our experiments, the *R51H2* transcript was readily detectable by RT–PCR in mRNA from several different mouse tissues. These data suggest that *MmR51H2* is expressed at highest levels in the testis (Fig. 5B), but has significant levels of expression in other tissues indicating a broader pattern of tissue expression than for *MmRAD51*. *R51H3* also has a relatively broad tissue expression pattern (Fig. 5A), but again is expressed at highest levels in testis. The elevated level of expression of both *R51H2* and *R51H3* in the testis, in particular, is suggestive of a role in meiotic recombination events, but its expression in some other tissues indicates an additional mitotic role.

In searching the protein databases with the full coding sequences of *R51H2* and *R51H3*, we find that they have no significant similarity to gene families other than the *RecA/RAD51* family. Using BLASTX, *R51H2* finds highly significant matches to regions of various Rad51 orthologues (probabilities of up to 4e–31), with ScRad51 at $P = 6e–26$ and HsRad51 at $P = 5e–25$. We did not find that R51h2 protein has significant similarity to the Rec2 gene product of *U.maydis*, which is approximately twice the size of other proteins in the RecA/Rad51 family (42), although as expected R51h2 does have significant similarity to the *U.maydis* Rad51 orthologue (43). In similar searches, *R51H3* finds ScRad51 at $P = 5e–10$, with HsRad51 at $P = 2e–8$. Thus, the predicted protein of *R51H2* is more closely related to Rad51 proteins than that of *R51H3*. However, if the amino acid sequences are directly compared along the whole length of the proteins, both R51h2 and R51h3 show a lower level of identity/similarity to RecA/Rad51 members than that shown by functional homologues of Rad51 from different species (see comparison of ScRad51 and HsRad51; Table 1). The values of identity/similarity for the new genes are close to those found

when comparing other accepted members of the family, such as Rad55 and Rad57, with each other and with Rad51 (Table 1). At present it is not clear whether orthologues of Rad55 and Rad57 exist in mammals, or how well conserved their structure will be if these exist. Therefore, until further functional data are available, these novel members of the RecA/Rad51 family are best described as structural homologues of *RAD51*, since on sequence criteria alone they cannot be considered simply as HsRad51 isoforms, or as orthologues of ScRad55 and ScRad57.

Table 1. Homology of members of the RecA/Rad51 family to each other and to the newly identified sequences, compared using GCG 'GAP' with default values

Protein comparison	Percent similarity	Percent identity
ScRad51:HsRad51	82.6	66.7
ScRad51:ScRad55	42.6	18.1
ScRad51:ScRad57	50.4	24.0
HsRad51:ScRad55	43.0	18.4
HsRad51:ScRad57	50.1	26.6
HsRad51:HsR51h2	48.9	30.0
HsRad51:HsR51h3	49.8	25.7
HsR51h2:ScRad55	45.8	22.1
HsR51h2:ScRad57	54.7	31.4
HsR51h3:ScRad55	49.0	25.3
HsR51h3:ScRad57	51.6	25.9

The presence of more than one structural homologue of *RAD51* in human tissues leads to questions of what their functions may be. There is clear evidence for conservation of specific residues in the central parts of these proteins, even in comparison with prokaryotic RecA proteins (residues underlined in the consensus of Fig. 3B). The crystal structure of *Escherichia coli* RecA has been solved (37), and from localization of the sequences of other members of the recA/Rad51 family onto the RecA structure it has been suggested that members will share common mechanisms for binding and hydrolysing ATP, leading to conformational change, as well as the ability to make filaments (44). We suggest from these structural analyses and our preliminary tissue expression data, with evidence for increased expression in germ cell tissue, that the new genes may also function in recombination-repair pathways. As found in *E.coli* (45,46), there may be several pathways to effect homologous recombination in human cells, involving groups of proteins with similar (and potentially overlapping) functions. These proteins would be organized to meet the diverse requirements of the cell for genetic exchange in a number of important processes, such as meiotic chromosome exchange and segregation and the repair of chromosome damage.

ACKNOWLEDGEMENTS

REFERENCES

1 West, S.C. (1992) *Annu. Rev. Biochem.* **61**, 603–640.
2 Roca, A.I. and Cox, M.M. (1997) *Prog. Nucleic Acid Res. Mol. Biol.* **56**, 129–223.
3 Basile, G., Aker, M. and Mortimer, R.K. (1992) *Mol. Cell. Biol.* **12**, 3235–3246.
4 Ogawa, T., Yu, X., Shinohara, A. and Egelman, E.H. (1993) *Science* **259**, 1896–1899.
5 Sung, P. (1994) *Science* **265**, 1241–1243.
6 Henikoff, S., Greene, E.A., Pietrokovski, S., Bork, P., Attwood, T.K. and Hood, L. (1997) *Science* **278**, 609–614.
7 Morita, T., Yoshimura, Y., Yamamoto, A., Murata, K., Mori, M., Yamamoto, H. and Matsushiro, A. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6577–6580.
8 Shinohara, A., Ogawa, H., Matsuda, Y., Ushio, N., Ikeo, K. and Ogawa, T. (1993) *Nature Genet.* **4**, 239–243.
9 Yoshimura, Y., Morita, T., Yamamoto, A. and Matsushiro, A. (1993) *Nucleic Acids Res.* **21**, 1665.
10 Benson, F.E., Stasiak, A. and West, S.C. (1994) *EMBO J.* **13**, 5764–5771.
11 Gupta, R.C., Bazemore, L.R., Golub, E.I. and Radding, C.M. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 463–468.
12 Tsuzuki, T., Fujii, Y., Sakumi, K., Tominaga, Y., Nakao, K., Sekiguchi, M., Matsushiro, A., Yoshimura, Y. and Morita, T. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 6236–6240.
13 Lim, D.-S. and Hasty, P. (1996) *Mol. Cell. Biol.* **16**, 7133–7143.
14 Sturzbecher, H.-W., Donzelmann, B., Henning, W., Knippschild, U. and Buchhop, S. (1996) *EMBO J.* **15**, 1992–2002.
15 Buchhop, S., Gibson, M.K., Wang, S.W., Wagner, P., Sturzbecher, H.-W. and Harris, C.C. (1997) *Nucleic Acids Res.* **25**, 3868–3874.
16 Scully, R., Chen, J., Plu, A., Xiao, Y., Weaver, D., Feunteun, J., Ashley, T. and Livingston, D.M. (1997) *Cell* **88**, 265–275.
17 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.
18 Game, J.C. (1993) *Semin. Cancer Biol.* **4**, 73–83.
19 Hays, S.L., Firmenich, A.A. and Berg, P. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 6925–6929.
20 Johnson, R.D. and Symington, L.S. (1995) *Mol. Cell. Biol.* **15**, 4843–4850.
21 Sung, P. (1997) *Genes Dev.* **11**, 1111–1121.
22 Tambini, C.E., George, A.M., Rommens, J.M., Tsui, L.-C., Scherer, S.W. and Thacker, J. (1997) *Genomics* **41**, 84–92.
23 Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) *J. Mol. Biol.* **215**, 403–410.
24 Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) *EMBO J.* **8**, 945–951.
25 Heyer, W.-D. (1994) *Experentia* **50**, 223–233.
26 Confalonieri, F. and Duguet, M. (1995) *Bioessays* **17**, 639–650.
27 Lennon, G., Auffray, C., Polymeropoulos, M. and Soares, M.B. (1996) *Genomics* **33**, 151–152.
28 Peterson, C. and Legerski, R. (1991) *Gene* **107**, 279–284.
29 Ioannou, P.A. and de Jong, P.J. (1996) In Dracapoli *et al.* (eds) *Current Protocols in Human Genetics*. John Wiley and Sons, New York, Unit 5.15.
30 Simpson, P., Morris, T., Savage, J. and Thacker, J. (1993) *Cytogenet. Cell Genet.* **64**, 39–45.
31 Kelsall, D.P., Rooke, L., Warne, D., Bonzyk, M., Cullin, L., Cox, S., West, L., Povey, S. and Spurr, N.K. (1995) *Ann. Hum. Genet.* **59**, 233–241.
32 Habu, T., Taki, T., West, A., Nishimune, Y. and Morita, T. (1996) *Nucleic Acids Res.* **24**, 470–477.
33 Rice, M.C., Smith, S.T., Bullrich, F., Havre, P. and Kmiec, E.B. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 7417–7422.
34 Kozak, M. (1991) *J. Biol. Chem.* **266**, 19867–19870.
35 Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) *Cell* **41**, 349–359.
36 Sandler, S.J., Satin, L.H., Samra, H.S. and Clark, A.J. (1996) *Nucleic Acids Res.* **24**, 2125–2132.
37 Storey, R.M., Weber, I.T. and Steitz, T.A. (1992) *Nature* **355**, 318–325.
38 Voloshin, O.N., Wang, L. and Camerini-Otero, R.D. (1996) *Science* **272**, 868–872.
39 Thacker, J., Tambini, C.E., Simpson, P.J., Tsui, L.-C. and Scherer, S.W. (1995) *Hum. Mol. Genet.* **4**, 113–120.
40 Jones, N.J., Zhao, Y., Siciliano, M.J. and Thompson, L.H. (1995) *Genomics*, **26**, 619–622.
41 Edwalds-Gilbert, G., Veraldi, K.L. and Milcarek, C. (1997) *Nucleic Acids Res.*, **25**, 2547–2561.
42 Rubin, B.P., Ferguson, D.O. and Holloman, W.K. (1994) *Mol. Cell. Biol.* **14**, 6287–6296.
43 Ferguson, D.O., Rice, M.C., Rendi, M.H., Kotani, H., Kmiec, E.B. and Holloman, W.K. (1997) *Genetics* **145**, 243–251.
44 Storey, R.M., Bishop, D.K., Kleckner, N. and Steitz, T.A. (1993) *Science* **259**, 1892–1896.
45 Smith, G.R. (1989) *Cell* **58**, 807–809.
46 Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D. and Rehrauer, W.M. (1994) *Microbiol. Rev.* **58**, 410–465.