...Functional motifs...

Sir — The *BRCA1* gene is linked to about one-half of familial breast cancers and over 80% of families with inherited breast and ovarian cancer[1]. Although somatic mutations of *BRCA1* in sporadic breast and ovarian cancers are rare[2–4], loss of one of the *BRCA1* alleles has been observed in many of these tumours[5]. Moreover, a transfected *BRCA1* gene inhibits the growth of breast and ovarian cancer cells but not of other malignant cells[6]. Thus BRCA1 appears to act as a tumour suppressor but the actual function of the protein is not known. BRCA1 consists of 1863 amino acids and contains a highly conserved, N-terminal RING finger domain[7]. The rest of the protein has not shown statistically significant similarity to other protein sequences.

In seeking to explore in detail potentially important but relatively weak sequence similarities for the BRCA1 protein, we examined the globular domains of the protein. Long regions with predicted non-globular structure in proteins frequently have a biased amino acid composition that can obscure functionally important conserved motifs in database search outputs[8]. Therefore, we partitioned the BRCA1 sequence into putative globular and non-globular domains using the SEG program, with parameters optimized for this purpose[9], and searched the non-redundant protein database at the National Center for Biotechnology Information (NIH, Bethesda) separately with each of the 6 putative globular domain sequences using the BLASTP program[10] and different amino acid substitution matrices of the BLOSUM series[11]. Apart from the previously described RING finger domain in the N-terminal globular domain, the remaining 5 globular domains did not show high similarity to any proteins in the database. We noticed, however, a moderate similarity between the 202 amino acid residue long C-terminal globular domain of BRCA1 and an analogously located region of a human protein, designated 53BP1 (ref. 12), that has been identified by its ability to bind p53, the universal tumour suppressor[13]. The statistical significance of the alignment was not strong (*P* value = 0.15 with the BLOSUM85 matrix). Nevertheless, when the database was searched with either the C-terminal domain of BRCA1 or the 53BP1 protein sequence as the query, the alignment of these two sequences had the highest score (excluding trivial hits to mouse BRCA1). This alignment included two distinct but compatible segments separated by 100 and 97 amino acid residues in BRCA1 and 53BP1, respectively. The aligned regions were positioned at almost precisely the same distance from the C termini of the proteins.

Given the biological plausibility of the similarity between BRCA1

a

```
2D prediction                          eeee              eeee            eEeEEE              eeee    eEEEEEEee              eee
                                                                                                **=***==*|=*===*===**================
BRCA1/HUMAN      1651                  SMVVSGLTPEEF-------MLVYKFARKHHIT-( 6)-EETTHVVMKTDAEFVCERTLKYFLGIAGGKWVVSYFWVTQSIKERKMLNEHDFEV-(6)-
BRCA1/MOUSE      1594                  SMVVSGLTPKEV-------MTVQKFAEKYRLT-( 6)-EETTHVIIKTDAEFVCERTLKYFLGIAGGKWIVSYSWVVRSIQERRLLNVHEFEV-(6)-
53BP1/HUMAN      (776)                 KTLFLGYAFLLTM-(39)-ESQLRAGAGYILE-( 9)-AYQCLLIADQHC-----RTRKYFLCLASGIPCVSHVWVHDSCHANQLQNYRNYLL-(2)-
KIAA0170/HUMAN   1893                  KVLFTGVVDARGE------RAVLALGGSLAGS-( 1)-AEASHLVTDRIR-----RTVKFLCALGRGIPILSLDWLHQSRKAGFFLPPDEYVV-(9)-
SPAC19G10_7/SCHPO 652                  RVYITGYDKKPSI-( 1)-NLKLDMSITSNP-------SKCTHLIAPRIL-----RTSKFLCSIPYGPCVVTMDWINSCLKTHEIVDEEPYLL-(9)-
RAD9/YEAST       1026                  TVIESGFSTLFNF-( 4)-AKSLVNKGNTDNI-(17)-CRFACLITKRHL-----RSLKYLETLALGWPTLHWKFISACIEKKRIVPHLIYQY-(4)-
consensus N                             uUuoGUo             u uo             o oOhUUo          R$uKUU OUo G  uUsu wU  Ou    Uu    U U

consensus C          G  uz       rore     uU  U Uuu o   !    Uu  ogoouu          UUUo p            o      u uuo BwUU Ouu  bo    d u
BRCA1/HUMAN          GRNHQGP----KRARESQDRKIFRGLEICCYG-(10)-EWMVQLCGASVVK-( 9)-GVHPIVVVQFDAWTEDNGFHAIG-( 4)-APVVTREWVLDSVALYQCQELDTYLI  8
BRCA1/MOUSE          GRNHQGP----RRSRESR-EKLFKGLQVYCCE-(10)-ERMLQLCGASVVK-( 9)-GAHLVVIVQPSAWTEDSNCPDIG-( 4)-ARLVMWDWVLDSLSSYRCRDLDAYLV 15
53BP1/HUMAN          GYSLEEQRILDWQPRE----NPFQNLKVLLVS-(10)-SEILMTGGAASVK-(13)-GVFDVVVTDPSCPASVLKCAEAL-( 1)-LPVVSQEWVIQCLIVGERIGFKQHPK  8
KIAA0170/HUMAN       GFSLQDAL---SRARE---RRLLEGYEIYVTP-( 9)-GEIISCCGGTYLP-( 6)-KPQRVVITCPQDFPHCSIPLRVG-( 0)-LPLLSPEFLLTGVLKQEAKPEAFVLS  7
SPAC19G10_7/SCHPO    GCTLESAL---KRARAQG-PSLLEDYVVYLTS-(11)-ISIVKSNGGVCST-(13)-DGNVVLITCNEDSHIWTNFLDNA-( 5)-IFLQNYDWLIKTVLRQEIDVNDRIAD  6
RAD9/YEAST           GESFRLSLDSPSKGGIIKSNNIFSFYTQFLRG-(33)-FAFACLSAGRMLT-(32)-KLKFLIYANENNGKSQMKLLERL-(14)-FHTESKEWLIQTIINEDTGFHDDITD 11
2D prediction                                EEEEEe        eeeeee  eEEE         EEEEE                      eeeeeeehhhhhhhhhh
```

b

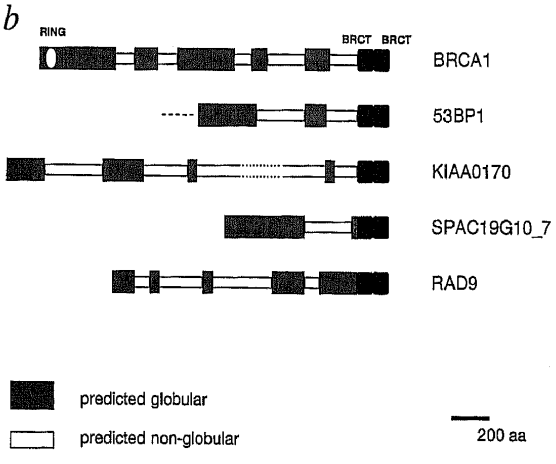

Fig. 1 The BRCT domain family. *a*, Multiple alignment of BRCT domains. We constructed the alignment using the MACAW program, with some modifications introduced on the basis of analysis with the CLUSTAL W program[27] and visual inspection. Distances from the protein termini to aligned regions and distances between the aligned blocks are indicated by numbers. Distance from the 53BP1 N terminus is shown in parentheses as the available sequence of this protein is incomplete. The two tandem BRCT domains are aligned to each other. The consensus lines for the two domains (consensus N and consensus C) include amino acid residues conserved in all of the 6 aligned sequences (upper case) or in at least 4 out of 6 sequences (lower case). U= bulky hydrophobic residue (I, L, V, M, F, Y, W); O = small residue (G, A, S, T, C); B = D or E; Z = E or Q; and $ = S or T. The residues conforming with the consensus are in bold. The most conserved motif, located at the C terminus of each of the two BRCT domains, is double-overlined. Amino acid residues comprising a unique signature derived from the N-terminal BRCT domain are indicated by asterisks. This signature can be presented as R[TS]xK[FY][ILVMF]xx[ILVMF]xxGxxx[ILVMF]xxx[FYW][ILVMF] (x = any residue; alternative residues are in square brackets. The pattern was expanded to include similar residues, even when only two were actually observed). The predicted secondary structure elements are shown below the alignment, with E/e indicating extended conformation (β-strand) and H/h indicating α-helix. Upper case indicates the most reliable prediction (>82% accuracy), and lower case indicates prediction with ~72% accuracy[18]. Exclamation marks indicate the positions of two missense mutations associated with familial breast cancer; one of these is located in a variable segment separating two aligned blocks. The BRCA1/HUMAN, BRCA1/MOUSE, and RAD9/YEAST sequences were from the SWISS-PROT database (Accession numbers P38398, P48754, and P14737, respectively); the 53BP1/HUMAN, KIAA0170/HUMAN, and SPAC19G10_7/SCHPO sequences were from GenBank (Accession numbers U09477, D79992, and Z69909, respectively). In addition to the aligned proteins, we detected a clear similarity to the predicted p53-binding domain in putative products of ESTs R24832 from human, D67484 from *C. elegans* , and N37100 from the nematode *Brugia malayi* . *b*, Organization of the large BRCT domain-containing proteins. The protein diagrams are aligned by the tandem BRCT domains. The predicted globular and non-globular domains are shown; the prediction was made using the SEG program with the following parameters: window size 45, trigger complexity 3.4, extension complexity 3.75 (ref. 9). The dashed line to the left of 53BP1 indicates that the sequence of this protein is incomplete from the N terminus. The dashed lines in the middle of KIAA0170 indicate an interruption introduced on the diagram into the central non-globular domain that includes over 950 amino acids. The diagrams are roughly to scale as shown at the bottom.

and a p53-binding protein, we investigated this potential relationship in detail. A 36 amino acid-long alignment of the most conserved regions of human and mouse BRCA1 and 53BP1 was converted to a position-dependent weight matrix. Using this to scan the database with the MoST program[14] and using a highly selective cut-off (ratio of 0.005 for the expected number of retrieved segments to the number actually observed), only two additional sequences were retrieved, those of uncharacterized, putative proteins from human and fission yeast (designated KIAA0170 and SPAC19G10_7, respectively) that showed a highly significant similarity to one another. MoST searches with a lower cut-off, as well as BLASTP searches with the detected sequences as queries also detected the motif in the yeast RAD9 protein, which is involved in the control of the DNA damage-induced cell cycle arrest at both the G_1 and G_2 checkpoints[15,16]. A reciprocal search — using aligned sequences of the conserved segments from 53BP1 and the uncharacterized human protein to derive a matrix that was run against the database — retrieved the human and mouse BRCA1 corresponding segments (probability of detection by chance below 10^{-4}). From the most conserved motif, we derived an amino acid signature that is not present in any sequence in the database except for the above 6 proteins (Fig. 1*a*). Based on these findings, we conclude that BRCA1, 53BP1, RAD9 and at least two uncharacterized proteins contain a new, conserved domain.

Multiple alignment analysis with the MACAW program[17] revealed additional, weaker motifs; all are in the C-terminal portion of the aligned proteins and contain ~180 residues in 5 proteins, but two inserts of ~30 residues each in RAD9 (Fig. 1*a*). Secondary structure prediction using the PHD method[18] suggests that the conserved domain consists of a β sheet (Fig. 1*a*). The domain could be the result of an internal duplication (Fig.1*a*); the presence of two tandem domains is supported by sequence conservation in a large superfamily of nuclear proteins mostly involved in cell cycle checkpoint functions (P.B. *et al.*, in preparation). We propose the designation BRCT (after BRCA1 C Terminus) for each of the tandem domains.

We define the family described here as a subset of a BRCT superfamily. All proteins in this proposed family are quite large and appear to have similar organization — BRCT domains located at the C terminus and upstream portions containing long non-globular regions (Fig. 1*b*). Proteins in this family may bind p53 or structurally similar proteins via the conserved C terminal domains. This is supported by the observation that the C-terminal 270 amino acids of 53BP1 are sufficient for p53 binding[12]. Truncation of BRCA1, even at position 1853, which removes only the 10 C-terminal residues, abolishes the ability of BRCA1 to suppress breast cancer cell growth[6]. Truncations in the *BRCA1* 3'-terminal portion also are among the most frequent mutations in familial breast cancers[19]. Furthermore, two missense mutations associated with such cancers mapped to the BRCA1 C terminus (refs 1,2; Fig. 1*a*). These data emphasize the potential importance of the BRCA1 C-terminal region, corresponding to the BRCT domain, for BRCA1-mediated breast cancer suppression. This domain may be required for cell cycle control in breast epithelial cells but not in other cell types.

Since there are remarkable analogies between the phenotypes of yeast RAD9 mutants and p53 mutant human cell lines[20,21], there is also the possibility that RAD9 binds a yeast protein structurally related to p53 even though there is currently no evidence of a yeast p53 protein.

The putative p53-binding domain, a RING finger domain and two putative nuclear localization signals in BRCA1 are all compatible with the reports on its nuclear localization[22–24]. Recently, however, it has been reported that BRCA1 is secreted from breast epithelial cells and contains an amino acid sequence pattern typical of granins, secreted proteins that are precursors to several active peptides[25]. The granin motif, however, has no statistical support. Given that the PROSITE database[26] contains 1179 motifs, one would expect to find by chance, within the 1863 residue BRCA1 sequence, approximately two motifs with an information content equivalent to that of the granin motif. Assuming a Poisson distribution, this gives a probability of approximately 0.87 that at least one such motif would be found. In contrast, an analogous calculation performed with the BRCT signature (Fig. 1*a*) yields a P value less than 3×10^{-4}.

In addition to granins themselves, the granin motif has been detected in 14 proteins in the complete non-redundant protein sequence database, including human BRCA1. None of these proteins produced an extended alignment with granins, and they did not appear to have any common structural or functional features. Furthermore, the murine BRCA1 protein contains a substitution in one of the motif's invariant positions, and if this substitution is allowed, the motif is found in a heterogeneous set of 28 sequences unrelated to granins.

The finding of a conserved domain in BRCA1, 53BP1 and RAD9 indicates that BRCA1 is likely to function in the cell nucleus and may be involved in one or more cell cycle checkpoints. The nature of the discrepancy between the experimental results on BRCA1 localization remains to be elucidated. The predictions made here can be tested in a variety of experiments. A direct demonstration of the predicted BRCA1 interaction with p53 or a similar protein and elucidation of the effect of point mutations in the BRCT domain on BRCA1 activity are only the most obvious of these.

Eugene V. Koonin
Stephen F. Altschul
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA

Peer Bork
European Molecular Biology Laboratory, Meyerhofstrasse 1, 690012 Heidelberg, Germany
Max Delbrück Center for Molecular Medicine, Berlin-Buch, Germany

Correspondence should be addressed to E.V.K.

1. Miki, Y. *et al. Science* **266**, 66–71 (1994).
2. Futreal, P. *et al. Science* **266**, 120–122 (1994).
3. Merajver, S. D. *et al. Nature Genet.* **9**, 439–443 (1995).
4. Hosking, L. *et al. Nature Genet.* **9**, 343–344 (1995)
5. Takahashi, H. *et al. Cancer Res.* **55**, 2998–3002 (1995).
6. Holt, J. T. *et al. Nature Genet.* **12**, 298–302 (1996).
7. Freemont, P. S. *Annal. NY Acad. Sci.* **684**, 174–192 (1993).
8. Altschul, S. F., Boguski, M. S., Gish, W. & Wootton, J. C. *Nature Genet.* **6**, 119–129 (1994).
9. Wootton, J. C. *Comput. Chem.* **18**, 269–285 (1994).
10. Altschul, S. F. *et al. J. Mol. Biol.* **215**, 403–410 (1990).
11. Henikoff, S. & Henikoff, J. *Proc. Natl. Acad. Sci. USA* **89**, 10915–10919 (1992).
12. Iwabuchi, K. *et al. Proc. Natl. Acad. Sci. USA* **91**, 6098–6102 (1994).
13. Holistein, M. *et al. Science* **253**, 49–53 (1991).
14. Tatusov, R. L., Altschul, S. F. & Koonin, E. V. *Proc. Natl. Acad. Sci. USA* **91**, 12091–12095 (1994).
15. Weinert, T. A. & Hartwell, L. H. *Science* **241**, 317–322 (1988).
16. Siede, W., Friedberg, A. S., Dianova, I. & Friedberg, E. C. *Genetics* **138**, 271–281 (1994).
17. Schuler, G. D., Altschul, S. F. & Lipman, D. J. *Prot. Struct. Funct. Genet.* **9**, 180–190 (1991).
18. Rost, B. & Sander, C. *Prot. Struct. Funct. Genet.* **9**, 55–72 (1994).
19. Hogervorst, F. B. *et al. Nature Genet.* **10**, 208–212 (1995).
20. Prives, C. *Curr. Opin. Cell Biol.* **5**, 214–218 (1993).
21. Cross, S. M. *et al. Science* **267**, 1353–1356 (1995).
22. Chen, Y. *et al. Science* **270**, 789–791 (1995).
23. Scully, R. *et al. Science* **272**, 123–125 (1996).
24. Chen, Y. *et al. Science* **272**, 125–126 (1996).
25. Jensen, R. A. *et al. Nature Genet.* **12**, 303–308 (1996).
26. Bairoch, A., Bucher, P. & Hofmann, K. *Nucl. Acids Res.* **24**, 189–196 (1996).
27. Thompson, J. D., Higgins, D. G. & Gibson, T. J. *Nucl. Acids Res.* **22**, 4673–4680 (1994).

...and secreted tumour suppressors

Sir — Recently, Jensen *et al.* suggested that BRCA1 and possibly BRCA2, two proteins encoded by the currently known inherited breast cancer genes, belong to the granin protein family[1]. They show that BRCA1 is secreted and others have reported that BRCA1 is upregulated during pregnancy[2,3], its expression is induced by estrogens[2–4] and it is present in breast milk. Jensen *et al.* speculated that BRCA1, like other granins[5] may undergo proteolytic cleavage leading to release of biologically active peptides. They further hypothesized that the protective role of pregnancy and lactation in breast cancer may be mediated by BRCA1 and possibly BRCA2. BRCA1 is a growth inhibitor in breast and ovarian cell lines[6].

Prostate specific antigen (PSA) is a serine protease of the kallikrein gene family that has recently been found in the female breast[7–9]. PSA is secreted by breast epithelial cells[10], its production is increased during pregnancy and it is also present in amniotic fluid[11]. PSA is found in milk of lactating women[12], and its presence in breast discharge fluid is associated with low risk for breast cancer. In women with a family history of breast cancer or with sporadic breast cancer, PSA levels in breast discharge fluid are dramatically reduced (Sauter, E.R. *et al.*, submitted manuscripts). Women with PSA in their tumours live longer and relapse less frequently[13]. Like BRCA1, PSA is regulated by steroid hormones[14]. Together, these data suggest that PSA and BRCA1 are two secreted protective factors in breast cancer. Their possible interactions and relationship have not been examined. Since PSA is a proteolytic enzyme and BRCA1 may be a precursor of bioactive peptides, it is tempting to speculate that they might interact as an enzyme-substrate pair for the release of peptides protective from breast cancer. The widespread availability of purified BRCA1 protein should allow such hypotheses to be tested.

E.P. Diamandis
Department of Pathology & Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, ON, M5G 1X5, Canada

1. Jensen, R.A. *et al. Nature Genet.* **12**, 303–308 (1996).
2. Marquis, S.T. *et al. Nature Genet.* **11**, 17–26 (1996).
3. Lane, T.F. *et al. Genes Dev.* **9**, 2712–2722 (1995).
4. Gudas, J., Nguyen, H., Li, T. & Cowan, K.H. *Cancer Res.* **55**, 4561–4565 (1995).
5. Huttner, W.B., Gerdes, H.-H. & Rosa, P. *Trends Biochem. Sci.* **16**, 27–30 (1991).
6. Hold, J.T. *et al. Nature Genet.* **12**, 298–302 (1996).
7. Diamandis, E.P. & Yu, H. *J. Clin. Endocrinol. Metab.* **80**, 1515–1517 (1995).
8. Yu, H., Diamandis, E.P. & Sutherland, D.J.A. *Clin. Biochem.* **27**, 75–79 (1994).
9. Diamandis, E.P., Yu, H. & Sutherland, D.J.A. *Breast Cancer Res. Treat.* **32**, 301–310 (1994).
10 Yu, H., Diamandis, E.P., Monne, M. & Croce, C.M. *J. Biol. Chem.* **270**, 6615–6618 (1995).
11. Yu, H. & Diamandis E.P. *Clin. Chem.* **41**, 204–210 (1995).
12. Yu, H. & Diamandis, E.P. *Clin. Chem.* **41**, 54–58 (1995).
13. Yu, H. *et al. Cancer Res.* **55**, 2104–2110 (1995).
14. Yu, H., Diamandis, E.P., Zarghami, N. & Grass, L. *Breast Cancer Res. Treat.* **32**, 291–300 (1994).

Sir — The recent article by Jensen *et al.*[1] describes some data that suggests that both BRCA1 and BRCA2 function as granins. This is based on several observations, sequence homologies to the granin consensus, cellular localization and biochemical properties. We should like to point out that in reaching this conclusion about BRCA1 and BRCA2, the authors have ignored two very important observations concerning genetics (both cellular and organismal) and evolutionary implications.

At the level of the organism, BRCA1 mutations, like most other mutations in tumour suppressor genes, cause a dominant cancer predisposition. At the mechanistic (cellular) level these mutations are recessive. Loss of the remaining wild-type allele (loss of heterozygosity) has been documented in many tumours arising in patients with *BRCA1* lesions[2]. This genetic event documents two important facts. First, the cell must be null for the normal gene product and second the mutation is cell autonomous. The fact that *BRCA1* and *BRCA2* lesions cause disease in the heterozygotes argues very strongly against these gene products being secreted, since loss of heterozygosity of a secreted gene product is highly