From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair

Isabelle Callebaut*, Jean-Paul Mornon

Systèmes Moléculaires & Biologie Structurale, Laboratoire de Minéralogie-Cristallographie, CNRS URA09, Universités Paris 6 and Paris 7, case 115, T.16, 4 place Jussieu, 75252 Paris Cedex 05, France

Received 26 September 1996

Abstract Inherited mutations in BRCA1 predispose to breast and ovarian cancer, but the biological function of the BRCA1 protein has remained largely elusive. The recent correspondence of Koonin et al. [Koonin, E.V., Altschul, S.F. and Bork, P. (1996) Nature Genet. 13, 266–267] has emphasized the potential importance of the BRCA1 C-terminal region for BRCA1-mediated breast cancer suppression, as this domain shows similarities with the C-terminal regions of a p53-binding protein (53BP1), the yeast RAD9 protein involved in DNA repair, and two uncharacterized, hypothetical proteins (KIAA0170 and SPAC19G10.7). The highlighted domain has been suggested to be the result of an internal duplication, each of the tandem domains being designated as a 'BRCT domain' (for BRCA1 C-terminus). Sequence analysis using hydrophobic cluster analysis reveals here the presence of 50 copies of the BRCT domain in 23 different proteins, including, in addition to BRCA1, 53BP1 and RAD9, XRCC1, RAD4, Ect2, REV1, Crb2, RAP1, terminal deoxynucleotidyltransferases (TdT) and three eukaryotic DNA ligases. Most of these proteins are known to be involved in DNA repair. The BRCT domain is not limited to the C-termini of protein sequences and can be found in multiple copies or in a single copy as in RAP1 and TdT, suggesting that it could well constitute an autonomous folding unit of approx. 90–100 amino acids.

Key words: Hydrophobic cluster analysis; Sequence analysis; BRCA1; DNA repair; Cancer; RAP1; DNA polymerase; DNA ligase

1. Introduction

The cloning of the familial breast and ovarian cancer susceptibility gene BRCA1 [1] was an important milestone in cancer research. Cancer-predisposing alleles of BRCA1, which generally behave as recessive alleles in somatic cells, typically carry mutations that cause loss or reduction of the gene function and the wild-type allele is lost in tumor tissue [1,2], suggesting that BRCA1, like many other genes involved in familial cancer, is a tumor suppressor gene. In sporadic tumors, somatic point mutations are very rare; complete somatic deletion of one allele of BRCA1 is often observed, with a decrease in BRCA1 mRNA expression [3–6]. Evidence of a role in tumor suppression is further supported by the observations of growth acceleration of both normal and malignant breast epithelial cells following inhibition of BRCA1 expression [7] and growth inhibition of tumor cell lines after transfection with wild-type, but not mutant, BRCA1 [8]. This activity appears to be tumor-type specific as BRCA1 cannot inhibit the growth of some other cancer cell lines, nor can it inhibit the growth of normal fibroblasts [8].

BRCA1 encodes a predicted protein of 1863 amino acids containing in its NH2-terminus a single C_3HC_4-type zinc finger domain, also referred to as the RING finger or A-box and found in various proteins showing transactivation activity for a number of viral and cellular genes [1,9]. The rest of the protein was initially reported to contain no significant similarities to any known genes. A recent report paradoxically suggests that the BRCA1 protein, whose sequence is found to match with a 'granin' consensus, might be secreted and so would function by a mechanism so far undescribed for tumor suppressor gene products [10]. On the other hand, several lines of evidence suggest that the C-terminal end of BRCA1 is essential to the normal function of the protein in breast epithelial cells. Patients inheriting 1853Stop were shown to develop very early onset breast cancer [11]. Moreover, truncations of the BRCA1 C-terminal region were shown to suppress the ability of BRCA1 to inhibit breast cancer cell growth [8]. Finally, this region of BRCA1 has recently been reported to act as a transcriptional transactivator when fused to the GAL4 DNA-binding domain [12].

In order to gain more insight into the structural and functional features of this essential region, we have used hydrophobic cluster analysis (HCA) [13,14] in combination with well-established linear methods of sequence analysis. HCA is indeed able to detect three-dimensional similarities between proteins showing very limited sequence relatedness. Its sensitivity at low levels of sequence identity (typically below the so-called twilight zone (25–30%)) stems from its ability to detect significantly secondary structure elements [15]. The effectiveness of the HCA method has been widely demonstrated (see, among others [16–20]).

The use of this method has led us to identify within the BRCA1 C-terminus a repeated motif which is widespread in several nuclear proteins closely related to cell cycle regulation and DNA repair. These findings complement the recent correspondence of Koonin and colleagues [21] in which they report the presence of this motif, which they named BRCT (BRCA1 C-terminus), in the repair protein RAD9 and in a p53-binding protein. Here we extend the retrieval of this module to XRCC1, RAD4, Ect2, REV1, Crb2, RAP1, terminal deoxynucleotidyltransferases (TdT) and three eukaryotic DNA ligases and emphasize its potential role in cell cycle control.

2. Materials and methods

Systematic searches of databanks [22,23] allow detection of sequences which could belong to the same functional and/or structural

*Corresponding author. Fax: (33) (1) 44 27 37 85.
E-mail: callebau@lmcp.jussieu.fr

Abbreviations: HCA, hydrophobic cluster analysis; TdT, terminal deoxynucleotidyltransferase.

PII S0014-5793(96)01312-9

family. However, at the low levels of sequence identity (< 25–30%) often observed, these automatic methods are unable to distinguish similarities due to structural relationships from background noise. The 'hydrophobic cluster analysis' method [13,14] is helpful in this regard insofar as it allows comparison of not only the sequences but also the protein secondary structures statistically centered on hydrophobic clusters, as well as their distribution [15]. Similar plots could therefore indicate similar three-dimensional folds. Guidelines to the use of this method are given in [13,14].

The HCA score is proportional to the hydrophobic amino acids which are topologically conserved (often not chemically identical), and therefore reflects the degree of conservation of the hydrophobic core. High HCA scores are associated with low root mean squares values between three-dimensional structures [14]. The accuracy of the alignments can be assessed by computing identity, similarity or HCA scores, as well as the corresponding Z scores as initially suggested by Doolittle [24]; these represent differences between the alignment score under consideration and the mean score of a distribution computed for alignment of sequence 1 versus a large number of randomly shuffled versions of sequence 2 (here 1000). These differences are expressed relative to the standard deviation (SD) of the random distribution. Scores that are 3.0 or more standard deviation above the scrambled mean scores can reasonably be expected to represent authentic relationships.

3. Results

The BRCT family members listed in Fig. 1 have been identified by first searching the sequence databases using standard 1D methods such as BlastP [23] and FastA [22] and then sorting and assessing the putative 3D relationships to the family through HCA [13,14] (see Section 2).

Conserved motifs (similar hydrophobic motifs often associated with sequence conservation) define five regions, designated *A–E*, which can be used to decipher the main features of the BRCT module (Figs. 2 and 3). The most highly conserved motif, motif *D*, is organized around a conserved aromatic residue (W, F or Y). The residue following it is always hydrophobic, as are usually the fourth and fifth residues preceding it. The conservation of this hydrophobic pattern can easily be visualized on the HCA plots (Fig. 3). Another striking feature of this motif *D* is that the fourth position after the conserved aromatic residue is usually occupied by a cysteine or a serine. Other positions also show some preferences for particular amino acids (often proline in positions −6 and −2 and hydrophobic residues in position +5). The region corresponding to this motif has been used by Koonin et al. [21] to derive a signature which, however, is too strict as it succeeds in 'picking up' only five proteins of the family. Moreover, in one of these five proteins, the SPAC19G10.7 hypothetical protein, the signature is not able to detect four additional BRCT domains which nonetheless harbor typical features of the module (Figs. 1 and 2). In particular, this pattern includes the region preceding the motif *D* which in fact appears less highly conserved than initially predicted. However, the use of a degenerate signature based on the unique motif *D* defined here is too permissive to describe BRCT domains. This motif should therefore be associated with several of the other motifs to assess the BRCT prediction.

Motif *B* is the second most highly conserved feature of the domain: it consists of two consecutive glycines preceded in positions −4 and −8 (relative to the first glycine) by hydrophobic residues. Although this pattern is not absolutely conserved in all of the domains, it can also be easily visualized on the plots as glycine is represented by a particular symbol (Fig. 3).

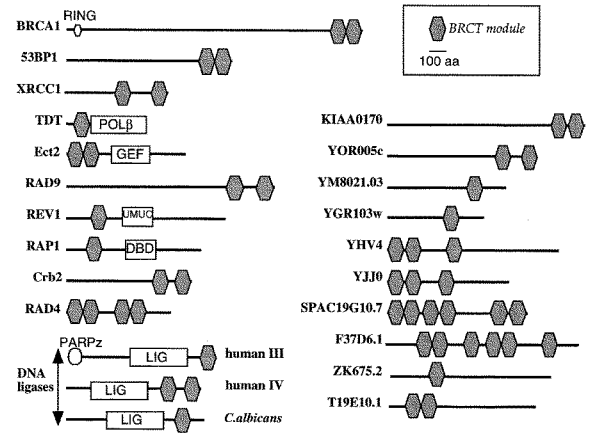

Fig. 1. Position of the BRCT domains within BRCT domain-containing proteins. Abbreviations, correspondences and sequence references are given in the legend to Fig. 2. Additional modules showing similarities with other proteins: RING, ring finger domain; POLβ, region similar to polymerase β; GEF, region similar to GTP-exchanging factor; UMUC, region similar to the bacterial DNA repair protein UmuC; DBD, DNA binding domain; LIG, region showing similarities with human DNA ligase I and corresponding to the minimal size of ATP-dependent bacterial ligases (Callebaut et al., in preparation); PARPz, region similar to the Zn fingers of human poly(ADP-ribose) polymerase (PARP).

Motif *C* corresponds to a continuous stretch of three or four hydrophobic amino acids (a vertical shape in the HCA plots) or, in several domains, to the sequence THΦΦ where Φ is a hydrophobic amino acid (V, I or L). This motif most probably corresponds to an internal β-strand. Two other β-strands probably constitute the motifs *A* and *E*, which begins and ends the domain, respectively, and whose shapes are also well retrieved within the family.

The limits of the domain can be well defined, especially since in some proteins such as RAP1, it is surrounded by non-globular regions mainly composed of non-hydrophobic amino acids or is located N- or C-terminal in the protein sequences. Moreover, in RAP1 as well as in TdT, the BRCT domain is found isolated in a single copy, suggesting that it could well constitute an autonomous folding unit. The minimal length of this domain can be fixed to approx. 100 amino acids but it appears to tolerate insertions of considerable length, especially between blocks A and B and blocks B and C (Fig. 2).

In conclusion, this investigation within a family of proteins sharing very low levels of sequence identity enables the identification of BRCT domains only if several of the 'BRCT clues' are brought together in a compatible way, excluding isolated motifs, even the stronger ones, which can occur by chance. When the BRCT pattern was found to be highly divergent, we verified the proposition by carrying out 1D searches with the candidate domain and retrieving in the output compatible alignments with established members of the family. For example, the fifth BRCT domain of the hypothetical protein F37D6.1 has a highly degenerate motif *D* (Figs. 2 and 3), but the relationship with the BRCT family is supported by the fact that it can be aligned with reasonable scores to SPAC19G10.7 and RAD4 (BLASP P values = 0.041 and 0.058 with the BLOSUM62 matrix, respectively). Good similarities also appear with BRCA1. In this example, the similarities are especially concentrated in the region of motifs

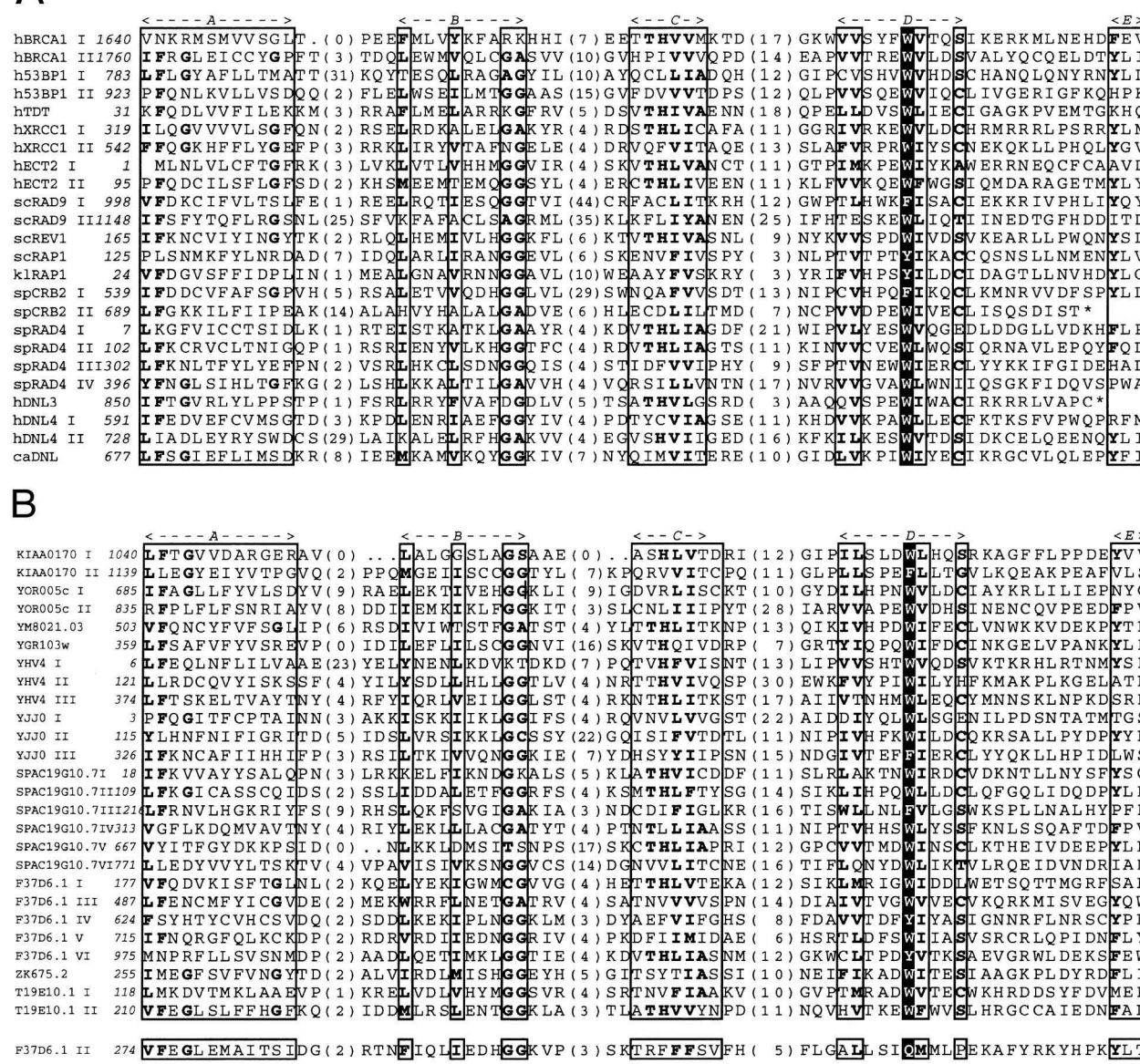

Fig. 2. HCA-based alignment of BRCT domains. The position of the first amino acid is given for each domain. Five aligned blocks (designated *A–E*) constituting the five conserved regions of the domain are separated from each other by more variable regions. Distances between the blocks are indicated by numbers within parentheses. Star indicates the end of a sequence. Highly conserved motifs are shown boxed. Alternative solutions to the proposed alignment of Koonin et al. [21] are given for the BRCA1 I and RAD9 I blocks A, as well as for the SPAC19G10.07-I block A (reported with a write error in [21]). The second BRCT domain of the *C. elegans* F37D6.1 hypothetical protein is shown separately as it possesses a highly divergent block D. Sequences were taken from the SWISS-PROT (sw), PIR (pir) and Genbank (gb) databases and are listed here with their accession numbers (AC): (A) hBRCA1 = human BRCA1 (sw AC: P38398, 1863 amino acids), h53BP1 = human 53BP1 (gb AC: U09477, 1027 amino acids), hXRCC1 = human XRCC1 (sw AC: P18887, 633 amino acids), hTDT = human terminal deoxynucleotidyltransferase (sw AC: P04053, 508 amino acids), mECT2 = mouse Ect2 oncogene (pir AC: S32372, 738 amino acids), scRAD9 = *S. cerevisiae* RAD9 (sw AC: P14737, 1309 amino acids), scREV1 = *S. cerevisiae* REV1 (sw AC: P12689, 985 amino acids), scRAP1 = *S. cerevisiae* RAP1 (sw AC: P11938, 827 amino acids), klRAP1 = *K. lactis* RAP1 (gb AC: X73629, 666 amino acids), spCRB2 = *S. pombe* Crb2 (gb AC: D86478, 778 amino acids), spRAD4 = *S.pombe* RAD4 (sw AC: P32372, 648 amino acids), hDNL3 = human DNA ligase III (gb AC: X84740, 922 amino acids), hDNL4 = human DNA ligase IV (gb AC: X84441, 844 amino acids), caDNLI = *C. albicans* DNA ligase (gb AC: X95001, 864 amino acids). (B) Uncharacterized proteins from human (h), *S. cerevisiae* (sc), *S.pombe* (sp) and *C. elegans* (ce): hKIAA0170 (gb AC: D79992, 2088 amino acids), scYOR005c (gb AC: Z74913, 944 amino acids), scYM8021.03 (pir AC: S54584, 732 amino acids), scYGR103w (gb AC: Z72888, 605 amino acids), scYHV4 (sw AC: P38850, 1070 amino acids), scYJJ0 (sw AC: P47027, 764 amino acids), spSPAC19G10.07 (gb AC: Z69909, 878 amino acids), ceF37D6.1 (gb AC: Z75540, 1214 amino acids), ceZK675.2 (gb AC: Z46812, 1027 amino acids), ceT19E10.1 (gb AC: Z46795, 932 amino acids). The sequence of the *S. cerevisiae* hypothetical protein UNE452 (gb AC: U43491, 452 amino acids) is identical to that of scYOR005c. The sequence of the *S. cerevisiae* hypothetical protein YJJ0 is identical to that of Dpb11, which interacts with DNA polymerase II (epsilon) and has a dual role in S-phase progression and a cell cycle checkpoint [54]. The sequences of the BRCT domains of mouse, bovine, chicken and *X. laevis* TdTs are not shown as they share more than 50% identity with the BRCT domains of human TdT (85, 86, 63 and 59%, respectively). The sequences of the two BRCT domains of mouse BRCA1, sharing 75 and 58% identity with those of human BRCA1, respectively, are also not shown.

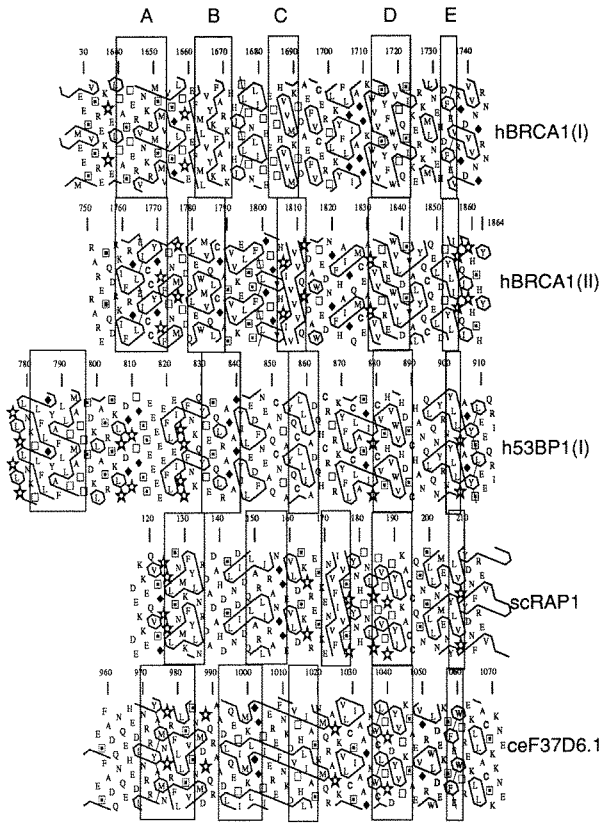

Fig. 3. Comparison of the HCA plots of BRCT domains of human BRCA1 (the two domains), human 53BP1 (first one), *S. cerevisiae* RAP1 and *C. elegans* F37D6.1 (fifth one). Conserved regions (A–E) described in Fig. 2 are shown within boxes. The protein sequences are written on a duplicated α-helical net and the contours of clusters of hydrophobic residues are automatically drawn. The standard one-letter code for amino acids is used except for proline (regular secondary structure breaker), glycine (the least constrained amino acid), serine and threonine (which can be accommodated either at the protein surface or in a hydrophobic environment masking their hydroxyl group through an H-bond with the polypeptide backbone) which are represented by (★), (♦), (▣) and (□), respectively.

B (the two glycines are present) and *C* (containing the sequence THΦΦ which is frequently retrieved and where Φ is an hydrophobic amino acid). Paradoxically, the use of this non-typical BRCT domain has allowed detection of the unique BRCT domain of the terminal deoxynucleotidyltransferase (TdT), which otherwise possesses a typical motif *D*. However, the presence of the motif was pointed out only through HCA, as the BlastP alignment (with a *P* value of 1.0) focused only on the *B* and *C* motifs.

As a general rule, pairwise identity and similarity scores leave little doubt about the relationships between the different domains described here. They were further assessed by the calculation of the 1225 pairwise *Z* scores (number of standard deviations that the score of pairwise optimally aligned sequences differs from the 'random' mean). The mean and standard deviation of identity *Z* scores are 3.03 and 1.80, respectively; those of similarity *Z* scores (using the Dayhoff matrix [25]) are 3.90 and 1.67, respectively (4.21 and 1.82 when using the BLOSUM62 matrix [26]).

Finally, profile searches with the multiple alignment given in Fig. 2 resulted in higher scores for the above-mentioned BRCT domains than for all other proteins of the sequence databases.

The BRCT domain was initially highlighted as a tandem duplicate in the C-terminal regions of BRCA1, the p53-binding protein 53BP1, and the RAD9 protein involved in DNA repair [21]. We show here that the BRCT domain can be retrieved widespread and arranged in different ways in the proteins shown in Fig. 1. Saka et al. [27] have already pointed out the presence of the two tandem repeats in RAD4 and have related them to similar regions found in the oncogene Ect2 [28] (two domains) and the repair proteins REV1 [29] (one domain) and XRCC1 [30] (one domain). However, they were not able to detect the second BRCT domain of XRCC1 which is separated from the first by an acidic region. This domain nonetheless possesses strong features of the family (Fig. 2). Nor was any relationship detected with the rest of the family, which includes terminal deoxynucleotidyltransferases (TdT), RAP1, three eukaryotic DNA ligases and Crb2 and several hypothetical proteins (Fig. 1).

4. Discussion

The discovery of a common domain between BRCA1 and a p53-binding protein has led to the hypothesis that BRCA1 may bind p53, the universal tumor suppressor [21]. The retrieval of this domain, named BRCT (for BRCA1 C-Terminus), in the yeast RAD9 protein [21] also indicates that BRCA1 may play a role in cell cycle checkpoints, the negative controls which impose delays in the eukaryotic cell cycle [31]. Following DNA damage, RAD9 is required to delay the cell cycle at the G1 and G2 checkpoints [32,33]. RAD9 is also involved in another signal related to DNA repair as it participates in the transcriptional response to DNA damage by controlling the induction of a large 'regulon' of repair, replication and recombination genes [34]. In this context, it is interesting to note that truncations of the BRCA1 C-terminal region, encompassing the BRCT domains, were shown to suppress the ability of BRCA1 to inhibit breast cancer cell growth, suggesting that it negatively regulates cell division [8]. Moreover, this region of BRCA1 has recently been reported to act as a transcriptional transactivator when fused to the GAL4 DNA-binding domain [12]. The investigation of the target genes regulated by BRCA1 would certainly enhance our understanding of the potential role of BRCA1 in cell cycle control and repair/replication-associated processes. Given the fact that p53 is known to trigger cell cycle delay at G1 [35,36], it is plausible to hypothesize that the potential activity of BRCA1 in cell cycle regulation occurs through an interaction with p53.

The hypothesis of a role of BRCT domains in cell cycle

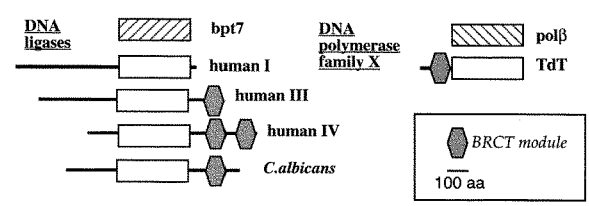

Fig. 4. Enzymes containing BRCT domains – relationship to related proteins. The BRCT domains are shown grey shaded. The structures of the enzymatic domains of the DNA ligase of bacteriophage T7 (bpt7) and rat polymerase β (polβ), represented by hatched boxes, have been solved experimentally ([55] and [47,48], respectively).

regulation and in repair mechanisms is strengthened here by the observation that this module is found to be widespread in other proteins also closely involved in these processes. Human XRCC1, which was shown here to possess two copies of the BRCT module, is required for repair of DNA single-strand breaks formed by exposure to ionizing radiation and alkylating agents [30]. The *S. cerevisiae* REV1 gene product functions in a cellular process required for mutagenesis caused by UV radiation and many chemical mutagens [29]. Mutations in *S. pombe* RAD4 are known to confer sensitivity to UV and ionizing radiation damage, as well as a temperature-sensitive phenotype [37,38]. For some of these proteins like for BRCA1, several lines of experimental evidence suggest a direct functional role of the BRCT module in cell cycle control. Indeed, the N-terminal domain of RAD4 (*cut5*), containing the first tandem repeat, is essential for complementation of temperature-sensitive *cut5* mutants [39] and its overexpression severely inhibits cell division [27]. Truncation of the N-terminal domain of the oncogene product Ect2 encompassing its two BRCT domains increases its transforming activity, suggesting that this region has a negative effect on cell division [28].

Three of the proteins reported here as members of the BRCT family are known to possess a catalytic activity. Terminal deoxynucleotidyltransferase (EC 2.7.7.31; TdT) is a template-independent DNA polymerase which catalyzes the elongation of polydeoxynucleotide chains by terminal addition [40]. One of the in vivo functions of this enzyme is the addition of nucleotides at the junction of rearranged Ig heavy chain and T cell receptor gene segments during the maturation of B and T cells [41–44]. It belongs to the family X of DNA polymerases which contains vertebrate polymerases β [45]. Both enzymes, TdT and polymerase β, share similarities within the catalytic domain [46–48] although TdT has an additional N-terminal domain reported to be non-essential but shown here to correspond to a BRCT domain (Fig. 4). The primary difference between the two enzymes is the utilization of template by polymerases β and not by TdT, but the reason for this difference is thought to lie in the divergence of the template binding domain included in the catalytic domain, rather than in the presence of an additional domain in TdTs [46].

Interestingly, the yeast REV1 protein, which also possesses a single copy of the BRCT module, has recently been reported to have a deoxycytidyl transferase activity which transfers a dCMP residue from dCTP to the 3′ end of a DNA primer in a template-dependent reaction [49]. As it uses only dCTP, at least with the template primers tested, and in view of its unusual template requirements, REV1 would represent a new category of nucleotide polymerizing enzymes. The presence of a common BRCT motif between two of these enzymes, TdT and REV1, is suggestive of a functional role in the polymerization event although it is apparently not directly linked to the enzymatic activity, as suggested by its non-essential role in TdT and its absence in polymerase β.

The BRCT domains found in three eukaryotic, ATP-dependent DNA ligases are also located outside the predicted catalytic domain of these enzymes. ATP-dependent ligases (EC 6.5.1.1) are found in eukaryotes, archaebacteria, viruses and phages whereas bacterial DNA ligases are NAD-dependent (EC 6.5.1.2). Four DNA ligases, I–IV, have been identified in mammalian cells. A short peptide sequence conserved close to the C-terminus of all known eukaryotic DNA ligases has allowed the identification of DNA ligases III and IV [50]. Compared to DNA ligase I, DNA ligases III and IV possess additional domains following a conserved catalytic domain which are shown here to correspond to BRCT domains (one and two domains for DNA ligase III and IV, respectively Fig. 4). A similar organization is found in the DNA ligase of *C. albicans*, with a single copy of the BRCT module following the catalytic domain (Fig. 4). Interestingly, DNA ligase III is known to form a characteristic, high-salt-resistant complex with XRCC1 [50].

A single BRCT module can also be retrieved in the yeast RAP1, a multifunctional protein which in addition to its role in regulating telomere length and initiating telomeric silencing, functions as a context-dependent transcriptional regulator at many other sites throughout the genome [51]. It contains a centrally located DNA-binding domain, a C-terminal domain including regulatory domains for both activation and repression of transcription and an N-terminal part which has no known biological function, except for its involvement in in vitro DNA bending [52]. It is this very region which is made of non-globular regions surrounding a single BRCT domain.

The non-essential character of the BRCT domain of RAP1 and enzymes such as TdT apparently contrasts with the fact that it is retrieved widespread within proteins which are related by their close involvement in DNA metabolism and with the important role that it may play in cell cycle control in RAD4, ECT2 and BRCA1. Given these data, it is tempting to speculate about the molecular role of BRCT domains. This module might serve to interact with other proteins involved in cell cycle regulation, as suggested by the fact that it is present in a p53-binding protein and that the human DNA ligase III tightly binds XRCC1 [50]. However, no experimental evidence exists for a direct involvement of the BRCT domains, rather than other domains contained in these large proteins, in protein-protein interaction. Another hypothesis is that the BRCT module might be a molecular sensor allowing direct or indirect recognition of particular DNA structures. This recognition could allow the activation of repair mechanisms or, on the contrary, protect the recognized region from the repair system. In this context, it is interesting to note that, in addition to its known binding activity for double-stranded DNA, RAP1 binds sequence in a sequence specific manner to the telomeric terminal GT tails [53]. However, this novel DNA binding activity involves regions of RAP1 located outside of the DNA binding domain. Whether or not it involves the BRCT domain therefore deserves some further investigation.

Acknowledgements: The authors acknowledge the ORGANIBIO-CM_2AO structural biology program for financial support. Sequence data were analysed using the genome facilities provided by P. Dessen and the ExPaSy and NCBI servers (http://expasy.hcuge.ch/ and http://ncbi.nlm.nih.gov/). A general HCA plot computing facility is available at the URL http://www.lmcp.jussieu.fr/~mornon/.

References

[1] Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bodgen, R., Dayananth, P., Ward, J., Tonin, P.,

Narod, S., Bristow, P.K., Norris, F.H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, J.C., Lewis, C., Neuhausen, S., Cannon-Albright, L., Goldgar, D., Wiseman, R., Kamb, A. and Skolnick, M.H. (1994) Science 66, 66–71.
[2] Smith, S.A., Easton, D.F., Evans, D.G.R. and Ponder, B.A.J. (1992) Nature Genet. 2, 128–131.
[3] Futreal, P.A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., Bennett, L.M., Haugen-Strano, A., Swensen, J., Miki, Y., Eddington, K., McClure, M., Frye, C., Weaver-Feldhaus, J., Ding, W., Gholami, Z., Sšderkvist, P., Terry, L., Jhanwar, S., Berchuck, A., Iglehart, J.D., Marks, J., Ballinger, D.G., Barrett, J.C., Skolnick, M.H., Kamb, A. and Wiseman, R. (1994) Science 266, 120–122.
[4] Merajver, S.D., Pham, T.M., Caduff, R.F., Chen, M., Poy, E.L., Cooney, K.A., Weber, B., Collins, F.S., Johnston, C. and Frank, T.S. (1995) Nature Genet. 9, 439–443.
[5] Hosking, L., Trowsdale, J., Nicolai, H., Solomon, E., Foulkes, W., Stamp, G., Signer, E. and Jeffreys, A. (1995) Nature Genet. 9, 439–443.
[6] Takahashi, H., Behbakht, K., McGovern, P.E., Chiu, H.-C., Couch, F.J., Weber, B.L., Friedman, L.S., King, M.-C., Furusato, M., LiVolsi, V.A., Menzin, A.W., Liu, P.C., Benjamin, I., Morgan, M.A., King, S.A., Rebane, B.A., Cardonick, A., Mikuta, J.J., Rubin, S.C. and Boyd, J. (1995) Cancer Res. 55, 2998–3002.
[7] Thompson, M.E., Jensen, R.A., Obermiller, P.S., Page, D.L. and Holt, J.T. (1995) Nature Genet. 9, 444–450.
[8] Holt, J.T., Thompson, M.E., Szabo, C., Robinson-Benion, C., Arteaga, C.L., King, M.-C. and Jensen, R.A. (1996) Nature Genet. 12, 298–302.
[9] Bienstock, R.J., Darden, T., Wiseman, R., Pedersen, L. and Barrett, J.C. (1996) Cancer Res. 56, 1539–2545.
[10] Jensen, R.A., Thompson, M.E., Jetton, T.L., Szabo, C.I., van der Meer, R., Helou, B., Tronick, S.R., Page, D.L., King, M.-C. and Holt, J.T. (1996) Nature Genet. 12, 303–311.
[11] Friedman, L.S., Ostermeyer, E.A., Szabo, C.S., Dowd, P., Lynch, E.D., Rowell, S.E. and King, M.-C. (1994) Nature Genet. 8, 399–404.
[12] Chapman, M.S. and Verma, I.M. (1996) Nature 382, 678–679.
[13] Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J.-P. (1987) FEBS Lett. 224, 149–155.
[14] Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. and Mornon, J.-P. (1990) Biochimie 72, 555–574.
[15] Woodcock, S., Mornon, J.-P. and Henrissat, B. (1992) Prot. Eng. 5, 629–635.
[16] Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.-P. and Davies, G. (1995) Proc. Natl. Acad. Sci. USA 92, 7090–7094.
[17] Saxena, I.R., Brown, R.M.J., Fevre, M., Geremia, R.A. and Henrissat, B. (1995) J. Bacteriol. 177, 1419–1424.
[18] Thoreau, E., Petridou, B., Kelly, P.A., Djiane, J. and Mornon, J.-P. (1991) FEBS Lett. 282, 26–31.
[19] Callebaut, I. and Mornon, J.P. (1995) FEBS Lett. 374, 211–215.
[20] Callebaut, I. and Mornon, J.-P. (1996) Biochem. J. (in press).
[21] Koonin, E.V., Altschul, S.F. and Bork, P. (1996) Nature Genet. 13, 266–267.
[22] Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448.
[23] Altschul, S., Gish, W., Miller, W., Myers, E. and Lipman, D. (1990) J. Mol. Biol. 215, 403–410.
[24] Doolittle, R.F. (1981) Science 214, 149–159.
[25] Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) Methods Enzymol. 91, 524–545.
[26] Henikoff, S. and Henikoff, J.G. (1992) Proc. Natl. Acad. Sci. USA 89, 10925–10919.
[27] Saka, Y., Fantes, P., Sutani, T., Mclnrtny, C., Creanor, J. and Yanagida, M. (1994) EMBO J. 13, 5319–5329.
[28] Miki, T., Smith, C.L., Long, J.E., Eva, A. and Fleming, T.P. (1993) Nature 362, 462–465.
[29] Larimer, F.W., Perry, J.R. and Hardigree, A.A. (1989) J. Bacteriol. 171, 230–237.
[30] Thompson, L.H., Brookman, K.W., Jones, J.J., Allen, S.A. and Carrano, A.V. (1990) Mol. Cell. Biol. 10, 6160–6171.
[31] Friedberg, E.C., Walker, G.C. and Siede, W. (1995) DNA Repair and Mutagenesis, ASM Press, Washington, DC.
[32] Siede, W., Friedberg, A.S., Dianova, I. and Friedberg, E.C. (1994) Genetics 138, 271–281.
[33] Weinert, T.A., Kiser, G.L. and Hartwell, L.H. (1994) Genes Dev. 8, 652–665.
[34] Aboussekhra, A., Vialard, J.E., Morisson, D.E., de la Torre-Ruiz, M.A., Cernakova, L., Fabre, F. and Lowndes, N.F. (1996) EMBO J. 15, 3912–3922.
[35] Kastan, M.B., Zhan, Q., El-Deiry, W.S., Carrier, F., Jacks, T., Walsh, W.V., Plunkett, B.S., Vogelstein, B. and Fornace, A.J., Jr (1992) Cell 71, 587–597.
[36] Kuerbitz, S.J., Plunkett, B.S., Walsh, W.V. and Kastan, M.B. (1992) Proc. Natl. Acad. Sci. USA 89, 7491–7495.
[37] Duck, P., Nasim, A. and James, A.P. (1976) J. Bacteriol. 128, 536–539.
[38] Fenech, M., Carr, A.M., Murray, J.M., Watts, F.Z. and Lehmann, A.R. (1991) Nucl. Acids Res. 19, 6737–6741.
[39] Saka, Y. and Yanagida, M. (1993) Cell 74, 383–393.
[40] Bollum, F.J. (1974) in: The Enzymes (Boyer, P.D. ed.), pp. 145–171, Academic Press, New York.
[41] Desiderio, S.V., Yancopoulos, G.D., Paskind, M., Thomas, E., Boss, M.A., Landau, N., Alt, F.W. and Baltimore, D. (1984) Nature 311, 752–755
[42] Yancopoulos, G.D., Blachwell, T.K., Suh, H., Hood, L. and Alt, F.W. (1986) Cell 44, 251–259.
[43] Lieber, M.R., Hesse, J.E., Mizunchi, K. and Gellert, M. (1987) Genes Dev. 1, 751–761.
[44] Schatz, D.G. and Baltimore, D. (1988) Cell 53, 107–115.
[45] Ito, J. and Braithwaite, D.K. (1991) Nucl. Acids Res. 19, 4045–4057.
[46] Anderson, R.S., Lawrence, C.B., Wilson, S.H. and Beattie, K.L. (1987) Gene 60, 163–173.
[47] Davies, J.F., Almassy, R.J., Hostomska, Z., Ferre, R.A. and Hostomsky, Z. (1994) Cell 76, 1123–1133.
[48] Sawaya, M.R., Pelletier, H., Kumar, A., Wilson, S.H. and Kraut, J. (1994) Science 264, 1930–1935.
[49] Nelson, J.R., Lawrence, C.W. and Hinkle, D.C. (1996) Nature 382, 729–731.
[50] Wei, Y.-F., Robins, P., Carter, K., Caldecott, K., Pappin, D.J.C., Yu, G.-L., Wang, R.-P., Shell, B.K., Nash, R.A., Schär, P., Barnes, D.E., Haseltine, W.A. and Lindahl, T. (1995) Mol. Cell. Biol. 15, 3206–3216.
[51] Shore, D. (1996) Nature Struct. Biol. 3, 491–493.
[52] Müller, T., Gilson, E., Schmidt, R., Giraldo, R., Sogo, J., Gross, H. and Gasser, S.M. (1994) J. Struct. Biol. 113, 1–12.
[53] Giraldo, R. and Rhodes, D. (1994) EMBO J. 13, 2411–2420.
[54] Miosga, T., Schaaff-Gerstenschlaeger, I., Chalwatzis, N., Baur, A., Boles, E., Fournier, C., Schmitt, S., Velten, C., Wilhelm, N. and Zimmermann, F.K. (1995) Yeast 11, 681–689.
[55] Subramanya, H.S., Doherty, A.J., Ashford, S.R. and Wigley, D.B. (1996) Cell 85, 607–615.