The EMBO Journal Vol.17 No.21 pp.6404–6411, 1998

Structure of an XRCC1 BRCT domain: a new protein–protein interaction module

Xiaodong Zhang[1], Solange Moréra[1,2], Paul A.Bates[3], Philip C.Whitehead[4], Arnold I.Coffer[4], Karl Hainbucher[1,5], Rachel A.Nash[6], Michael J.E.Sternberg[3], Tomas Lindahl[6] and Paul S.Freemont[1,7]

[1]Molecular Structure and Function, [3]Biomolecular Modelling, and [4]Protein Isolation and Cloning Laboratories, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX and [6]ICRF Clare Hall Laboratories, South Mimms, Herts EN6 3LD, UK

[2]Present address: Laboratoire d'Enzymologie et de Biochimie Structurales, UPR9063 CNRS-Université Paris-Sud, Bât. 34, 1 Avenue de la Terrasse, 91198-Gif-sur-Yvette, France
[5]Present address: Institut für Physikalische Chemie, Universität Graz, Heinrichstraße 28, 8010 Graz, Austria

[7]Corresponding author
e-mail: freemont@icrf.icnet.uk

X.Zhang and S.Moréra contributed equally to this work

The BRCT domain (BRCA1 C-terminus), first identified in the breast cancer suppressor protein BRCA1, is an evolutionarily conserved protein–protein interaction region of ~95 amino acids found in a large number of proteins involved in DNA repair, recombination and cell cycle control. Here we describe the first three-dimensional structure and fold of a BRCT domain determined by X-ray crystallography at 3.2 Å resolution. The structure has been obtained from the C-terminal region of the human DNA repair protein XRCC1, and comprises a four-stranded parallel β-sheet surrounded by three α-helices, which form an autonomously folded domain. The compact XRCC1 structure explains the observed sequence homology between different BRCT motifs and provides a framework for modelling other BRCT domains. Furthermore, the established structure of an XRCC1 BRCT homodimer suggests potential protein–protein interaction sites for the complementary BRCT domain in DNA ligase III, since these two domains form a stable heterodimeric complex. Based on the XRCC1 BRCT structure, we have constructed a model for the C-terminal BRCT domain of BRCA1, which frequently is mutated in familial breast and ovarian cancer. The model allows insights into the effects of such mutations on the fold of the BRCT domain.

Keywords: BRCA1/BRCT/protein–protein interaction/X-ray crystallography/XRCC1

Introduction

The BRCT domain is defined by distinct hydrophobic clusters of amino acids and is believed to occur as an autonomous folding unit of ~95 amino acids. This domain is found in proteins involved in DNA repair, recombination and cell cycle control (Koonin *et al.*, 1996; Bork *et al.*, 1997; Callebaut and Mornon, 1997). It was first identified in the C-terminal region of the breast cancer suppressor protein BRCA1 and was thus named BRCT domain (Koonin *et al.*, 1996). Two mammalian proteins which contain BRCT domains and have functions in DNA repair, DNA ligase III and XRCC1 protein, bind each other strongly (Ljungquist *et al.*, 1994; Caldecott *et al.*, 1995) to form a heterodimer through specific interactions between their C-terminal BRCT domains (Nash *et al.*, 1997). Thus, the C-terminal stretch of 96 amino acids of XRCC1 is necessary and sufficient to bind DNA ligase III efficiently. A second, more N-terminal BRCT domain in XRCC1 interacts with poly(ADP-ribose) polymerase (Masson *et al.*, 1998). Similar results have been obtained for the C-terminal BRCT region of DNA ligase IV, which forms a heterodimer with the XRCC4 protein and is involved in DNA double-strand break repair (Critchlow *et al.*, 1997). These data identify BRCT domains as protein–protein interaction entities, which can either bind different BRCT domains specifically or interact with other unknown protein folds.

XRCC1 (633 amino acids) has no known enzymatic activity (Thompson *et al.*, 1990) but apparently functions as a scaffolding protein in the mammalian base excision-repair pathway. Thus, XRCC1 promotes the efficiency of the repair process and serves to bring together DNA polymerase β and DNA ligase III, since these two enzymes do not interact directly (Kubota *et al.*, 1996; Cappelli *et al.*, 1997). As observed for several proteins involved in the correction of abasic sites in DNA, deletion of the *XRCC1* gene in mice results in an embryonic lethal phenotype (Tebbs *et al.*, 1996). This is consistent with an essential role in removal of endogenous DNA damage. However, two Chinese hamster ovary cell lines with mutations in XRCC1 have been isolated; these cells show reduced ability to join single-strand breaks in DNA, with concomitant cellular hypersensitivity to ionizing radiation and alkylating agents (Thompson *et al.*, 1990; Cappelli *et al.*, 1997; Shen *et al.*, 1998).

The BRCT domain was first identified by a database search, comparing regions of the BRCA1 protein with other available protein sequences (Koonin *et al.*, 1996). The *BRCA1* gene encodes a 220 kDa nuclear phosphoprotein in which structural changes confer susceptibility to familial breast and ovarian cancer (Miki *et al.*, 1994). Thus, inherited mutations associated with loss of activity of BRCA1 result in a greatly increased risk of women developing breast cancer (Easton, 1997). The detailed molecular functions of BRCA1 are unknown, but recently BRCA1 has been implicated in transcriptional regulation and DNA repair (reviewed in Bertwistle and Ashworth, 1998). Specifically, the C-terminal region of BRCA1 containing two BRCT domains acts as a transcriptional

Table I. Crystallographic statistics

Diffraction data sets	Resolution (Å)	No. of unique reflections	Redundancy	I/σ	Completeness (%)	R_{sym}[a] (outer shell)	Mosaicity (°)
Sel-Met	30–3.2	7526	4.7	6.2	99.5	7.2 (28.6)	0.2
Sel-Met + Pt	30–3.5	5570	10.5	7.9	100	12.7 (30.2)	0.6

Phasing statistics

Sites	R_{cullis}[b] (centric/acentric)	Phasing powers[c] (centric/acentric)	SIR FOM (20–3.5 Å)[d]	FOM after DM (20–3.2 Å)[e]
2	0.66/0.63	1.44/1.65	0.285	0.81

Refinement statistics

Resolution	No. of reflections 2σ cutoff (work/free)	Atoms	R_{cryst} (%)	R_{free} (%)[f]	r.m.s. bond length (Å)	r.m.s.bond angle (°)
15–3.2 Å	6209/712	808	21.7	26.5	0.012	1.684

[a] $R_{sym} = \Sigma_i\Sigma_i|I_i - <I>|/\Sigma<I>$, where I_i and $<I>$ are the observed and average intensity.

[b] $R_{cullis} = \dfrac{|F_{PH} - |F_P + F_H||}{|F_{PH} - F_P|}$, where F_{PH} and F_P are the derivative and native protein structure factors.

[c] Phasing powers $= \left\langle \dfrac{F_P}{F_{PH} - |F_P + F_H|} \right\rangle$, where F_H is the calculated heavy-atom structure factor.

[d] SIR FOM: single isomorphous replacement figure of merit.
[e] FOM after DM: figure of merit after density modifications using real space 2-fold NCS averaging and histogram matching.
[f] 10% of the data at 3.2 Å were set aside for free *R*-factor calculation.

activation region in reporter assay systems (Chapman and Verma, 1996). BRCA1 co-purifies with the RNA polymerase II holoenzyme complex (Scully *et al.*, 1997a), and a direct interaction between the RNA helicase A component of the holoenzyme and BRCA1 has been detected (Anderson *et al.*, 1998). Moreover, BRCA1 is phosphorylated in response to DNA damage and associates with hRAD51 (Scully *et al.*, 1997b), the human homologue of the *Escherichia coli* RecA protein, which plays a key role in homologous recombination and post-replication repair. Recent data indicate that BRCA1 co-immunoprecipitates with p53, and can regulate p53-mediated gene expression with an absolute requirement for the most C-terminal BRCT domain, suggesting that BRCA1 functions as a coactivator of p53 (Ouchi *et al.*, 1998). A unique involvement of the second of two C-terminal BRCT domains in heterodimer formation and tight binding of its protein partner has also been observed for DNA ligase IV (Herrmann *et al.*, 1998). Many different mutations in BRCA1 have been described [see the Breast Cancer Information Core (BIC) databases on the World Wide Web: http://www.nhgri.nih.gov/Intramural_research/Lab_transfer/Bic], the positions of which have been correlated with a predisposition to breast and/or ovarian cancer (Gayther *et al.*, 1995).

Here we report the first three-dimensional structure and fold of a BRCT domain. This provides a structural basis to explain the observed sequence conservation of the BRCT family as well as giving insights into the function of the XRCC1 BRCT domain in terms of specific protein–protein interactions leading to intracellular stabilization of its DNA ligase III partner. Using the XRCC1 fold, we have constructed a model for the homologous BRCA1 C-terminal BRCT domain to predict the structural consequences of cancer-predisposing mutations found within the domain.

Results and discussion

The structure of the XRCC1 C-terminal BRCT domain

The tertiary structure of the C-terminal BRCT domain from human XRCC1 (residues 538–633; 96 residues) was solved by X-ray crystallography to 3.2 Å resolution (see Table I and Materials and methods for details), using phases obtained from single isomorphous replacement, further improved through density averaging, and refined to a free *R*-factor of 26.5% (Table I). The relatively large unit cell and highly hydrated nature of the crystals obtained were limiting factors in the resolution of the data. However, the final electron density map improved by 2-fold non-crystallographic-symmetry averaging and solvent flattening showed clear density for main chain and most side chains. This allowed an unambiguous trace of a model containing residues 538–633. The structure forms a compact globular α/β domain (~36 Å×26 Å×23 Å) consisting of a four-stranded parallel β-sheet (with strand order β2β1β3β4) surrounded by three α-helices (α1-α3; Figure 1A). The overall topology is β1α1β2β3α2β4α3, with two α-helices (α1 and α3) on one side of the β-sheet and the third helix (α2) on the other (Figure 1A). The β-sheet forms the core of the structure, with helix α1 forming hydrophobic interactions with residues from β1 and β2. Helix α2 interacts with β4, also through hydrophobic interactions, and is stabilized further by a salt bridge (Glu52 from loop c3 to Arg71 from β4, residues numbered

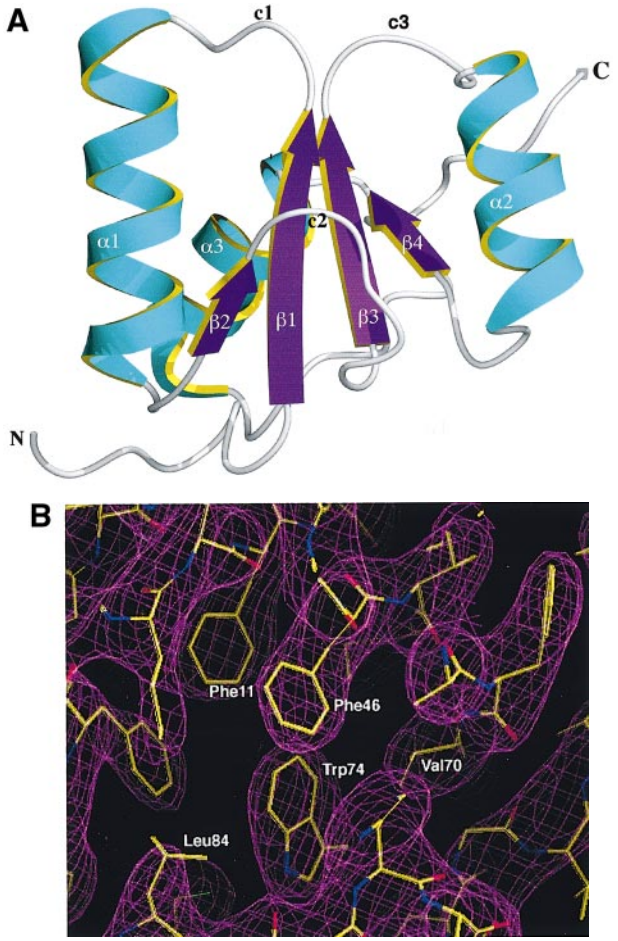

Fig. 1. XRCC1 BRCT domain structure. (**A**) Ribbon representation of the C-terminal BRCT domain of human XRCC1. The domain consists of a central β-sheet comprising four parallel β-strands (arrrows; magenta) flanked with helices α1 and α3 on one side and α2 on the other (cyan). The β-strands and α-helices are numbered and labelled as in the text, as are the N- and C-termini and the three loop regions c1, c2 and c3. (**B**) Electron density map around the conserved hydrophobic core superimposed on the current XRCC1 BRCT model. The map is calculated using SIGMAA-weighted 2*F*o – *F*c coefficients and is contoured at 1.0σ at 3.2 Å. Clear electron density was observed for all residues. The conserved Trp74 is labelled, as are other conserved hydrophobic residues (see Figure 2).

within the BRCT domain; Figure 2). Helix α3 contains the highly conserved core residue Trp74, which interacts with other conserved residues from β1, β3 and β4 as well as the C-terminal segment (Figure 1B). Helices α1 and α3 form a two-stranded helical bundle through interactions between residues Leu25, Tyr28, Val29 (α1) and Ile75, Tyr76 (α3). A search of the Dali database (Holm and Sander, 1993) identified several proteins, including bacterial chemotaxis factor CheY (PDB accession code 1Chy) and the energy-coupling IIB enzyme (PDB accession code 1iib-A), which have some similarity to the XRCC1 C-terminal BRCT domain. However, there is no evidence for sequence homology between the XRCC1 BRCT domain and these structurally similar folds.

Sequence conservation among different BRCT domains

A multiple sequence alignment of a representative subset of BRCT sequences is shown in Figure 2, along with the secondary structure of the XRCC1 C-terminal BRCT domain. Using the secondary structure as a guide, we have defined five conserved regions, N-terminus/β1, α1/β2, β3, β4/α3 and the C-terminus, with boundaries indicated by the coloured lines below the sequences (Figure 2). This slightly modifies the boundary definitions by Koonin *et al.* (1996) which were based on sequence homology within the BRCT family, although the overall secondary structure elements were predicted correctly for the core of the motif. The conserved hydrophobic amino acid clusters which define the family (Bork *et al.*, 1997; Callebaut and Mornon, 1997) are located within the central β-sheet, on α1 and α3 and at the N- and C-termini of the domain (Figures 2 and 3). Residues which form helix α3 are highly conserved both on the surface of the structure and with regard to those residues which provide the interface with α1. Likewise, residues on α1 which interact with α3 (Leu25, Tyr28 and Val29) are conserved (Figure 2), suggesting that the two-stranded helical bundle is an essential element of the BRCT domain. Trp74 is one of the three residues (Trp74, Asn33 and Gly34) which are the most invariant in the BRCT family (Figure 2 and Bork *et al.*, 1997; Callebaut and Mornon, 1997). Trp74 is located on α3 at the centre of a highly conserved hydrophobic pocket (comprising Trp74, Phe6, Phe11, Phe46, Val70 and Leu84; Figure 1B) and forms interactions with Phe11 (β1), Phe46 (β3), Val70 (β4) and Leu84 (Figure 1B). Cys78 is another relevant residue (Figure 2; Bork *et al.*, 1997), which specifically interacts with Trp74 (S-Nε 3.3 Å) and the main chain amide of Leu85 (S-N 3.4 Å), positioning the C-terminal segment onto the core of the structure (Figure 3). It is interesting to note that a mutational change of this conserved Cys residue to Tyr in the other BRCT domain of the XRCC1 protein causes functional inactivation *in vivo* (Shen *et al.*, 1998).

Sequences corresponding to the α2 region are the least conserved within the BRCT family, varying both in length (none to >20 residues) and amino acid composition (Figure 2; Bork *et al.*, 1997; Callebaut and Mornon, 1997). In the domains which have deletions in α2, such as the DNA ligase III and RAP1 BRCT domains, Phe69 on β4, which is buried by Leu61 on α2 in XRCC1, is replaced by a hydrophilic residue (Gln in DNA ligase III or Thr in RAP1; Figure 2). Apart from α2, several other sequence insertions or deletions occur within the superfamily, which map to the surface loops connecting β1 and α1 (c1) and β2 to β3 (c2) respectively (Figures 1 and 2). None of the sequence variability within the BRCT family is likely to alter the overall core fold of the domain, comprising the central β-sheet, α1 and α3, which is highly conserved.

A recent study by Caldecott and co-workers (Taylor *et al.*, 1998), using site-specific mutagenesis and deletion analysis, aimed to identify regions within the XRCC1 BRCT domain required for interaction with DNA ligase III. Most of their results can be explained easily, since deleting parts of the BRCT domain would profoundly affect the overall fold and structure of the domain. Furthermore, substitution of Val47 Ile48 in β3 by Asp (Taylor *et al.*, 1998; Figure 2) would probably prevent the correct folding of the domain, which correlates with the observation that this mutant cannot interact with DNA ligase III. However, Taylor *et al.* (1998) also reported the puzzling result that substitution of the centrally located

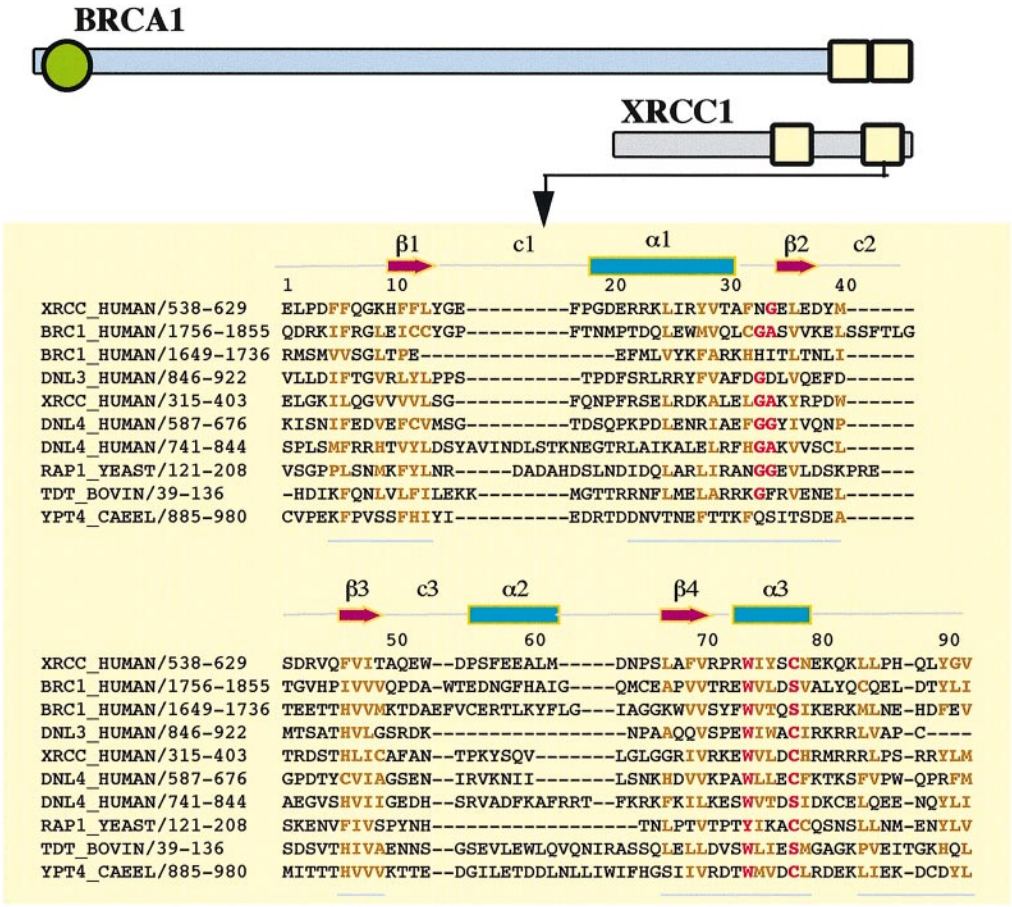

Fig. 2. Multiple sequence alignment of some BRCT family members as obtained from Pfam v3.0 (Sonnhammer *et al.*, 1997). The secondary structure of the XRCC1 BRCT domain is shown on top, labelled and coloured according to Figure 1A. Most of the sequence variations occur within helix α2 and the two surface loops c1 and c2 (Figure 1A). For reference, the relative locations of the BRCT domains in BRCA1 and XRCC1 are indicated in the schematic diagram (top). The green sphere represents the BRCA1 RING finger. Residues are coloured: most invariant in the BRCT family (red); conserved hydrophobic residues (orange). Five conserved regions are indicated by cyan lines under the sequences. The residue numbers within the XRCC1 BRCT domain are indicated above the sequences.

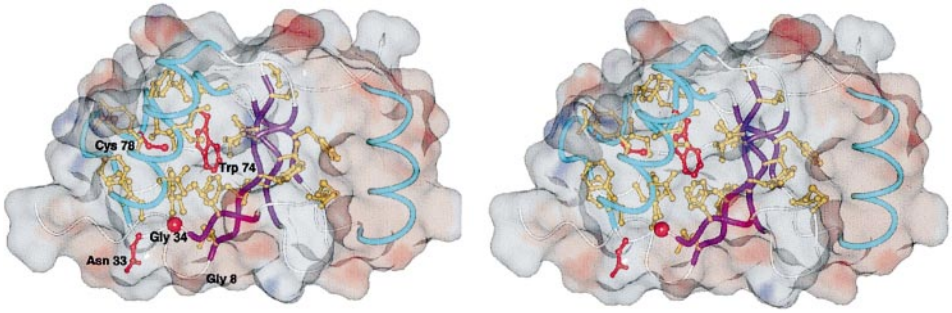

Fig. 3. Stereo view of conserved residues of the BRCT family mapped onto the XRCC1 BRCT domain structure. Most of the conserved residues cluster within the central β-sheet and helices α1 and α3. Helix α2 is the most variable region within the family, with no apparent sequence conservation. Trp74 is the most invariant residue in the family and is located at the centre of a highly conserved hydrophobic pocket (see Figure 1B.). The XRCC1 BRCT structure is represented as a coil superimposed on an electrostatic surface potential map (partially transparent) generated by GRASP (Nicholls *et al.*, 1991). β-Strands and α-helices are coloured as in Figure 1A. Conserved residue side chains are coloured according to Figure 2.

and highly conserved Trp residue (Trp74, Figures 1B and 3) by Asp did not seem to affect the interaction with DNA ligase III. Given that Trp74 is at the centre of a highly conserved hydrophobic pocket at the core of the BRCT domain, the substitution with a charged residue is likely to have severe consequences on the folding of the domain, and further experimental evidence will be required to confirm and clarify the structural and functional effects of mutations at this site.

Dimer interface and potential protein–protein interaction sites

In the XRCC1 crystal, there are two BRCT domains in the asymmetric unit forming a dimer primarily through

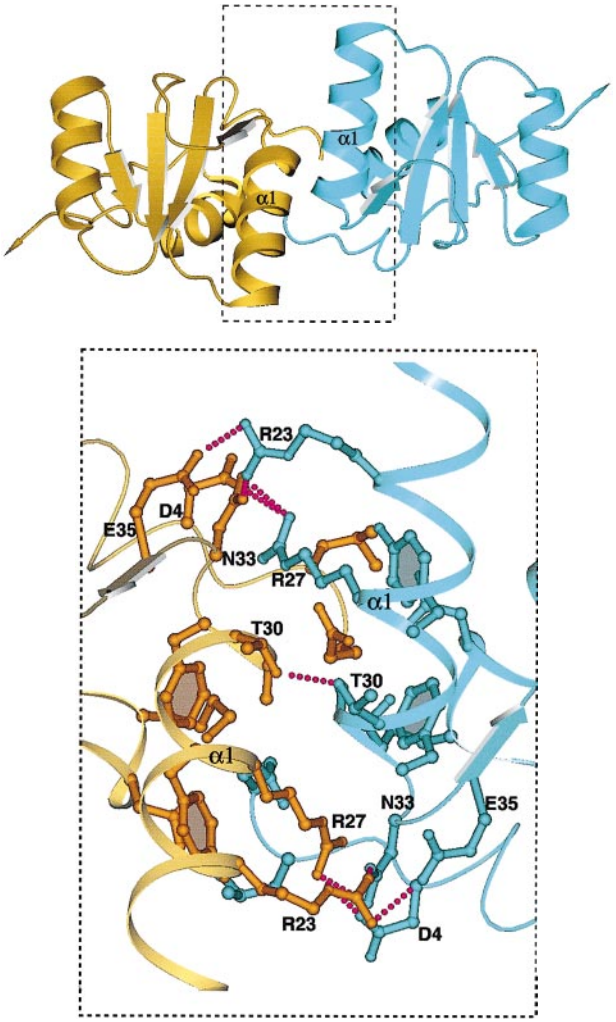

Fig. 4. XRCC1 BRCT non-crystallographic dimer. Side chains at the dimer interface which form intermolecular interactions are shown, and coloured according to each monomer. Two salt bridges (Arg23–Glu35; Arg27–Asp4) at the dimer interface are labelled.

interactions between α1 and the N-terminal region (Figure 4). Solvent-accessible surface areas (SAAs) for the non-crystallographic dimer were calculated, using the program ASA (A.Lesk, Cambridge; probe size of 1.4 Å), in order to assess the significance of the dimer interface. Five residues in the N-terminal segment provide ~30% of the total SAA at the dimer interface, whereas six residues in helix α1 (residues 23–31) followed by a turn of four residues (residues 32–35) provide ~59% of the dimer SAA. The remaining area (~11%) is provided by three residues after α3 (residues 79–81). The α1–α1 contact involves extensive interactions between residues from the monomers (designated A and B to distinguish between them), especially the salt bridge between $Arg23^{A}$ and $Glu35^{B}$ across the dimer interface (Figure 4). Another significant interaction in the dimer interface is the salt bridge between $Asp4^{A}$ and $Arg27^{B}$ (Figure 4).

The XRCC1 C-terminal BRCT domain forms a specific heterodimer *in vitro* with the BRCT domain of DNA ligase III (Nash *et al.*, 1997). Of the residues which make significant contributions to the XRCC1 non-crystallographic dimer interface, Asp4 and Arg23 are conserved

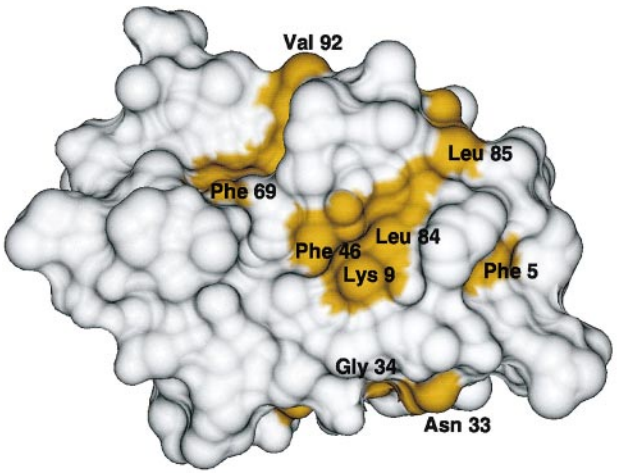

Fig. 5. Surface representation of the XRCC1 BRCT domain. The surface is coloured white and gold (conserved residues), and was calculated with a probe radius of 1.4 Å. Some conserved residues are labelled (see Figure 2). The view is the same as Figure 3 except for a 180° vertical rotation.

in the DNA ligase III BRCT domain (see Figure 2). Furthermore, Arg27 and Glu35 in DNA ligase III (Glu is replaced by an Asp) are shifted one residue towards the N-terminus relative to XRCC1. In viewing these similarities, it is reasonable to propose that some of the interactions in the XRCC1 dimer interface could be retained in a heterodimeric BRCT complex between XRCC1 and DNA ligase III. It is also notable that the subunit interface in the XRCC1 non-crystallographic dimer covers 1306 $Å^2$ (653 $Å^2$ per subunit representing ~11% of the total SAA), a value which is reported to be significant for protein–protein interfaces (Janin, 1997). Thus, the XRCC1 non-crystallographic dimer interface which we observe in the crystal structure may represent some aspects of the XRCC1–DNA ligase III BRCT interface, although further mutational and structural studies are necessary to confirm this notion.

The other highly conserved residues within the BRCT family comprise a double Gly–Gly motif (Asn33–Gly34 in XRCC1) which is located in a short loop/turn connecting α1 and β2, thereby allowing the main chain to reverse direction. There is also sequence preference at positions Gly8 (preceding β1) and Ser66 (turn between α2 and β4), suggesting that the geometry of these turns is retained. These residues form part of a relatively flat surface of the domain (Figures 3 and 5, bottom surface) which could be of functional importance for interaction with other proteins. Some of the conserved hydrophobic residues also form part of the BRCT domain surface (Figure 5). Noteworthy is the surface area comprising residues Lys9 (hydrophobic residue in most other BRCT domains; Figure 2), Phe46, Leu84 and Leu85. Since BRCT domains are found in proteins with diverse functions, it is possible that in addition to the observed XRCC1 BRCT dimer interface, BRCT domains may contain other protein–protein interaction sites specific to individual proteins.

BRCA1 BRCT model and the potential structural consequences of BRCT mutations

In this present project, we have studied the BRCT domain from XRCC1, since attempts at producing the correspond-

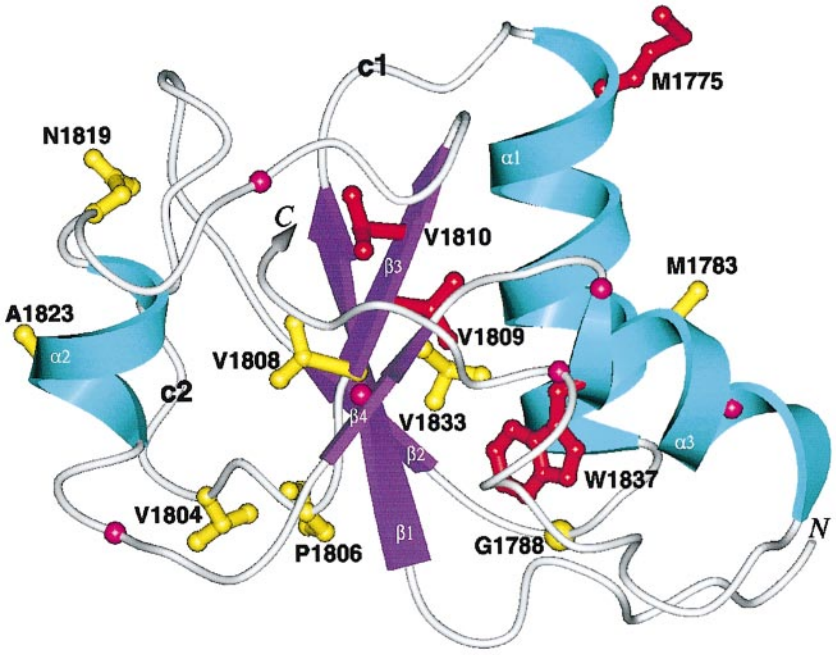

Fig. 6. BRCA1 predisposing mutations mapped onto a homology model of the most C-terminal BRCA1 BRCT domain. The BRCT domain is represented as a ribbon coloured and labelled as in Figure 1A. Residues are labelled and coloured as follows: red represents missense mutations [Trp1837(74)→Arg; Met1775(20)→Arg and an in-frame double deletion Val1809(48)–Val1810(49)]; yellow, unclassified variants [Met1783(28)→Thr; Gly1788(33)→Val; Val1804(43)→Asp; Pro1806(45)→Ala; Val1808(47)→Ala; Asn1819(57)→Ser; Ala1823(61)→Thr; Val1833(70)→Met]. Spheres (magenta) correspond to translational stop positions introduced by either nonsense or frameshift mutations (BIC databases). The mutation of Trp1837(74) and deletion of Val1809(48)–Val1810(49) would probaly prevent the correct folding of the BRCT domain.

Table II. BRCA1 BRCT mutations and predicted structural consequences

BRCA1 mutations[a]	Location in XRCC1 BRCT structure	Sequence conservation in BRCT family[b]	Predicted structural consequences
Met1775(20)[c] to Arg	helix α1/surface	variable	disrupts interface
Trp1837(74) to Arg	helix α3/core	conserved Trp	incorrect folding
ΔVal1809(48), Val1810(49)	strand β3/core	conserved hydrophobics	incorrect folding
Unclassified[d]			
Met1783(28) to Thr	helix α1/intrahelical	conserved hydrophobic	destabilize the fold
Gly1788(33) to Val	loop between α1 and β2	conserved Gly	incorrect folding
Val1804(43) to Asp	loop between β2 and β3	variable	–
Pro1806(45) to Ala	loop between β2 and β3	variable	–
Val1808(47) to Ala	β3/core	conserved hydrophobic	destabilize the fold
Asn1819(57) to Ser	α2	variable	–
Ala1823(61) to Thr	α2/exposed	variable	–
Val1833(70) to Met	β4/core	conserved hydrophobic	destabilize the fold

[a]BRCA1 mutations were obtained from the Breast Cancer Information Core (BIC) databases on the World Wide Web: http://www.nhgri.nih.gov/Intramural_research/ Lab_transfer/Bic.
[b]See Figure 2.
[c]Numbers in parentheses refer to the residue numbering of XRCC1 (see Figure 2).
[d]Unclassified mutations in terms of cancer predisposition.
–, no obvious structural effect.

ing BRCT domain from BRCA1 in quantities suitable for structural studies have so far been unsuccessful (data not shown). However, since a large number of BRCA1 single-site mutations have been mapped to the BRCT sequences and the most C-terminal BRCT domain appears to be of functional importance, a three-dimensional model for this BRCT domain of BRCA1 was constructed based on its partial sequence identity to the C-terminus of XRCC1 (see Materials and methods for details). As expected for this low level of sequence identity (Figure 2; 12% identity, 35% similarity), there are regions in BRCA1 that were difficult to model, in particular the alignment of residues within loop c2 (Lys1793–Pro1806) and α2 (Gln1814–Pro1831). However, the nature of the sequence conservation for the entire family implies that the overall predicted fold and the core area described here (Figure 6) are correct.

The location of mutations in this BRCA1 BRCT domain (obtained from BIC) are mapped onto the predicted structure (Figure 6) and shown in Table II. In addition, there are several other predisposing mutations that cause premature termination of the molecule due to frameshift and nonsense mutations which would clearly prevent formation of a fully folded BRCT domain (Figure 6). Two missense point mutations are known to be predisposing

to cancer, and from their location in the model we can interpret the structural consequences of these substitutions (Table II). A Trp1837 (corresponding to residue 74 in the XRCC1 structure) to Arg mutation leads to this predominantly buried tryptophan in the protein core (conserved in nearly all BRCT family members; Figure 2) being replaced by a charged residue, and this substitution would probably prevent formation of a properly folded BRCT domain. The second point mutation, a substitution of an exposed Met1775(20) by Arg, is unlikely to prevent correct folding of the domain. Instead, this mutation, which is close to the dimer interface in the XRCC1 BRCT structure, may be involved in recognition of another protein molecule. A recent report on the association of BRCA1 with RNA helicase A supports this interpretation, since a Met1775(20) to Glu mutation reduces the binding stability of BRCA1 to RNA helicase A (Anderson *et al.*, 1998). The predisposing in-frame double residue deletion of conserved Val1809(48) and Val1810(49) within β3 (Figure 6) would also be likely to prevent correct BRCT domain folding. A number of unclassified mutant variants (in terms of cancer predisposition) of BRCA1 have also been identified (BIC database; Table II). Among these (Table II), at least four mutations would be expected to affect the integrity of the folding pattern of the BRCT domain and, consequently, would be likely candidates for cancer-predisposing mutants. The Met1783(28) to Thr substitution changes a conserved hydrophobic residue mediating intramolecular helix–helix packing (α1–α3), and this alteration should prove detrimental to the protein fold. Furthermore, a Gly1788(33) to Val mutation would affect residues at a conserved turn region, and could result in incorrect folding. Two other mutations (Table II) may also have structurally deleterious consequences.

The structure of the C-terminal XRCC1 BRCT domain reported here defines for the first time the overall fold of a BRCT protein module. The structure also provides a framework both to explain the sequence conservation within the BRCT family and to model other BRCT domains. Such a model for the C-terminal BRCT domain of BRCA1 has allowed us to interpret a number of BRCA1 predisposing mutations in terms of their effects on the structure and folding of the BRCT domain. Furthermore, there are several unclassified BRCA1 BRCT mutations whose effects on cancer risk are unknown (BIC databases). We can now at least provide some assessment of the structural consequences of these mutations based on our BRCA1 BRCT model. An obvious next step in these investigations is to define the region of interaction between the two complementary BRCT domains of XRCC1 and DNA ligase III. To address this question, we are pursuing two parallel approaches, namely extensive site-specific mutagenesis of XRCC1 BRCT surface residues, to identify the changes that disrupt the interaction with DNA ligase III, and co-crystallization of the BRCT heterodimer complex. The present availability of the three-dimensional structure of a member of the conserved BRCT domain family is an essential step in the unravelling of functional roles for these domains.

Materials and methods

Crystallization

The BRCT domain of the C-terminal 96 amino acids of XRCC1 (residues 538–633) was overexpressed in *E.coli* fused to a C-terminal FLAG octapeptide (Nash *et al.*, 1997), purified by binding to an anti-FLAG affinity column and eluted with FLAG peptide (Sigma Chemical Co.). Selenium–methionine (Se-Met)-labelled protein was expressed in *E.coli* strain B834 grown in minimal media supplemented with selenomethionine. The incorporation of selenium was confirmed by mass spectrometry. Se-Met and native protein crystals were grown by vapour diffusion from solutions containing 4 M sodium formate, using protein concentrations of 3.5 mg/ml. The crystals belong to space group $P3_121$ ($a = b = 100.8$ Å, $c = 72.5$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$), with two molecules in the asymmetric unit.

Data collection and processing

All diffraction data were measured on single crystals. A Se-Met 'native' data set was collected to 3.2 Å resolution ($\lambda = 0.97$ Å) at room temperature on a Mar image plate using beam line 9.5 at the Daresbury Synchrotron Radiation Laboratory. A platinum heavy-atom derivative was prepared by soaking Se-Met crystals (8 h) in 1 mM $K_2Pt(NO_2)_4$. The platinum data set was measured to 3.5 Å resolution ($\lambda = 0.97$Å) at 100 K on a Mar image plate using beam line W21 at LURE. Cryoprotectant conditions were achieved by adding and increasing the concentration of glycerol from 5% (v/v) to 20% (v/v). Data were processed with DENZO (Otwinowski and Minor, 1997) and SCALEPACK (Otwinowski and Minor, 1997), and most subsequent calculations were carried out with the CCP4 program suite (CCP4, 1994). Data statistics are summarized in Table I.

Structural determination and refinement

Two Pt heavy-atom sites were first located from manual inspections of isomorphous difference Patterson maps, and their parameters were refined using the maximum likelihood method as implemented in SHARP (De La Fortelle and Bricogne, 1997). Residual difference Fourier analyses revealed no further significant Pt sites. Single isomorphous replacement phases were calculated and improved by solvent flattening with SOLOMON (CCP4, 1994) using a solvent content of 68%. Subsequent electron density maps allowed an initial backbone trace to be fitted with the program O (Jones *et al.*, 1991). Extra electron density was identified and the backbone model was fitted manually into these densities which form a heterodimer with the initial backbone model. The non-crystallographic symmetry (NCS) 2-fold axis and its transformation matrix were then obtained using the backbone heterodimer model and the function Lsq_explicit in program O (Jones *et al.*, 1991). The electron density map was improved further by a combination of real space 2-fold NCS averaging and histogram matching with DM (CCP4, 1994). The phases were then extended from 3.5 to 3.2 Å resolution using real space 2-fold NCS averaing and solvent flattenting. A model containing residues 538–633 of XRCC1 was then built into the resulting electron density map using program O (Jones *et al.*, 1991). The model was refined using least square minimization as implemented in X-PLOR (version 3.84) (Brünger, 1996). This includes 150 steps of positional refinement followed by simulated annealing refinement with slow cooling, during which the temperature was decreased from 3000 to 300 K (Brünger, 1996). The temperature factors were refined with the restriction that adjacent atoms do not vary more than 2σ. Strict NCS restraints were enforced throughout the refinement, which gives a data to parameter ratio of 1.9:1. Subsequent ($2F_o - F_c$) maps showed clear electron density for all residues, except the FLAG octapeptide which was not visible. The final model has a free *R*-factor of 26.5% with good geometry, as detailed in Table I; 76% of the residues in the final model are in the most favoured region in the Ramachandran plot, with no residues found in disallowed regions. Coordinates are being deposited in the Protein Data Bank.

Comparative modelling of the BRCA1 C-terminal BRCT domain

The model of the C-terminal BRCA1 BRCT domain was constructed from a single template, namely the XRCC1 BRCT structure. Loops and regions with incompatible φ/ψ angles to the template were replaced by database searches as described in Bates *et al.* (1997). Manual intervention was needed if candidate fragments could not be found to cover a region. A number of fragment conformations were selected for each gap, and the best candidate was chosen from the ensemble by a modification of the self-consistent mean field approach to gap closure (Koehl and Delarue, 1995). Side chain rotamers were assigned initially by tracing the path of the template side chain as far as possible (Bates *et al.*, 1997). After the replacement of all side chains, extra conformers from a side chain rotamer library were built at each residue position. The best conformer was then selected via a second mean field refinement, where

each conformer feels the average environment due to conformers of other residues weighted by their respective probabilities (Koehl and Delarue, 1994). Energy parameters were taken from Lee and Subbiah (1991). To remove the small number of steric clashes remaining in the model, 100 steps of steepest descents were run using program CHARMM (v 3.3; Molecular Simulations Inc. 200 Fifth Avenue, Waltham, MA). The Protein health checks option in the program QUANTA (v 3.3; Molecular Simulations Inc.) was used to check the general packing quality of the protein core. The model passed all filters such as excess volume within the core and close contacts. The program PROCHECK (Laskowski *et al.*, 1993) was also used to check stereochemical quality of the model (see Martin *et al.*, 1997, for a detailed assessment of comparative modelling).

Acknowledgements

We are grateful to the staff of LURE (Orsay) for making station W21 of LURE-DCI available to us. The Se-Met data sets were collected at Daresbury as part of the Daresbury's protein crystallography data collection trial service. We thank Graeme Card, Pawel Dokurno, Mike Gorman, Suhail Islam, Richard Bowater and John Sgouros for helping with this project. We also thank Dinah Raman and Darryl Pappin for mass spectrometry and sequencing analyses.

References

Anderson,S.F., Schlegel,B.P., Nakajima,T., Wolpin,E.S. and Parvin,J.D. (1998) BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A. *Nature Genet.*, **19**, 254–256.

Bates,P.A., Jackson,R.M. and Sternberg,M.J.E. (1997) Model building by comparison: a combination of expert knowledge and computer automation. *Proteins*, Suppl. **1**, 59–67.

Bertwistle,D. and Ashworth,A. (1998) Functions of the *BRCA1* and *BRCA2* genes. *Curr. Opin. Genet. Dev.*, **8**, 14–20.

Bork,P., Hofmann,K., Bucher,P., Neuwald,A.F., Altschul,S.F. and Koonin,E.V. (1997) A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J.*, **11**, 68–76.

Brünger,A. (1996) *X-PLOR (Version 3.8) A System for X-ray Crystallography and NMR*. Yale University Press, New Haven. CT.

Caldecott,K.W., Tucker,J.D., Stanker,L.H. and Thompson,L.H. (1995) Characterization of the XRCC1–DNA ligase III complex *in vitro* and its absence from mutant hamster cells. *Nucleic Acids Res.*, **23**, 4836–4843.

Callebaut,I. and Mornon,J.-P. (1997) From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair. *FEBS Lett.*, **400**, 25–30.

Cappelli,E., Taylor,R., Cevasco,M., Abbondandolo,A., Caldecott,K.and Frosina,G. (1997) Involvement of XRCC1 and DNA ligase III gene products in DNA base excision repair. *J. Biol. Chem.*, **272**, 23970–23975.

Chapman,M.S. and Verma,I.M. (1996) Transcriptional activation by Brca1. *Nature*, **382**, 678–679.

Collaborative Computational Project No. 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Critchlow,S.E., Bowater,R.P. and Jackson,S.P. (1997) Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Curr. Biol.*, **7**, 588–598.

De la Fortelle,E. and Bricogne,G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.*, **276**, 472–494.

Easton,D. (1997) Breast cancer genes—what are the real risks? *Nature Genet.*, **16**, 210–211.

Gayther,S.A. *et al*. (1995) Germline mutations of the *BRCA1* gene in breast and ovarian cancer families provide evidence for a genotype–phenotype correlation. *Nature Genet.*, **11**, 428–433.

Herrmann,G., Lindahl,T. and Schär,P. (1998) *Saccharomyces cerevisiae* LIF1: a function involved in double strand break repair related to XRCC4. *EMBO J.*, **17**, 4188–4198.

Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

Janin,J. (1997) Specific versus non-specific contacts in protein crystals. *Nature Struct. Biol.*, **12**, 973–974.

Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.

Koonin,E.V., Altschul,S.F. and Bork,P. (1996) BRCA1 protein products...functional motifs. *Nature Genet.*, **13**, 266–267.

Koehl,P. and Delarue,M. (1994) Application of a self-consistent mean field theory to predict protein side chains conformations and estimate their conformational entropy. *J. Mol. Biol.*, **239**, 249–275.

Koehl,P. and Delarue,M. (1995) A self consistent mean field approach to simultaneous gap closure and side chain positioning in homology modelling. *Nature Struct. Biol.*, **2**, 163–170

Kubota,Y., Nash,R.A., Klungland,A., Schar,P., Barnes,D.E. and Lindahl,T. (1996) Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase β and the XRCC1 protein. *EMBO J.*, **15**, 6662–6670.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Lee,C. and Subbiah,S. (1991) Prediction of protein side chain conformation by packing optimization. *J. Mol. Biol.*, **217**, 373–388.

Ljungquist,S., Kenne,K., Olsson,L. and Sandström,M. (1994) Altered DNA ligase III activity in the CHO EM9 mutant. *Mutat. Res.*, **314**, 177–186.

Martin,A.C.R., MacArthur,M.W. and Thornton,J.M. (1997) Assessment of comparative modeling in CASP2. *Proteins*, **1**, 14–28.

Masson,M., Niedergang,C., Schreiber,V., Muller,S., Menissier-de Murcia,J. and de Murcia,G. (1998) XRCC1 is specifically associated with poly(ADP-ribose) polymerase and negatively regulates its activity following DNA damage. *Mol. Cell. Biol.*, **18**, 3563–3571.

Miki,Y. *et al*. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science*, **266**, 66–71.

Nash,R.A., Caldecott,K.W., Barnes,D.E. and Lindahl,T. (1997) XRCC1 protein interacts with one of two distinct forms of DNA ligase III. *Biochemistry*, **36**, 5207–5211.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.

Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.

Ouchi,T., Monteiro,A.N., August,A., Aaronson,S.A. and Hanafusa,H. (1998) BRCA1 regulates p53-dependent gene expression. *Proc. Natl Acad. Sci. USA*, **95**, 2302–2306.

Scully,R., Anderson,S.F., Chao,D.M., Wei,W., Ye,L., Young,R.A., Livingston,D.M. and Parvin,J.D. (1997a) BRCA1 is a component of the RNA polymerase II holoenzyme. *Proc. Natl Acad. Sci. USA*, **94**, 5605–5610.

Scully,R., Chen,J., Plug,A., Xiao,Y., Weaver,D., Feunteun,J., Ashley,T. and Livingston,D.M. (1997b) Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell*, **88**, 265–275.

Shen,M.R., Zdzienicka,M.Z., Mohrenweiser,H., Thompson,L.H. and Thelen,M.P. (1998) Mutations in hamster single-strand break repair gene *XRCC1* causing defective DNA repair. *Nucleic Acids Res.*, **26**,1032–1037.

Sonnhammer,E.L.L., Eddy,S.R. and Durbin,R. (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins*, **28**, 405–420.

Taylor,R.M., Wickstead,B., Cronin,S. and Caldecott,K.W. (1998) Role of a BRCT domain in the interaction of DNA ligase III-α with the DNA repair protein XRCC1. *Curr. Biol.*, **8**, 877–880.

Tebbs,R.S., Meneses,J.J., Pedersen,R.A., Thompson,L.H. and Cleaver,J.E. (1996) XRCC1 knockout mice are nonviable. *Environ. Mol. Mutagen.*, **27**, 68.

Thompson,L.H., Brookman,K.W., Jones,N.J., Allen,S.A. and Carrano,A.V. (1990) Molecular cloning of the human *XRCC1* gene, which corrects defective DNA strand break repair and sister chromatid exchange. *Mol. Cell. Biol.*, **10**, 6160–6171.

Received August 21, 1998; revised September 4, 1998; accepted September 8, 1998