Structural basis of phospho-peptide recognition by the BRCT domain of BRCA1

R. Scott Williams, Megan S. Lee, D. Duong Hau, and J. N. Mark Glover*

Department of Biochemistry, University of Alberta, Edmonton, AB, T6G 2H7, Canada

*To whom correspondence should be addressed

Abstract

The BRCT repeats in BRCA1 are essential for its tumour suppressor activity and function to interact with phosphorylated protein targets containing the sequence, pSer-X-X-Phe. The structure of the tandem BRCA1 BRCT repeats bound to an optimized phospho-peptide reveals the N-terminal repeat harbors a conserved BRCT pSer binding pocket, while the interface between the repeats forms a hydrophobic groove which recognizes the Phe. Crystallographic and biochemical data suggest that the structural integrity of both binding sites is essential for peptide recognition. A diminished peptide binding capacity observed for cancer associated BRCA1-BRCT variants may explain the enhanced cancer risks associated with these mutations.

A large proportion of BRCA1-linked breast and ovarian cancers can be traced to truncation or missense mutations within the C-terminal BRCT domain of BRCA1 (reviewed in ¹). The integrity of this region of the protein is essential for the normal functioning of BRCA1 in the repair of DNA double strand breaks (DSBs) and homologous recombination².

The BRCA1 BRCT domain interacts with a diverse complement of proteins, suggesting a multifaceted role for this protein in the DNA damage response (reviewed in ³). Several of its binding partners have been implicated in transcriptional regulation, such as the transcriptional co-repressor CtIP, histone deacetylases, and the RNA polymerase holoenzyme. This is consistent with the idea that BRCA1 may impact the DNA damage response through regulation of DNA damage responsive genes such as p21 and GADD45. A more direct role in DNA repair is indicated by the interaction of the BRCA1 BRCT domain with the DNA helicase BACH1, an interaction that is required for efficient repair of DSBs⁴.

The BRCT domain of BRCA1 consists of a pair of conserved BRCT repeats approximately 90-100 amino acids in length. BRCT repeats are found not only in BRCA1, but in a large number of other proteins that respond to DNA damage⁵⁻⁷. A number of structural studies have revealed a conserved structure for the repeat, consisting of a 4-stranded parallel β sheet flanked by a pair of α helices on one face ($\alpha 1$ and $\alpha 3$), and a single α helix ($\alpha 2$) on the opposite face⁸⁻¹³. The two BRCT repeats of BRCA1 pack in a head-to-tail manner involving $\alpha 2$ of the N-terminal repeat and $\alpha 1$ and $\alpha 3$ of the C-terminal repeat, as well as inter-repeat linker^{10,11}. The interface between the two repeats is highly conserved in a number of other proteins that contain tandem BRCT repeats, and

an essentially identical packing arrangement has been observed in the dual BRCT repeats of the DNA damage-responsive p53 binding protein, 53BP1^{11,13}. Several BRCA1 missense mutations that have been linked to an enhanced risk for breast and ovarian cancer are localized at the interface between the two BRCT repeats and indicate that the correct packing of the two repeats is essential to BRCA1 function and tumour suppression^{10,14,15}.

It has recently been shown that the tandem BRCT repeats of BRCA1, but not the isolated, individual repeats, function as phospho-peptide binding modules¹⁶⁻¹⁸. The BRCT repeats of several other proteins involved in the cellular response to DNA damage have also been suggested to bind phosphorylated peptide targets, indicating that this may be a common property of the BRCT protein family. The BRCA1 BRCT binds with high affinity to serine-phosphorylated peptides that contain phenylalanine three residues C-terminal to the phospho-serine^{17,18}. A pSer-X-X-Phe motif within BACH1 has been shown to be essential for the interaction of BRCA1 and BACH1 in human cells and is required for the activation of the G2/M cell cycle checkpoint in response to DNA damage¹⁷. This may represent a primary means by which the BRCA1 BRCT domains, as well as many other BRCT proteins, interact with binding partners in a way that is responsive to DNA damage-induced cellular signaling.

Here we report the structure of a complex of the human BRCA1 BRCT domain bound to a high affinity peptide target containing the recognition motif, pSer-X-X-Phe. The structure reveals the nature of binding pockets for both the target phospho-serine and phenylalanine residues. The conservation of residues that line these pockets indicates that this function is conserved not only in BRCA1 homologs, but also in a number of

other BRCT proteins involved in DNA repair. The phospho-peptide binding properties of a large set of clinically-derived BRCA1 BRCT variants, and the structures of missense variants that are defective in peptide binding, highlight the importance of the integrity of the peptide-binding surface to BRCA1 function and tumor suppression.

RESULTS

Structure Determination. To define the structural basis for the recognition of serine-phosphorylated peptides by the BRCA1 BRCT, we crystallized and determined the structure of the human BRCA1 BRCT domain (encompassing residues 1649-1859) bound to an optimized pSer-X-X-Phe containing peptide selected from an oriented peptide library¹⁸ (**Figs. 1a, 1b, Table 1**). The peptide, Gly-Ala-Ala-Tyr-Asp-Ile-pSer-Gln-Val-Phe-Pro-Phe-Ala-Lys-Lys-Lys (termed BRCTide), binds the BRCT domain tightly with a dissociation constant of 400 nM¹⁸. The structure was solved by molecular replacement with the structure of the unbound human BRCA1 BRCT domain. The crystallographic asymmetric unit contains five BRCT domain-peptide complexes. Non-crystallographic symmetry enhanced the quality of the electron density maps and facilitated crystallographic refinement (see Methods).

Overall Structure. The five copies of the BRCT-peptide complex are arranged as 2 similar dimers, with a third, equivalent dimer generated by the rotation of the fifth BRCT-peptide complex about a crystallographic 2-fold axis (**Fig. 1c**). Each dimer buries a large solvent accessible surface (approximately 2000 Å² buried in total in the two BRCT-peptide complexes in each dimer). Molecular contacts between the BRCT repeats

in the dimer are largely mediated through the bound peptides, which are sandwiched between the two domains. At the center of the dimer, a cage of hydrogen bonds is formed between the Gln(+1) residues from the peptides and Asn1774 of the BRCT domains. However, fixed-angle light scattering experiments and gel filtration chromatography indicates that the dimerization in solution is very weak, with a dissociation constant significantly above 10 μ M (data not shown).

The refined structures of the five peptide-BRCT complexes show that each peptide is positioned within a groove that extends across both BRCT repeats. The N-terminal portion of each peptide (residues Gly(−6)–pSer) lies over the N-terminal BRCT repeat, while the C-terminal portion (residues Asn(+1)–Lys(+7)) lies over the C-terminal BRCT repeat (**Figs. 1a** and **2a**). Tight interactions between the peptide and the BRCT domain involve residues from the pSer to the Phe at the +3 position; residues further N- and C-terminal are in less intimate contact with the BRCT domain and display conformational variability in the 5 independent complexes (**Fig. 2b** and **supplementary Fig. 1**). For instance, pairwise comparisons of the relative positions of peptide main chain atoms between Gly(−6) and pSer show displacements of up to 3.5Å between Ca atoms. These distances are significant at the resolution of this of up to 3.5Å distances

RMS deviations from

For example, significant deviations, greater than six times the overall mean expected coordinate error for the structure as estimated from likelihood, are observed from , which vary as much 3.x .

More dramatic differences in the peptide conformation are observed for residues C-terminal to Phe(+3). This region of the peptide exhibits varied degrees of ordering in

the complexes (**supplementary Fig. s1**), and in one BRCT-peptide complex, the direction of the peptide chain sharply bends back upon itself. These observations are consistent with the peptide selection studies which have shown that positions other than +3 within this peptide scaffold are not strongly selected¹⁸. The pSer is recognized by a rather shallow pocket composed of residues from the N-terminal BRCT repeat, while the Phe residue at +3 is recognized by a deeper and largely hydrophobic pocket at the interface between the N- and C-terminal repeats (**Fig. 2c**). This explains why both repeats are needed for efficient peptide binding¹⁸. In total, approximately 1100 Å² are buried in the peptide-protein interface.

pSer Recognition. A comparison of the structures of the BRCT domain both free and bound to peptide reveals that the orientation of the hydrogen bond donors that constitute the phospho-serine binding pocket are pre-aligned for recognition of the phosphate (not shown). The main chain NH of Gly1656, and the hydroxyl group of Ser1655, both donate H-bonds to the phosphate (**Fig. 2a**). The Ser hydroxyl is supported by further hydrogen bonding interactions with Thr1700. The phosphate is also recognized by a salt-bridging interaction with Lys1702, which in turn is buttressed by hydrogen bonds to the main chain carbonyls of Val1654 and Asn1678. The orientation of the phosphate is further supported by a hydrogen bond between the phosphate O γ and the main NH of the pSer. The coordination of the phosphate oxygens by main chain NH, Ser OH, and positively charged side chains is reminiscent of the way in which other protein modules, such as the FHA and 14-3-3 families, recognize phosphorylated residues¹⁹.

The importance of the phosphate-contacting residues to overall peptide binding was confirmed by assessing the ability of BRCA1 BRCT variants containing mutations in the phosphate binding pocket to bind a biotinylated peptide library containing the Ser-X-X-Phe motif, immobilized on streptavidin beads, in a phosphorylation-dependent manner (**Fig. 3a**). Wild-type BRCA1 BRCT bound tightly to the phospho-peptide library such that essentially all the in vitro transcribed/translated protein was bound to the beads. This interaction was absolutely dependent on phosphorylation of the serine. In contrast, mutation of Ser1655 to alanine completely abolished the interaction, indicating that the hydroxyl group of Ser1655 is essential for peptide binding in solution. The patient-derived missense mutation, S1655F, also shows no detectable binding to the phospho-peptide (**Fig. 3c**), indicating that this previously unclassified variant may be associated with an increased risk for cancer. Mutation of Lys1702, either to alanine or the iso-steric methionine, also obliterated binding, verifying the importance of the side chain amino group for phospho-peptide recognition.

Structure of the +3 binding pocket. The hydrophobic +3 binding pocket occupies the interface between the N- and C-terminal BRCT repeats (**Figs. 1a, 1b and 2a**). The floor of the pocket is composed of Leu1701, Phe1704, Met1775, and Leu1839. The sides of the pocket are formed by Arg1699, Asn1774 and Arg1835. Arg1699 also plays an important role in positioning the main chain of the Phe(+3). The main chain carbonyl of Arg1699 hydrogen bonds to the main chain NH of the +3 residue, and the guanidinium group of side chain makes contacts the main chain carbonyl of the +3 residue.

BRCT mutations affect peptide binding. Truncation and missense mutations within the BRCA1 BRCT domain that have been isolated from patients in breast and ovarian cancer screening programs are recorded in a large database (available through the Breast Cancer Information Database <http://research.nhgri.nih.gov/bic/>, and reviewed in¹⁵). All of the truncation mutations and a large proportion of the missense mutations destabilize the protein fold^{15,20}. We probed the ability of a large panel of these BRCA1 BRCT variants to bind biotinylated pSer-X-X-Phe peptides as described above (**Figs. 3b and 3c**). The missense mutations which cause little or no folding defects displayed a range of phospho-peptide binding activities. None of the highly destabilizing truncation or missense mutations specifically bound the phosphorylated peptides, demonstrating that correct folding of the BRCT domain is essential for specific recognition of the phospho-peptide target.

Several mutants (F1695L, T1720A, R1751Q, M1783T, and M1652I) showed peptide binding properties that were essentially indistinguishable from wild type. M1652I has been shown to be a polymorphism not associated with an increased cancer risk²¹ while there is insufficient pedigree data for the other mutations to directly assess their cancer risks. One of the moderately destabilizing mutants, D1692Y, displays a slight but reproducible defect in peptide binding, which may account for its association with disease. Most interestingly, several of these mutations (R1699Q, R1699W, V1696L, M1775R, and V1809F) displayed a complete loss of phospho-peptide binding activity. R1699W would impair the ability of the BRCT domain to hydrogen bond with the backbone of the peptide at the +3 position and this mutation has been suggested to be associated with breast/ovarian cancer²². Somewhat surprisingly, R1699Q shows an

equally profound peptide binding defect, even though the Gln side chain might be expected to fulfill a similar hydrogen bonding role to the wild type Arg. M1775R, V1696L and V1809F are also unable to specifically bind the phospho-peptide.

Structures of BRCT missense variants

Crystallization screens. We set out to crystallize BRCT missense variants to gain structural insights into the nature of the mutation-induced peptide binding defects. Of the five intermediately destabilizing BRCT variants which were defective in peptide binding (D1692Y, R1699W, V1696L, M1775R and V1809F), four (R1699W, V1696L, M1775R and V1809F) could be produced in soluble form in *E. coli*, three (R1699W, M1775R and V1809F) yielded crystals, and two (M1775R and V1809F) produced usable diffraction data beyond 3.0 Å with synchrotron radiation. The crystal structures of the M1775R and V1809F variants provide an explanation for their inability to bind pSer-X-X-Phe targets.

V1809F. To understand the way in which this mutation blocks peptide binding, we determined the crystal structure of this variant to 2.8 Å resolution (**Fig. 4** and **Table 1**). The structure of this mutant reveals that a set of rearrangements in the protein hydrophobic core lead to the disruption of the +3 binding pocket (**Figs. 4a, 5b**). In this mutant, the larger substituted Phe side chain contacts Leu1780, causing an adjustment in this side chain that in turn brings it into contact with Met1775. This contact pushes the Met1775 side chain out into the +3 pocket, explaining the loss of peptide binding in this mutant. Interestingly, in the unbound structures of the V1809F variant, there is a strong electron density peak that overlays with the position of the phosphate from the pSer in the

peptide-bound structure (**Fig. 4**). This feature is consistent with a bound sulfate anion, present in the crystallization solution, mimicking phosphate binding, and suggests that the mutation does not disrupt phosphate anion coordination, but only the +3 binding pocket.

M1775R. The M1775R mutation blocks the ability BRCA1 to interact with various protein partners including BACH1⁴, histone deacetylases²³, and CtIP²⁴, is associated with defects in DNA repair², and is linked to an enhanced cancer risk²⁵. The M1775R mutation has previously been shown to cause a significant destabilization of the BRCA1 fold^{14,20}. The structure of M1775R, described previously¹⁴ (RCSB: 1N5O), revealed a subtle alteration in the protein structure at the site of the substitution. Superimposition of the structure of the BRCT-peptide complex on that of the M1775R variant determined in the absence of bound peptide, reveals that the guanidinium group of the substituted arginine occludes the +3 pocket, thereby blocking access by the Phe(+3) (**Fig. 5c**). Thus, the combined effects of protein destabilization and direct obstruction of the peptide interaction surface by the M1775R substitution likely contribute to BRCA1 tumour suppressor inactivation and ultimately, to disease predisposition in individuals carrying this mutation.

DISCUSSION

A Conserved pSer Binding Motif. The structures of BRCT containing proteins have revealed that the conserved set of BRCT residues, which contribute to the hydrophobic core and overall fold of these domains, are found within the central β -sheet, in helix α 3,

and at the C-terminus of helix $\alpha 1$ ^{8,10-12}. In general, there are few invariant residues amongst the BRCTs, and the sequences diverge greatly in regions that form the BRCT $\alpha 2$, and in two hypervariable insertion regions that are found at the $\beta 1$ - $\alpha 1$ and $\beta 2$ - $\beta 3$ junctions^{6,7,9}. These variable regions were originally hypothesized to form adaptable protein-protein interaction surfaces⁸. Indeed, both the $\beta 1$ - $\alpha 1$ loop and helix $\alpha 2$ participate in phospho-serine recognition in BRCA1.

Several BRCT proteins have now been tested for phospho-peptide binding, although the results found to date are controversial¹⁶⁻¹⁸. The structure of the BRCT domain – peptide complex reveals BRCT residues that play a key role in these interactions and provides a means of further identifying BRCT domains that may function as phospho-peptide binding modules. Sequence analysis suggests residues comprising a conserved pSer binding pocket include a (S/T)G motif found in the variable $\beta 1$ - $\alpha 1$ connecting loop (corresponding to Ser1655 and Gly1656 of hBRCA1) and a (T/S)XK motif (where X indicates any residue) at the N-terminus of $\alpha 2$ of the same repeat (residues 1700–1702 of hBRCA1)(**Figs. 1b** and **supplementary Fig. s2**). Strikingly, the motif, which has not been included in previous descriptions of BRCT domain homology, is present in 57 of 518 unique BRCT sequences in the current Pfam database (www.sanger.ac.uk/cgi-bin/Pfam) , and in the N-terminal BRCT of dual repeat containing BRCT proteins reported to bind phosphopeptides, including all *bona fide* BRCA1 homologues, hMDC1¹⁷, hPTIP¹⁸ and hBARD1^{16,17} (**Fig. 1b**). Moreover, single BRCT-containing proteins with apparent phospho-serine binding activity, such as FCP1¹⁷ also contain this motif. Thus, we suggest that pSer binding is a common function of

BRCT domains, which plays a critical role in regulating the assembly of protein complexes during the DNA damage response.

The dual repeat, BRCT domains of both BRCA1 and the transcriptional activator, PTIP, have been shown to bind not only phospho-serine, but also phospho-threonine containing peptides, albeit with reduced affinity¹⁸. The γ -methyl group of a pThr residue would point away from the BRCA1 BRCT domain, however, it would sterically clash with a phosphate oxygen which is in a similar position in each of the five complexes in the asymmetric unit. Thus, replacement of pSer for pThr in the phospho-peptide would require a readjustment of the conformation of the peptide which may result in a reduced binding affinity. Phospho-tyrosine peptides do not appear to bind to these proteins¹⁸, and indeed, it would be difficult to accommodate this large side chain within the phosphate binding pocket, without large changes in the conformation of either the phospho-peptide backbone or phosphate-binding pocket structure.

+3 Binding Specificity. The crystal structures of BRCT protein complexes have revealed a role for dual repeat domains as a scaffold for variable BRCT-BRCT linker structures which participate directly in ligand binding^{11,13,26}. For example, the dual repeat BRCT domain of 53BP1 anchors a β -hairpin BRCT-linker structure which contacts the DNA binding domain of p53^{11,13}. In an analogous manner, specific recognition of a bound phospho-peptide by the BRCA1 BRCT domain is dependent on both the dual-repeat scaffold structure and a BRCT variable loop ($\beta 1'$ - $\alpha 1'$) which together create a hydrophobic groove which binds the Phe(+3).

As expected, the interface between the two BRCT repeats and the +3 binding pocket are highly conserved in BRCA1 homologs^{10,13}. Residues predicted to form the repeat interface are also conserved in other dual BRCT proteins known to bind pSer-containing peptides (**Fig. 1b**). However, the amino acids lining the walls of this pocket, formed by the variable $\beta 1'$ - $\alpha 1'$ loop and the N-terminus of $\alpha 3'$ from the C-terminal BRCT of a tandem repeat, appear less well conserved. Such sequence variability suggests that different BRCT domains may have diverse binding specificities.

In an independent study presented in this issue of *Nature Structural and Molecular Biology*, the structure of the BRCA1 BRCT domain bound to the specific BACH1 phospho-peptide is reported (Clapperton et al. *ibid*). Descriptions of the key determinants of phospho-peptide recognition, including the pSer and Phe(+3) binding pockets, are in good agreement with the observations reported here. The mode of phospho-serine recognition observed for BRCA1 is likely to be conserved in other BRCT containing proteins which participate in the coordination of the DNA damage response. Molecular snapshots of cancer associated mutant BRCT proteins provide a structural explanation for the catastrophic inactivation of the BRCA1 tumour suppressor. The observation that the majority of mutations which map to the BRCT of BRCA1 impair peptide binding helps to explain the elevated cancer risks associated with these variants.

METHODS

Cloning, Protein Expression, and Purification. Mutations S1655A, S1655F, K1702A, and K1702M were introduced into the human BRCA1 coding sequence for residues 1646-1859 using PCR methods and cloned into the EcoRI and BamHI sites of pLM1. Vector construction of all other pLM1-based T7 expression vectors for BRCA1-BRCT variants was as described previously¹⁵. For structural studies, hBRCA1(1646-1859)-WT or hBRCA1(1646-1859) variants were over-expressed in *E. coli* BL21(DE3) at 25 °C (WT) or 20 °C (missense variants) for 12-15 hours and purified as described previously¹⁰.

Crystallization and Data Collection

BRCT-peptide complex Protein stocks were dialyzed into protein buffer (400 mM NaCl, 10mM TrisHCl pH 7.5, 1mM DTT) and concentrated to 20 mg/mL for use in crystallization trials. For peptide complex formation, hBRCA1(1646-1859)-WT at 20 mg/mL, was mixed with the optimized BRCT-binding phospho-peptide¹⁸ (BRCTtide: Ac-GAAYDIpSQVFPEAKKK-NH₂) at a 1.5:1 phospho-peptide:protein molar ratio. Crystals were grown at 20-22 °C using the hanging drop vapor diffusion method by mixing 1 µL of complex in protein buffer with 1 µL of well solution 1 (0.2 M ammonium acetate, 0.1 M tri-sodium citrate dehydrate pH 5.6, 30% (w/v) PEG 4000). Small crystals, which were observed after 3-4 weeks, grew to a maximal size of ~30x30x30 µm within 5-6 weeks.

BRCT V1809F. Hexagonal crystals of BRCT missense variant V1809F, with unit cell dimensions similar to the wild type protein crystals, crystallized at 4 °C within 2-4 days by mixing 2 µL of 20 mg/mL hBRCA1(1646-1859)-V1809F in protein buffer with well

solution 2 (1.2-1.4 M ammonium sulfate, 50 mM MES pH 6.7-6.8, 10 mM CoCl₂). For cryo-preservation, single crystals were soaked in the appropriate well solution supplemented with 26% (v/v) glycerol and then flash frozen in liquid nitrogen. All data were collected at beamline 8.3.1 of the Advanced Light Source, Lawrence Berkeley National Laboratory. Data reduction and scaling was performed using the HKL package²⁷ (**Table 1**).

Structure determination

BRCT-peptide complex. Molecular replacement trials using MOLREP (v7.3)²⁸ were performed with coordinates of the unliganded human BRCT dual repeat structure¹⁰ (RCSB: 1JNX). An initial 4 molecule solution was obtained from a rotation and translation search using data in the range 10.0-4.0 Å. Following rigid body refinement of this solution, inspection of sigma-A weighted 2Fo-Fc and Fo-Fc electron density maps revealed clear density for a fifth BRCT domain which was fit manually into the electron density using O(v7.0)²⁹. After a second round of rigid body fitting, electron density in 2Fo-Fc and Fo-Fc maps for the corresponding five phospho-peptide chains was apparent, but discontinuous in some regions. Subsequent 5-fold NCS averaging and density modification using RESOLVE³⁰ improved the overall quality of the 2Fo-Fc model phased experimental electron density map and allowed for the unambiguous trace of the phospho-peptide chains. Further model building and refinement was carried out using O and REFMAC³¹ (v5.0.32). 5-fold NCS restraints applied to the BRCA1 protein chains allowed for positional refinement of coordinate parameters using data to 3.3 Å. Differences between the BRCA1 protein chains were limited to regions which were

ordered or disordered depending on crystallographic packing interactions. To allow for conformational variability of the peptides observed between the different complexes, NCS restraints were not used for the phospho-peptide chains. Refinement of thermal parameters was limited to overall B-factor refinement with a separate group TLS anisotropic thermal parameter model applied to each of the five protein-peptide complexes. The final model contains 1101 amino acid residues, with good geometry. Refinement statistics are summarized in **Table 1**.

BRCT-V1809F. Structure solution and refinement of BRCA1 variant V1809F was similar to that described for variant M1775R¹⁴. Residue 1809 was mutated to alanine prior to rigid body fitting of the model to the V1809F hexagonal dataset. Sigma-A weighted Fo-Fc difference density contained two large 4σ peaks corresponding to Phe1809 and a sulfate bound in the phospho-serine binding pocket near Lys1702 and Ser1655. Further iterative model building and refinement was done using O and REFMAC and the model was assessed with PROCHECK³². Molecular surfaces were drawn with GRASP³³. Structural figures were created using BOBSCRIPT³⁴ and rendered with POVray (www.povray.org).

Peptide binding assays. 0.1 to 0.4 μ g of pLM1 plasmid encoding mutant BRCT domains was used directly as template for coupled transcription and translation reactions in a reticulocyte lysate. For binding assays, the BRCT variants were translated and labeled with ³⁵S-methionine for 2 h at 30 °C using the TNT-Quick system (Promega). Bead immobilized peptide affinity resin was prepared in binding buffer (150 mM NaCl, 50 mM TrisHCl pH 7.5, 0.1% (v/v) NP40, 1 mM EDTA, 1 mM DTT) by incubating a

ten-fold molar excess of a partially selected biotinylated phospho-peptide library (pSXXF: biotin-ZGZGGAXXXpSXXFXXAYKKK, Z: aminohexanoic acid, X: all amino acids excluding Cys, pS: phospho-serine) and the corresponding dephosphorylated peptide (SXXF: biotin-ZGZGGAXXXSXXFXXAYKKK) with streptavidin agarose beads (25 pmol/ μ L stock, Sigma-Aldrich) for 30 minutes at 4 °C. Excess peptide was removed by washing five times with ten bed volumes of binding buffer. 1 μ L of 35 S-methionine labeled BRCT-variant containing lysate was added to 20 μ L of affinity resin in a total of 150 μ L binding buffer supplemented with 0.3 mg/mL BSA which was used to reduce background resin binding observed for highly destabilizing BRCT variants. Following incubation for two hours at 4 °C, the resin was washed three times with 200 μ L binding buffer. Bound BRCT variants were eluted from the resin with the addition of 20 μ L SDS-page loading buffer, run on a 15% SDS-PAGE gel, and visualized using a Molecular Dynamics phosphorimager.

Coordinates. Coordinates have been deposited in the RCSB protein data bank (RCSB ID: XXXX.PDB).

Acknowledgements

We are grateful to Jo Parrish, Ernst Bergmann, Trevor Moraes, James Holton, and the BCSB staff for discussions and technical support during data collection. Data was collected on ALS beamline 8.3.1 with funding from the Alberta Synchrotron Institute as part of a PRT. We thank Michael Yaffe and Steve Smerdon for sharing unpublished data and the generous gift of biotinylated peptide libraries, and Robert Boyko for help with BRCT sequence alignments. This work was supported by a grant from the Canadian Breast Cancer Research Alliance. J.N.M.G. is a Canada Research Chair in Structural Molecular Biology.

Figure legends

Figure 1. Overview of the BRCA1 BRCT – peptide complex. (a) The BRCT domain is colored with β -strands green, and α -helices in yellow. The peptide is in blue, with the critical pSer and Phe residues drawn. The BRCT residues that recognize the pSer and Phe(+3) residues are displayed with transparent surfaces. (b) The amino acid sequence of the human BRCA1 BRCT domain is shown, with the secondary structure. The residues that contact the pSer are shaded blue, and the residues that form the Phe(+3) binding pocket are marked by red circles. Missense mutations which have been assayed for interactions with the peptide are displayed above the sequence. Residues involved in inter-repeat BRCT interactions are boxed. (c) Structural arrangement of one of the three similar dimers in the crystallographic asymmetric unit.

Figure 2. Details of BRCA1 BRCT – peptide interactions. (a) Stereoview showing key residues that recognize the pSer and Phe(+3) residues of the peptide. Hydrogen bonds/salt bridges are indicated by dotted lines. (b) Conformational variability of phospho-peptide binding. Overlay of the 5 independent peptide structures within the crystallographic asymmetric unit. The BRCT C α atoms of each of the BRCT-peptide complexes were superimposed to generate the overlay. Note that the structures of the 5 complexes are essentially identical between pSer and the Phe(+3), however the structures of the peptides deviate significantly outside this region. (c) Electrostatic surface representation of the BRCT domain displaying the pSer- and Phe-binding pockets, with the peptide drawn in grey.

Figure 3. Ability of mutant forms of the BRCA1 BRCT domain to bind pSer-containing peptides. (a) Effect of mutations in residues which directly contact pSer. Each of the mutants was produced by in vitro transcription/translation and was assayed for binding to biotinylated peptides containing Ser-X-X-Phe (SXXF) or pSer-X-X-Phe (pSXXF) motifs (see Methods). The left-most lane of each set of three reactions shows 100% of the load material. The stability of the protein fold for most of the mutants were tested using a proteolytic assay¹⁵, where ++ indicates a severe folding defect, + indicates a more modest defect, and – indicates that the stability of the mutant is indistinguishable from wild type. (b) Effect of truncation mutations that are associated with hereditary breast and ovarian cancer. (c) Effect of missense mutations derived from breast cancer screening programs.

Figure 4. Structure determination of variant V1809F. (a) Sigma-A weighted Fo-Fc difference electron density at 4.0 σ is displayed with a structural overlay of wild type (yellow) and V1809F (grey) structures. Electron density peaks for Phe1809 (grey) and a bound sulfate (gold ball and sticks) are observed. S δ of Met1775 in the V1809F structure moves ~1.8 Å into the Phe(+3) binding pocket. (b) Sigma-A weighted 2Fo-Fc density for the final model at 2.8 Å is contoured at 1.0 σ .

Figure 5. Occlusion of the Phe(+3) binding pocket by BRCA1 missense mutations V1809F and M1775R. (a) Wild type amino acid packing environment near the Phe(+3) binding pocket. Complementary van der Waals packing is observed between the protein (grey surfaces) and the Phe(+3) (blue surface). (b) The V1809F substitution (red surface) causes a relay of van der Waals packing rearrangements which permeate to the

protein surface. The +3 binding pocket is blocked by Met1775 atoms C γ , S δ , and C ϵ , which are found within 2-3 Å of the overlayed Phe(+3) aromatic ring. Black arrows and dotted lines indicate the BRCT surface incompatible with peptide binding. (c) Cancer causing mutation M1775R (red surface) fills the +3 binding pocket and blocks peptide binding.

Supplementary Figure S1. (a) RESOLVE density modified 2Fo-Fc electron density at 3.3 Å resolution, which was used to trace the peptide chains, is displayed for the five phospho-peptides and contoured at 1 σ . (b) Stereoview of a least squares structural overlay of peptide bound structures (grey BRCTs with blue peptides) and an unbound BRCT domain (RCSB: 1JNX) (green) is displayed.

Supplementary Figure S2. Multiple sequence alignment of BRCT domains which contain a conserved (S/T)G...(T/S)XK phospho-serine binding motif. 57 unique sequences from the BRCT PFAM database are shown with the motif highlighted in blue. Conserved BRCT fold residues are highlighted in grey.

1. Nathanson, K.L., Wooster, R., Weber, B.L., & Nathanson, K.N. Breast cancer genetics: what we know and what we need. *Nat. Med.* **7**, 552-556 (2001).
2. Scully, R. *et al.* Genetic analysis of BRCA1 function in a defined tumor cell line. *Mol. Cell* **4**, 1093-1099 (1999).
3. Venkitaraman, A.R. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* **108**, 171-182 (2002).
4. Cantor, S.B., Bell, D.W., Ganesan, S., Kass, E.M., Drapkin, R., Grossman, S., Wahrer, D.C., Sgroi, D.C., Lane, W.S., Haber, D.A., & Livingston, D.M. Bach1, a novel helicase-like protein, interacts directly with brca1 and contributes to its DNA repair function. *Cell* **105**, 149-160 (2001).
5. Bork, P., Hofmann, K., Bucher, P., Neuwald, A.F., Altschul, S.F., & Koonin, E.V. A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J.* **11**, 68-76 (1997).
6. Koonin, E.V., Altschul, S.F., & Bork, P. BRCA1 protein products ... Functional motifs... *Nature Genet.* **13**, 266-268 (1996).
7. Callebaut, I. & Mornon, J.P. From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair. *FEBS Lett.* **400**, 25-30 (1997).
8. Zhang, X., Morera, S., Bates, P.A., Whitehead, P.C., Coffey, A.I., Hainbucher, K., Nash, R.A., Sternberg, M.J., Lindahl, T., & Freemont, P.S. Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. *EMBO J.* **17**, 6404-6511 (1998).
9. Huyton, T., Bates, P.A., Zhang, X., Sternberg, M.J., & Freemont, P.S. The BRCA1 C-terminal domain: structure and function. *Mutat. Res.* **460**, 319-332 (2000).
10. Williams, R.S., Green, R., & Glover, J.N. Crystal structure of the BRCT repeat region from the breast cancer- associated protein BRCA1. *Nat. Struct. Biol.* **8**, 838-842 (2001).
11. Joo, W.S. *et al.* Structure of the 53BP1 BRCT region bound to p53 and its comparison to the Brcal BRCT structure. *Genes. Dev.* **16**, 583-593 (2002).
12. Krishnan, V.V., Thornton, K.H., Thelen, M.P., & Cosman, M. Solution structure and backbone dynamics of the human DNA ligase IIIalpha BRCT domain. *Biochemistry* **40**, 13158-13166 (2001).
13. Derbyshire, D.J. *et al.* Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor. *EMBO J.* **21**, 3863-3872 (2002).
14. Williams, R.S. & Glover, J.N. Structural consequences of a cancer-causing BRCA1-BRCT missense mutation. *J. Biol. Chem.* **278**, 2630-2635 (2003).

15. Williams, R.S. *et al.* Detection of protein folding defects caused by BRCA1-BRCT truncation and missense mutations. *J. Biol. Chem.* **278**, 53007-53016 (2003).
16. Rodriguez, M., Yu, X., Chen, J., & Songyang, Z. Phosphopeptide binding specificities of BRCA1 COOH-terminal (BRCT) domains. *J Biol Chem* **278**, 52914-8 (2003).
17. Yu, X., Chini, C.C., He, M., Mer, G., & Chen, J. The BRCT domain is a phospho-protein binding domain. *Science* **302**, 639-642 (2003).
18. Manke, I.A., Lowery, D.M., Nguyen, A., & Yaffe, M.B. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science* **302**, 636-639 (2003).
19. Yaffe, M.B. & Smerdon, S.J. PhosphoSerine/threonine binding domains: you can't pSERious? *Structure (Camb)* **9**, R33-8 (2001).
20. Ekblad, C.M. *et al.* Characterisation of the BRCT domains of the breast cancer susceptibility gene product BRCA1. *J. Mol. Biol.* **320**, 431-442 (2002).
21. Deffenbaugh, A.M., Frank, T.S., Hoffman, M., Cannon-Albright, L., & Neuhausen, S.L. Characterization of common BRCA1 and BRCA2 variants. *Genet. Test.* **6**, 119-121 (2002).
22. Vallon-Christersson, J. *et al.* Functional analysis of BRCA1 C-terminal missense mutations identified in breast and ovarian cancer families. *Hum. Mol. Genet.* **10**, 353-360 (2001).
23. Yarden, R.I. & Brody, L.C. BRCA1 interacts with components of the histone deacetylase complex. *Proc. Natl. Acad. Sci. U S A* **96**, 4983-4988 (1999).
24. Yu, X., Wu, L.C., Bowcock, A.M., Aronheim, A., & Baer, R. The C-terminal (BRCT) domains of BRCA1 interact in vivo with CtIP, a protein implicated in the CtBP pathway of transcriptional repression. *J. Biol. Chem.* **273**, 25388-25392 (1998).
25. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W. and others. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66-71 (1994).
26. Sibanda, B.L. *et al.* Crystal structure of an Xrcc4-DNA ligase IV complex. *Nat. Struct. Biol.* **8**, 1015-1019 (2001)
27. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276**, 307-325 (1997).
28. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular

- replacement. *J. Appl. Cryst.* **30**, 1022-1025 (1997).
29. Jones, T.A., Zou, J.Y., Cowan, S.W., & Kjeldgaard. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta. Crystallogr. A* **47**, 110-119 (1991).
 30. Terwilliger T.C. SOLVE and RESOLVE: automated structure solution and density modification. *Meth. Enzymol.* **374**, 22-37 (2003).
 31. Winn, M.D., Isupov, M.N., & Murshudov, G.N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta. Crystallogr. D Biol. Crystallogr.* **57**, 122-133 (2001).
 32. Laskowski, R.A., MacArthur, M.W., & Thornton, J.M. Validation of protein models derived from experiment. *Curr. Opin. Struct. Biol.* **8**, 631-639 (1998).
 33. Merritt, E. A. and Bacon, D. J. Raster3D: Photorealistic Molecular Graphics. *Meth. Enzymol.* **277**, 505-524. (1997).
 34. Esnouf, R.M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 112-113, 132-134. (1997).

Table 1: Crystallographic Data and Refinement Statistics

		BRCT-peptide complex	BRCT-V1809F
Space group		C222 ₁	P6 ₁ 22
Unit cell	a (Å)	97.12	113.92
	b (Å)	138.36	113.92
	c (Å)	198.27	120.91
Resolution Range (Å)		50-3.3	20-2.8
No. of observed reflections		81261	116289
No. of unique reflections		20317	11828
Completeness (%)		99.4 (99.9) ^a	99.3 (96.3)
R _{sym} (%) ^b		12.5 (50.0)	5.1 (43.4)
Overall I / σ I		11.1 (2.7)	37.9 (3.0)
R _{cryst} / R _{free} (%) ^c		25.8 (30.2)	27.5 (29.4)
No. of Atoms/ H ₂ O / SO ₄ ²⁻ / Co ²⁺		8380/0/0/0	1647/51/10/1
Ramachandran plot (%):			
Most Favored / Allowed /Generous		90.0/9.7/0.3	85.7/14.3/0
RMSD Bonds (Å) / Angles (°)		0.008/1.252	0.019/1.636

^a Values in parentheses are statistics for the highest resolution shell, 3.42-3.30 Å for the BRCT-peptide complex, and 2.90 - 2.80 Å for BRCT-V1809F.

^b $R_{sym} = 100 \sum_{hkl} | \langle I \rangle - I | / \sum_n \langle I \rangle$

^c $R_{cryst} = \sum_n |F_o(h) - F_c(h)| / \sum_n |F_o(h)|$; Where $F_o(h)$ and $F_c(h)$ are observed and calculated structure factors for the resolution range 50-3.3 Å (BRCT-peptide complex) or 20 – 2.8 Å (BRCT-V1809F). R_{free} calculated with 5% (BRCT-peptide complex) or 7 % (BRCT-V1809F) of all reflections excluded from refinement.

Fig. 1

Fig 2

Fig. 3

Fig. 4

A WT

B V1809F

C M1775R

Fig. S2

A

B

Brct
Domain
Number

■ Phospho-Serine Binding Pocket
■ BRCT hydrophobic repeat residues