The BRCT domain from the large subunit of human Replication Factor C

Kobayashi, Masakazu

Chapter 6

General Discussions and future prospects

The main goal of this project was to elucidate the three dimensional structure of the complex comprised of 106 amino acids including the BRCT domain (aa 403-480) of the human RFC p140 subunit and dsDNA. Previous studies had shown that the N-terminal half of the human RFC p140 subunit contains a region with dsDNA binding activity (1), which does not contribute to the 3' primer-template DNA binding essential for the RFC complex to perform PCNA loading. This region shares sequence homology with the superfamily of BRCT domains, which are found in both prokaryotic and eukaryotic proteins associated with cell cycle-checkpoint regulations and DNA replication/repair. The domains consist of roughly 70-80 amino acids and have been characterized to act as scaffolds for protein complexes via homo/hetero-BRCT dimerization and BRCT-nonBRCT interactions, and to bind specifically to the phosphorylated serine (pS) of target proteins (Chapter 1). As no other members of the BRCT superfamily had been shown to bind DNA, the orthologous human RFC BRCT region was studied by deletion, point mutagenesis and NMR spectroscopy to understand better how the molecular basis of this unique interaction is achieved.

DNA binding by the N-terminus of the p140 subunit of RFC

Since there was no information on the amino acid sequence boundaries required for the DNA binding, first we cloned several truncated genes encoding for amino acid residues encompassing around the BRCT domain (403-480) of the human p140 subunit, which were expressed and purified from *E. coli*. All of the purified proteins were soluble and subsequently characterized for their DNA binding activities. This work revealed that the BRCT domain alone cannot bind dsDNA but requires approximately 28 amino acids N-terminal to the BRCT domain (Chapter 3) for binding activity. Furthermore DNA binding defective mutants suggested potential residues that come in contact with the DNA (Chapter 2 & 3). This functional region, p140(375-480) was shown to specifically bind double-stranded DNA dependent on 5'-phosphate termini (Chapter 2). Similar specificities and binding characteristics had been reported for the homologous region of p140 from *D. melanogaster* (2). Our study also revealed a yet to be fully characterized, non-specific

dsDNA binding activity by a protein consisting of p140(375-545) (Chapter 2). Unlike p140(375-480), the construct p140(375-545) formed two distinctive protein-DNA complexes separable in gel-retardation assay (Chapter 2). Similar protein-DNA complexes had been observed in earlier studies (2;3), in which the N-terminal half (roughly 550 amino acids) of human RFC p140 was reported to bind indiscriminatly to 5' non- or phosphorylated dsDNA (3). In contrast to this report, p140(375-545) bound tighter to 5' phosphorylated dsDNA than to non-phosphorylated dsDNA (Chapter 2). Further studies are clearly necessary to characterize how the 5'-phosphate specific binding is modulated in p140(375-545).

Structure of p140(375-480) bound to dsDNA

Despite the relative lack of overall sequence similarity, the comparison of different BRCT structures determined to date shows a common three-dimensional fold consisting of a central four stranded, parallel β-sheet surrounded by three α-helices. Helices α1 and α3 are more conserved in sequence and in length (4). Loop 3 and helix α2 are the most variable both in sequence and size. Here we present the first structure of a BRCT protein bound to DNA (Chapter 5). The NMR structure of p140(375-480) bound to dsDNA consists of an extra helix (α1') at the N-terminus, separated by a long loop (L1') from the rest of the BRCT domain. The fold of the BRCT domain is stabilized by a small number of key hydrophobic residues, which are relatively well conserved among the BRCT family members. Although the stretch of sequence N-terminal to the BRCT domain is essential for p140(375-480) binding to DNA, this region is not at all conserved among the orthologous p140 proteins (Chapter 3). However secondary structure prediction by PSIpred (5) shows the presence of one or more helices in over 80% of orthologous sequences immediately before the BRCT domain. It is not yet clear whether these helices should be considered as conservation of structure rather than of amino acid sequence in relation to the DNA binding function. It definitely warrants further investigation on the 5' phosphorylation dependent DNA binding activity in other eukaryotic p140 subunits.

Protein-nucleic acid docking based on biochemical and structural information

In spite of the poor sequence homology between the two, structural alignment of the BRCT-n domain from BRCA1 and RFC p140(375-480) revealed remarkable

conservation in the spatial location of the functional residues involved in BRCA1 phosphoserine binding (S1655, G1656 and K1702) and the conserved residues of the p140 (T415, G416 and K458 respectively) (Chapter 5). The residues that make hydrogen bonds to oxygen atoms of the phosphate are conserved among the tandem BRCT domains with known binding to phosphoserine containing peptides (6). Furthermore these residues are also shared among the BRCT domains from bacterial NAD^+ dependent DNA ligases, which were recently shown to have DNA binding activity (7;8). To date we have no direct evidence to suggest that these three conserved residues are involved in the interaction with 5' phosphate of dsDNA in the p140(375-480)-dsDNA complex, however both the mutagenesis and the NMR data indicate an important role for these residues in DNA binding. Our current model of the p140(375-480)-dsDNA complex therefore heavily relies on the assumption that those conserved residues interact with the 5'-phosphate of dsDNA specifically. The detection of intermolecular NOEs at the interface of the complex, which would experimentally confirm the assumption, proved to be difficult, first due to the reduced sensitivity of experiments by the low concentration of the sample (~ 0.5 mM) and second by the dynamic behaviour of the complex.

A number of alternative methods are available that could help to better define the protein-DNA interface. One such alternative using NMR is to irradiate the ^1H resonances of the bound DNA (H1' sugar, H5 base and imino proton resonances), which do not overlap with those of the protein. The resulting spin saturation can be transferred to neighbouring spins, including the amide proton of the bound protein, which can then be identified using the [^1H, ^{15}N]-HSQC experiment(9). We have used this cross-saturation method in an attempt to map the interface, but no saturation transfer from the DNA to the protein was observed (result not shown). Thus we were limited to using only information derived from the few observable intermolecular NOEs and the mutagenesis data for docking. However, no nucleotide specific information was available (Chapter 5). The inclusion of such information obtained by biochemical methods has been shown to improve correctness of empirical protein-DNA docking (10). Various biochemical assays, such as methylation or ethylation interference, can be used to identify which nucleotides contact the protein portion of a complex. Dimethyl sulfate methylates predominantly the 7-nitrogen of guanine in the major groove and the 3-nitrogen of adenine in the minor groove. Therefore interferences of methylation by interacting amino acids identify the binding sites in the major or minor groove of dsDNA (11). On the other hand, ethylation occurs on the

phosphate backbone of DNA and thus its interference can identify critical DNA phosphate backbone contacts for protein interactions (12). Furthermore, since the phosphate backbone is relatively uniform regardless of the sequence of bases, ethylation interference is one of the few techniques available to investigate non-sequence specific interactions.

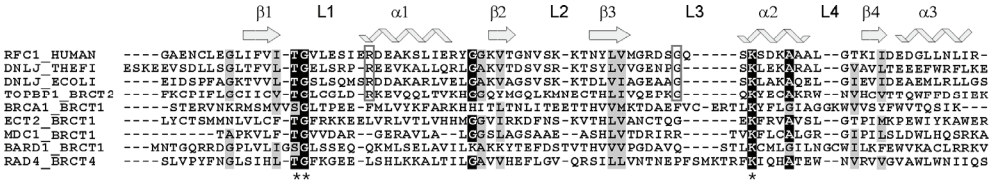
Because of the dynamic behaviour often observed in non-specific protein-DNA complexes, NMR analysis using short distance interactions such as the NOE (< 5Å) are not sufficient to completely characterize the behaviour of the complex. Long-range distance information (up to 35Å) can however be obtained from paramagnetic relaxation enhancement (PRE) which relies on magnetic interaction between unpaired electrons and nuclei. The introduction of a paramagnetic probe consisting of Mn^{+2} – chelating EDTA attached to dT, which is located at either side of dsDNA sequence allowed observation of PRE induced line broadening of the ^{1}H amide in the [^{1}H, ^{15}N]-HSQC spectra of HMGB, which revealed HMGB binding non-specifically to multiple sites on dsDNA (13). Similarly, using florescence energy transfer (FRET) between two fluorescence probes could provide long distance information on between two fluorophore labelled pairs (e.g. protein and DNA). The large structural changes associated with DNA upon protein binding have been successfully monitored by fluorescence labelled DNA, which allowed atomic structure based modelling of e.g. FEN1-DNA complex (14)

Structural conservation among pS binding/5'PO_4 binding

In the various structures of BRCT-pS peptide complexes that are presently available, the hydrogen bonds to the phosphate oxygens are weak and additional stabilizing interactions are required between other amino acids of the bound peptide and the BRCT. Similarly, our model of the p140(375-480)-dsDNA complex suggests that additional non-5' phosphate directed interactions are present, which indeed correlate with our mutagenesis and NMR data (Chapter 5).

The remarkable conservation of the mechanism of recognition of the 5'-phospate of DNA and the phosphoserine moiety of the peptide noted in Chapter 5 may help us to identify other BRCT bearing proteins that are phosphate-specific binders. However, it is rather difficult to identify these key residues simply by amino acid sequence alignment, as there is a lack of sequence homology among the BRCT family members, which is further diverged by various insertions and deletions within the sequence. In order to improve the accuracy of a sequence alignment, one could use predicted secondary structures derived

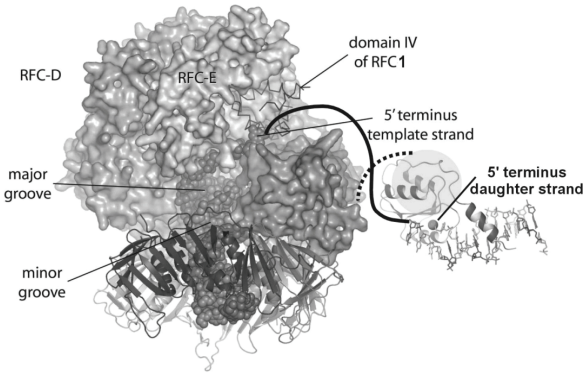
from the primary structure as a guide. The secondary structure of BRCT domains known to bind either peptides (15) or DNA (7;8) was predicted using PSIpred (5) and used to manually adjust the sequence alignment generated by Clustal W(16) (Figure 6.1). The resulting sequence alignment clearly indicates the conservation of three key residues (indicated by asterisks), which form hydrogen bonds to the oxygen atoms of the phosphate in the BRCA1-pS structures (6). Among the DNA binding BRCT domains (RFC1_HUMAN, DNLJ_THEFI and ENLJ_ECOLI), two other characteristic conservations are observed for R in helix α1 and G in the loop 3 (L3) (both residues are boxed), of which R was directly shown important for DNA binding activity (Chapter 3). From the sequence alignment, one can therefore define two conserved amino acids motifs shared by phosphate binding BRCT domains as sequences (T/S)-G within Loop 1 (L1) and K-x-x-x-A/G within helix α2 (Figure 6.1). However a special attention must be paid in identifying the conserved K (in helix α2), which can fall in either L3 or helix α2 depending on the level of confidence in the structure prediction.



(Figure 6.1) Sequence alignment of BRCT domains with known DNA or pS binding activities. Sequence alignment was first performed using ClustalW, and then the alignment was further adjusted manually according to the alignment of the secondary structure. DNA binding BRCTs are RFC1_BRCT (RFC p140), DNLJ_THEFI and DNLJ_ECOL (NAD+ dependent DNA ligase from *Thermus filiformis* and *Escherichia coli*). Phospho-serine binding BRCTs are (second BRCT domain of human Topoisomerase Binding Protein1), BRCA1_BRCT (the first BRCT domain from human BRCA1), ECT2_BRCT1 (the first BRCT from human Epithelial Cell Transforming sequence 2), MDC1_BRCT1 (the first BRCT from human Mediator of DNA damage checkpoint protein 1), BARD1_BRCT1 (the first BRCT from human BRCA1-associated RING domain protein 1) and RAD4_BRCT4 (fourth BRCT from *Schizosaccharomyces pombe* S-M checkpoint control protein Rad4). The secondary structure, an arrow for a β-strand and a ribbon for an α-helix, were derived from the three dimensional structure of BRCT domain of RFC1_BRCT (Chapter 5).

Possible Cellular Roles for the BRCT region of RFC p140

The cellular function of the 5' terminus binding by BRCT domain of RFC is not yet understood. In contrast, the cellular role of RFC complex binding directed at the 3' end of primer-template DNA has been well characterized and known to be crucial in PCNA loading and subsequent recruitment of PCNA associated DNA modifying enzymes. Accordingly, the crystal structure of the complex between BRCT-truncated RFC complex (trRFC) and PCNA implied that the heteropentameric subunits of trRFC caps itself onto the 3' end of the primer-template DNA blocking the 3' elongation while allowing the protruding template to snake out between the subunits (see details Figure 1.1). The predicted exit route of the single stranded template DNA has been suggested to pass directly over the Rfc1 subunit (Figure 6.2, green and solid line) and come out exactly where the BRCT region would be expected in the model where truncated BRCT in the crystal structure of trRFC was substituted with our NMR structure of the BRCT domain. An important implication of this combined model is that binding by the BRCT region of a 5' dsDNA terminus of, for instance, a previously synthesized Okazaki fragment, would place the clamp loader portion of the complex in the correct position to interact with PCNA that is involved in synthesizing a downstream Okazaki fragment. Indeed involvement of the RFC p140 in completion of Okazaki fragment synthesis has recently been implied (see below).

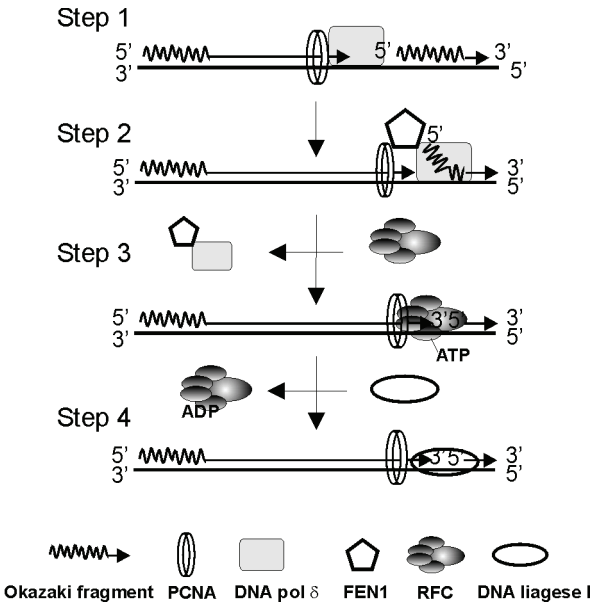


(Figure 6.2). Schematic model of the complete RFC-PCNA-DNA complex. The model is formed from our NMR structure of the RFC p140 BRCT region bound to 5' P dsDNA (right side ribbon view). The purple sphere behind it represents the 60 aa's C-terminal to the BRCT region we believe form a single domain. The broken line indicates the protein connection to the N-terminal portion of RFC1 in the crystal structure (purple surface). The template DNA strand (green) is schematically connected to the BRCT complex via the solid line. We estimate a minimum of 15 bases of ssDNA are required to bridge the 3' and 5' termini.

Completion of Okazaki Fragment Synthesis

At present the cellular role of 5'-phosphate DNA binding by the BRCT region of RFC remains elusive. The most probable function of this activity is to localize the RFC

complex at the 5' end of dsDNA during DNA repair and replication. In both processes, DNA synthesis eventually meets the 5'end of the down stream Okazaki fragment or reinitiated DNA synthesis. Efficient completion of the Okazaki fragment maturation requires coordinated activities of DNA pol δ, FEN1, DNA ligase I, PCNA and RFC (Chapter 1). Therefore we could speculate that localization of RFC at the 5' terminus may allow RFC to coordinate release of incoming nick translation enzymes, pol δ/FEN1 from PCNA and recruitment of DNA ligase I to PCNA as a result (Figure 6.3). Such coordinated interaction among RFC, PCNA and DNA ligase I in Okazaki fragment maturation has been shown (17), in which it was postulated that the N-terminus half of RFC p140 subunit, most likely the conserved BRCT domain, is involved in protein-protein interaction with the catalytic domain of DNA ligase I. Mediation of DNA ligase I activity by RFC is alleviated upon interaction of DNA ligase I with PCNA indicating that RFC regulates the activities of DNA modifying enzymes to function at appropriate places and times.



(Figure 6.3) Schematic presentation of Okazaki fragment maturation and potential role of 3' and 5' phosphate DNA ends by RFC at the ligatable nick. Steps: 1) Incoming 3' elongation by pol δ. 2) strand displacement of the Okazaki fragment and subsequent removal of RNA molecules by FEN1. 3) pol δ and FEN has to be displaced. The displacement could be achieved via RFC binding at 3' and 5' ends of the ligatable nick releasing the pol δ-FEN from PCNA or off-loading of PCNA-polδ-FEN1 complex by unknown off-loader complex, e.g. Ctf18-RFC (Chapter 1) followed by reloading of PCNA at the nick site by RFC. 4) Interaction between loaded PCNA and RFC results in release of RFC from the DNA only then allowing access of DNA ligase I to the nick for ligation.

Additional Role in Repair of DNA

As mentioned in Chapter 1, the BRCT domain is dispensable for both *in vivo* (18) and *in vitro* (19) replication activity of RFC. The budding yeast *rfc1-1* strain, carrying a point mutation (D513N) in the conserved RFC box II, accumulated short single strand DNA during DNA synthesis *in vivo* showing the phenotype expected for defective Okazaki

fragment maturation and displayed synthetic lethality in a *rad52* null mutant (20). It appears in yeast that replication induced DNA damage, such as single strand breaks caused by defective replication proteins, *rad27* (FEN1), *rad30* (PCNA) and *pol3* (Pol δ), are converted into double strand breaks which are then efficiently repaired via homologous recombination mediated by Rad52 during the S phase of the cell cycle (reviewed in (21)). Therefore the lack of obvious phenotypic defects in the *rfc* mutants is likely due to efficient DNA repair *via* the Rad52 mediated pathway. The fact that the yeast strain *rfc1::Tn3,* expressing an N-terminally truncated RFC1 (p140), is inviable when present in a *rad52* null mutant (22), clearly indicates a role for the BRCT region of RFC1 in one of the DNA replication or repair pathways.

It is not known if the observed mutator phenotype of *rfc1::Tn3* strain is due to the direct consequence of the defective 5' end DNA binding, since there is no evidence to show the occurrence of this specific binding *in vivo*. Our site directed mutagenesis study revealed several mutations that could abrogate directly 5'-PO_4 dsDNA binding without interfering with the overall fold of the BRCT domain (Chapter 3). These point mutations should therefore be introduced into yeast RFC1 to see if the mutants would display similar phenotypes observed in *rfc1:Tn3*. These studies however, urgently require development of an assay that can detect binding of RFC at the 5'-phosphorylated end of dsDNA *in vivo*.

In vitro studies have suggested a role for the BRCT region of RFC1 in base excision repair process since sensitivity towards DNA damage by Methyl-Methane Sulfonate (MMS) was observed in a yeast strain carrying trRFC (18;22). MMS causes DNA damage through methylation of DNA bases, which are subsequently removed by DNA *N*-glycosylase (23) to generate apurinic/apyrimidinic (AP) sites. AP sites stall replication forks and in yeast, are repaired prior to replication through base excision repair (BER) initiated by a single strand break on the 5' side of the AP site by AP endonuclease. The 5' incision generates 5' deoxyribosephosphate, which is subsequently removed by Rad27 (FEN1) (reviewed in (24)). The resulting gap is filled by Pol2 (Pol ε) and the nick is sealed by DNA ligase I (25;26). Importantly, the efficiency of BER appears to be strongly dependent on the presence of PCNA and its loader, RFC, *in vitro* (27). Alternatively AP lesions can also be bypassed via insertion of preferably dCMP or dGMP opposite the AP sites, in a mutagenic process called translesion synthesis (TLS) requiring REV1, Pol ζ and Pol η.

The data currently available are not sufficient to define the role of the BRCT region of RFC in a specific repair or replication pathway. To do so will require further studies using genetic crossing between yeast strains bearing enzymes defective in either BER or TLS and the BRCT defective RFC1 strain, followed by analysis of growth in the presence of specific DNA damaging reagents in order to pinpoint which pathway(s) the BRCT region of RFC1 is involved in.

Reference list

1. Fotedar, R., Mossi, R., Fitzgerald, P., Rousselle, T., Maga, G., Brickner, H., Messier, H., Kasibhatla, S., Hubscher, U., and Fotedar, A. (1996) *EMBO J.* **15,** 4423-4433

2. Allen, B. L., Uhlmann, F., Gaur, L. K., Mulder, B. A., Posey, K. L., Jones, L. B., and Hardin, S. H. (1998) *Nucleic Acids Res.* **26,** 3877-3882

3. Mossi, R., Keller, R. C., Ferrari, E., and Hubscher, U. (2000) *J.Mol.Biol.* **295,** 803-814

4. Dulic, A., Bates, P. A., Zhang, X., Martin, S. R., Freemont, P. S., Lindahl, T., and Barnes, D. E. (2001) *Biochemistry* **40,** 5906-5913

5. Jones, D. T. (1999) *J.Mol.Biol.* **292,** 195-202

6. Williams, R. S., Lee, M. S., Hau, D. D., and Glover, J. N. (2004) *Nat.Struct.Mol.Biol.* **11,** 519-525

7. Feng, H., Parker, J. M., Lu, J., and Cao, W. (2004) *Biochemistry* **43,** 12648-12659

8. Wilkinson, A., Smith, A., Bullard, D., Lavesa-Curto, M., Sayer, H., Bonner, A., Hemmings, A., and Bowater, R. (2005) *Biochim.Biophys.Acta* **1749,** 113-122

9. Ramos, A., Kelly, G., Hollingworth, D., Pastore, A., and Frenkiel, T. (2001) *J.Am.Chem.Soc* **122,** 11311-11314

10. Knegtel, R. M., Boelens, R., and Kaptein, R. (1994) *Protein Eng* **7,** 761-767

11. Chandrashekaran, S., Manjunatha, U. H., and Nagaraja, V. (2004) *Nucleic Acids Res.* **32,** 3148-3155

12. Manfield, I. and Stockley, P. G. (1994) *Methods Mol.Biol.* **30,** 125-139

13. Iwahara, J., Schwieters, C. D., and Clore, G. M. (2004) *J.Am.Chem.Soc* **126,** 12800-12808

14. Chapados, B. R., Hosfield, D. J., Han, S., Qiu, J., Yelent, B., Shen, B., and Tainer, J. A. (2004) *Cell* **116,** 39-50

15. Yu, X., Chini, C. C., He, M., Mer, G., and Chen, J. (2003) *Science* **302,** 639-642

16. Higgins, D. G., Thompson, J. D., and Gibson, T. J. (1996) *Methods Enzymol.* **266,** 383-402

17. Levin, D. S., Vijayakumar, S., Liu, X., Bermudez, V. P., Hurwitz, J., and Tomkinson, A. E. (2004) *J.Biol.Chem.* **279,** 55196-55201

18. Gomes, X. V., Gary, S. L., and Burgers, P. M. (2000) *J.Biol.Chem.* **275,** 14541-14549

19. Uhlmann, F., Cai, J., Gibbs, E., O'Donnell, M., and Hurwitz, J. (1997) *J.Biol.Chem.* **272,** 10058-10064

20. Merrill, B. J. and Holm, C. (1998) *Genetics* **148,** 611-624

21. Kuzminov, A. (2001) *Proc.Natl.Acad.Sci.U.S.A* **98,** 8461-8468

22. Xie, Y., Counter, C., and Alani, E. (1999) *Genetics* **151,** 499-509

23. Guillet, M. and Boiteux, S. (2003) *Mol.Cell Biol.* **23,** 8386-8394

24. Boiteux, S. and Guillet, M. (2004) *DNA Repair (Amst)* **3,** 1-12

25. Blank, A., Kim, B., and Loeb, L. A. (1994) *Proc.Natl.Acad.Sci.U.S.A* **91,** 9047-9051

26. Jin, Y. H., Ayyagari, R., Resnick, M. A., Gordenin, D. A., and Burgers, P. M. (2003) *J.Biol.Chem.* **278,** 1626-1633

27. Matsumoto, Y., Kim, K., Hurwitz, J., Gary, R., Levin, D. S., Tomkinson, A. E., and Park, M. S. (1999) *J.Biol.Chem.* **274,** 33703-33708