

## Functional Evolution of BRCT Domains from Binding DNA to Protein

Zi-Zhang Sheng<sup>1,2</sup>, Yu-Qi Zhao<sup>1,2</sup> and Jing-Fei Huang<sup>1,3</sup>

<sup>1</sup>State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. <sup>2</sup>Graduate School of Chinese Academy of Sciences, Beijing, China. <sup>3</sup>Kunming Institute of Zoology, Chinese University of Hong Kong Joint Research Center for Bioresources and Human Disease Mechanisms, Kunming, Peoples' Republic of China. Corresponding author email: [huangjf@mail.kiz.ac.cn](mailto:huangjf@mail.kiz.ac.cn)

**Abstract:** The BRCT domain (BRCA1 C-terminal domain) is an important signaling and protein targeting motif in the DNA damage response system. The BRCT domain, which mainly occurs as a singleton (single BRCT) or tandem pair (double BRCT), contains a phosphate-binding pocket that can bind the phosphate from either the DNA end or a phosphopeptide. In this work, we performed a database search, phylogeny reconstruction, and phosphate-binding pocket comparison to analyze the functional evolution of the BRCT domain. We identified new BRCT-containing proteins in bacteria and eukaryotes, and found that the number of BRCT-containing proteins per genome is correlated with genome complexity. Phylogeny analyses revealed that there are two groups of single BRCT domains (sGroup I and sGroup II) and double BRCT domains (dGroup I and dGroup II). These four BRCT groups differ in their phosphate-binding pockets. In eukaryotes, the evolution of the BRCT domain can be divided into three phases. In the first phase, the sGroup I BRCT domain with the phosphate-binding pocket that can bind the phosphate of nicked DNA invaded eukaryotic genome. In the second phase, the phosphate-binding pocket changed from a DNA-binding type to a protein-binding type in sGroup II. The tandem duplication of sGroup II BRCT domain gave birth to double BRCT domain, from which two structurally and functionally distinct groups were evolved. The third phase is after the divergence between animals and plants. Both sGroup I and sGroup II BRCT domains originating in this phase lost the phosphate-binding pocket and many evolved protein-binding sites. Many dGroup I members were evolved in this stage but few dGroup II members were observed. The results further suggested that the BRCT domain expansion and functional change in eukaryote may be driven by the evolution of the DNA damage response system.

**Keywords:** BRCT domain, evolution, DNA damage response, superfamily

*Evolutionary Bioinformatics* 2011:7 87–97

doi: [10.4137/EBO.S7084](https://doi.org/10.4137/EBO.S7084)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



## Introduction

The DNA damage response is an essential system used by all cellular organisms to detect, signal, and repair DNA damage.<sup>1</sup> The DNA damage response proteins are highly diverse structurally, but many of them contain a conserved globular domain, first identified in BRCA1, the breast cancer tumor suppressor protein, and thus designated BRCT (BRCA1 C-terminus).<sup>2–4</sup> The BRCT domains can either transmit the signals generated by DNA damage detectors to the repair machinery or target diverse proteins (eg, DNA polymerase  $\lambda$ , DNA ligase III, and DNA ligase IV) into the repair complexes.<sup>5–8</sup> BRCT domains in some proteins (eg, BRCA1, microcephalin protein MCPH1, and topoisomerase [DNA] II-binding protein, TOPBP1) can also interact directly or indirectly with components of other cell processes, such as cell cycle checkpoint, DNA transcription, DNA replication, and apoptosis.<sup>4,8–12</sup> It has been proposed that the BRCT domain may be a link between DNA damage response and other cell processes (eg, cell cycle checkpoint and DNA transcription) which cooperate to repair the damaged DNA.<sup>13</sup> Therefore, it is not surprising that the mutations impairing BRCT domain structures or functions in some proteins, such as BRCA1, may lead to tumors or cancers.<sup>4,8,14</sup>

The BRCT domain is a folding unit of approximately 95 residues, which consists of a four-stranded parallel  $\beta$ -sheet surrounded by three  $\alpha$ -helices ( $\alpha$ 1,  $\alpha$ 3-helix on one side and  $\alpha$ 2-helix on the other side).<sup>2,5</sup> Previous studies have uncovered two kinds of BRCT modules, ie, a single BRCT such as the C-terminal domain of XRCC1,<sup>5</sup> and a double or tandem pair BRCT in proteins such as BRCA1 and 53BP1. The N-terminal BRCT (BRCTa) and C-terminal BRCT (BRCTb) of double BRCT are closely linked in sequence (about 30 residues) and in three-dimensional structure through the interaction between  $\alpha$ 2 of BRCTa and  $\alpha$ 1 and  $\alpha$ 3 of BRCTb.<sup>9,14,15</sup> Both single and double BRCT domains are mainly observed in eukaryote proteins, while the single BRCT domain can also be observed in the NAD<sup>+</sup>-dependent DNA ligase of bacteria and viruses.<sup>2</sup> The sequence identity between the BRCT domains is very low (average identity approximately 14%), but five conserved hydrophobic motifs are observed, ie, motif A in  $\beta$ 1, motif B in  $\alpha$ 1 and the loop between

$\alpha$ 1 and  $\beta$ 2, motif C in  $\beta$ 3, motif D in  $\alpha$ 3, and motif E at the C-terminal of the BRCT domain.<sup>3</sup>

To date, biochemical and cellular studies have shown that BRCT domains can interact in various ways with proteins or DNA. For example, the single BRCT domain in XRCC1 and DNA ligase III can form a heterodimer, which has demonstrated that the BRCT domain can function in protein-protein interactions.<sup>5,16</sup> The interaction between 53BP1 and p53 suggests that the double BRCT domain can also bind to protein.<sup>9</sup> Yu et al reported that many single and double BRCT domains can bind to phosphopeptides *in vitro*, and the phosphorylated protein-binding partners of many double BRCT domains have since been characterized.<sup>6–8,17,18</sup> Williams et al showed that the double BRCT in BRCA1 can bind to a phosphoserine-X-X-phenylalanine (X represents any residue) motif.<sup>19</sup> The phosphoserine is recognized by a phosphate-binding pocket (human BRCA1: serine 1655 and glycine 1656 in the loop between  $\beta$ 1 and  $\alpha$ 1, lysine 1702 at  $\alpha$ 2) in BRCTa, whereas the phenylalanine is recognized by a hydrophobic pocket at the interface between BRCTa and BRCTb. Interestingly, the single BRCT domains in NAD<sup>+</sup>-dependent DNA ligase and the replication factor C large subunit, RFC1, also contain a similar phosphate-binding pocket that can bind the phosphate of a DNA nick or end.<sup>20–23</sup> However, little is known about the processes that have led to these functional divergences. Therefore, we investigated the evolution of the BRCT domain and its functions using the increasing amount of information on crystal structure data and function. Our analyses indicate that the evolution of the eukaryote BRCT domain can be divided into three phases, and that the function is changed from binding DNA to binding protein. This functional change may be influenced by the evolution of the eukaryote DNA damage response interaction network.

## Materials and Methods

### Data collection

We retrieved 11 BRCT domain-hidden Markov model (HMM) profiles from the Pfam (<ftp://ftp.sanger.ac.uk/pub/databases/Pfam/releases/Pfam24.0/>) and superfamily (<http://supfam.org/SUPERFAMILY/>) databases. These profiles were used to search for new BRCT domains against the



redundant protein database (NR) using the HMM search program in HMMER.<sup>24</sup> If a new hit was predicted to be a BRCT domain with an *E* value <0.001 by any HMM profile, and the hit region did not overlap with another domain in the Pfam database (lower *E* value) by more than 10 amino acids, PSIPRED<sup>25</sup> was used to predict its secondary structure and ensure that it contained a  $\beta\alpha\beta\beta\alpha\beta\alpha$  topological architecture. If not, this new hit was discarded. All members of the new BRCT-containing protein family were retrieved from the NR database using BLASTp,<sup>26</sup> and a HMM profile was built with this new BRCT domain. These new HMM profiles were used to search the NR database with HMM search for the second round. This procedure continued until no new BRCT domain was found. The Mafft program was used for sequence alignment.<sup>27</sup> The default parameters were used for all programs in this procedure, except that the maximum target sequence was set to 10,000 in BLASTp.

Twelve eukaryotic genomes were then searched using the HMM search program with default parameters and 14 HMM profiles which could recognize all the BRCT domains in the NR database. The genome sequences of *Homo sapiens* (v.36.50), *Xenopus tropicalis* (v.4.1.50), *Takifugu rubripes* (v.4.50), *Gallus gallus* (v.2.50), *Danio rerio* (v.7.50), *Drosophila melanogaster* (v.5.4.50), *Aedes aegypti* (v.1.50), *Caenorhabditis elegans* (v.190.50) and *Saccharomyces cerevisiae* (v.1.01.50) were derived from the Ensemble database (<http://www.ensembl.org>). The genome sequences of *Arabidopsis thaliana* (v.5), *Oryza sativa* (v.5) and *Apis mellifera* (pre-release 2) were derived from the TIGR *A. thaliana* database (<http://www.tigr.org/tdb/e2k1/-ath1/>), Rice Genome Annotation Project website (<http://rice.plantbiology.msu.edu/-index.shtml>), and Human Genome Sequencing Center at Baylor College of Medicine (<http://www.hgsc.bcm.tmc.edu/projects/honeybee/>).

### Domain architecture and homolog relationship analyses

Domain architectures were predicted by searching the Pfam database plus the 14 HMM profiles of BRCT domain using HMM pfam with default parameters. The start and end positions of the BRCT domains were predicted by the 14 BRCT HMM profiles, and

the most frequently predicted N-terminal position and the most frequently predicted C-terminal position were used. Domain arrangement along sequences and reciprocal blast best hit were used to determine protein homolog relationships among species.

### Phylogenetic analyses

In this work, BRCT domains in nine representative eukaryote species, ie, *H. sapiens*, *X. tropicalis*, *T. rubripes*, *G. gallus*, *D. melanogaster*, *C. elegans*, *S. cerevisiae*, *A. thaliana*, and *O. sativa*, were used for the phylogenetic analyses. Because some lineage-specific BRCT domains were lost in *D. melanogaster*, *C. elegans*, and *S. cerevisiae*, the homolog BRCT domains in species close to them (*Ixodes scapularis* or *A. aegypti* for *D. melanogaster*; *Trichoplax adhaerens*, *Nematostella vectensis* or *Saccoglossus kowalevskii* for *C. elegans*; *Dictyostelium discoideum* or *Monosiga brevicollis* MX1 for *S. cerevisiae*) were used. Species-specific BRCT domains were excluded from these analyses. *S. cerevisiae* RAP1 and *C. elegans* BRCA1, the BRCT domains in which are diverged too much from their respective homologs, were substituted by *M. brevicollis* MX1 RAP1 and *N. vectensis* BRCA1, respectively.

The sequences were aligned using Mafft (Ginsi strategy for S-tree alignment, and Linsi strategy for D-tree and O-tree alignments),<sup>27</sup> and were manually adjusted according to their structures. Columns with more than one-third of gap characters were removed. The linker region between BRCTa and BRCTb in the D-tree alignment was also removed because it was highly diverged between the double BRCT domains. Bayesian trees were reconstructed using Mr Bayes version 3.1 for 200,000 generations.<sup>28</sup> The likelihood plot and potential scale reduction factor were used to check run convergence. The evolution rate among the sites was set to gamma distribution; the mixed protein substitution model was used, but the WAG model showed 1.0 posterior probability. The maximum-likelihood trees were reconstructed using Treefinder with the WAG model.<sup>29</sup> The expected-likelihood weights test was used to measure the confidence of tree topology with 1000 replications.

### Structural conservation analyses

The homologs of each BRCT family were derived in the 12 eukaryotic genomes and the nucleotide



NR database held by the National Center for Biotechnological Information. Blastp was used to search the NR database with default parameters. Sequences were listed in Fasta format in the supplementary materials. Sequences of every BRCT family were aligned separately using Mafft with the Linsi strategy and manually edited. Sequence conservation information for each BRCT family was then calculated on the ConSurf server using the maximum-likelihood method,<sup>30</sup> and mapped to their respective Protein Databank structures (Table S2) or homology models. The homology models of TOPBP1 BRCT4-5, PTIP BRCT1-2, and PTIP BRCT3-4 were generated using HHpred and Modeller on the Bioinformatics Toolkit website (<http://toolkit.tuebingen.mpg.de/sections/tertstruct>),<sup>31,32</sup> and multiple Protein Databank structures were automatically selected for templates. DaliLit version 3 was used for structure superimposition using default parameters.<sup>33</sup> The BRCT domain structures, 2EBU and 1T15, were used for RFC1 and BRCA1, respectively, for superimposition of structure.

## Results and Discussion

### Distribution of the BRCT domain

Database search results showed that BRCT-containing proteins are mainly observed in bacteria and eukaryotes. In bacteria, eight further species-specific single BRCT-containing proteins were identified (Fig. 1 and Table S1), the BRCT domains of which show high sequence identity with that of bacteria NAD<sup>+</sup>-dependent DNA ligase (average identity approximately 40%). In eukaryotes, new BRCT-containing proteins were also identified, including BRG1-associated factor BAF155, BAF170, ANK-containing protein ANKRD32, and GCN5-related N-acetyltransferase (Fig. 1 and Fig. S1). Because most archaea lineages do not contain BRCT-containing protein, except for the homolog of bacteria NAD<sup>+</sup>-dependent DNA ligase in some species of euryarchaeotes, these results are consistent with the hypothesis that the eukaryotic BRCT domain may be from bacteria through horizontal gene transfer.<sup>13</sup>

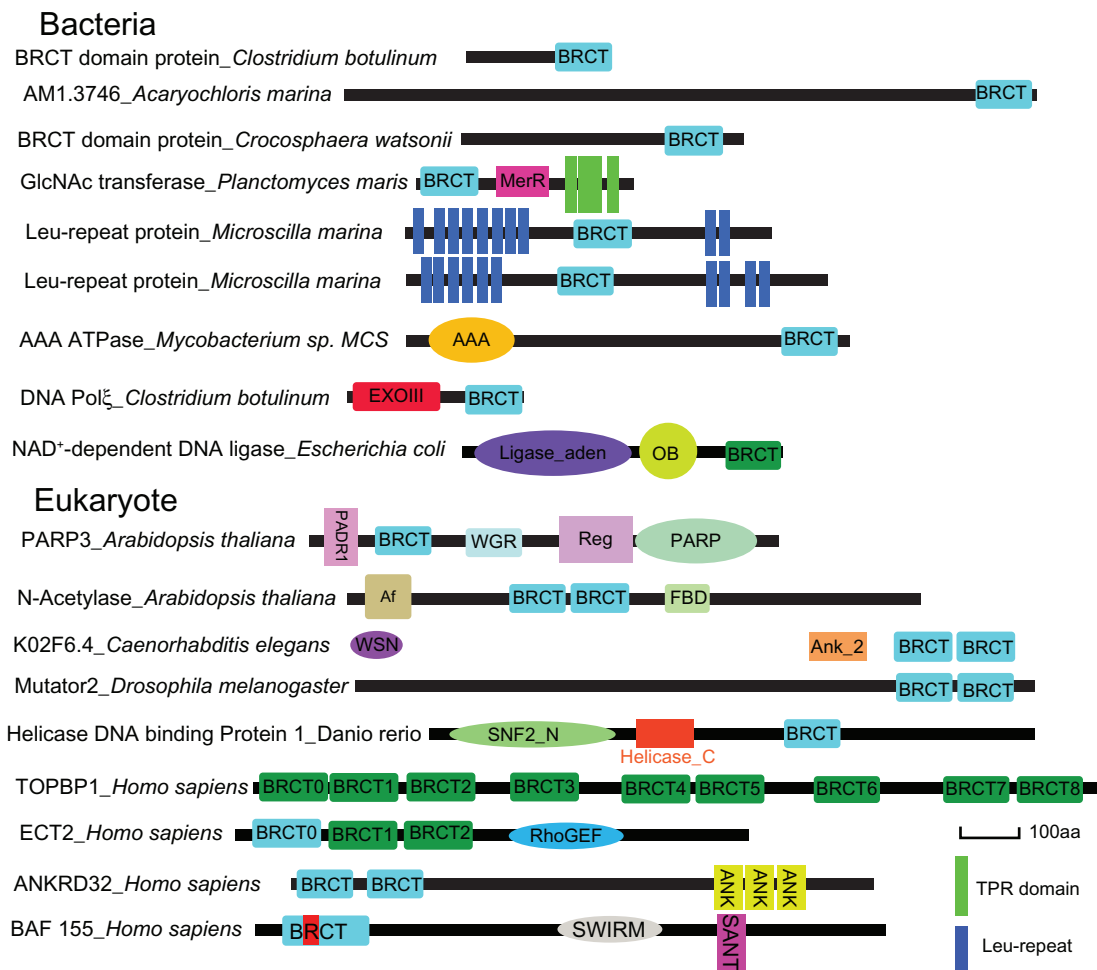
The number of BRCT-containing proteins per genome is correlated with genome complexity (one in *Escherichia coli*, 12 in *S. cerevisiae*, 27 in *H. sapiens*, and 28 in *O. sativa*, see Table S1). However, this correlation is not significant between

bacteria species, because most have only one or two BRCT-containing proteins, ie, NAD<sup>+</sup>-dependent DNA ligase and the DNA polymerase III  $\epsilon$  subunit. Two explanations can be offered for this correlation in eukaryotes. Firstly, more BRCT-containing proteins have evolved in species with high genome complexity. For example, in *A. thaliana* and *O. sativa*, two or more homologs of some animal BRCT-containing proteins, such as CTD phosphatase FCP1, BRCA1 associated RING domain-containing protein BARD1, and TOPBP1, have been identified, and there are also lineage-specific proteins containing 1–4 BRCT domains. In vertebrates, five BRCT-containing proteins, ie, DNA polymerase  $\mu$ , terminal deoxynucleotidyl transferase TdT, BAF170, regulatory subunit of Cdc7 kinase DBF4b, and ANKRD32, have evolved. Secondly, some BRCT-containing proteins are less conserved in species with low genome complexity. For example, there are nine single BRCT-containing proteins (RFC1, deoxycytidyl transferase REV1, DNA Pol  $\lambda$ , poly(ADP-ribose) polymerase PARP1, XRCC1, FCP1, TOPBP1, BAF155, and pescadillo protein PES1) and six double BRCT-containing proteins (BRCA1, BARD1, DNA ligase IV, nijmegen breakage syndrome protein NBS1, PAX transcription activation domain interacting protein PTIP, and TOPBP1). These can be observed in many eukaryote lineages, possibly having evolved in a eukaryote ancestor, but some have lost their specific lineage, including PARP1, XRCC1, BRCA1, and BARD1 in *S. cerevisiae*, and RFC1, DNA polymerase  $\lambda$ , and NBS1 in *C. elegans*. In addition, *C. elegans* has 30 BRCT-containing proteins, which is more than that expected according to the correlation. However, 19 of these have a similar domain arrangement of WSN, ANK, and BRCT, suggesting that they may be the result of one gene duplication (Fig. 1).

### Phylogeny of BRCT domains

To examine the evolutionary relationships of the BRCT domains, we performed three phylogenetic analyses with BRCT domains, mainly from nine eukaryote species. The Bayesian and maximum-likelihood methods were used to reconstruct the phylogenetic trees. Twenty single BRCT domains from 17 eukaryote proteins and bacteria NAD<sup>+</sup>-dependent DNA ligase were aligned to reconstruct single





**Figure 1.** Domain architectures of bacteria and eukaryote BRCT-containing proteins. New observed BRCT domains are colored with blue rectangles, while known ones are colored with green rectangles. The red color region in the middle of BRCT domain represents insertion. Domain architectures were predicted by HMMER. Protein names are linked with species name underlined. The species or clades distribution of these new observed BRCT-containing proteins are: DNA polymerase III  $\epsilon$  subunit in species of Actinobacteria, Firmicutes, Proteobacteria, and Planctomycetes; PARP3 in clade Magnoliophyta; N-acetylase in Viridiplantae; CHS5 in Fungi; TOPBP1 in eukaryotes; ECT2 in metazoan; MUTATOR2 in Drosophila; ANKRD32 in vertebrates; BAF155 in eukaryotes. Other new observed BRCT-containing proteins are only identified in one species.

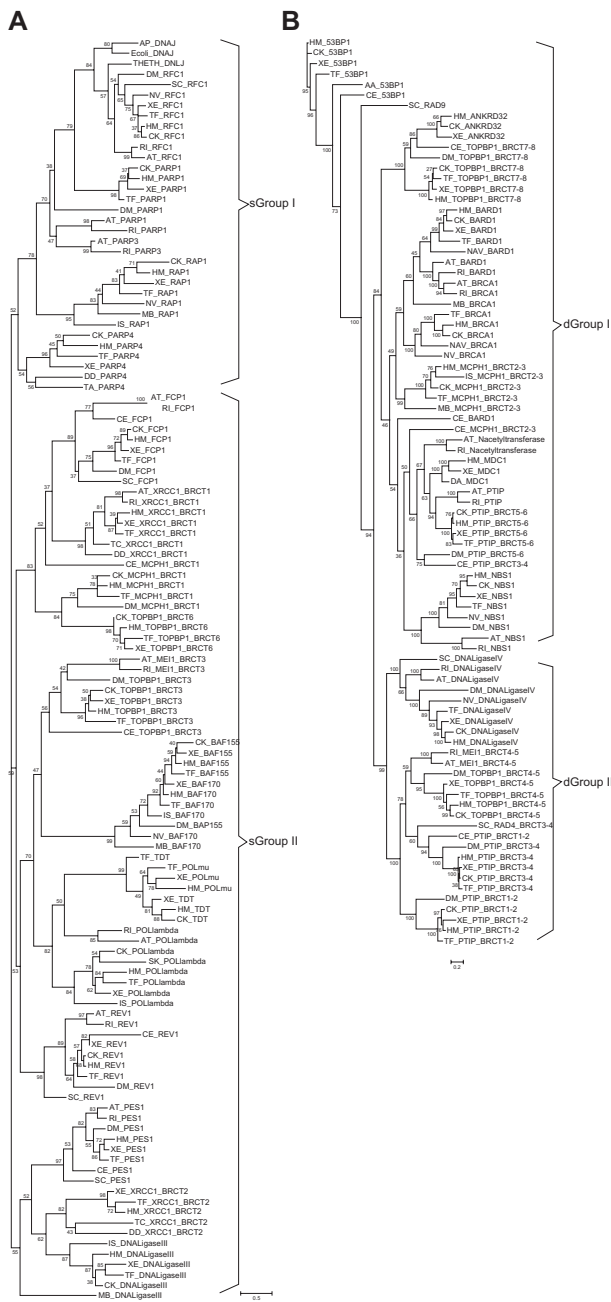
BRCT phylogenetic trees (S-tree, see Figs. 2A, S2A, and S4). Fourteen double BRCT domains from 11 eukaryote proteins were aligned to reconstruct double BRCT phylogenetic trees (D-tree, see Figs. 2B, S2B, and S5). To gain insight into the origin of double BRCT, we reconstructed the O-tree (Figs. S3 and S6), with seven single BRCT, and BRCTa and BRCTb of four double BRCT, all of which are distributed in both animals and plants.

The Bayesian S-tree and maximum-likelihood S-tree are consistent in their clustering of single BRCT domains into two major groups (ie, sGroup I and sGroup II) because of the low statistical support between the two groups, although the two trees differ in the placement of some branches within sGroup II. This clustering is further confirmed by

sequence analysis. The histidine in motif C and tryptophan in motif D, which locate in the hydrophobic core of the BRCT domain and play important roles in structure stability<sup>5,34</sup> (Fig. 3A), are observed in most sGroup II members, but are absent in most sGroup I members (Fig. S4).

Two double BRCT groups (dGroup I and dGroup II) are observed in the D-tree, with high statistic support (Figs. 2B, S4 and S5). The two groups may share a common origin because TOPBP1 and PTIP contain both dGroup I and dGroup II BRCT domains. Some members of the two groups are distributed across different eukaryote lineages, implying that the two groups are diverged very early.

The O-tree reconstructed by the two methods shows similar topology (Fig. S3). sGroup II BRCT



**Figure 2.** Phylogenetic trees of single BRCT domains (S-tree) and double BRCT domains (D-tree). **(A)** The unrooted S-tree was constructed with the maximum-likelihood method using the WAG model (see Materials and Methods section). The confidence values of branches were calculated using the expected-likelihood weights test with 1000 replications. **(B)** The unrooted D-tree was constructed using the Bayesian method and the mixed amino acid substitution model (see Materials and Methods section). Numbers above or below the branches indicate the percentage posterior probability. Taxons are displayed in the order of species name abbreviation, protein name, and the BRCT number, which are linked by underlining. The BRCT domains are numbered from the N-terminus to the C-terminus in proteins containing multiple BRCT domains.

**Abbreviations:** HM, Homo sapiens; XE, *Xenopus tropicalis*; CK, *Gallus gallus*; TF, *Takifugu rubripes*; DA, *Danio rerio*; IS, *Ixodes scapularis*; AA, *Aedes aegypti*; TA, *Trichoplax adhaerens*; DM, *Drosophila melanogaster*; NV, *Nematostella vectensis*; SK, *Saccoglossus kowalevskii*; CE, *Caenorhabditis elegans*; DD, *Dictyostelium discoideum*; MB, *Monosiga brevicollis* MX1; SC, *Saccharomyces cerevisiae*; AT, *Arabidopsis thaliana*; RI, *Oryza sativa*.

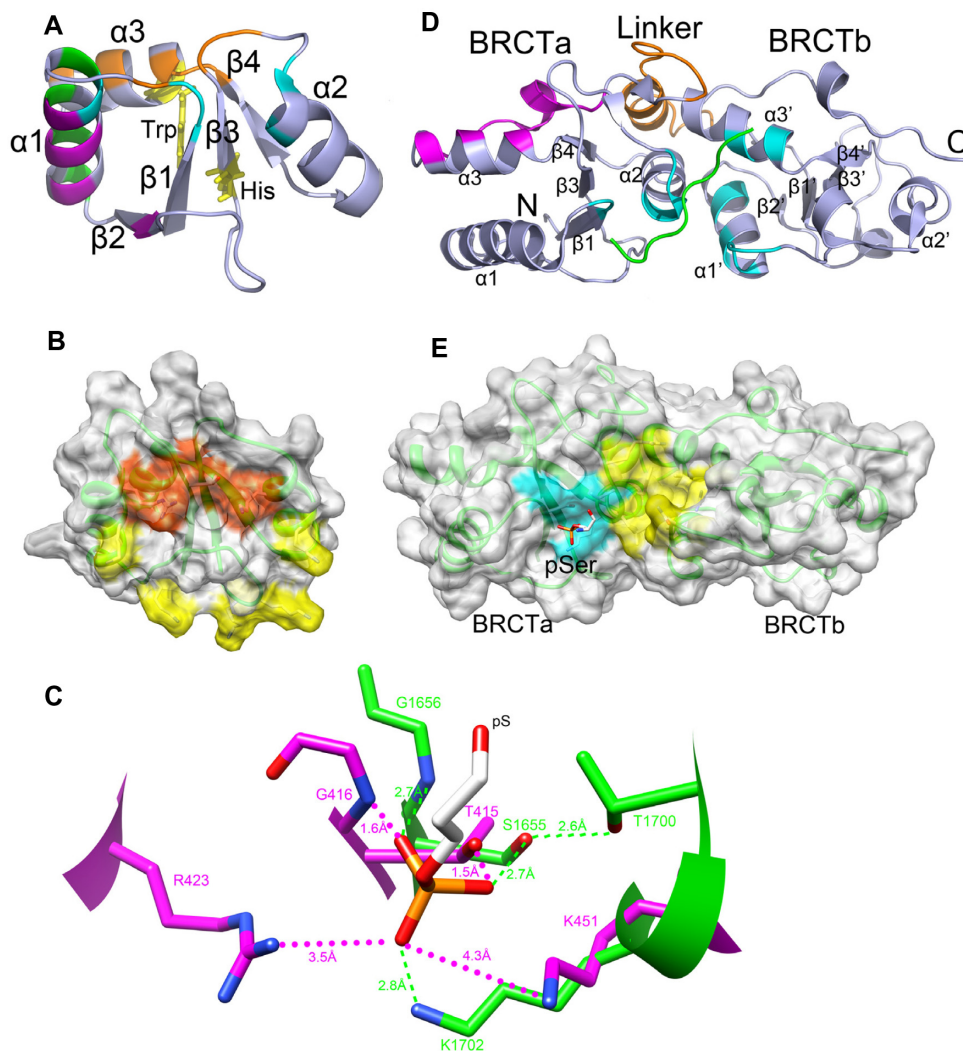
clustered together with either BRCTa or BRCTb of double BRCT, suggesting that double BRCT may have originated by tandem duplication of sGroup II BRCT. This is further confirmed by the fact that the histidine in motif C and tryptophan in motif D, which is common in sGroup II, are also observed in BRCTa and BRCTb of double BRCT (Figs. S5 and S6).

## Functional evolution of the BRCT domain

Because the sequence similarity between BRCT domains is low, the phylogenetic analyses may contain noise coming from multiple sequence alignments. Thus, we compared the functional sites between the four BRCT domain groups because their functional sites are more conserved than their sequences. One functional site is the phosphate-binding pocket in both single and double BRCT domains (Figs. 3A–3E). In this study, we first analyzed the conservation of this pocket in each protein family, and then compared this pocket in the different groups.

In eukaryote RFC1, the phosphate-binding pocket is composed of threonine 415 (or serine) and glycine 416 at the N-terminal of loop 1, arginine 423 at the N-terminal of  $\alpha 1$ , and lysine 458 at the N-terminal of  $\alpha 2$  (Figs. 3A, 3B, and 3C). Previous studies have revealed that this conserved pocket may bind the phosphate of nicked double-stranded DNA.<sup>20,21</sup> In bacteria NAD<sup>+</sup>-dependent ligase, this pocket is also highly conserved, and plays an important role in the formation of the AMP-ligase-DNA complex.<sup>22,23</sup> Our analyses have demonstrated that this pocket can be observed in the BRCT domain of animal but not plant PARP1, and the threonine 415 is substituted by a hydrophobic residue (human PARP1: leucine 398) which may influence phosphate binding.<sup>19</sup> In PARP3 and PARP4, and telomeric repeat binding factor interacting protein RAPI, two or more residues in this pocket are diverged, implying that these BRCT domains cannot bind nicked double-stranded DNA. Nonetheless, the functional similarity between RFC1 and NAD<sup>+</sup>-dependent ligase suggests that DNA binding is an important function of the original BRCT fold.

In sGroup II, the phosphate-binding pocket can only be observed in the BRCT domains of XRCC1



**Figure 3.** Single and double BRCT domain functional sites. **(A)** Model of functional sites in single BRCT (PDB ID: 2D8M). The cyan color residues form the phosphate-binding site; the green color residues represent the DNA polymerase  $\lambda$  protein-binding site; the magenta color residues form the XRCC1 BRCT2 and DNA ligase III protein-binding site; the orange color residues represent the functional sites of MCPH1 BRCT1. The histidine in motif C and tryptophan in motif D are colored yellow. **(B)** Transparent surface and ribbon diagram of RFC1 BRCT (PDB ID: 2EBU). The phosphate-binding pocket residues are colored orange-red, and the possible DNA-binding residues are colored yellow. **(C)** Closeup view of the superimposed phosphate-binding site in RFC1 (purple, PDB ID: 2EBU) and BRCTa of BRCA1 (green, PDB ID: 1T15). The distance between phosphate-binding residues and phosphate from phosphoserine are labeled with a dashed line. **(D)** Model of functional sites of the double BRCT domain (PDB ID: 1T15). The cyan color regions form the phosphopeptide-binding pocket. The regions colored orange and magenta represent the 53BP1 and Crb2 protein-binding sites, respectively. The phosphorylated peptide is colored green. **(E)** Transparent surface and ribbon diagram for double BRCT phosphopeptide-binding pocket. The side chains of phosphate-binding residues and the phosphoserine from phosphopeptide were shown. The cyan color represents the phosphate-binding pocket while the yellow color represents the specificity-determining pocket.

(BRCT1), FCP1, and REV1. However, there are differences between the pockets in these families and that of sGroup I. Arginine 423 is lost in FCP1 and REV1, and threonine 415 is substituted by asparagine in animal REV1 (human REV1: asparagine 57) and aspartate in plant REV1 (*A. thaliana* REV1: aspartate 96, Fig. S4). Nonetheless, other residues in the pocket are highly conserved, suggesting that this pocket is still functional. In addition, other BRCT domains in sGroup II have lost the phosphate-binding pocket

and evolved various other protein-binding sites (Fig. 3A).<sup>5,34,35</sup>

In all members of dGroup I, the phosphate-binding pocket is observed in BRCTa but not in BRCTb. Unlike single BRCT domains, the main function of this pocket is binding to the phosphate of phosphopeptides (Figs. 3D and 3E).<sup>6,7,12,19,28,36,37</sup> In order to compare the phosphate-binding pocket between the single and double BRCT domains, we superimposed the BRCT domain of RFC1 and BRCTa of BRCA1



(DaliLite: Z score 8.9, root mean square deviation 3.2). The superimposition showed that threonine 415, glycine 416, and lysine 458 of RFC1 were superimposed well with serine 1655, glycine 1656, and lysine 1702 of BRCA1 (DaliLite relative entropy  $> 3.456$  bit, Fig. 3C). However, differences are also observed between them. The phosphate-binding pocket in the single BRCT domain contains a highly conserved arginine 423 which is not observed in most BRCTa of dGroup I, and the BRCTa of dGroup I contains conserved threonine/serine (threonine 1700 in BRCA1, Fig. 3C) that interacts with serine 1655.<sup>19</sup> We therefore named the sGroup I pocket DNA-binding type and dGroup I pocket protein-binding type, but the differences in the phosphate recognition mechanism require further study. Interestingly, both the arginine (arginine 423 in RFC1) and the threonine (threonine 1700 in BRCA1) are conserved in the phosphate-binding pocket of XRCC1 BRCT1 (human XRCC1: arginine 335 and threonine 368), and threonine 1700 is also conserved in the phosphate-binding pocket of the FCP1 BRCT domain (human FCP1: threonine 691). Thus, the phosphate-binding pocket changed from the DNA-binding type to the protein-binding type in sGroup II, and the pocket in BRCTa may be evolved from that of sGroup II. This hypothesis is supported by the fact that the BRCT domain in FCP1 can bind phosphopeptide.<sup>7</sup>

In order to compare the phosphate-binding pocket in dGroup II with that in other groups, we modeled the structure of TOPBP1 BRCT4-5, PTIP BRCT1-2, and PTIP BRCT3-4 using dGroup I BRCT structures as templates (data not shown). The analysis showed that the phosphate binding pocket was observed in BRCTb but not BRCTa of this group except DNA ligase IV. The glycine 146 is substituted by glutamine in TOPBP1 BRCT5 (human TOPBP1: glutamine 655) and PTIP BRCT2 (human PTIP: glutamine 108). Both arginine 423 and threonine 1700 are not observed in this group. Whether these BRCTb domains can bind to phosphate needs further investigation. In DNA ligase IV, the phosphate-binding pocket is only observed in vertebrate BRCTa, the lysine 458 is substituted by arginine (human DNA ligase IV: arginine 708). In addition, our analyses demonstrated that the BRCTa and BRCTb of dGroup II members may not interact through the  $\alpha 2$  of BRCTa and  $\alpha 1$  and  $\alpha 3$

of BRCTb, because the charged residues in the middle of BRCTa  $\alpha 2$  (data not shown) cannot interact in a stable manner with the hydrophobic  $\alpha 1$  and  $\alpha 3$  regions of BRCTb. This was confirmed by a study showing that TOPBP1 BRCT1 and BRCT2, which showed high sequence similarity to the BRCTa and BRCTb of this group, respectively, interact with each other through the linker between BRCT1 and BRCT2, the  $\alpha 2$  region of BRCT1, and the C-terminal of BRCT2.<sup>38</sup>

Taken together, the intragroup similarity and intergroup differences in the phosphate-binding pocket is consistent with the phylogenetic analyses. In the D-tree, the high statistical support between the two groups may be mainly due to the different interaction patterns between BRCTa and BRCTb.

### Evolution of BRCT functions

The DNA-binding function may be conserved in the BRCT domains of bacteria because the DNA-binding pocket residues are conserved in them. However, the functions of eukaryote BRCT domains are diverged. The species distribution, phylogeny, and functional site comparison analyses suggest that evolution of the eukaryote BRCT domain can be divided into three phases. In the first phase, sGroup I members, such as RFC1 and PARP1, obtained the original BRCT fold with the DNA-binding function in the eukaryote ancestor. In the second phase, sGroup II BRCT domains, which contain histidine in motif C and tryptophan in motif D, were evolved. Although the phosphate-binding pocket is retained in XRCC1 (BRCT1), FCP1, and REV1, the functions of single BRCT domains originating in this phase began to diverge, such as the protein-binding function in XRCC1 (BRCT1), REV1, and DNA Pol  $\lambda$ .<sup>5,34,35</sup> Double BRCT domains were also evolved in this phase by tandem duplication of the sGroup II BRCT domain containing the phosphate-binding pocket. However, the phosphate-binding pocket was selectively lost in BRCTb of dGroup I and BRCTa of dGroup II. The specific phosphopeptide-binding function may also have evolved in the double BRCT ancestor because phosphopeptide-binding activity has been detected in sGroup II members.<sup>7</sup> The third phase may be after the divergence between animals and plants. Both sGroup I and sGroup





II BRCT domains originating in this phase lost the phosphate-binding pocket and many of them evolved protein-binding functions, such as TOPBP1 BRCT6, MCPH1 BRCT1, and XRCC1 BRCT2.<sup>5,34,35</sup> Lineage-specific dGroup I members were evolved, but few dGroup II members were observed. In addition to the phosphopeptide-binding function, some dGroup I members have also evolved other protein-binding sites (Fig. 3D).

In addition, the closely linked structure and high sequence similarities between the N-terminal three BRCT domains of TOPBP1 and ECT2 (identities: BRCT0 23%, BRCT1 40%, and BRCT2 30%) suggest that the BRCT domain may have evolved with three BRCT domains as a unit. This is supported by a recent crystal structure and function study of TOPBP1 BRCT0-2.<sup>38</sup> The central  $\beta$ -sheets in the three BRCT domains are perpendicular to each other rather than parallel as in dGroup I BRCT structures. The phosphate-binding pocket is conserved in ECT2 BRCT1, TOPBP1 BRCT1, and BRCT2, but arginine 423 and threonine 1700 are not observed in these pockets. TOPBP1 BRCT1 can bind phosphoserine 387 of RAD9.<sup>38</sup> Furthermore, there are reports that multiple BRCT domains function together, such as the phosphorylation-dependent interaction between PTIP3-6 and 53BP1.<sup>39</sup> Thus, there are other functional and evolutionary modules with multiple BRCT domains which may link multiple proteins together.

### Evolution of eukaryote BRCT domain in DNA damage response

Because most eukaryote BRCT domains function as signaling and protein targeting motifs in the DNA damage response system,<sup>2-4,8,18</sup> the question arises as to how the BRCT domain evolved in this system. Because the evolution of the DNA damage response system is affected by pressures from both the external and internal environments, such as lifestyle and genome complexity, the components of this system may be coevolved under these pressures.<sup>13</sup> The DNA repair functions of eukaryote proteins that contain the first-phase and second-phase BRCT domains, such as RFC1, PARP1, REV1, and XRCC1,<sup>20,35,40,41</sup> suggest that the BRCT domain may function firstly in the DNA repair process. This can be explained

by the fact that in the early stage of evolution of the eukaryote DNA repair process, the DNA repair proteins may be targeted to the exact site of damage independently because no sophisticated protein interaction network had evolved. Thus, the DNA end or nick binding function of the BRCT domain may be essential for these proteins. However, as the protein interaction network became more sophisticated, especially after the divergence of the metazoan/fungi group from plants, more proteins were targeted to sites of DNA damage by protein complexes which are more rapid and more efficient than targeted independently by DNA end-binding motifs.<sup>42</sup> Thus, the protein-binding function is more necessary than the DNA end-targeting function. This may explain why the single BRCT domains originating in the third phase changed their function from binding DNA to binding protein.

In the second phase of BRCT domain evolution, the specific phosphomotif-binding function of the double BRCT domain, which can recognize phosphomotifs generated by DNA damage detectors,<sup>6,7,10,36,37</sup> such as ATR and ATM, extends the BRCT domain to DNA damage signaling. The double BRCT domain then expands in DNA damage-signaling proteins accompanying changes in the internal environment. For example, the higher order chromatin structure formed by histones and DNA create barriers to DNA damage signaling and recruitment of repair enzymes for damaged DNA; the rise of multicellularity may have led to expression of tissue-specific genes in response to DNA damage.<sup>13</sup> Therefore, more proteins and signaling motifs may be required to communicate between components of the DNA damage system or with other cellular processes. The expansion of the BRCT domain, especially dGroup I members, may partially satisfy these requirements.

Taken together, the evolution of the eukaryote BRCT domain may be influenced by the evolution of the DNA damage response system. In organisms living in complex environments, for which a highly efficient DNA damage response system is essential, more BRCT domains may be required to recruit proteins to the damage sites and for signaling between cell processes, which may explain the correlation between BRCT domain number and genome complexity.



## Acknowledgments

We thank the staff of the Huang Laboratory and Professor Peng Shi for valuable comments. This work was supported by Grants to JFH from the National Basic Research Program of China (Grant No. 2007CB815705), the National Natural Science Foundation of China (Grant No. 30623007 and 30621092), and Chinese Academy of Sciences (Grant No. 2007211311091).

## Disclosure

This manuscript has been read and approved by all authors. This paper is unique and is not under consideration by any other publication and has not been published elsewhere. The authors and peer reviewers of this paper report no conflicts of interest. The authors confirm that they have permission to reproduce any copyrighted material.

## References

- Rouse J, Jackson SP. Interfaces between the detection, signaling, and repair of DNA damage. *Science*. 2002;297:547–51.
- Bork P, Hofmann K, Bucher P, Neuwald AF, Altschul SF, Koonin EV. A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J*. 1997;11:68–76.
- Callebaut I, Morion JP. From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair. *FEBS Lett*. 1997;400:25–30.
- Huyton T, Bates PA, Zhang XD, Sternberg MJE, Freemont PS. The BRCA1 C-terminal domain: structure and function. *Mutat Res*. 2000;460:319–32.
- Zhang X, Morera S, Bates PA, et al. Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. *EMBO J*. 1998;17:6404–11.
- Manke IA, Lowery DM, Nguyen A, Yaffe MB. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science*. 2003;302:636–9.
- Yu X, Chini CC, He M, Mer G, Chen J. The BRCT domain is a phospho-protein binding domain. *Science*. 2003;302:639–42.
- Glover JNM, Williams RS, Lee MS. Interactions between BRCT repeats and phosphoproteins: tangled up in two. *Trends Biochem Sci*. 2004;29:579–85.
- Joo WS, Jeffrey PD, Cantor SB, Finnin MS, Livingston DM, Pavletich NP. Structure of the 53BP1 BRCT region bound to p53 and its comparison to the Brcal BRCT structure. *Genes Dev*. 2002;16:583–93.
- Mohammad DH, Yaffe MB. 14-3-3 proteins, FHA domains and BRCT domains in the DNA damage response. *DNA Repair (Amst)*. 2009;8:1009–17.
- Garcia V, Furuya K, Carr AM. Identification and functional analysis of TopBP1 and its homologs. *DNA Repair*. 2005;4:1227–39.
- Shiozaki EN, Gu L, Yan N, Shi Y. Structure of the BRCT repeats of BRCA1 bound to a BACH1 phosphopeptide: implications for signaling. *Mol Cell*. 2004;14:405–12.
- Aravind L, Walker DR, Koonin EV. Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res*. 1999;27:1223–42.
- Williams RS, Green R, Glover JN. Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1. *Nat Struct Biol*. 2001;8:838–42.
- Derbyshire DJ, Basu BP, Serpell LC, et al. Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor. *EMBO J*. 2002;21:3863–72.
- Dulic A, Bates PA, Zhang X, et al. BRCT domain interactions in the heterodimeric DNA repair protein XRCC1-DNA ligase III. *Biochemistry*. 2001;40:5906–13.
- Rodriguez M, Yu X, Chen J, Songyang Z. Phosphopeptide binding specificities of BRCA1 COOH-terminal (BRCT) domains. *J Biol Chem*. 2003;278:52914–8.
- Rodriguez MC, Songyang Z. BRCT domains: phosphopeptide binding and signaling modules. *Front Biosci*. 2008;13:5905–15.
- Williams RS, Lee MS, Hau DD, Glover JN. Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. *Nat Struct Mol Biol*. 2004;11:519–25.
- Kobayashi M, Figaroa F, Meeuwenoord N, Jansen LE, Siegal G. Characterization of the DNA binding and structural properties of the BRCT region of human replication factor C p140 subunit. *J Biol Chem*. 2006;281:4308–17.
- Kobayashi M, Ab E, Bonvin AM, Siegal G. Structure of the DNA-bound BRCA1 C-terminal region from human replication factor C p140 and model of the protein-DNA Complex. *J Biol Chem*. 2010;285:10087–97.
- Wang LK, Nair PA, Shuman S. Structure-guided mutational analysis of the OB, HhH, and BRCT domains of *Escherichia coli* DNA ligase. *J Biol Chem*. 2008;283:23343–52.
- Wilkinson A, Smith A, Bullard D, et al. Analysis of ligation and DNA binding by *Escherichia coli* DNA ligase (LigA). *Biochim Biophys Acta*. 2005;1749:113–22.
- Eddy SR. Profile hidden Markov models. *Bioinformatics*. 1998;14:755–63.
- McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics*. 2000;16:404–5.
- Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–402.
- Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30:3059–66.
- Ronquist F, Huelsenbeck JP. MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19:1572–4.
- Jobb G. TREEFINDER October 2008. Available at: [www.treefinder.de](http://www.treefinder.de). Accessed April 20, 2011.
- Landau M, Mayrose I, Rosenberg Y, et al. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res*. 2005;33:W299–302.
- Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. 2005;33:W244–8.
- Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M. Evaluation of comparative protein modeling by MODELLER. *Proteins*. 1995;23:318–26.
- Holm L, Park J. DaliLite workbench for protein structure comparison. *Bioinformatics*. 1995;16:566–7.
- Mueller GA, Moon AF, Derose EF. A comparison of BRCT domains involved in nonhomologous end-joining: introducing the solution structure of the BRCT domain of polymerase lambda. *DNA Repair (Amst)*. 2008;7:1340–51.
- Guo C, Sonoda E, Tang TS, et al. REV1 protein interacts with PCNA: significance of the REV1 BRCT domain in vitro and in vivo. *Mol Cell*. 2006;23:265–71.
- Wood JL, Singh N, Mer G, Chen J. MCPH1 functions in an H2AX-dependent but MDC1-independent pathway in response to DNA damage. *J Biol Chem*. 2007;282:35416–23.
- Eliezer Y, Argaman L, Rhie A, Doherty AJ, Goldberg M. The direct interaction between 53BP1 and MDC1 is required for the recruitment of 53BP1 to sites of damage. *J Biol Chem*. 2009;284:426–35.
- Rappas M, Oliver AW, Pearl LH. Structure and function of the Rad9-binding region of the DNA-damage checkpoint adaptor TopBP1. *Nucleic Acids Res*. 2010;39:313–24.
- Munoz IM, Jowsey PA, Toth R, Rouse J. Phospho-epitope binding by the BRCT domains of hPTIP controls multiple aspects of the cellular response to DNA damage. *Nucleic Acids Res*. 2007;35:5312–22.



40. Hassa PO, Hottiger MO. The diverse biological roles of mammalian PARPs, a small but powerful family of poly-ADP-ribose polymerases. *Front Biosci.* 2008;13:3046–82.
41. Masson M, Niedergang C, Schreiber V, Muller S, Menissier-de Murcia J, de Murcia G. XRCC1 is specifically associated with poly(ADP-ribose) polymerase and negatively regulates its activity following DNA damage. *Mol Cell Biol.* 1998;18:3563–71.
42. Lisby M, Barlow JH, Burgess RC, Rothstein R. Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. *Cell.* 2004;118:699–713.

**Publish with Libertas Academica and every scientist working in your field can read your article**

*“I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely.”*

*“The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I’ve never had such complete communication with a journal.”*

*“LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought.”*

**Your paper will be:**

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

**<http://www.la-press.com>**