



Universiteit  
Leiden  
The Netherlands

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

Kobayashi, Masakazu

### Citation

Kobayashi, M. (2006, September 6). *The BRCT domain from the large subunit of human Replication Factor C*. Retrieved from <https://hdl.handle.net/1887/4546>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4546>

**Note:** To cite this publication please use the final published version (if applicable).

The BRCT domain from the large subunit of human Replication Factor C: Protein-DNA complex determined by NMR and mutagenesis.

**PROEFSCHRIFT**

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR AAN DE  
UNIVERSITEIT LEIDEN, OP GEZAG VAN DE RECTOR  
MAGNIFICUS DR. D. D. BREIMER, HOOGLERAAR IN DE  
FACULTEIT DER WISKUNDE EN NATUURWETENSCHAPPEN EN  
DIE DER GENEESKUNDE, VOLGENS BESLUIT VAN HET COLLEGE  
VOOR PROMOTIES TE VERDEDIGEN OP WOENSDAG 6  
SEPTEMBER 2006 TE KLOKKE 16:15 UUR

Door

Masakazu Kobayashi

Geboren te Kofu-city (Japan) in 1977



Promotie commissie

Promotor : Prof. Dr. G.W. Canters

Co-Promoter: Dr. G. Siegal

Referent : Prof. Dr. P. M. Burgers

Overige Leden :

Prof. Dr. J. Brouwer

Prof. Dr. R. Boelens

Prof. Dr. L. H. F. Mullenders

*For my wife and daughter,  
and my parents*

**Front and rear covers:** The layout was designed by Emanuela Lonardi using the original painting “*CICADA*” by Fumikazu Kobayashi

# Contents

<b>Chapter 1</b>	General Introduction	<b>7</b>
<b>Chapter 2</b>	Characterization of the DNA binding and Structural Properties of the BRCT region of the p140 subunit of human replication Factor C	<b>39</b>
<b>Chapter 3</b>	Amino acid determinants for DNA binding by the BRCT region of human RFC p140	<b>65</b>
<b>Chapter 4</b>	$^1\text{H}$ , $^{15}\text{N}$ and $^{13}\text{C}$ resonance assignments and secondary structure determination of the BRCT Region of the large subunit of human Replication Factor C, bound to DNA	<b>81</b>
<b>Chapter 5</b>	Structure of the BRCT domain from RFC p140: A model Protein-DNA complex determined by NMR and mutagenesis data	<b>93</b>
<b>Chapter 6</b>	General discussions and future prospective	<b>133</b>
	<b>Summary in English</b>	<b>143</b>
	<b>Samenvatting</b>	<b>146</b>
	<b>Appendices: Colour figures</b>	<b>148</b>
	<b>List of publications</b>	<b>155</b>
	<b>Curriculum Vitae</b>	<b>156</b>
	<b>Acknowledgement</b>	<b>157</b>



# Chapter 1

The roles of Replication factor C (RFC) in Eukaryotic DNA replication, and the unique DNA binding mediated by the BRCT domain of the p140 subunit of RFC.

---

## Abstract

---

This chapter begins with an introduction to eukaryotic DNA synthesis, which is followed by detailed descriptions of the structural and functional properties of Replication Factor C (RFC) in DNA replication. The potential functions of the BRCT (*BRAC1 C-Terminus*) domain of RFC, which have not been well characterized, are also discussed. The second half of the chapter describes structural and functional properties of BRCT domains, which in recent years, have become increasingly recognized as an important player in cellular signal transduction. At the end of the chapter, a brief description of structure determination using NMR is given, and the chapter concludes with an overview of this thesis.

---

## Replication Factor C and DNA replication

---

### *General introduction to Eukaryotic DNA synthesis*

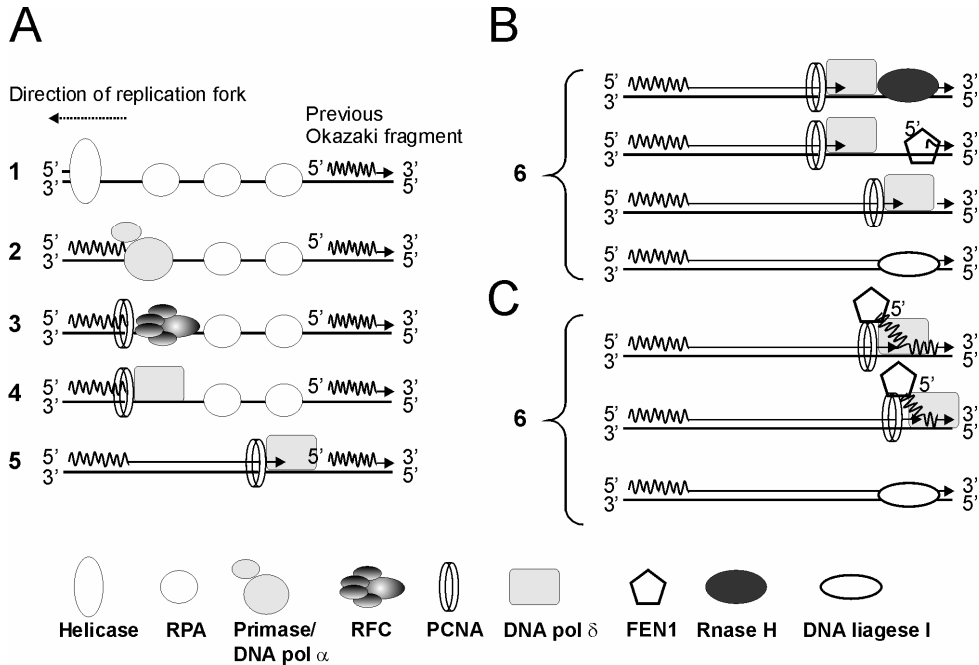
Every eukaryotic cell that divides must pass on two equal complements of the genetic material to the two daughter cells. The discovery of the duplex structure of DNA by Watson and Crick (1;2) led to the notion that each strand of the duplex is used as a template for the synthesis of new DNA, e.g. “DNA replication”. DNA replication in eukaryotes does not occur throughout the life of a cell, but rather it occurs at a specific time during the cell cycle which consists of  $G_1$  (the first growth), S (DNA synthesis),  $G_2$  (the second growth) and M (Mitotic, cell division) phases. During S phase, the genetic material must be duplicated with great precision, therefore multiple protein complexes are involved in order to perform this task.

The synthesis of new DNA is closely coupled to the unwinding of the parental strands. The initiation of DNA replication requires encircling of unreplicated DNA by MCM (Mini-Chromosome Maintenance) proteins. By twisting DNA from a distance, MCM unwind the strands at the constraint site of DNA synthesis called replication fork (3). At the replication fork, both unwound strands called parental strands serve as templates for the synthesis of new DNA. Due to the anti-parallel nature of dsDNA and the unique  $5' \rightarrow 3'$  directionality of DNA synthesis, a new daughter DNA strand must be either continuously synthesized in the direction of replication fork movement (leading strand), or in a direction opposite to fork movement (lagging strand). On the lagging strand, DNA is synthesized as discontinuous, small fragments, called Okazaki fragments. As replication proceeds, these fragments are joined to complete lagging strand synthesis.

### *Eukaryotic DNA synthesis in detail*

The identification of essential DNA replication components in eukaryotic cells was led by the development of a cell-free system for replication of Simian Virus 40 (SV40) origin containing DNA. DNA replication in the in vitro system is dependent on the SV40 large T antigen and human cell extracts (4). In this system, unwinding of dsDNA is initiated by the  $3'-5'$  helicase activity of the SV40 large T antigen, which is stimulated by the ssDNA binding protein, Replication Protein A (RPA)(Figure 1.1A Step 1). DNA synthesis is initiated on the template DNA by an RNA primer, which is synthesized by the primase subunit of DNA pol  $\alpha$  (pol  $\alpha$ ). Subsequently a stretch of DNA is synthesized by the polymerase subunit of DNA pol  $\alpha$  (Figure 1.1A Step2). DNA synthesis by pol  $\alpha$  is

tightly limited to approximately 20 deoxynucleotides, by Replication Factor C (RFC), which binds the 3' end of the primer/template displacing pol  $\alpha$  from DNA and initiating the so-called “polymerase switch” (Figure 1A Step3) (4-6).



(Figure 1.1) Proposed model of eukaryotic DNA replication. A. Model was adopted from ref (7). Okazaki fragment maturation models (B) and (C) were adopted from ref (8;9). Each step is numbered as referenced in the text.

Simultaneously, the RFC complex loads proliferating cell nuclear antigen (PCNA) onto DNA at the primer/template junction in an ATP-dependent manner. Upon PCNA loading, the bound ATP is hydrolyzed by the ATPase activity of RFC which results in release of RFC from the DNA (Figure 1.1A Step3 to 4). PCNA is a toroidally shaped homotrimer which encircles the double helix and slides freely, hence the name “sliding clamp”(10-12). The sliding clamp recruits DNA pol  $\delta$  to the 3' primer/template junction. Although DNA pol  $\delta$  is a poorly processive enzyme by itself, due to its weak interaction with DNA, PCNA bound DNA pol  $\delta$  forms a stable DNA/pol complex (Figure 1.1A Step 4) resulting in a highly processive polymerase, which continues to add nucleotides without dissociating from the DNA (4) (Figure 1.1A Step 5). There is growing evidence to suggest that another highly processive polymerase, DNA pol  $\epsilon$ , plays as important a role as DNA pol  $\delta$  in DNA



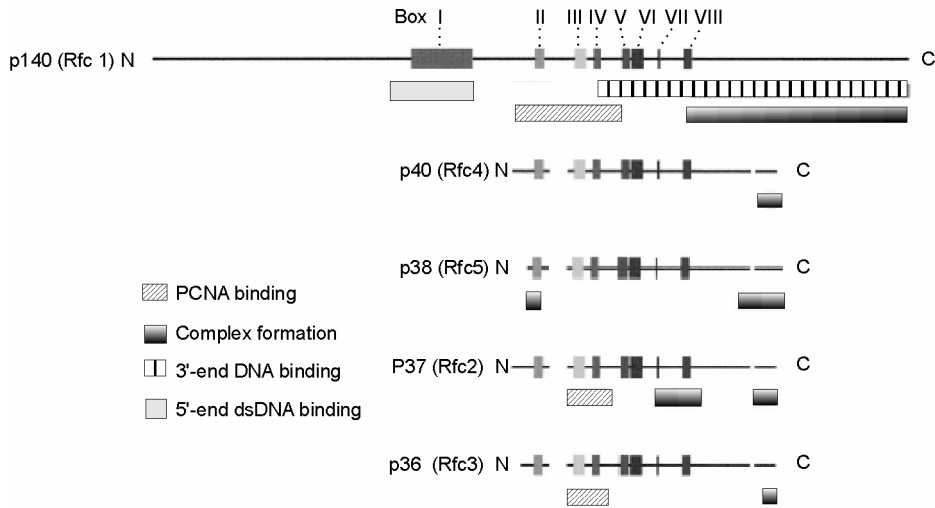
replication. During S-phase of the cell cycle, the association of pol  $\epsilon$  at the initiation site of DNA replication has been shown in yeast cells (13;14). The use of mutator DNA polymerases in Yeast suggests that the pol  $\delta$  and  $\epsilon$  replicate opposite template strands at the replication fork (15). Recent studies using immunodepletion of either DNA pol  $\delta$  or  $\epsilon$  in *Xenopus* egg extracts indicate that DNA pol  $\delta$  is essential for the completion of lagging strand synthesis (16) leaving pol  $\epsilon$  as a potential leading strand polymerase.

Generation of Okazaki fragments on the lagging strand is initiated by essentially the same series of events as on the leading strand (Figure 1.1A); initiation of RNA/DNA priming by DNA pol  $\alpha$  followed by the RFC/PCNA mediated polymerase switch to DNA pol  $\delta$ , which then adds nucleotides until the 5'- end of the previous Okazaki fragment, which has to be processed. There are two models to suggest how the RNA/DNA primer synthesized pol  $\alpha$ /primase is removed during the process of Okazaki fragment maturation. One model suggests that RNase H is responsible for digestion and removal of the RNA primer (17), leaving a single ribonucleotide at the RNA/DNA junction, which is then removed by the 5'- 3' exonuclease activity of Flap EndoNuclease -1 (FEN1) (Figure 1.1B Step 6). The resulting gap is filled by DNA pol  $\delta$  and the nick is sealed by DNA ligase I. The second model involves strand displacement synthesis by pol  $\delta$  through the RNA/DNA primer of the previous Okazaki fragment (Figure 1.1C step 6) (18). The strand displacement activity is unique to DNA pol  $\delta$  and not present in pol  $\epsilon$ , further supporting a role for pol  $\delta$  in lagging strand DNA synthesis (9). This unique activity is enhanced in the presence of PCNA/RFC (18) and the displaced strand can be readily cleaved by FEN1 (9). When pol  $\delta$  encounters the previous Okazaki fragment, it performs a process called “idling”, which is successive cycles of strand displacement synthesis (2-3 nucleotides), followed by removal of the newly synthesized replacement nucleotides by the 3'-exonuclease activity of pol  $\delta$  (9). This action prevents pol  $\delta$  from larger strand displacement while keeping the polymerase at the site of the ligatable nick (9). Digestion of the displaced strand by FEN1 releases predominantly mononucleotides, indicating repeated cycles of strand displacement by pol  $\delta$  and FEN1 nucleolytic digestion (9). This process needs to proceed past the RNA-DNA junction since DNA ligase cannot catalyze a RNA-DNA nick ligation. The length of DNA displaced seems to be regulated by the DNA ligase, and the nick closure can occur as early as few nucleotide past the RNA-DNA junction (19). On contrary, pol  $\epsilon$  has very weak “idling” activity and no functional interaction with FEN1

to perform a nick-filling (9). These observations lend supports for the previously mentioned role of pol  $\epsilon$  in leading strand synthesis, while that of pol  $\sigma$  in lagging strand synthesis. In the both models of lagging strand synthesis, most of the DNA portion made by Pol  $\alpha$  is likely not excised, as follows the observation that the mutant Pol  $\alpha$  with the specific activity of the wildtype but exhibiting a lower fidelity DNA synthesis than the wildtype in vitro caused the mutator phenotype in *S. cerevisiae* (20).

### *Replication Factor C (RFC)*

As described earlier, RFC plays an essential role in tight regulation of pol  $\alpha$  activity through polymerase switching in order to promote high fidelity DNA replication by pol  $\delta$ . Replication Factor C is a multisubunit complex consisting of one large and four small subunits with similar size. The subunits of human RFC (hRFC) are p140, p40, p38, p37 and p36, all named for their migration position (in kDa) on SDS-PAGE (21). In yeast RFC (yRFC), the equivalent subunits are respectively Rfc1, Rfc4, Rfc5, Rfc2, and Rfc3 (22). Yeast genetic studies showed that all subunits of RFC are essential for viability (22). Sequence alignment of the eukaryotic and some prokaryotic RFC subunits reveals a stretch of highly conserved amino acid sequences, where the small subunits align with the central part of the largest subunit, p140 (22). These conserved regions called RFC-boxes II – VIII (Figure 1.3A) with similarity to ATP-binding/ATPase proteins (22). The most notable conservation is found in box III and box V, which are ATP-binding sites respectively called Walker A and B (23). The Walker A motif forms a structure called the P-loop specifically designed to bind the  $\beta$ - and  $\gamma$ -phosphates of a nucleotide (23), while the residues from the Walker B motif chelate  $Mg^{2+}$  for ATP hydrolysis (23;24). Box VII carries the SRC (serine-arginine-cystine) motif which is highly conserved, but only within the small subunits.



(Figure 1.2) Summary of the human and yeast RFC subunits. Human RFC subunits are labeled as p140, p40, p38, p37, and p36. Equivalent yeast subunits are indicated in the bracket. The conserved regions among the RFC subunits are indicated in boxes numbered I to VIII. See text for details. The region essential for molecular interactions are indicated with colored boxes and their interactions are described on the left. These regions were identified by deletion studies (25;26).

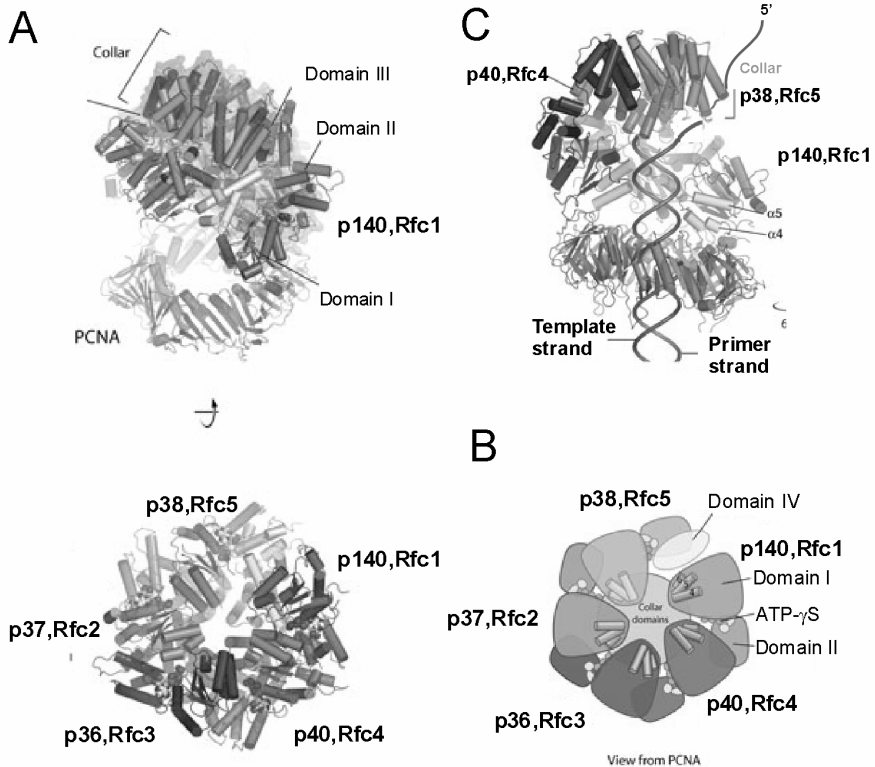
The eukaryotic p140 subunit, contains amino acid sequences in the N-terminal portion that are unique to this subunit including the Box I region of roughly 80 amino acids. Box I contains sequences of high similarity to prokaryotic  $\text{NAD}^+$  dependent DNA ligases and some eukaryotic poly (ADP-ribose) polymerases. This region was later defined as a member of the distinct class of the BRCT domain superfamily (27). Interestingly, this region has a dsDNA binding activity (28;29) which does not play a role in either RFC complex formation or clamp loader function (26;29;30). Deletion of the N-terminal fragment, including the BRCT domain, enhanced *in vitro* DNA replication ability of hRFC/PCNA mediated pol  $\delta$  by 2 – 5 times (26;29). A reconstituted yeast trRFC, with Rfc1 (rfc1- $\Delta\text{N}$ ) deleted for the coding region (including the BRCT domain) of amino acids 3-273 (approximately equivalent to amino acids 1 – 555 in human RFC p140) bears five times more replication activity *in vitro* (30) and yeast expressing the genomic mutant rfc1- $\Delta\text{N}$  are not only viable, but have no apparent replication defects (30). Isolation of the soluble form of the native yeast complex by recombinant expression was difficult, however the trRFC complex, which has equivalent activity to the wild-type, was readily purified for biochemical studies (30) and crystallization (31).

*Crystal structure of yeast RFC-PCNA complex and implications for 3' primer-template recognition*

Both native and reconstituted RFC require ATP to support PCNA-pol  $\delta$  mediated DNA replication (32;33). Furthermore, the lack of one of the subunits can not be substituted by other subunits despite the overall sequence homology. Elucidation of the trRFC structure revealed the basis for these observations (Figure 1.3). Crystallization of yeast trRFC-PCNA complex was achieved by using BRCT deleted Rfc1 and reducing ATPase activity, known to weaken trRFC-PCNA interaction, by arginine to glutamic acid substitution at the SRC motifs of each RFC subunit. Binding of PCNA and complex stability was further enhanced by the use of the non-hydrolyzable ATP analogue, ATP- $\gamma$ S (31). The resulting crystal structure of the trRFC-PCNA complex reveals that each subunit of trRFC folds into three distinct domains. Domain I and domain II are structurally conserved domains that together comprise an ATPase module of the AAA+ family (Figure 1.3A & B). This module corresponds to Boxes II –VIII in the sequence where all subunit share significant similarities (Figure 1.2) (22). AAA+ ATPases are a diverse class of oligomeric proteins (typically hexameric) that couple ATP-hydrolysis and protein-protein interactions. The third helical domain (domain III) is formed by the unique sequences at the C-terminus of each subunit (Figure 1.2). Domain III of each subunit is tightly packed against the homologous domain from the neighboring subunits to form a cylindrical structure referred to as “collar” in Figure 1.3 A (top). This structure is in accordance with a mutational study, in which the C-termini were shown to stabilize complex formation (Figure 1.2) (25). In contrast to the highly symmetric hexameric structures found in other AAA+ ATPases, the five subunits of RFC are arranged in a right handed spiral, leaving a wedge shaped gap between Rfc1 and Rfc5. The spiral arrangement allows only three subunits (Rfc1, Rfc4 and Rfc3) to interact with PCNA (Figure 1.3A, top). Each subunit contains ATP- $\gamma$ S in the ATP binding pocket which is formed by contributions from the AAA+ ATPase module (domain I and II) of one subunit and domain II of other subunit (Figure 1.3A and B) (31). Deletion of the ATP binding site belonging to any of the subunits, severely impaired DNA replication due to lack of clamp loading activity (25). The biochemical data is clearly understandable in light of the structure in which ATP- $\gamma$ S molecules are essential components that hold the spiral assembly of the AAA+ATPase modules together by anchoring inter-subunit interactions through hydrogen bonds to the phosphate groups.

The efficient operation of the polymerase switch requires RFC to bind the 3' end of the primer/template junction and load PCNA. Footprinting experiments using human RFC, have demonstrated that RFC recognizes DNA structures with a recessed 3' end and interacts with both double and single stranded DNA at the primer/template junction in a sequence independent manner (21). Other studies have shown that RFC, not only binds primer-template DNA, but also single- and double-stranded DNA (ssDNA and dsDNA respectively)(6;34;35). Despite high affinity for all three of these DNA structures, in the presence of ATP- $\gamma$ S, the RFC complex preferentially binds primer-template DNA over ssDNA and dsDNA to form a stable complex (34). This specificity of RFC for the primer/template junction can be explained by a model of DNA binding based on the trRFC-PCNA structure. The model is based on three key observations from the trRFC-PCNA complex: the screw-cap like threading of the RFC spiral onto the last turn of the DNA helix, the need to terminate the primer-template helix within the RFC spiral, and the non-specific binding of the single stranded extension of the primer/template (31). In the proposed DNA binding model of RFC (31), the primed DNA goes through PCNA and into the RFC spiral of AAA+ ATPase modules (Figure 1.3C), and bumps into the physical barrier imposed by the C-terminal collar (Figure 1.3C).

The right handed spiral arrangement of the five subunits of RFC displays roughly the same pitch as that of double stranded B-form DNA, allowing each subunit of RFC to track the minor groove of the double helix (~11 base pairs). RFC binds the negatively charged phosphate backbone of DNA using several conserved lysine and arginine residues. The primer strand of the DNA duplex, runs in the 5' to 3' direction towards the RFC spiral and terminates at the interior wall of the RFC complex. The template strand runs in the 3' to 5' direction and faces into the wedge shaped gap between Rfc1 and Rfc5, where there is sufficient room for the 5' single stranded extension to snake out of the interior complex (Figure 1.3C). In the wedge shaped gap, the unique domain IV of Rfc1 would most likely interact with the 5' single stranded extension as suggested by deletion studies in which this region of human (p140) was shown to have primer/template DNA binding activity (26). This model describing the RFC-DNA interaction was also observed in an archaea bacterial clamp-loading complex (RFC-PCNA-DNA) determined by single-particle electron microscopy (36).



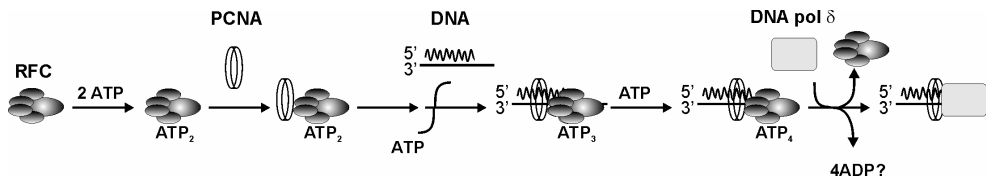
(Figure 1.3) Yeast trRFC-PCNA complex. N-terminally truncated Rfc1 was used to form the trRFC complex. (A) Spiral assembly of AAA+ ATPase modules of RFC subunits (bottom). Each AAA+ ATPase module is formed by Domain I and II (B). The ATP binding site (where ATP- $\gamma$ S is bound) is located at the subunit interface which is comprised of the Walker A and B motifs of one subunit and the SRC motif of the adjacent subunit. (C) Domain III of one subunit packs against domain II from its neighbours to form a “Collar” (Top). (C) The proposed model of primer-template recognition by RFC. The 3' end of the primer strand (orange) is physically blocked by the “Collar” of RFC while the 5' template ssDNA (green) can escape through the wedge shaped gap between p38 and p140. Figures were adopted from the original publication by Bowman et al (31).

### Clamp loading Pathway

In order to ensure highly processive synthesis by pol  $\delta$  and pol  $\epsilon$  during DNA replication, efficient loading of PCNA at the primer/template DNA is crucial. Although PCNA can be loaded onto linearized DNA by diffusion (37), efficient PCNA loading at primer/template DNA junctions and subsequent processive pol  $\delta$  synthesis are dependent on the presence of RFC (21;38). Deletions studies, which specifically interfered with the interaction between RFC and PCNA resulted in a significant reduction of *in vitro* DNA synthesis (25;26;39).

Extensive studies with the reconstituted yeast trRFC carrying an N-terminally truncated Rfc1 (40;41) have been performed to delineate the order of clamp loading. As a

clamp loader, trRFC functions as efficiently as the wild type complex (30). Productive loading of PCNA leading to DNA synthesis is an ordered process in which the complex of RFC-PCNA is preformed prior to binding of the primer/template DNA (Figure 1.4) (40;41). First, RFC binds 2 ATP molecules thereby increasing its affinity for PCNA (40). After the RFC-PCNA-2ATP complex is formed, an additional ATP molecule is bound to RFC for a total of three ATP molecules bound during the loading of PCNA onto primer/template DNA (41). However, it is not clear whether the third ATP binds to a preexisting RFC-PCNA-2ATP complex with PCNA in an open form or whether binding results in the opening of PCNA. In either case, the resulting RFC/PCNA/DNA-3ATP complex binds one more ATP (41). RFC itself has a very weak ATPase activity, which is greatly stimulated in the presence of both PCNA and primer/template DNA (40). It seems that hydrolysis of at least one of the three ATP initially bound to RFC occurs during the steps of PCNA loading. Hydrolysis of the fourth ATP is crucial for the release of RFC (40;41), which must occur in order to proceed to productive DNA synthesis by DNA pol  $\delta$  (40).



(Figure 1.4) Schematic PCNA loading by yRFC. At least four ATP are bound to the yRFC to perform a productive PCNA loading. Binding of 2ATP increases the affinity of yRFC for PCNA. Primer-template binding by yRFC-PCNA occurs in the presence of third ATP but productive loading of PCNA onto DNA occurs only when the yRFC is released upon hydrolysis of fourth and one of three previously bound ATP (40;41).

### Alternative RFC-like complexes

There is increasing evidence to suggest that the 4 small subunits form a core complex (RFC<sub>core</sub>) that can associate with a variety of larger subunits with different functions; with p140 for DNA replication, with human Rad17 (yeast Rad24) for cell cycle checkpoint control and with yeast Elg1 for genome integrity. A brief summary of alternative large subunits and the function of the complex is provided in Table 1.1. Rad17/RFC<sub>core</sub> has been shown to have in vitro clamp loading activity specific for a heterotrimeric complex of Rad9-Rad1-Hus1 (9-1-1 complex), which has structural similarity to PCNA. Rad17/RFC<sub>core</sub> and the 9-1-1 complex co-localize together in nuclear foci after the induction of DNA damage in vivo (42). It is thought that recruitment of 9-1-1 complex to a DNA lesion by Rad17/RFC<sub>core</sub> may serve as a platform for checkpoint kinases to phosphorylate other proteins that have been recruited to the damage site (reviewed in ref

(43)). The mechanism of loading of the 9-1-1 complex is similar to that of PCNA loading by RFC in that a preformed Rad17/RFC<sub>core</sub>- 9-1-1 complex is required prior to DNA binding and that it is dependent on ATP binding/hydrolysis (44). In contrast to the yeast RFC complex carrying an ATPase defective RFC1 mutation which exhibited wildtype loading activity (45), the ATPase defective yeast homologue Rad17/RFC<sub>core</sub> (Rad24/RFC<sub>core</sub>) could not interact or load the yeast counterpart Rad17-Mec3-Ddc1 clamp (46). These studies suggest that despite the overall similarity, there are however, some mechanistic differences between the loading of PCNA and the 9-1-1 (Rad17-Mec3-Ddc1) complex.

Yeast genetic studies indicate that the Elg1/ RFC<sub>core</sub> complex acts in a redundant pathway with Rad24 in DNA damage response and activation of the checkpoint kinase Rad53 (homolog of human Chk2) in the intra S phase checkpoint (47;48). Yeast lacking Elg1 exhibit increased DNA double strand breaks (DSB), which is often observed in cells with inhibited Okazaki fragment maturation due to stalled DNA replication (47;48). Accordingly, Elg1 mutants display synthetic lethality with genes involved in the repair of DSB by homologous recombination (47). It has therefore been suggested that the Elg1/ RFC<sub>core</sub> complex also takes part in genome stability by regulating replication pathways (47;48).

(Table 1.1) Alternative Clamps and clamp loaders and their cellular functions

Clamp loader	Clamp	Functions
RFC p140 (RFC1 yeast)	PCNA	DNA replication
Rad17 (yeast Rad24)	Rad9-Rad1-Hus1 (Rad17-Mec3-Ddc1)	Damage checkpoint
Yeast Elg1	?	Genome integrity/DNA replication
Yeast Ctf18, Dcc1, Ctf8	PCNA	Sister Chromatid cohesion

In addition to the alternative complexes involved in DNA repair and in checkpoints, another class of alternative RFC complexes has emerged which play an important role in sister chromatid cohesion during S-phase. Sister chromatid cohesion refers to the physical association of replicated sister chromatids mediated by a multisubunit cohesin complex and is established from the end of G<sub>1</sub> to S phase at discrete sites along the chromosome (49). Upon mitosis, chromatid cohesion is rapidly disrupted when the chromatids separate to opposite poles. Yeast genetics and in vivo co-immunoprecipitation



assays revealed that the RFC<sub>core</sub> forms a seven subunit complex with Ctf18, Dcc1 and Ctf8 linked to sister chromatid cohesions (50). Although  $\Delta ctf18$ ,  $\Delta dcc1$  and  $\Delta ctf8$  yeast strains are all viable, they exhibit chromatid cohesion defects resulting in chromosome-loss phenotype and synthetic lethal with mutant proteins involved in both establishment and maintenance of sister chromatid cohesions. In human, two alternative complexes consisting of seven subunits (Ctf18-Dcc1-Ctf8-RFC<sub>core</sub>) and five subunits (Ctf18-RFC<sub>core</sub>) were reconstituted, both of which are capable of loading PCNA however much less efficiently in comparison to the replicative RFC (51). It is not known which factor is targeted as a result of this new PCNA loading (51;52). Similarly loading activity of yeast Ctf18-Dcc1-Ctf8-RFC<sub>core</sub> is also poor and furthermore inhibited by RPA interaction via the Rfc4 of the seven subunits complex. In contrast to the weak unloading activity of replicative RFC, the yeast Ctf18-Dcc1-Ctf8-RFC<sub>core</sub> efficiently unloads PCNA from the primer-template DNA coated with RPA in ATP dependent manner (53). During the end of G<sub>1</sub> phase, at discrete sites of chromosomes, ring-like structure cohesins are loaded encircling the chromosome and appear to trap the both sister strands during the S phase inside one ring with 50 nm diameter (54). In coordinated leading- and lagging strand synthesis, the lagging strand is proposed to fold back on itself forming a protruding loop (53). The physical size of the replication fork would be larger than the cohesin diameter, therefore trapping two sister chromatids by one cohesin ring would require dissociation and rebinding of a cohesin to the replicating chromosome prior and after passage of the replication fork. Alternatively the fold back loop structure maybe collapsed allowing passage of the remaining replication fork through the cohesin complex, and this could be potentially achieved by the Ctf18-Dcc1-Ctf8-RFC<sub>core</sub> unloading PCNA, which forms a structural organization of the fork (53).

### *Regulation of RFC*

DNA replication is carried out by a network of enzymes and proteins which work together to ensure the accurate duplication of the genetic material. It is therefore crucial to regulate precisely the activity of these proteins during specific stages of the cell cycle. Regulation of many replication proteins is achieved through specific phosphorylation by cyclin dependent kinases (Cdks), which are a family of regulatory kinases that control transition from one phase to the next in the cell cycle. Cdks become active upon binding proteins called cyclins (type A, B, D and E). At the G<sub>1</sub>/S transition, mitogen activated cdk4-cyclin D and cdk6-cyclin D complexes phosphorylate the retinoblastoma protein (RB),

which in turn dissociates from complexes with members of the E2F family of transcription factors allowing transcription of genes required in S phase (55). Similarly, phosphorylation of RFC by regulatory kinases appears to regulate the activity of RFC at specific phases of the cell cycle (Table 1.2), despite its constant high-level of expression throughout G<sub>1</sub>, S and G<sub>2</sub> phases (56). In addition to its role in replication, PCNA appears to act as a platform for regulatory proteins as shown by a quaternary complex formed with the kinase inhibitor p21, the cyclin-dependent kinases (cdks) and the cyclins *in vivo* (57). Although the physiological significance is not yet well understood, the studies indicate that the phosphorylation of RFC p140 by cdk kinase and CaMKII regulate activity of RFC in DNA replication by destabilizing the RFC complex or the PCNA-RFC interaction in a cell cycle-dependent manner (Table 1.2).

(Table 1.2) Regulatory kinases and phosphorylation of p140 subunit of RFC during specific phases of the cell cycle.					
Regulatory kinase	phosphorylation site			Effects of phosphorylation	Cell cycle ref
cdk2-cyclin A	PCNA binding domain of p140			???	G <sub>1</sub> /S (58)
CaMKII	PCNA binding domain of p140			Inhibits RFC-PCNA interaction <i>in vitro</i>	S/G <sub>2</sub> (59)
unknown	Thr406 of p140			Inhibits RFC-PCNA interaction <i>in vitro</i>	S (60)
Cdc2-cyclin B	p140			Dissociation of RFC complex	G <sub>2</sub> /M (61)

Cell cycle specific regulation of RFC activity in DNA replication appears also to be regulated by direct interaction with various other proteins. The bromo domain of Brd4, which belongs to the BET family of nuclear proteins, binds to RFC p140 thereby interacting with the entire RFC complex (62). The result of this interaction inhibits *in vitro* RFC-dependent DNA replication and *in vivo* entry into S phase (62). Similarly, involvement of RFC in a cell-cycle checkpoint pathway is implied by the interaction between RFC and Rb observed in co-immunoprecipitation experiments, which has an important role in the promotion of DNA repair. Involvement of RFC in Rb mediated mammalian cell survival was shown where cells expressing exogenously introduced RFC p140 subunit had increased survival after UV induced DNA damage. The protein-protein interaction was shown to be dependent on the LxCxE motif of p140 (63). This suggests a potential role of RFC in Rb-mediated damage checkpoint control (at G<sub>1</sub>) which arrests the cell cycle prior to S phase in order to promote DNA repair and cell survival. Alternatively, as mentioned earlier, phosphorylation of Rb by cdk2-cyclin D and cdk4- cyclinD serves to inactivate Rb allowing cells to enter S phase.

*Uncharacterized functions associated with the 5' dsDNA binding of the BRCT domain of RFC p140*

Early deletion studies of each subunit of RFC revealed that the large subunit of RFC is capable of binding various types of DNA (26;28;34;64-66). *In vitro* studies of mammalian (26;28;29;65) and insect RFC p140 (64) revealed that the N-terminal region including the BRCT domain binds dsDNA in a nonsequence specific manner (26;28;65). This interaction is strictly dependent on the presence of a 5' phosphate (64). Binding to this substrate was not competed by 5' phosphorylated single strand DNA or 3' recessed dsDNA (64), indicating that binding is structure specific. In contrast to the 3' specific primer/template binding observed, 5' end binding by p140 does not contribute to DNA replication (26;30;34). So far no definite physiological function has been assigned to this 5' phosphorylation dependent dsDNA binding.

One possible function of 5' phosphate dsDNA binding has been suggested by a recent *in vitro* study on the involvement of RFC in PCNA modulated human exonuclease 1 activity in mismatch repair. Mismatches are caused by nucleotide miss-incorporations and DNA slippage errors during DNA replication and are corrected by the mismatch repair (MMR) machinery. The mechanism of MMR is thought to be similar in both eukaryotes and prokaryotes. MMR is initiated by the binding of MutS $\alpha$  to the mismatch and subsequent recruitment of MutL $\alpha$  (reviewed in (67)). In *E. coli*, the MutH endonuclease is then activated, creating a nick in the newly synthesized strand either upstream or downstream of the mismatch. A specific exonuclease then excises the intervening DNA to remove the mismatch (67). In human MMR, excision is performed by EXO1, which is a single strand DNA exonuclease with bi-directional hydrolysis activities from a single strand nick to the site of mismatch. In the presence of MutS and RPA, EXO1 excises 5'-3' from the nick. Excision in the 3' to 5' direction by EXO1 to the nearest mismatch requires the addition of PCNA and RFC, which serve to suppress inappropriate 5' to 3' excision away from the mismatch. RFC is essential for PCNA loading which is required 3' to 5' excision by EXO1 (68). Discrimination between nascent and template DNA strands by EXO1 is thought to be mediated by the specific orientation of loaded PCNA which interacts strongly with EXO1 (68). Meanwhile the suppression of 5' to 3' hydrolysis activity of EXO1 depends on the integrity of the BRCT domain of RFC (68). A possible explanation is that binding of the BRCT domain to the 5' end (64) is responsible for suppression of unproductive 5' to 3' hydrolysis by EXO1 when the nick is located 3' to the mismatched pair (68). As mentioned before, yeast expressing the genomic mutant *rfc1- $\Delta$ N* lacking the

BRCT domain are not only viable but also show no apparent phenotype other than a slight sensitivity to the DNA alkylating agent methyl methanesulfonate (MMS) (30), thus BRCT dependent suppression of 5' to 3' hydrolysis of EXOI may not be the most important in vivo role of RFC. Further investigation to identify cellular role of the RFC BRCT domain is clearly warranted.

---

*BRCT domain and RFC*

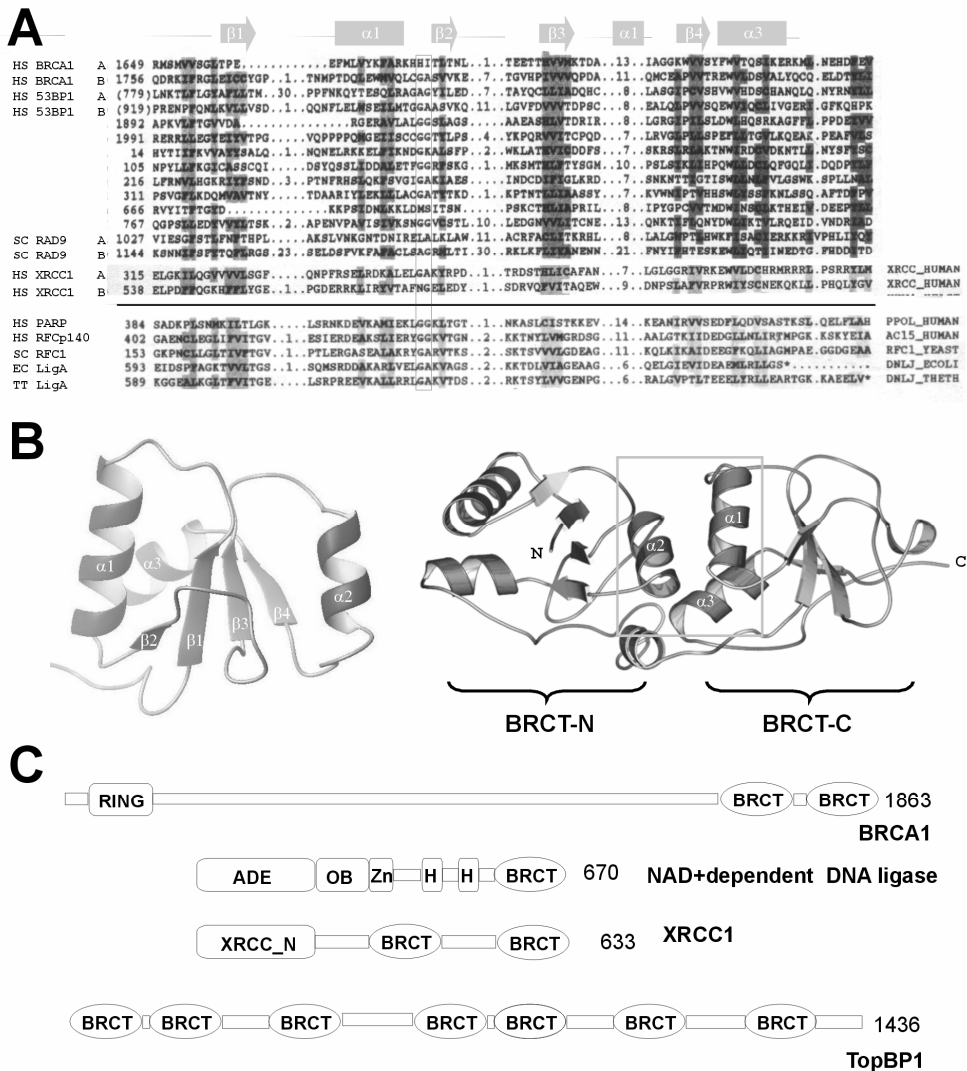
---

*The BRCT domain super family*

The region of RFC p140 which binds 5' dsDNA contains sequences that are related to BRCT domains. The BRCT domain (**BRCA1 C-Terminus**) was first identified as a tandem repeat of roughly 90 amino acids at the C- terminus of the Brcal (Breast Cancer susceptibility 1) protein (69). Extensive amino acid sequence profiling led to the discovery of a vast number of proteins (currently 915 open reading frames deposited in Pfam) carrying BRCT domains (27;70), and strikingly most of those characterized are either directly or indirectly associated with various aspects of DNA metabolism; including DNA repair, DNA replication or cell cycle-checkpoint regulation. Apart from the protein TopBP1, which consists solely of BRCT domains, most BRCT domains are found in large, multidomain proteins carrying other functional domains i.e. DNA ligase IV, RFC, etc. As represented in the scheme of Figure 1.5C, BRCT domains can be categorized as single, multiple, and tandem pairs. In general, tandem pairs of BRCT domains are separated by a short inter-domain linker (of roughly 20 amino acids) and form one structural unit (71;72). In contrast, multiple copies of BRCTs within one protein have variable but larger separation and often function independently. Its small size and distribution in multidomain proteins strongly suggest that the BRCT domain may function in protein-protein interactions for cellular signal transduction linking components essential for DNA metabolisms.

While some BRCTs form distinct sub-families with significant similarities (> 25 %) (Figure 1.5A), the overall sequence similarity amongst the superfamily is very low (average identity 17 % Pfam) implying diverse functions (27;70). For example, the BRCT domain of eukaryotic RFC belongs to a distinct subclass of the BRCT family, which is shared amongst the bacterial NAD<sup>+</sup> dependent DNA ligases and eukaryotic PARP (Poly ADP-Ribose Polymerase) (Figure 1.5A). In contrast to the low sequence similarity, the secondary structure of BRCTs is well conserved (Figure 1.5A)(27;70). The three-dimensional structure determination of BRCT domains from XRCC1(73), BRCA1(72), 53BP1(71;74), DNA ligase III (75) and NAD<sup>+</sup> dependent DNA ligase (76) confirmed conservation of the same overall protein fold; a compact architecture composed of four parallel  $\beta$  - strands forming the core, flanked on one side by two  $\alpha$  - helices and on the other by a single  $\alpha$  - helix. The architecture of the  $\beta$ -sheet and helices  $\alpha$ 1/ $\alpha$ 3 is maintained by the packing of a limited number of conserved hydrophobic residues in the core of the

BRCT fold. The structures of BRCT domains, which are discussed in this chapter, are summarized on Table 1.3.



(Figure 1.5) **(A)** Amino acid sequence alignment of BRCT domains. HS (*Homo sapiens*), Conserved residues are shaded in dark. The conserved GG repeat is highlighted in a box. SC (*Saccharomyces cerevisiae*), EC (*Escherichia coli*) and TT (*Thermus thermophilus*). Predicted secondary structures are indicated on the top of the alignment. **(B)** Ribbon representation of the XRCC1 C-terminal BRCT (left) and the tandem BRCT domains of BRCA1 (right). Each BRCT unit is consists of a  $\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4\alpha 3$  topology. The interface between the tandem BRCT domains (-N and -C denoting N- and C-terminus) is highlighted with a box (right). **(C)** Domain architectures of BRCT family proteins. BRCA1 (BRCA1\_HUMAN) contains tandem BRCT domains at the C-terminus and RING domain at N-terminus. NAD<sup>+</sup> dependent DNA ligase (DNLJ\_THFE) contains single BRCT domain, ADE (adenylation domain), OB-fold, Zn-finger domain, and Helix-loop-helix. XRCC1 (XRCC1\_BRCT) contains two BRCT domains and N-terminal DNA binding domain (XRCC1\_N). TOPBP1 (Q7LGC1\_HUMAN) carries seven BRCT domains.

A notable feature of BRCT domains is that each secondary element is generally connected by long flexible loops. The most conserved sequence element, a glycine repeat (Figure 1.5B), forms a tight turn between the  $\alpha 1$  - helix and the  $\beta 2$  - strand. This turn is structurally important as substitution of glycine by large bulky residues has been shown to result in proteolytic sensitivity in BRCA1(72). In contrast to the conserved regions, the  $\alpha 2$  helix and the preceding loop are the least conserved in terms of size and amino acid composition, such a local structural variability may reflect differences in their biological functions of each protein in the list (Table 1.4).

(Table 1.3) Structures of BRCT domains

PDB code	Brief Descriptions	Type	Method	Ref.
1CDZ	XRCC1	Single	X-ray	(73)
1IMO	DNA ligase III	Single	NMR	(75)
1L7B/1DGS	NAD <sup>+</sup> dependent ligase	Single	NMR/X-ray	NP/(76)
1T29/1T2V/1T15	BRCA1 complex with phosphoserine peptide	Tandem	X-ray	(77-79)
1KZY/1LOB	53BP complex with p53	Tandem	X-ray	(71;74)
1JNX	BRCA1	Tandem	X-ray	(72)

NP = not published

Although single isolated BRCT domains from XRCC1, DNA ligase III and NAD<sup>+</sup> dependent DNA ligase can exist as a stably folded unit, many BRCT domains are found to fold in tandem pairs (71;74). A close inspection of the first tandem BRCT structure to be elucidated, the BRCT domains from BRCA1, shows that the two domains interact in a head-tail orientation. The interaction of the two domains is stabilized by hydrophobic interactions between the  $\alpha 2$  – helix of the N-terminal domain and the  $\alpha 1$  –  $\alpha 3$  – helices of C-terminal domain (Figure 1.5B, right indicated in a box). Essentially identical packing is also observed in the tandem BRCT repeats of 53BP1, where the inter-domain interactions are again mediated by conserved hydrophobic residues from the  $\alpha 1$  – and  $\alpha 3$  – helices. These two structures suggest that this kind of intra-molecular packing is common among the tandem BRCTs. The only notable structural difference is found in the inter-domain linkers, which folds into loop-helix-loop in BRCA1 and in contrast forms a  $\beta$ -ribbon structure in 53BP1 (71;72).

(Table 1.4) Examples of BRCT interactions					
Pairs	BRCT carrier	Partner	Biological function	Description of specific Interactions known	Ref.
BRCT- non BRCT	BRCA1 (tandem)	BACH1	G2/M check point	BACH1 S990 phosphorylation dependent	(77;78;80-83)
	S. p. Rad4 (tandem)	Rad9 (PCNA like Rad9-Rad1-Hus1)	Damage response/ intra-S phase checkpoint	Rad9 T412 phosphorylation dependent	
				Likely similar to S. p. Rad4 – Rad9 interaction	(84-86)
	S. c. Dpb11	Ddc1 (PCNA like Rad17-Mec3-Ddc1)	Damage response response/c intra S-phase checkpoint	Likely similar to S. p. Rad4 – Rad9 interaction	
	MDC1 DNA ligase IV BIRD1	Potential targets in Peptide library		Target serine phosphorylation dependent	(82)
	S. c. RAD9 (tandem)				
	DNA ligase IV (tandem)	XRCC4	NHEJ	The inter-domain linker interact with XRCC1	(87;88)
	BRCA1 (tandem)	Acetyl-CoA Carboxylase	Fatty acid metabolism		(89)
	BRCA1	TRAP220 CtIP LMO4	Transcription regulation		(90-93)
	TopBP1	E2F1	DNA repair/checkpoint		(94)
BRCT- BRCT heterodimer	53BP1 (tandem)	p53	DNA repair	The inter-domain linker interact with p53	(71;74)
	BRCT carrier	BRCT partner			
	XRCC1 (single)	PARP (single)	Strand break repair	Zn Finger/BRCT domain of PARP with N-terminus	(95;96)
	XRCC1 (single)	DNA ligase III (single)	Base Excision repair/ strand break repair	Hydrophobic and salt-bridge interaction between the $\alpha 1$ - $\alpha 1$ .	(97;98)
	S. p. Ctrb2	N.A	Damage response	unknown	(99;100)
BRCT- BRCT homodimer	S. c. Rad9				

### *BRCT domains as a protein-protein interaction module and their cellular roles*

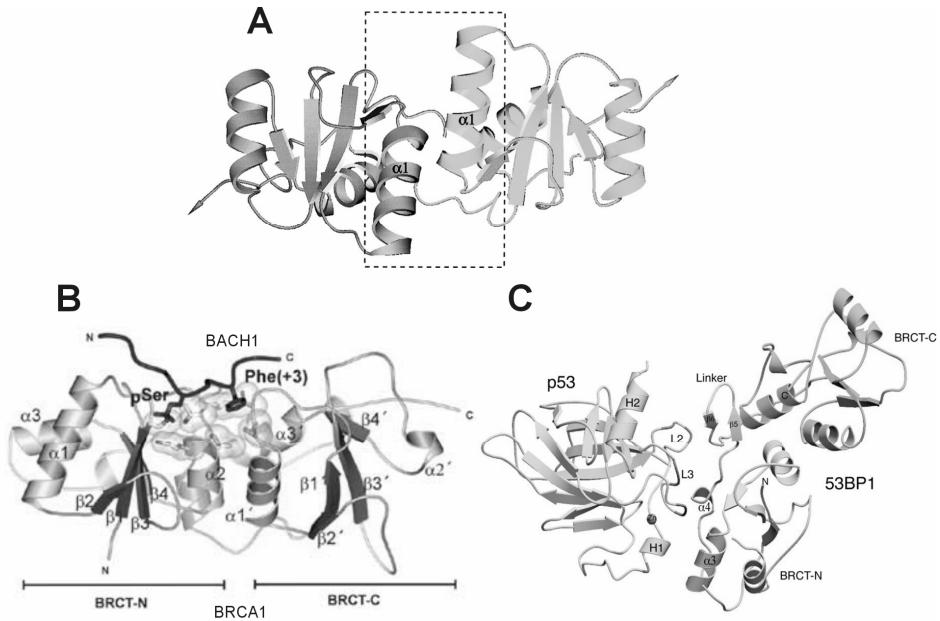
Several biochemical and genetic studies indicate that BRCT domains are important mediators of protein – protein interactions, which can be separated into either BRCT - BRCT or BRCT – non BRCT pairs (Table 1.2). The following examples of BRCT mediated interactions illustrate that although the individual BRCT units have the same fold, the ways in which they execute their function differs from one BRCT to the other.



Human XRCC1, which has no catalytic activity, is known to act as a scaffold protein that anchors DNA ligase III, DNA pol  $\beta$  and PARP (poly ADP-ribose polymerase) to the site of damage during Base Excision Repair (BER). Defective XRCC1 results in an increased frequency of single strand breaks which are formed as intermediates during base excision repair. XRCC1 contains two BRCTs, of which the amino-terminal BRCT is used to interact with the Zn finger and BRCT domains of PARP (95). The XRCC1-PARP interaction down-regulates the activity of PARP, which modifies nuclear proteins involved in chromatin architecture as a result of DNA damage. The carboxyl-terminal BRCT of XRCC1 interacts with the BRCT domain of DNA ligase III (97) and this interaction is essential for single strand break repair during the  $G_1$  phase of the cell cycle (101). The role of the carboxyl-terminal BRCT in protein-protein interaction was first implied by the crystal structure of XRCC1 BRCT, in which the two BRCT monomers are arranged in a 2 fold axis of symmetry and interact through both hydrophobic and salt-bridge interactions between N-termini and the  $\alpha 1$  -  $\alpha 1$  helices (Figure 1.6 A) (73). The residues involved in the homodimer interface of XRCC1 BRCT are the most conserved between the heterodimer BRCT partners XRCC1 and DNA ligase III in comparison to the rest of their sequences suggesting that the homodimer interaction might be a relevant model for the heterodimer interaction (97). In light of this observation, the conserved amino acids in the N-termini and  $\alpha 1$  – helices substituted providing further support for the idea that the interface involved in the homodimer also mediates heterodimer formation between XRCC1 and DNA ligase BRCTs (73;97). A notable difference between BRCT dimers formed by two isolated BRCTs and tandem BRCT pairs is that the former dimerize via the N-termini and  $\alpha 1$  helices of the two domains (Figure 1.6A) while the latter pairs together by interactions between  $\alpha 2$  of the N-terminal BRCT and the  $\alpha 1/\alpha 3$  of the C-terminal BRCT (Figure 1.5B, right).

The BRCT domains of *S. cerevisiae* Rad9 provide yet a further example of homotypic interactions. Homo dimerisation has been shown to activate Rad53, a kinase required for cell cycle arrest in response to DNA damage. Rad9 is hyperphosphorylated in a normal cell, and forms a BRCT-dependent homo-dimer upon further phosphorylation by Mec1/Ddc2, a DNA damage sensing complex (100;102). Rad53 (103) binds specifically to the phosphorylated Ser residues of each dimeric Rad9 bringing a pair of Rad53 together for trans-autophosphorylation. This phosphorylation of Rad53 results in activation and release of Rad53 for further regulation of the downstream checkpoint pathway (104). Although

the molecular mechanism of Rad9 BRCT dimer formation is not yet understood, its function is clear.



(Figure 1.6) BRCT domain interactions. **(A)** BRCT-BRCT interaction. Non-crystallographic dimer of the XRCC1 BRCT domain (1CDZ). The dimer-interface is created by the N-termini and  $\alpha 1$  helices of the BRCT domains (73). **(B)** Tandem BRCTs – PhosphoSerine (pS) peptide interaction. The pS peptide (in blue) resembling BACH1 is bound by the tandem BRCT domains from BRCA1 (1T2V) (105). Phospho-moiety (pS) is bound by the BRCT-N while phenolalanine (Phe) is accommodated in the hydrophobic groove between the two BRCTs **(C)** The interaction between inter-domain linker (green) of 53BP1 and DNA binding domain of p53 (1KZY) (71).

It is clear that protein phosphorylation by protein kinases may generate docking sites for other proteins allowing the assembly of signaling complexes in response to kinase activation. Relevant to our studies, the binding of BRCA1 to BACH1 (BRCA1 associated Carboxyl-terminal Helicase), which is mediated by the tandem BRCT repeat of BRCA1, is dependent on the phosphorylation state of BACH1 (81-83). This interaction is cell cycle-regulated and plays a critical role in the maintenance of the G2/M checkpoint by arresting the cell cycle until DNA repair is completed. Cells bearing a truncated BRCA1 that lacks the BRCT domains are deficient for this checkpoint (83). Further analysis of the BRCA1-BACH1 interaction demonstrated that the BRCT domains specifically recognize phosphorylated Ser990 (ISRSTS<sup>990</sup>PTFNK) of BACH1 and the aa position +2 relative to Ser990 to yield a binding motif of pSer-x-x-Phe (where x can be any amino acid)(81;106). Additional tandem BRCT domains including BIRD1, yeast Rad9, DNA ligase IV and MDC1 as well as single BRCT domains from Fcp1, TopBP1 (BRCT6), TDT, REV1 and

DNA ligase III have been identified as binding phosphoserine-peptides (83). In contrast to single domains such as SH2 or FHA, an isolated BRCT domain from the tandem repeat is not sufficient for phosphoserine peptide binding (81;83). The reason for this notable difference became clear when three crystallographic studies described how the recognition of a phospho-peptide is achieved by the BRCA1 tandem BRCT domains (77-79). In the crystal structure of the phospho-peptide complex, the conformation of the BRCT is unperturbed from that of the native form (Figure 1.6B). The phosphate-moiety of pSer990 is bound by network of hydrogen bonds to three residues in the N – terminal BRCT (BRCT-n in Figure 1.6B). Meanwhile Phe993 of BACH1 is bound in a hydrophobic groove, which is created by the interface between the tandem BRCTs revealing the essence of the tandem domain for pS peptide binding. Interestingly, the residues involved in pSer binding are conserved amongst the tandem BRCT domains that have been identified as binding to phospho-peptides suggesting a conserved recognition mechanism(79;105;106). A variation on phosphorylation dependent peptide binding has been noted for the tandem BRCT domains at the C-terminus of *Schizosaccharomyces pombe* Rad4 (TopBP1 like). In this model, the BRCT domains of Rad4 specifically recognize a phosphothreonine residue of the PCNA-like complex (Rad9- Rad1-Hus1) (84). Sequence comparison between BRCA1 and Rad4 indicates that the residues involved in binding to the phosphate moiety in BRCA1 are not shared by Rad4, which may reflect the specificity for pT over pS. It will therefore be interesting to see how the pT peptide is recognized by the tandem BRCTs from Rad4.

In addition to the BRCT domains themselves, the inter-domain linker can play a direct role in protein – protein interactions. 53BP1 is a multidomain protein that, like BRCA1, carries a tandem BRCT repeat at the C-terminus. The tandem BRCTs bind the DNA binding domain of p53 to enhance p53 mediated transcriptional activation (71;74). The crystal structure of the heterodimer between p53 and 53BP1 revealed that the tandem BRCT domains of 53BP1 pack together in essentially the same way as described for BRCA1, except that the inter-domain linker adopts a somewhat more complicated structure that includes a  $\beta$  - hairpin (Figure 1.6C.) (71;74). The interaction with the DNA binding domain of p53 is mediated by residues in the inter-domain linker and in the  $\alpha 3$  - helix of the N - terminal BRCT (Figure 1.6C) (71;74). An even more extreme example of this type is provided by the complex formed between the inter-domain linker of DNA ligase IV and the protein Xrcc4. In this case the isolated inter-domain linker from DNA ligase IV forms a stable complex with the coiled-coil domain of Xrcc4 even in the absence of any BRCT

domains (88). However this particular linker is slightly longer than those found in the BRCA1 and 53BP1, and the BRCT repeats of DNA ligase IV has not been shown to form a tandem unit as seen in the BRCA1 and 53BP1.

### *The BRCT as a DNA binding module*

Although a large number of genetic and biochemical studies indicate that the primary role of BRCT domains is in protein-protein interactions, there is growing evidence to suggest that some BRCT domains are involved in DNA recognition. For example, the BRCT domain from the bacterial NAD<sup>+</sup> dependent DNA ligases has been implicated in DNA binding ((107-111). DNA ligases are essential components of DNA replication, repair and recombination and catalyze the phosphodiester bond formation of single stranded nicks in double stranded DNA. Ligases can be classified into two categories depending on their requirement for NAD<sup>+</sup> or ATP. NAD<sup>+</sup> dependent DNA ligases are found in eubacteria. The ligation reaction proceeds in three steps. First, adenylation of the ligase occurs via the adenylate moiety of NAD<sup>+</sup>. Second, the adenylate moiety is transferred to the 5'-terminal phosphate of the nick. Third, the phosphodiester bond is formed via nucleophilic attack of the 3' hydroxyl terminus on the other side of the nick. In a recent study of the mechanism of the bacterial NAD<sup>+</sup> dependent ligase, deletion or mutation of the BRCT domain resulted in reduced nick binding (109-111), and in a severe reduction of the adenylate-moiety transfer to the 5'-terminal phosphate, while adenylation of the ligase itself was not affected (110). The authors conclude that the loss of stable nick binding reduces the subsequent adenyl transfer reaction.

The best documented case of DNA binding is by the BRCT domain of RFC p140. A study of insect p140 BRCT domain revealed that it binds specifically to the 5' phosphorylated end of dsDNA. The BRCT domains from both RFC p140 and the group of NAD<sup>+</sup> dependent DNA ligase belong to the distinct class of the BRCT superfamily and share significant amino acid homology (> 30 %) (27). The conserved residues within the BRCT domain that affect DNA binding of NAD<sup>+</sup> dependent DNA ligase (110) are also found in RFC p140. Although it has not been shown yet, it seems logical that the 5' phosphate dsDNA binding function of RFC p140 (64) is mechanistically similar to the 5' end nick recognition by the BRCT domain of the NAD<sup>+</sup> dependent DNA ligases.

DNA binding remains a specialized function of a subset of members of the BRCT superfamily. Beyond a potential regulatory role in mismatch repair, nothing is known about the function of the conserved BRCT region of RFC p140. Furthermore, the structure

specific, 5' phosphate dsDNA binding by members of this BRCT class is unprecedented and, therefore, worthy of further investigation. Since the 3D structure of a BRCT-DNA complex is currently available, we set out to characterize this unique interaction using biochemical and structural analysis. Understanding the mechanism of DNA recognition could help to identify more members of BRCT family with potential DNA binding functions.

---

#### NMR as a tool for Structure determination

---

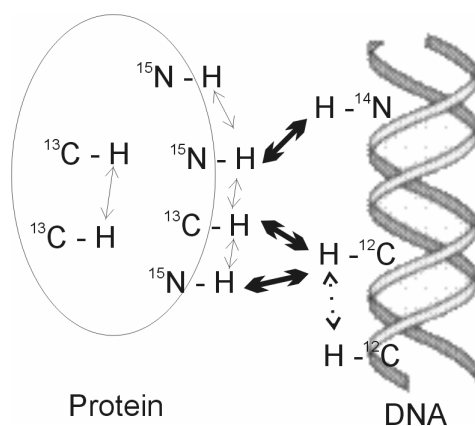
At present X-ray (or neutron) diffraction and Nuclear Magnetic Resonance spectroscopy are the only means to determine the atomic resolution structure of biomacromolecules. For now, NMR can not compete with the accomplishments of X-ray diffraction in the structure determination of supramolecular assemblies such as the ribosome and proteosome. However, often proteins may not crystallize or co-crystallize as a complex due to the dynamic nature of the interaction. NMR can not only be used to study the structure of a protein or a complex in solution, but also to derive information on dynamics aspects of a molecule and between interacting molecules, providing additional parameters such as binding constants.

The method of 3D structure elucidation of biomacromolecules is still performed essentially as developed by Wüthrich and coworkers (112). In this process three distinct stages can be defined. First, the sequential assignment of the resonances is performed. Typically this involves labeling the molecule with the stable isotopes  $^{15}\text{N}$  and  $^{13}\text{C}$ . Triple resonance, through bond NMR experiments are recorded to provide correlations between the amino acids of a protein that allow the resonances from each to be assigned. In the second stage, NOESY spectra are recorded to generate an interaction map between pairs of protons in the molecule. The peaks in the NOESY spectrum are integrated and assigned to generate pair-wise distance restraints (112). The NOE measures interproton distances with a sensitivity that falls off with  $r^{-6}$  where  $r$  is the distance between the protons. As a result, the best modern NMR spectrometers are capable of determining an upper distance limit of approximately 6 Å. (112). In the third stage, one generates a list of structural restraints, which may include interproton distances, ranges of dihedral angles and more recently, orientation constraints derived from residual dipolar couplings. Starting from a randomized conformation of the peptide (or nucleic acid) the structure is folded in silico using, typically, a simulated annealing protocol. A potential describing energy penalties for

restraint violations is used to drive the simulated annealing procedure. Typically a number of conformers are calculated that approximately equally well satisfy the experimentally derived restraints, and used to represent the final 3D structure.

#### *Studying intermolecular interaction by NMR*

The most unambiguous way to determine a full three-dimensional structure of a complex is to use distance information between the interacting molecules derived from intermolecular NOEs. This method is applicable generally when the interaction between two molecules is relatively tight ( $K_d \leq 10^{-5}M$ ) but also when the exchange dynamics are appropriate to allow the build up of intermolecular NOEs. Isotope-editing and filtering are the commonly used technique to discriminate intermolecular NOEs that arise between interacting molecules from those that rise from within one of the components of a complex (113). This technique requires the two components of the complex to have different isotopic labeling. Typically a DNA binding protein is isotope labeled (with  $^{13}C$  or  $^{15}N$ , or both) while the target DNA contains  $^{12}C$  and  $^{14}N$  at natural abundance (Figure 1.7). NOE correlations within the labeled molecule (protein) can be selectively observed using isotope-edited experiments (in thin arrows), while correlations within the unlabeled molecule (DNA) can be selectively observed by filtering out  $^{13}C$  or  $^{15}N$  attached protons (in dashed arrow). Intermolecular NOEs between the protein and the DNA may be selectively observed using experiments that are isotope filtered with respect to one proton dimension and isotope-edited with respect to the other (in thick arrows). Although technically highly demanding, this method has been successfully applied to a number of DNA-protein complexes (114) and protein-protein complexes (115).



(Figure 1.7) Observable NOEs in a complex between isotopically labeled protein and unlabeled DNA. Thin arrows indicate observable intramolecular NOE in isotope edited NOE experiment. Dashed arrow indicates observable intramolecular NOE in DNA by isotope-filter experiments. The thick arrows indicate the intermolecular NOE may be identified (113).

A second widely used NMR method for studying inter-molecular interactions is referred to as the chemical shift perturbation method. Complex formation changes the local electronic environment of protons at the interaction surface which can be monitored by observing changes in the chemical shift of these protons. The chemical shift change is generally observed by heteronuclear correlation spectra such as the [ $^{15}\text{N}$ ,  $^1\text{N}$ ]-HSQC. This experiment monitors primarily backbone amides of a labeled protein upon titration with an unlabeled partner allowing one to follow the resonances to the “bound” position. This method generally works well with molecular interaction with modest affinity ( $10^{-5}\text{M}$ ) where free and bound forms are in fast exchange (116). Chemical shift perturbation analysis allows one to define the molecular surface of the isotope-labeled protein involved in interaction with the unlabeled partner (117). Molecular interactions of high affinity ( $K_d \leq 10^{-5}\text{M}$ ) in slow exchange exhibit one set of resonances for free protein and one set for the bound protein. During the titration, the set of resonances belonging to the free will disappear and the new set belonging to the bound replaces. If the interaction does not the chemical environment around the interacting protein too much, the majority of the resonances of the two sets will overlap and the differences will therefore make up the interaction interface. In the case of molecular interactions in intermediate exchange, the changing resonance frequency becomes poorly defined resulting in line-broadening and often broad enough to disappear during the titration (117).

---

## Outline of the thesis

---

In Chapter 2, the specificity of DNA binding by the BRCT region of RFC p140 is discussed. Several gene fragments encoding the region including the BRCT domain of p140, were generated, and the protein products were tested for DNA binding activity using various DNA structures. Surprisingly, a polypeptide comprising only the conserved BRCT domain (403-480 a.a) did not bind dsDNA but required an additional 28 amino acids at the N-terminus. DNA binding was non-sequence specific but 5' phosphate dependent as reported earlier. When the C-terminus of the protein is extended to residue 545, the peptide (375-545 a.a) binds dsDNA in 5' phosphate independent manner.

Chapter 3 describes the results of mutagenesis studies on the DNA binding activity of the BRCT region of RFC p140 (375-480). Conserved amino acid residues on the surface of the protein were substituted by residues that changed the electrostatic potential.

Mutations affecting DNA binding were localized on one molecular surface of the BRCT domain of p140.

In Chapter 4, the methodology used to collect the NMR data on isotope labeled protein bound to DNA is described and the chemical shifts assignment of the protein is reported. The secondary structure of the BRCT domain (403 - 480) predicted from the NMR data was in good agreement with homologous proteins with known structures. The data also indicates that there is an extra  $\alpha$  - helix near the N – terminus.

Chapter 5 describes the structure of the BRCT region of human RFC p140 which was calculated based on the NOE-derived distance restraints. A surprising resemblance to the structure of the phospho-peptide binding of the BRCA1 BRCT domains was found. A model of the DNA-protein complex was generated based on the mutation data, intermolecular NOEs and residue conservation with the phospho-peptide bound structure.

---

## Reference list

---

1. Watson, J. D. and Crick, F. H. (1953) *Nature* **171**, 964-967
2. Watson, J. D. and Crick, F. H. (1953) *Nature* **171**, 737-738
3. Laskey, R. A. and Madine, M. A. (2003) *EMBO Rep.* **4**, 26-30
4. Waga, S. and Stillman, B. (1994) *Nature* **369**, 207-212
5. Maga, G., Stucki, M., Spadari, S., and Hubscher, U. (2000) *J.Mol.Biol.* **295**, 791-801
6. Mossi, R., Keller, R. C., Ferrari, E., and Hubscher, U. (2000) *J.Mol.Biol.* **295**, 803-814
7. Waga, S., Bauer, G., and Stillman, B. (1994) *J.Biol.Chem.* **269**, 10923-10934
8. Maga, G., Villani, G., Tillement, V., Stucki, M., Locatelli, G. A., Frouin, I., Spadari, S., and Hubscher, U. (2001) *Proc.Natl.Acad.Sci.U.S.A* **98**, 14298-14303
9. Garg, P., Stith, C. M., Sabouri, N., Johansson, E., and Burgers, P. M. (2004) *Genes Dev.* **18**, 2764-2773
10. Krishna, T. S., Kong, X. P., Gary, S., Burgers, P. M., and Kuriyan, J. (1994) *Cell* **79**, 1233-1243
11. Krishna, T. S., Fenyo, D., Kong, X. P., Gary, S., Chait, B. T., Burgers, P., and Kuriyan, J. (1994) *J.Mol.Biol.* **241**, 265-268
12. Kuriyan, J. and O'Donnell, M. (1993) *J.Mol.Biol.* **234**, 915-925
13. Masumoto, H., Sugino, A., and Araki, H. (2000) *Mol.Cell Biol.* **20**, 2809-2817
14. Araki, H., Leem, S. H., Phongdara, A., and Sugino, A. (1995) *Proc.Natl.Acad.Sci.U.S.A* **92**, 11791-11795



15. Karthikeyan, R., Vonarx, E. J., Straffon, A. F., Simon, M., Faye, G., and Kunz, B. A. (2000) *J.Mol.Biol.* **299**, 405-419
16. Fukui, T., Yamauchi, K., Muroya, T., Akiyama, M., Maki, H., Sugino, A., and Waga, S. (2004) *Genes Cells* **9**, 179-191
17. Murante, R. S., Henricksen, L. A., and Bambara, R. A. (1998) *Proc.Natl.Acad.Sci.U.S.A* **95**, 2244-2249
18. Maga, G., Villani, G., Tillement, V., Stucki, M., Locatelli, G. A., Frouin, I., Spadari, S., and Hubscher, U. (2001) *Proc.Natl.Acad.Sci.U.S.A* **98**, 14298-14303
19. Ayyagari, R., Gomes, X. V., Gordenin, D. A., and Burgers, P. M. (2003) *J.Biol.Chem.* **278**, 1618-1625
20. Niimi, A., Limsirichaikul, S., Yoshida, S., Iwai, S., Masutani, C., Hanaoka, F., Kool, E. T., Nishiyama, Y., and Suzuki, M. (2004) *Mol.Cell Biol.* **24**, 2734-2746
21. Tsurimoto, T. and Stillman, B. (1991) *J.Biol.Chem.* **266**, 1950-1960
22. Cullmann, G., Fien, K., Kobayashi, R., and Stillman, B. (1995) *Mol.Cell Biol.* **15**, 4661-4671
23. Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982) *EMBO J.* **1**, 945-951
24. Fry, D. C., Kuby, S. A., and Mildvan, A. S. (1986) *Proc.Natl.Acad.Sci.U.S.A* **83**, 907-911
25. Uhlmann, F., Gibbs, E., Cai, J., O'Donnell, M., and Hurwitz, J. (1997) *J.Biol.Chem.* **272**, 10065-10071
26. Uhlmann, F., Cai, J., Gibbs, E., O'Donnell, M., and Hurwitz, J. (1997) *J.Biol.Chem.* **272**, 10058-10064
27. Bork, P., Hofmann, K., Bucher, P., Neuwald, A. F., Altschul, S. F., and Koonin, E. V. (1997) *FASEB J.* **11**, 68-76
28. Bunz, F., Kobayashi, R., and Stillman, B. (1993) *Proc.Natl.Acad.Sci.U.S.A* **90**, 11014-11018
29. 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
30. Gomes, X. V., Gary, S. L., and Burgers, P. M. (2000) *J.Biol.Chem.* **275**, 14541-14549
31. Bowman, G. D., O'Donnell, M., and Kuriyan, J. (2004) *Nature* **429**, 724-730
32. Ellison, V. and Stillman, B. (1998) *J.Biol.Chem.* **273**, 5979-5987
33. Uhlmann, F., Cai, J., Flores-Rozas, H., Dean, F. B., Finkelstein, J., O'Donnell, M., and Hurwitz, J. (1996) *Proc.Natl.Acad.Sci.U.S.A* **93**, 6521-6526
34. Hingorani, M. M. and Coman, M. M. (2002) *J.Biol.Chem.* **277**, 47213-47224
35. Keller, R. C., Mossi, R., Maga, G., Wellinger, R. E., Hubscher, U., and Sogo, J. M. (1999) *Nucleic Acids Res.* **27**, 3433-3437
36. Miyata, T., Oyama, T., Mayanagi, K., Ishino, S., Ishino, Y., and Morikawa, K. (2004) *Nat.Struct.Mol.Biol.* **11**, 632-636
37. Bauer, G. A. and Burgers, P. M. (1988) *Biochim.Biophys.Acta* **951**, 274-279

38. Tsurimoto, T. and Stillman, B. (1990) *Proc.Natl.Acad.Sci.U.S.A* **87**, 1023-1027
39. Zhang, G., Gibbs, E., Kelman, Z., O'Donnell, M., and Hurwitz, J. (1999) *Proc.Natl.Acad.Sci.U.S.A* **96**, 1869-1874
40. Gomes, X. V. and Burgers, P. M. (2001) *J.Biol.Chem.* **276**, 34768-34775
41. Gomes, X. V., Schmidt, S. L., and Burgers, P. M. (2001) *J.Biol.Chem.* **276**, 34776-34783
42. Melo, J. A., Cohen, J., and Toczyski, D. P. (2001) *Genes Dev.* **15**, 2809-2821
43. Melo, J. and Toczyski, D. (2002) *Curr.Opin.Cell Biol.* **14**, 237-245
44. Rauen, M., Burtelow, M. A., Dufault, V. M., and Karnitz, L. M. (2000) *J.Biol.Chem.* **275**, 29767-29771
45. Schmidt, S. L., Gomes, X. V., and Burgers, P. M. (2001) *J.Biol.Chem.* **276**, 34784-34791
46. Majka, J., Chung, B. Y., and Burgers, P. M. (2004) *J.Biol.Chem.* **279**, 20921-20926
47. Bellaoui, M., Chang, M., Ou, J., Xu, H., Boone, C., and Brown, G. W. (2003) *EMBO J.* **22**, 4304-4313
48. Ben Aroya, S., Koren, A., Liefshitz, B., Steinlauf, R., and Kupiec, M. (2003) *Proc.Natl.Acad.Sci.U.S.A* **100**, 9906-9911
49. Uhlmann, F. (2004) *Exp.Cell Res.* **296**, 80-85
50. Mayer, M. L., Gygi, S. P., Aebersold, R., and Hieter, P. (2001) *Mol.Cell* **7**, 959-970
51. Bermudez, V. P., Maniwa, Y., Tappin, I., Ozato, K., Yokomori, K., and Hurwitz, J. (2003) *Proc.Natl.Acad.Sci.U.S.A* **100**, 10237-10242
52. Mayer, M. L., Gygi, S. P., Aebersold, R., and Hieter, P. (2001) *Mol.Cell* **7**, 959-970
53. Bylund, G. O. and Burgers, P. M. (2005) *Mol.Cell Biol.* **25**, 5445-5455
54. Uhlmann, F. (2004) *Exp.Cell Res.* **296**, 80-85
55. Deshpande, A., Sicinski, P., and Hinds, P. W. (2005) *Oncogene* **24**, 2909-2915
56. van der, K. H., Carius, B., Haque, S. J., Williams, B. R., Huber, C., and Fischer, T. (1999) *J.Mol.Med.* **77**, 386-392
57. Xiong, Y., Zhang, H., and Beach, D. (1992) *Cell* **71**, 505-514
58. Koundrioukoff, S., Jonsson, Z. O., Hasan, S., de Jong, R. N., van der Vliet, P. C., Hottiger, M. O., and Hubscher, U. (2000) *J.Biol.Chem.* **275**, 22882-22887
59. Maga, G., Mossi, R., Fischer, R., Berchtold, M. W., and Hubscher, U. (1997) *Biochemistry* **36**, 5300-5310
60. Salles-Passador, I., Munshi, A., Cannella, D., Pennaneach, V., Koundrioukoff, S., Jaquinod, M., Forest, E., Podust, V., Fotedar, A., and Fotedar, R. (2003) *Nucleic Acids Res.* **31**, 5202-5211

61. Munshi, A., Cannella, D., Brickner, H., Salles-Passador, I., Podust, V., Fotedar, R., and Fotedar, A. (2003) *J.Biol.Chem.* **278**, 48467-48473
62. Maruyama, T., Farina, A., Dey, A., Cheong, J., Bermudez, V. P., Tamura, T., Sciortino, S., Shuman, J., Hurwitz, J., and Ozato, K. (2002) *Mol.Cell Biol.* **22**, 6509-6520
63. Pennaneach, V., Salles-Passador, I., Munshi, A., Brickner, H., Regazzoni, K., Dick, F., Dyson, N., Chen, T. T., Wang, J. Y., Fotedar, R., and Fotedar, A. (2001) *Mol.Cell* **7**, 715-727
64. 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
65. Luckow, B., Bunz, F., Stillman, B., Lichter, P., and Schutz, G. (1994) *Mol.Cell Biol.* **14**, 1626-1634
66. Uchiumi, F., Ohta, T., and Tanuma, S. (1996) *Biochem.Biophys.Res.Comm.* **229**, 310-315
67. Surtees, J. A. and Alani, E. (2004) *Mol.Cell* **15**, 164-166
68. Dzantiev, L., Constantin, N., Genschel, J., Iyer, R. R., Burgers, P. M., and Modrich, P. (2004) *Mol.Cell* **15**, 31-41
69. Koonin, E. V., Altschul, S. F., and Bork, P. (1996) *Nat.Genet.* **13**, 266-268
70. Callebaut, I. and Mornon, J. P. (1997) *FEBS Lett.* **400**, 25-30
71. Derbyshire, D. J., Basu, B. P., Serpell, L. C., Joo, W. S., Date, T., Iwabuchi, K., and Doherty, A. J. (2002) *EMBO J.* **21**, 3863-3872
72. Williams, R. S., Green, R., and Glover, J. N. (2001) *Nat.Struct.Biol.* **8**, 838-842
73. Zhang, X., Morera, S., Bates, P. A., Whitehead, P. C., Coffey, A. I., Hainbucher, K., Nash, R. A., Sternberg, M. J., Lindahl, T., and Freemont, P. S. (1998) *EMBO J.* **17**, 6404-6411
74. Joo, W. S., Jeffrey, P. D., Cantor, S. B., Finnin, M. S., Livingston, D. M., and Pavletich, N. P. (2002) *Genes Dev.* **16**, 583-593
75. Krishnan, V. V., Thornton, K. H., Thelen, M. P., and Cosman, M. (2001) *Biochemistry* **40**, 13158-13166
76. Lee, J. Y., Chang, C., Song, H. K., Moon, J., Yang, J. K., Kim, H. K., Kwon, S. T., and Suh, S. W. (2000) *EMBO J.* **19**, 1119-1129
77. Clapperton, J. A., Manke, I. A., Lowery, D. M., Ho, T., Haire, L. F., Yaffe, M. B., and Smerdon, S. J. (2004) *Nat.Struct.Mol.Biol.* **11**, 512-518
78. Shiozaki, E. N., Gu, L., Yan, N., and Shi, Y. (2004) *Mol.Cell* **14**, 405-412
79. Williams, R. S., Lee, M. S., Hau, D. D., and Glover, J. N. (2004) *Nat.Struct.Mol.Biol.* **11**, 519-525
80. Botuyan, M. V., Nomine, Y., Yu, X., Juranic, N., Macura, S., Chen, J., and Mer, G. (2004) *Structure.(Camb.)* **12**, 1137-1146
81. Manke, I. A., Lowery, D. M., Nguyen, A., and Yaffe, M. B. (2003) *Science* **302**, 636-639
82. Rodriguez, M., Yu, X., Chen, J., and Songyang, Z. (2003) *J.Biol.Chem.* **278**, 52914-52918

83. Yu, X., Chini, C. C., He, M., Mer, G., and Chen, J. (2003) *Science* **302**, 639-642
84. Furuya, K., Poitelea, M., Guo, L., Caspari, T., and Carr, A. M. (2004) *Genes Dev.* **18**, 1154-1164
85. Makiniemi, M., Hillukkala, T., Tuusa, J., Reini, K., Vaara, M., Huang, D., Pospiech, H., Majuri, I., Westerling, T., Makela, T. P., and Syvaioja, J. E. (2001) *J.Biol.Chem.* **276**, 30399-30406
86. Wang, H. and Elledge, S. J. (2002) *Genetics* **160**, 1295-1304
87. Critchlow, S. E., Bowater, R. P., and Jackson, S. P. (1997) *Curr.Biol.* **7**, 588-598
88. Sibanda, B. L., Critchlow, S. E., Begun, J., Pei, X. Y., Jackson, S. P., Blundell, T. L., and Pellegrini, L. (2001) *Nat.Struct.Biol.* **8**, 1015-1019
89. Magnard, C., Bachelier, R., Vincent, A., Jaquinod, M., Kieffer, S., Lenoir, G. M., and Venezia, N. D. (2002) *Oncogene* **21**, 6729-6739
90. Li, S., Chen, P. L., Subramanian, T., Chinnadurai, G., Tomlinson, G., Osborne, C. K., Sharp, Z. D., and Lee, W. H. (1999) *J.Biol.Chem.* **274**, 11334-11338
91. Sum, E. Y., Peng, B., Yu, X., Chen, J., Byrne, J., Lindeman, G. J., and Visvader, J. E. (2002) *J.Biol.Chem.* **277**, 7849-7856
92. Wada, O., Oishi, H., Takada, I., Yanagisawa, J., Yano, T., and Kato, S. (2004) *Oncogene* **23**, 6000-6005
93. Yu, X., Wu, L. C., Bowcock, A. M., Aronheim, A., and Baer, R. (1998) *J.Biol.Chem.* **273**, 25388-25392
94. Liu, K., Lin, F. T., Ruppert, J. M., and Lin, W. C. (2003) *Mol.Cell Biol.* **23**, 3287-3304
95. Masson, M., Niedergang, C., Schreiber, V., Muller, S., Menissier-De Murcia, J., and de Murcia, G. (1998) *Mol.Cell Biol.* **18**, 3563-3571
96. El Khamisy, S. F., Masutani, M., Suzuki, H., and Caldecott, K. W. (2003) *Nucleic Acids Res.* **31**, 5526-5533
97. Dulic, A., Bates, P. A., Zhang, X., Martin, S. R., Freemont, P. S., Lindahl, T., and Barnes, D. E. (2001) *Biochemistry* **40**, 5906-5913
98. Taylor, R. M., Wickstead, B., Cronin, S., and Caldecott, K. W. (1998) *Curr.Biol.* **8**, 877-880
99. Du, L. L., Moser, B. A., and Russell, P. (2004) *J.Biol.Chem.* **279**, 38409-38414
100. Soulier, J. and Lowndes, N. F. (1999) *Curr.Biol.* **9**, 551-554
101. Taylor, R. M., Moore, D. J., Whitehouse, J., Johnson, P., and Caldecott, K. W. (2000) *Mol.Cell Biol.* **20**, 735-740
102. Gilbert, C. S., Green, C. M., and Lowndes, N. F. (2001) *Mol.Cell* **8**, 129-136
103. Liao, H., Yuan, C., Su, M. I., Yongkiettrakul, S., Qin, D., Li, H., Byeon, I. J., Pei, D., and Tsai, M. D. (2000) *J.Mol.Biol.* **304**, 941-951
104. Gilbert, C. S., Green, C. M., and Lowndes, N. F. (2001) *Mol.Cell* **8**, 129-136

105. Glover, J. N., Williams, R. S., and Lee, M. S. (2004) *Trends Biochem.Sci.* **29**, 579-585
106. Rodriguez, J. A., Au, W. W., and Henderson, B. R. (2004) *Exp.Cell Res.* **293**, 14-21
107. Kaczmarek, F. S., Zaniewski, R. P., Gootz, T. D., Danley, D. E., Mansour, M. N., Griffor, M., Kamath, A. V., Cronan, M., Mueller, J., Sun, D., Martin, P. K., Benton, B., McDowell, L., Biek, D., and Schmid, M. B. (2001) *J.Bacteriol.* **183**, 3016-3024
108. Lim, J. H., Choi, J., Kim, W., Ahn, B. Y., and Han, Y. S. (2001) *Arch.Biochem.Biophys.* **388**, 253-260
109. Feng, H., Parker, J. M., Lu, J., and Cao, W. (2004) *Biochemistry* **43**, 12648-12659
110. Jeon, H. J., Shin, H. J., Choi, J. J., Hoe, H. S., Kim, H. K., Suh, S. W., and Kwon, S. T. (2004) *FEMS Microbiol.Lett.* **237**, 111-118
111. 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
112. Wuthrich, K. (1986) *NMR of Proteins and Nucleic acids*, Wiley, New York,
113. Breeze, A. L. (2000) *Progress in Nuclear Magnetic Resonance Spectroscopy* **36**, 323-372
114. Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R., and Kaptein, R. (2004) *Science* **305**, 386-389
115. Burgering, M. J., Boelens, R., Caffrey, M., Breg, J. N., and Kaptein, R. (1993) *FEBS Lett.* **330**, 105-109
116. Worrall, J. A., Reinle, W., Bernhardt, R., and Ubbink, M. (2003) *Biochemistry* **42**, 7068-7076
117. Zuiderweg, E. R. (2002) *Biochemistry* **41**, 1-7

## Chapter 2

# Characterization of the DNA Binding and Structural Properties of the BRCT region of the p140 subunit of human Replication Factor C

---

### Abstract

---

BRCT domains, present in a large number of proteins that are involved in cell-cycle regulation and/or DNA replication or repair, are primarily thought to be involved in protein-protein interactions. The large (p140) subunit of Replication Factor C contains roughly ~100 amino acids in the N-terminal half, which has been shown to bind DNA. This domain also contains sequences distantly related to known BRCT domains. Here we demonstrate that a region of p140 that encompasses the BRCT domain contains two distinct DNA binding activities. Residues 375 to 480, which include 28 amino acids N-terminal to the conserved BRCT domain, contain a binding activity specific for 5'-phosphorylated dsDNA while a non-sequence or structure specific dsDNA binding activity requires an additional 65 amino acids C-terminal to the BRCT domain (residues 480-545). The N-terminal extension provides an extra  $\alpha$ -helix to an otherwise conserved BRCT domain. However, the extended domain is flexible in the absence of a DNA ligand. Although the 65 amino acid region C-terminal to the BRCT domain is unstructured when expressed by itself, a larger polypeptide incorporating these sequences (375-545) is well ordered even in the absence of the DNA ligand. This suggests that the complete structural unit in the intact protein includes amino acids 375-545.

M.Kobayashi, M. C. van Wijk, L. Jansen and G. Siegal (Parts of this chapter and of chapter 3 will be submitted for publication)

---

Introduction

---

The BRCA1 C-terminal homology (BRCT) domain is an abundant structural unit found in more than 50 nonorthologous proteins (1;2). Nearly all of these proteins are involved in the cell cycle checkpoint response to damaged DNA and/or more directly in DNA replication or repair. The BRCT superfamily has been further classified into three subsets. The first consists of a core of highly conserved domains found in proteins such as BRCA1 itself, the *Saccharomyces cerevisiae* Rad9 protein and the p53 binding protein 53BP1. A second distinct and more distantly related set can be found in DNA binding enzymes such as the bacterial NAD-dependent ligases and poly (ADP-ribose) polymerase (1). Finally, the retinoblastoma tumor suppressor and related proteins may contain a very distant member of the BRCT family(1).

BRCT domains are responsible for a number of important homo- and heterotypic protein-protein interactions. For instance, XRCC1, a protein involved in repair of single-stranded DNA breaks, binds DNA ligase III *via* its C-terminal BRCT domain (3) while its N-terminal BRCT domain specifically binds poly-ADP Ribose polymerase (4). Other BRCT domains, such as those in the *S. cerevisiae* protein Rad9, function in homotypic (self) protein binding (5). Recently, it has been shown that many BRCT domains possess a phosphoserine specific, protein binding function (6-8). The serine phosphorylations that have been investigated to date are performed by either ATM or ATR, two protein kinases involved in the early steps of signal transduction generated by damaged DNA (6). Although there are reports that some BRCT domains can bind DNA, the binding requirements have not been defined and it therefore remains unclear whether binding is a property of the BRCT domain itself in these cases (20).

BRCT domains consist of approximately 95 amino acid residues with a strictly conserved fold. The first 3D structure elucidation was that of the C-terminal BRCT domain of XRCC1 (9). The structure consists of a central 4-stranded parallel  $\beta$ -sheet surrounded by three  $\alpha$ -helices. Recently the structures of the BRCT domain from DNA ligase III (10), and the tandem BRCT domains from BRCA1 (11) and from the p53 binding protein p53BP1 (12) have also been solved. In addition, a low-resolution structure of the BRCT domain from the DNA-ligase of *Thermus filiformis* (13), a member of the second class of BRCT domains has been solved. The structural conservation from bacteria to humans indicates that this is likely an ancient fold.

Replication Factor C (RFC) is a five protein complex involved in both the replication and repair of chromosomal DNA (14-16). The primary function of RFC appears to be to open the “sliding clamp” protein PCNA and “load” it onto DNA where it serves as a binding platform for a multitude of enzymes and regulatory proteins involved in the replication and repair of DNA. RFC consists of four subunits of between 35 and 40 kDa that share homology with a central region of the fifth subunit, which has a molecular mass of 140 kDa in mammals (referred to as p140). The N-terminal half of the p140 subunit contains sequences unique to RFC, including a region with DNA binding activity (17-21) that is not required for clamp loading (22;23) mediated by the DNA binding specific to the 3' end of the primer/template DNA. This region contains amino acid sequences which form part of the second distinct class of BRCT domains (1;2). Here we demonstrate that this region of p140 that includes the BRCT homologous sequences binds double-stranded DNA in both a 5' phosphate specific and a non-specific manner.

---

## Materials and Methods

---

### *Gene Cloning*

The plasmid containing a cDNA clone coding for residues 369 - 480 RFC p140 was a kind gift of Prof. Ullrich Hübscher (Universität Zürich, Switzerland). The three different constructs were generated using standard PCR based methods and were cloned into pET20b (Novagen) with a C-terminal 6-His tag (Fig 1A). The plasmid containing a cDNA clone encoding full length RFC p140 was a kind gift of Prof. Bruce Stillman (Cold Spring Harbor Laboratory). The two constructs encoding for residues 374-545 and 480-545 of RFC p140 were generated using PCR based methods and were cloned into pET20b(Novagen) with a C-terminal 6-His tag. A plasmid containing a cDNA clone of full length human C/EBP $\alpha$  was used to generate a DNA construct coding for residues 281 - 358 of C/EBP $\alpha$  in pET20b (Novagen) with a C-terminal 6-His tag. All PCR generated clones were sequenced bidirectionally.

### *Protein expression and purification*

Proteins were expressed for 3 hours at 37 °C in BL21/DE3/pLysS cells (Novagen). Lysed cells were centrifuged at 20,000 x g for 30' at 4 °C and the supernatant applied to a 6 ml metal-chelate column (Novagen) charged with Ni<sup>2+</sup>. The column was developed as per the manufacturers recommendations. Fractions containing the His- tagged



protein, as judged by SDS-PAGE electrophoresis, were pooled and EDTA was added to the protein solution to a final concentration of 5 mM to chelate Ni<sup>2+</sup> leached from the column. The protein solution was concentrated through an Amicon ultrafiltration device (YM10 membrane) to 5 ml volume. The protein was further purified to apparent homogeneity using a 140 ml (1.6 x 75 cm) column of Superose 12 (Amersham Pharmacia) equilibrated with 20 mM Tris pH7.5, 50 mM NaCl and 1 mM DTT. All proteins were stably stored at 4 °C. All three purified RFC p140 proteins (Figure 2.1A) were exchanged into the PBS buffer. Subsequently the secondary structure contents were evaluated by circular dichroism spectra of 10 µM of each proteins in PBS solution by using Jobin Yvon CD6 instrument at 1 nm intervals over 190-240 nm wavelength at 21 °C in a 0.5 mm path-length cell. All three RFCp140 proteins were judged to be folded based on the circular dichroism spectra.

#### *Stoichiometric titration*

10 µM of the oligo 1 (in 25mM Tris-HCl pH 7.5, 50mM NaCl and 1mM DTT) was mixed with increasing amount of RFC p140 (375-480) (in the identical buffer condition) in the final volume of 250 µl. The DNA-protein complex was separated from the free DNA on the Superdex™ 10/300 GL Columns (Tricorn) which had been pre-equilibrated with 25mM Tris-HCl pH 7.5, 50mM NaCl and 1mM DTT. The volume of the absorbance peaks corresponding to the complex and free DNA were integrated and the fraction of the complex was calculated by dividing the volume of the complex peak by the sum of the volumes of the complex and the free DNA peaks.

#### *Preparation of phosphorylated oligonucleotides*

The sequences of all oligonucleotides used in this study are presented in Figure 2.1. The oligonucleotides were synthesized using standard solid-state methods and contained a 5' PO<sub>4</sub> (with the exception noted in Figure 2.1). Two pmol of 5' phosphorylated oligonucleotide were used in an exchange reaction employing the phage T4 polynucleotide kinase and 20 pmol of [ $\gamma$ -<sup>32</sup>P] ATP as substrate. In order to ensure a very high percentage of double stranded DNA, the hairpin oligonucleotides were subsequently denatured at 100 °C for 10 minutes and then slowly cooled to room temperature. All non-phosphorylated oligonucleotides (including the CREB oligonucleotide) were phosphorylated using a standard protocol for the forward reaction and the same amount of substrate and enzyme as described above.

*Detection of protein-DNA complexes*

DNA binding was detected using a gel retardation assay. The indicated amount of the various RFC p140 proteins were diluted in a buffer of 10 mM HEPES pH 7.8, 2 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 100 µg/ml BSA, 15 % glycerol, 0.8 µg/ml poly (dI-dC) and 2 mM DTT. Indicated amount and appropriate volume of 5' [32P] oligonucleotide was added to the mixture to the final volume of 12 µl, which was subsequently incubated on ice for 15 minutes, then applied to a non-denaturing 8% Tris-Glycine, acrylamide gel and electrophoresed at 150 V for 30 minutes at 4 °C in 25 mM Tris-HCl pH 8.5, 200 mM glycine, 1mM EDTA (TGE) buffer. Radioactivity was detected by exposing to a film (X-OMAT, Kodak) or to a phosphor-imager (BIORAD) after the gels were dried. Gel retardation was also used to determine the ligand (DNA) requirements for binding but in the form of a competition assay. In these experiments both labeled and unlabeled DNA were premixed and then added to the protein solution. The amounts of the competing oligonucleotides used in excess over the labeled oligonucleotide in the assay are indicated in the Figures.

*Southern blotting and ECL-detection of protein-DNA complexes*

The 5' biotinylated CREB oligonucleotide (CRE binding consensus sequence) was purchased from Microsynth GmbH, Switzerland. The CREB oligonucleotide was dissolved in the same buffer described above, heated to 100 °C and annealed by slow cooling. DNA binding of the RFC p140 proteins and/or C/EBPα DBD (DNA binding domain) was detected using 100 fmol of the CREB oligonucleotide in a gel retardation assay using the same buffer conditions described earlier (25). The biotinylated CREB oligonucleotide was then transferred to Hybond-N+ nylon membranes (Amersham Bioscience) at 4 °C at 200mV for 1 hour in 90 mM Tris - 90 mM borate - 2 mM EDTA (TBE) buffer. The wet membrane was exposed to UV (254 nm) for 3 minutes in order to cross-link DNA to the membrane. The membrane was incubated in 25 ml of TBE (with 5 % dry milk) for 1 hour and subsequently 15 µg of Streptavidin-HRP (horse radish peroxidase) conjugate (Zymed laboratories inc.) was added. CREB oligonucleotides were subsequently detected using the ECL reagent as described by the manufacturer (Amersham Bioscience).

### *NMR Measurements*

All three proteins were concentrated to between 0.5 and 1 mM in 25 mM Tris-HCl, 100 mM NaCl, 1 mM DTT and NaN<sub>3</sub> 0.01 % (w/v) in 5% D<sub>2</sub>O where the final pH was adjusted to 7.5 (direct meter reading). The sample of 1:1 DNA-RFC p140(375-480) complex had a final concentration of approximately 0.5 mM in 25 mM Tris-HCl, 100 mM NaCl, 1 mM DTT and NaN<sub>3</sub> 0.01 % (w/v) in 5% D<sub>2</sub>O with the final pH adjusted to 7.5. All NMR spectra were recorded at 298 K on a 600 MHz Bruker DMX spectrometer. A standard [15N,1H] HSQC pulse sequence employing the WATERGATE solvent suppression method (24) was used. Spectra typically consisted of 128 increments of 64 transients each and were processed using TOPSPIN software. The sequential assignment was accomplished using <sup>13</sup>C, <sup>15</sup>N labeled 1:1 complex of RFC p140(375-480) and oligo 1 using standard through bond, triple resonance NMR experiments. Details will be published elsewhere. The consensus CSI score was generated with the program CSI version 1.0 (25) from <sup>13</sup>C $\alpha$ , <sup>13</sup>C $\beta$  and <sup>1</sup>H $\alpha$  chemical shifts values as inputs.

---

## Results

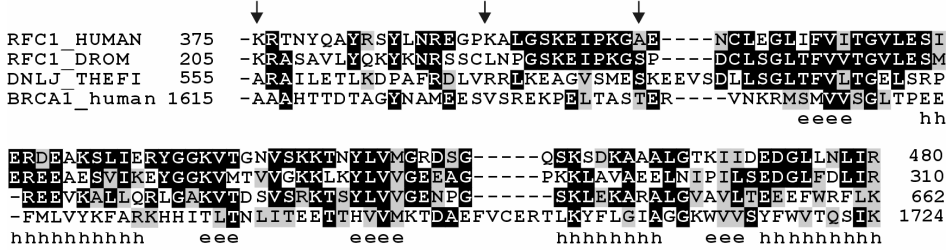
---

### *Identification of protein domains required for DNA binding activity*

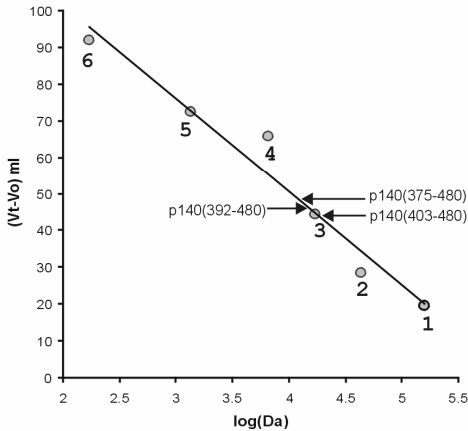
The DNA binding domain of human RFC p140 was initially defined as consisting of residues 369 – 480 (17). Since this region includes sequences outside the conserved BRCT domain we were interested to find out whether just the BRCT domain (403-480) itself was sufficient for DNA binding. Based on the published sequence alignments (1;2) we made three N-terminal deletion constructs whose C-termini all coincide with the expected C-terminus of the BRCT domain (see arrows in Figure 2.1A). All three-deletion constructs were C-terminally tagged with six histidines, expressed in *E. coli* and subsequently purified using metal chelate and gel-filtration chromatography. A Superose 12 gel filtration column was calibrated with the proteins with known molecular weight. The resulting calibration curve was used to estimate the molecular weight of the purified p140 constructs (Figure 2.1 B). The estimated molecular mass of the p140(375-480) derived from the gel-filtration experiment, 14 kDa, was very close to the expected value of a monomer (12 kDa). p140 (403-480) eluted as a very broad peak and the estimated molecular weight (18 kDa) was twice the expected value. It is known that many isolated BRCT domains tend to form homodimers or multimers. These observations imply that this

construct is a mix of monomer and dimer or has an elongated shape. The elution of p140 (392-480) correlates with a molecular weight of (15 kDa) which is again slightly larger than the expected value (10 kDa). The secondary structure of the three constructs was analyzed using circular dichroism (CD) spectroscopy. Prediction of secondary structures from the CD spectra using the K2d program (26) shows that all of the peptide constructs contained a mixture of  $\alpha$  and  $\beta$  - structures (Figure 2.1C).

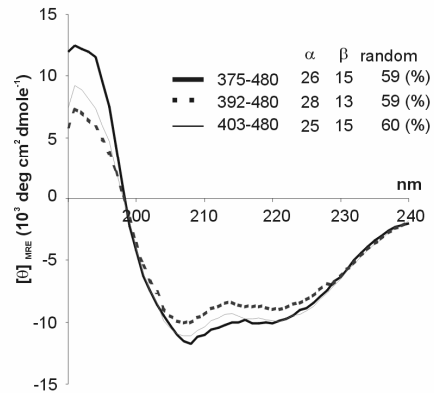
**A**



**B**

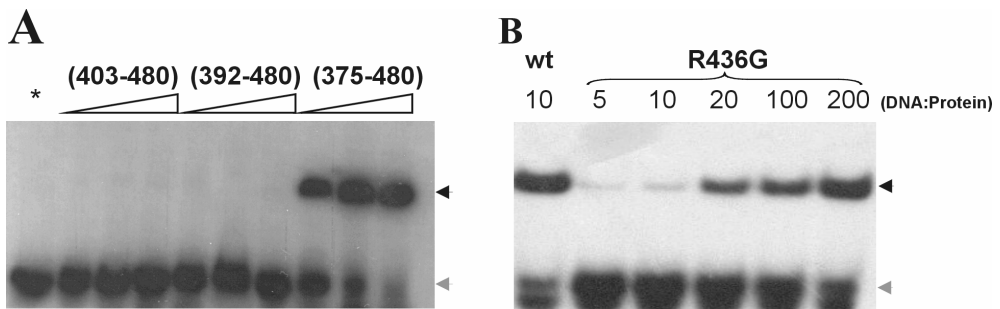


**C**



(Figure 2.1) Characterization of the three purified constructs, p140(375-480), p140(392-480) and p140(403-480). **(A)** Alignment of the amino acid sequences of various BRCT domains including: the N-terminal BRCT domain of human BRCA1 (BRCA1\_HUMAN1), the BRCT domain of NAD<sup>+</sup> dependent DNA ligase from *Thermus filiformis*, the BRCT domain of the p140 subunit of *Drosophila melanogaster* RFC (RFC1\_DROM) and the BRCT domain of the p140 subunit of human RFC (RFC\_HUMAN). The alignment is recreated according to the original alignment described by Bork *et al.*, 1997 (1) by using Clustal W. Identical residues are highlighted in black while similar residues are shown in grey. Gaps are shown as dashes. At the bottom of the alignment, the secondary structures found in the crystal structure of human BRCA1 BRCT1 are indicated in e ( $\beta$ -strand) and h ( $\alpha$ -helix). The vertical arrows indicate the first amino acid of each of the three N-terminal deletion constructs. All three constructs included sequences up to residue 480. **(B)** The calibration curve for the Superose 12 column was generated with 1.  $\gamma$ -globulin (670 kDa), 2. ovalbumin (44 kDa), 3. myoglobin (17 kDa), 4. aprotinin (6.5 kDa), 5. vitamin E (1.35 kDa) and 6. tryptophan (0.17 kDa). The elution position of each protein is indicated with arrows. The estimated molecular mass of each is given in the text. **(C)** CD spectra were measured for the three constructs. The far UV-CD spectra of all three constructs have characteristic minima at 208 nm and 220 nm, reflecting  $\alpha$ -helical and  $\beta$ -strand contents. The percentages of each secondary structure were predicted by the K2d program (26).

Protein-DNA complex formation was detected using a gel retardation assay in which purified proteins were titrated into a constant amount of a 5' [<sup>32</sup>P] labeled, 10 bp, hairpin oligodeoxynucleotide (Table I, Figure 2.2A). Only the protein containing 28 residues N-terminal to the predicted boundary of the BRCT domain (p140(375-480)) demonstrated DNA binding activity. The CD spectra of both RFC p140(403-480) and p140(392-480) were similar to that of the p140(375-480) protein indicating that the lack of DNA binding is likely not due to gross structural rearrangement or inability to fold (Figure 2.1C).



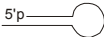
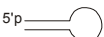
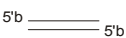

(Figure 2.2) Extent of the protein domain required for DNA binding.

**(A)** Gel retardation experiments were used to detect DNA binding to the 5' <sup>32</sup>P labeled oligodeoxynucleotide (oligo 1, Figure 2.1). Increasing amounts (40fmoles, 200fmoles, and 400fmoles) of the indicated constructs of hRFC p140 were titrated into 40fmoles of DNA. The dash indicates a no protein control. The black arrow indicates the DNA-protein complex and the gray arrow indicates the unbound DNA. **(B)** Amino acid substitution at the conserved glycine 435 to arginine (G435R) reduced the DNA binding activity of the p140(375-480). The numbers above the gel lanes indicate the protein to DNA ratio. The amount of DNA used was 40fmoles. To achieve a similar amount of DNA-protein complex observed for wt in G435R mutants, roughly 20 fold higher protein:DNA ratio was required (lane DNA:protein ratio of 200).

An amino acid substitution at the conserved G435 to R in p140(375-480) reduced DNA binding by almost 20 fold in comparison to the wildtype (wt) protein (Figure 2.2B). This reduced DNA binding due to the mutation suggests that dsDNA binding is contributed by the both BRCT domain and the N-terminal sequence outside of the domain. The mutated glycine occurs within the most conserved GG repeat found in the BRCT family and forms a tight turn between the  $\alpha 1$  and  $\beta 2$  structures of the X-ray crystal structure of the BRCT domain from XRCC1 (27). The substitution of a bulky, charged residue like arginine could potentially result in destabilization of the fold of the protein, which may have been indicated by the precipitation behavior of the G435R mutant observed at room

temperature and the lower yield during purification<sup>1</sup> compared to the wild type (data not shown).

(Table 2.1) The sequence of oligonucleotides used in this study.

Name	Duplex size	DNA sequence	Structure
Oligo 1	10 bp	pCTCGAGGTCG <b>TCAT</b> CGACCTCGAGATCA	
	9 bp	pCTCGAGTCG <b>TCAT</b> CGACTCGAGATCA	
	8 bp	pCTCAGTCG <b>TCAT</b> CGACTCAGATCA	
	7 bp	pCTAGTCG <b>TCAT</b> CGACTAGATCA	
	6 bp	pCTGTCG <b>TCAT</b> CGACAGATCA	
Oligo 2	9 bp	pCTCGATGTCG <b>TCAT</b> CGACCTCGAGATCA	
5'OH	10 bp	CTCGAGGTCG <b>TCAT</b> CGACCTCGAGATCA	
5'Pb	10 bp	pCTCGAGGTCG <b>TCAT</b> CGACCTCGAG	
5'B	20 bp	bTGCAGATTGCGCAATCTGCA	
Ss	Single stranded	TGGGGTGGGGT	
5'Ps	Single stranded	pTGGGGTGGGGT	

<sup>1</sup> For hairpin oligonucleotides, the hairpin sequence is highlighted with bold letters. The unpaired bases of oligo 2 are indicated in italics. A "p" indicates 5' PO<sub>4</sub> while "b" indicates a 5' biotinylated. <sup>2</sup>The consensus sequence for the C/EBP DBD (bZIP) binding site.

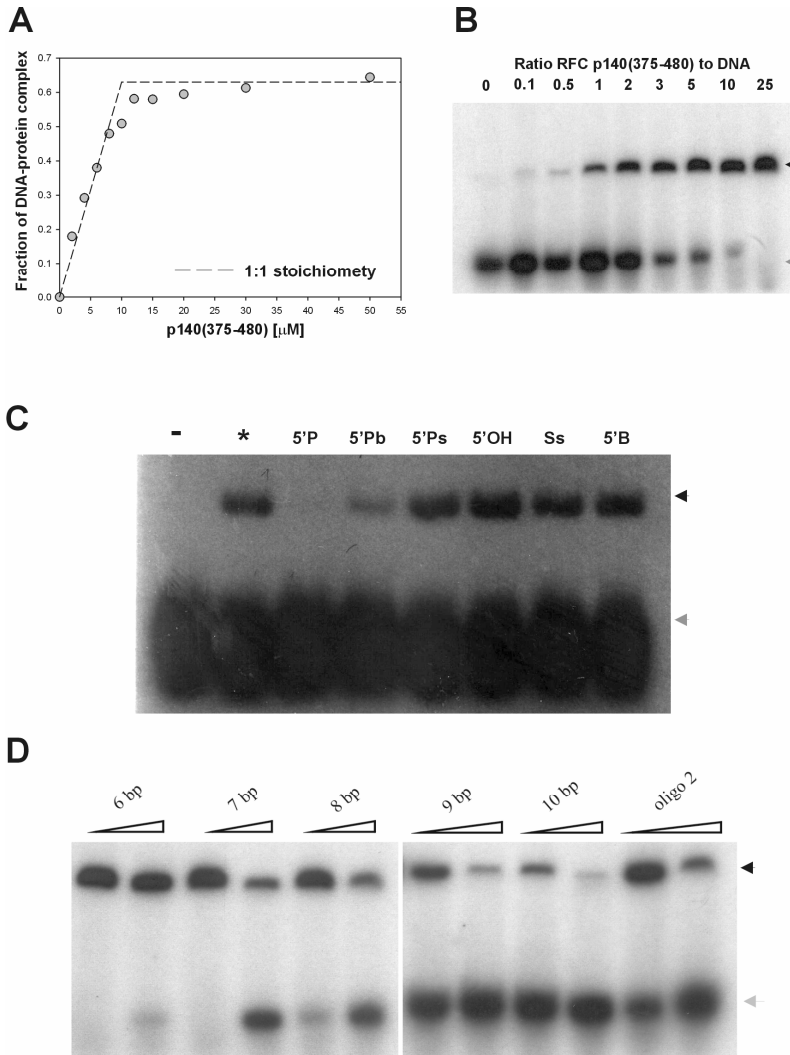
### Specificity of DNA binding

To determine the stoichiometry of the DNA-protein complex, a titration experiment was performed by adding an increasing amount of p140(375-480) to a constant amount of oligo 1 (Figure 2.1). The protein-DNA complex was separated from unbound DNA through an analytical scale gel filtration column and the absorbance peaks corresponding to the free DNA and DNA-protein complex were integrated and the fraction of the DNA-protein complex was calculated. The fraction of DNA bound to protein

<sup>1</sup> The quantity of the mutant G435R was carefully measured by using both the UV absorbance and the Bradford protein assay prior to use. Due to the poor stability, the mutant could not be reached to the suitable concentration for CD measurement.

increased almost linearly until a DNA-protein ratio of 1.25:1 and starts to saturate after a ratio of 1:1.2 (Figure 2.3A). The experimental titration curve closely resembles an ideal titration curve of a DNA-protein complex ratio of 1:1. We also analyzed the 5' PO<sub>4</sub> specific DNA binding activity using the gel shift assay (Figure 2.3B). A unique DNA-protein complex was formed between p140(375-480) and oligo 1 under all titration conditions tested. Relative amounts of free and bound DNA were quantified using a phosphor-imager and the data was analyzed to yield a  $K_D \sim 10$  nM. The fit of the binding isotherm was performed by assuming that one p140(375-480) binds to a single DNA binding site (See Methods). Using a protein concentration close to  $K_D$ , competition binding between two different oligonucleotide ligands was performed by mixing the 5'[<sup>32</sup>P] labeled oligo 1 (10 base pairs) with an excess of various oligonucleotides (Figure 2.1) prior to the addition of p140(375-480). Competitive binding to the protein is detected as a reduction in the amount of protein-DNA complex observed in the gel retardation assay (Figure 2.3C). As expected from previous studies (17;18), binding to the 5' phosphorylated hairpin oligonucleotide could not be competed with either a 5' OH hairpin oligonucleotide, a ds 5' biotinylated, or 5' OH single stranded oligodeoxynucleotide (Figure 2.3C). Competition was slightly more effective with dsDNA with a recessed 5'-phosphate end than with blunt ended DNA (Figure 2.3C) correlating with the earlier report that the *Drosophila* RFC p140 (18) binds 5' end recessed dsDNA better than blunt ended DNA. DNA binding was also insensitive to salt up to 500 mM NaCl and not dependent on Mg<sup>2+</sup> (data not shown).

In order to precisely determine the binding determinants of the DNA, we synthesized a series of 5' phosphorylated oligonucleotides in which the ds region was successively shortened by one base pair. As seen in Figure 2.3D, 8-10 base pairs compete effectively to the protein binding to the oligo 1 indicating that 8-9 base pairs are required for optimal binding of RFC p140 (375-480). Binding to oligonucleotides with duplex regions longer than 10 base pairs was not more efficient than binding to oligo 1 (data not shown). In addition, fully base-paired DNA forms the best ligand since a mismatched-Watson-Crick base pair 6 nucleotides downstream from the 5' phosphate (oligo 2) reduced the affinity of the protein-DNA interaction (Figure 2.3D).

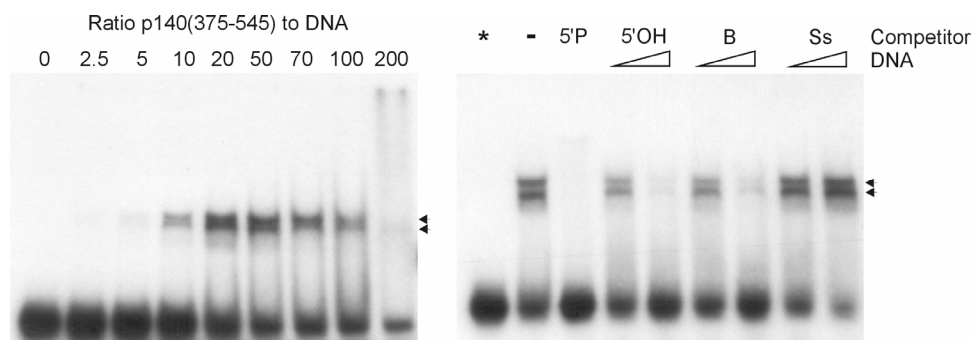


(Figure 2.3) Characterization of the DNA binding requirements of RFC p140(375-480). **(A)** Stoichiometric titration of protein-DNA complex. A constant amount of oligo 1 (10  $\mu$ M) was titrated with increasing amount of RFC p140(375-480) and the resulting DNA-protein complex was separated from the unbound DNA through an analytical gel filtration column. **(B)** Titration of protein-DNA complex. A constant amount 40fmol of 5'  $^{32}$ P labeled oligo 1 was titrated with increasing amount of RFC p140(375-480) in a gel retardation assay. The protein:DNA ratio is indicated at the top of each lane. Only one DNA-protein complex (black arrow) was observed over the titration range. **(C)** A competition gel retardation assay was used to delineate the chemical features of DNA required for binding by RFC p140(375-480). 40fmol of 5'  $^{32}$ P labeled oligo 1 was mixed with a 50 -fold excess of the competing DNA and subsequently 200fmol of protein was added. The competing DNA was a unlabeled oligo 1 (5'P), 5' PO<sub>4</sub> blunt end (5'Pb) hairpin-oligonucleotide, 5' PO<sub>4</sub> single stranded (5'Ps) oligonucleotide, 5'OH version of oligo 1 (5'OH), a single stranded (Ss) and the 5' biotinylated C/EBP oligonucleotide (5'B) oligonucleotide (see details on the Table I). **(D)** Competition gel retardation assay to characterize the extent of dsDNA required for binding. A constant amount 40fmol of  $^{32}$ P labeled oligo 1 was mixed with a 10 or 60 fold excess of the indicated unlabeled DNA (see Figure 2.1) and allowed to bind to 200fmol of protein. The black arrow indicates DNA-protein complexes and the grey arrow indicates unbound DNAs.



*Altered DNA binding activity in p140(375-545)*

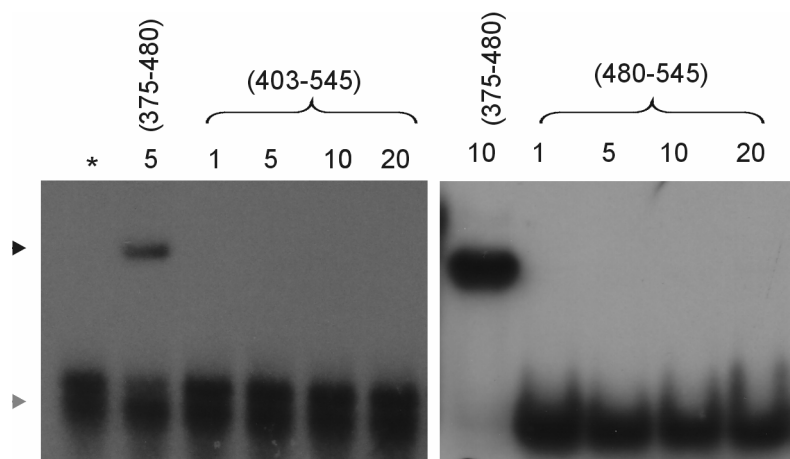
Since sequences N-terminal to the conserved BRCT domain were required for DNA binding, it was logical to determine the effect of sequences C-terminal as well. Accordingly a fragment of RFC p140 including amino acids 375 to 545 was cloned. The p140(375-545) contained a c-terminal six histidine tag and was purified to apparent homogeneity in the same manner as the wildtype protein. Using the gel shift assay, we titrated increasing amounts of the protein into the same 5' phosphorylated, hairpin oligonucleotide. As shown in Figure 2.4A, the C-terminally extended RFC p140(375-545) behaves differently than the shorter RFC p140(375-480). First, the affinity for DNA binding as a whole is reduced. Second, a single complex is never observed, rather two complexes with slightly different mobility (Figure 2.4, arrows), yet with similar apparent DNA binding affinities can be observed. The sum amount of the two complexes reached a maximum ~25% of the total labeled oligo1 when the p140(375-545) concentration was 150nM (Figure 2.4A, lane marked 20). As the ratio of protein to DNA is further increased, the amount of these two complexes decreases and broad bands migrating significantly slower are observed (see Figure 2.4A lane marked 200). In order to better understand the basis of these differences we determined which characteristics of the DNA ligand were necessary for binding using the competition gel shift assay (Figure 2.4B).



(Figure 2.4) Characterization of DNA binding requirements of RFCp140 (375-545). **(A)** Titration of DNA-protein complex. 90fmol (7.5 nM) of 5'  $^{32}$ P labeled oligo 1 was titrated with increasing amount of p140 (375-545). The numbers on the lanes indicate the ratio of protein :DNA. **(B)** A competition gel retardation assay was used to delineate the chemical features of DNA required for binding by RFC p140(375-545). 90fmol of 5'  $^{32}$ P labeled oligo 1 was mixed with a 10 or 20 fold excess (only 10 fold for oligo 1) of the competing DNA and subsequently 1.8pmol of protein was added.

In contrast to the RFC p140(375-480) fragment, in the absence of poly(dI-dC)<sup>2</sup> in the reactions, a 5' OH hairpin oligonucleotide competed with the 5' phosphorylated hairpin oligonucleotide for protein binding in both complexes. Similarly, a double stranded 5' biotinylated oligonucleotide competed as efficiently as the 5' OH hairpin oligonucleotide. However competition by these alternative ligands was not as effective as that by the 5' phosphorylated hairpin, where only a 10 fold excess was necessary to completely block binding to the labeled oligonucleotide (Figure 2.4B).

In order to assess whether the non-specific dsDNA binding by p140(375-545) is due to an independent binding activity of the C-terminal aa's 480-545 alone or requires the BRCT domain, two further proteins consisting of RFC p140 residues 480-545 and 403-545 were expressed and purified. These two soluble proteins, p140(403-545) and p140(480-545) were purified but did not bind oligo1 in the gel shift assay (Figure 2.5). This observation suggests that p140(480-545) does not bind DNA independently but that an alternative DNA binding site maybe created by a combination of the p140(375-480) and the aa's 480-545.



(Figure 2.5) Analysis on DNA binding activity of p140(403-545) and p140(480-545). Constant amount, 40 fmoles of 5' [<sup>32</sup>P] oligo 1 was titrated with various peptide constructs of p140 (the residue numbers are described on the lanes). The number on each lane indicates the ratio of protein : DNA concentrations in each lane. Asterisk indicates lane with only oligo 1. DNA-protein complex formed by p140(375-480) and unbound oligo1 are indicated by black and grey arrows respectively. No apparent formation of DNA-protein complexes were observed for p140(480-545) and (403-545).

<sup>2</sup> An excess amount of Poly(dI-dC) was used to reduce non-specific protein-DNA interaction and maintain a constant ratio of protein to DNA (see Materials and Methods).

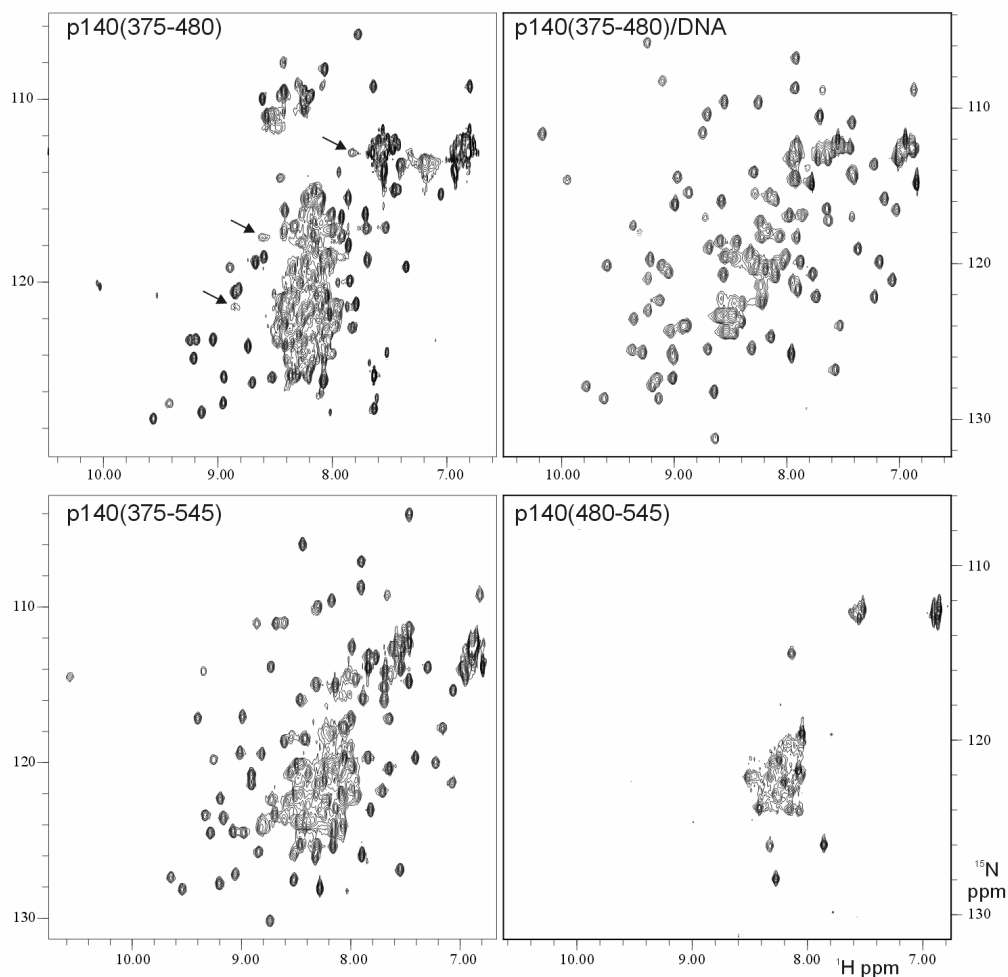
As expected, p140(375-545) was also capable of binding to the 5' biotinylated ds oligonucleotide (5'B in Figure 2.1) when the DNA-protein complex was detected with streptavidin-HRP (Supplementary materials Figure S1). This biotinylated oligonucleotide contained a consensus binding sequence for the bZIP domain of C/EBP- $\alpha$ , which is also known to interact with the p140(375-480) while binding to its cognate DNA to form a ternary complex (28). However no such complex were observed in the presence of bZIP and p140(375-545), and instead we observed two separate DNA-protein complexes that appear to be the result of independent DNA binding by bZIP or p140(375-545) (Supplementary materials Figure S1A). The ternary complex was only observed when the 5'-biotin was replaced by a phosphate. The simplest explanation for the ternary complex is that bZIP binds its cognate sequence simultaneously to p140(375-545) binding to the 5' phosphorylated terminus (Figure S1B).

### *Structural Characterization of the RFC p140 BRCT Region*

We used NMR to characterize the folded state of the three different fragments of RFC p140. The [ $^{15}\text{N}$ ,  $^1\text{H}$ ] HSQC experiment, which detects the 1 bond correlation within an amide moiety, is particularly useful for this purpose. Since 1 peak should be observed for every non-prolyl residue, the folded state of a protein can rapidly be determined by analyzing the dispersion and linewidth of the peaks in the spectrum (Figure 2.6). The spectrum in Figure 2.5 implies that large portions of the p140(375-480) protein are conformationally flexible in the absence of DNA as indicated by the poor dispersion of resonances (see the overlapped peaks in the region from 7.8 to 8.6 ppm) and by the presence of very broad peaks indicative of exchange that is intermediate on the NMR time scale (e.g.  $k_{\text{ex}} \sim \Delta\delta$  (Hz))<sup>3</sup>, indicated by arrows in Figure 2.6).

---

<sup>3</sup> When a nucleus experiences two or more states with different chemical properties, it is in exchange. For example a nucleus experiencing two different environments due to conformational exchange between A and B states would have two different chemical shifts ( $\delta_A$  and  $\delta_B$ ). Then the NMR signal(s) of the nucleus depend on the exchange rate ( $k$ ), relative to the difference in the chemical shifts of the two states ( $\Delta\delta = \delta_A - \delta_B$  in frequency unit  $\text{s}^{-1}$ ). The exchange rate can be used to describe three extreme regimes; fast exchange ( $k \ll \Delta\delta$ ), intermediate ( $k \sim \Delta\delta$ ), and slow ( $k \gg \Delta\delta$ ). In the slow exchange regime, two distinctive resonances representing separate conformational states can be observed. In the fast exchange regime, a single chemical shift, representing the stoichiometrically averaged resonance, is observed. In the intermediate state, the two resonances coalesce resulting in a single broad, often undetectable, resonance.



(Figure 2.6)  $^{15}\text{N}$ ,  $^1\text{H}$  HSQC NMR spectra of the various constructs of RFC p140. Across the top the spectra of the free protein and the complex formed between p140(375-480) and oligo 1 are shown. The HSQC spectra of RFC p140 (375-545) and p140(480-545) are presented across the bottom. NMR spectra were measured at 298 K at 600 MHz. Examples of some of the resonances with unusually broad lines indicating intermediate exchange behavior are shown with arrows. The spectra of p140(480-545) is clearly indicative of an unstructured protein.

Sharp peaks deriving from the backbone amide moiety of the unstructured 6 His tag can also be seen ( $^1\text{H}$  7.4 –7.7 ppm,  $^{15}\text{N}$  124-128 ppm). In contrast, the spectrum of the DNA bound protein in Figure 2.6 shows good peak dispersion and more uniform linewidth. The  $^{15}\text{N}$ ,  $^1\text{H}$  HSQC spectrum of RFC p140(375-545), with its good peak dispersion and uniform linewidth, indicates a well ordered protein, even in the absence of DNA (Figure 2.6). However, the  $^{15}\text{N}$ ,  $^1\text{H}$  HSQC spectrum of RFC p140(480-545), characterized by very poor peak dispersion and narrow linewidths, suggests that this protein fragment is completely unstructured in solution (Figure 2.6).

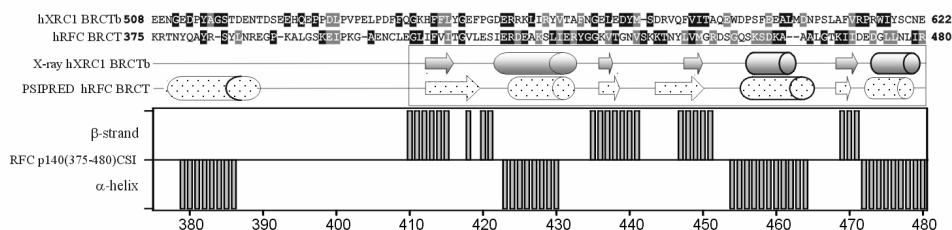
*Structure of DNA bound p140(375-480)*

Since the minimal DNA binding domain includes the N-terminal sequences outside of the conserved BRCT domain, we wished to ascertain whether or not it contained elements of regular secondary structure. As a first approach we used the PSIPRED secondary structure prediction service (29). The N-terminal sequences of the BRCT region were predicted to form an  $\alpha$ -helix with a high degree of confidence (Figure 2.7). As a control we performed similar analysis on other BRCT domain sequences with known structures and obtained a close correlation between the predicted and experimentally determined elements of secondary structure (Figure 2.7). As an example we show the alignment of the N-terminal BRCT domain of XRCC1, which forms specific, heterologous protein-protein complexes. The PSIPRED (v. 2.4) predicted secondary structure of the XRCC1 BRCT domain closely matches the secondary structure elements defined by the crystal structure (not shown). However, the region N-terminal to the BRCT domain is predicted to be in a random coil conformation.

Once the sequential chemical shift assignment of a protein is known, the chemical shift index (CSI) is a useful means of correlating the so-called secondary chemical shift (the difference between the observed chemical shift and that expected for the same residue in random coil) with the secondary structure of a protein (30). We have used  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeled protein in conjunction with standard triple resonance NMR experiments, to determine the essentially complete sequential assignment of the p140(375-480)-DNA complex (Chapter 4 supplementary materials Table 1S) (31) and have obtained a preliminary backbone assignment of the free protein. CSI analysis of the NMR data (Chapter 4 supplementary materials Figure 1S) was used to determine the secondary structure of the DNA bound RFC p140(375-480) (Figure 2.7). The secondary structure of residues 410 to 480 is consistent with that of other BRCT domains whose 3D structure has been determined by X-ray crystallography and closely matches that predicted by PSIPRED. Importantly, the  $\alpha$ -helix between residues 379 and 386 that is predicted by PSIPRED is experimentally confirmed by the CSI analysis of the backbone chemical shifts.

Our preliminary backbone resonance assignment of RFC p140(375-480) in the absence of DNA indicates that the secondary structural elements are retained but they are shorter, that is all 4 helices are present but each is reduced by one or two residues at either end as are the strands of the central  $\beta$ -sheet (data not shown). This analysis however is

complicated by the dynamic behavior of the unbound protein, a significant portion of which is either in intermediate conformational exchange on the NMR time scale or in chemical exchange with 1H's of water. As a result, there are gaps in the sequential assignment of the isolated p140 (375-480). The reduced secondary structure contents in the free protein suggested by NMR is also supported by the secondary structure contents derived from the CD of the free p140(375-480) (Figure 2.1C), which predicts 26% and 15% for  $\alpha$ -helices and  $\beta$ -strands respectively in comparison to the CSI predicted values of 31 % and of 21% for  $\alpha$ -helices and  $\beta$ -strands respectively in the DNA bound p140(375-480) (Figure 2.7).



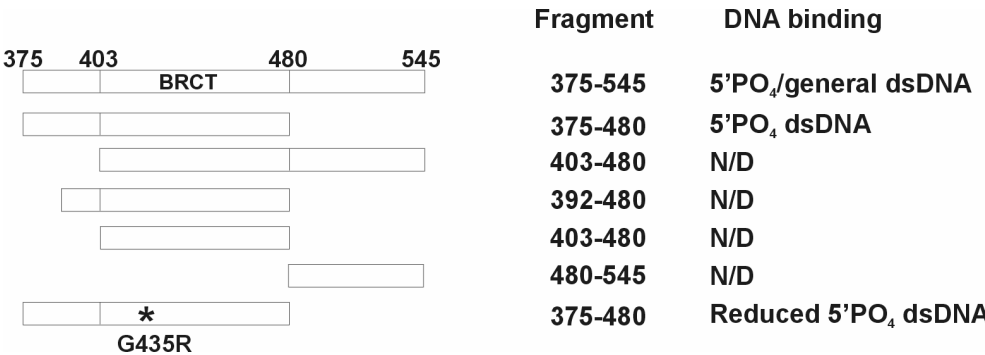
(Figure 2.7) Secondary structure analysis of RFC p140(375-480). The sequence alignment of human XRCC1 BRCTb and human RFC p140 BRCT region was performed as described in Figure 2.1A. The secondary structure elements, arrows for  $\beta$ -strands and cylinders for  $\alpha$  helices, are aligned with the corresponding amino acid sequences based on the X-ray crystal structure of XRCC1 (11) and PSIPRED (26) prediction of hRFCp140 (375-480). The consensus secondary structure predictions of hRFC p140 (375-480) by the chemical shift index method was generated from NMR Assignment data (31) using the C $\alpha$ , C $\beta$ , and H $\alpha$  chemical shifts as input.

## Discussion

### *Domain Boundaries required for DNA binding*

The results presented here (summarized in Figure 2.8) confirm and extend the work of others (1;2;17;18) concerning this interesting region of the p140 subunit of RFC. A minimal DNA binding domain appears to comprise residues 375 to 480 of human RFC p140, a region that contains 28 more residues at the N-terminus than the conserved BRCT domain. The importance of the N-terminal extension was indicated by the complete absence of DNA binding of N-terminally truncated proteins (Figure 2.8). Furthermore a mutation at one of the highly conserved glycine repeats found in the BRCT domain resulted in a reduction in DNA binding activity of p140(375-480) indicating that the binding necessitates both the intact N-terminal sequence and the BRCT domain. In yeast,

an equivalent single amino acid substitution at the conserved G193 to R in the BRCT domain of Rev1 reduced the efficiency of the lesion T–T pyrimidine pyrimidinone adduct bypass *in vivo* (32). Rev1 is a deoxycytidyl transferase, which adds dCMP opposite an abasic (apyrimidinic/apurinic) site generated by glycosylase activity at sites of DNA damage to bypass the damaged site. In addition, participation of Rev1 in bypass of a T–T pyrimidine pyrimidinone adduct has been reported (32). While the dCMP transferase activity is not required for bypass of T-T lesions, the G193R mutation suggests that the BRCT domain is (32).



(Figure 2.8) Summary of DNA binding study. N/D means “Not Detected”. \* indicates the amino acid substitution. The data of DNA binding by 403-545 and 480-545 are presented in the Appendix A1.

*Specificities of DNA binding by p140(375-480)*

Secondary structure based on NMR data indicates that the DNA bound p140(375-480) consists of an extra  $\alpha$ -helix at the N-terminus followed by a loop that connects to the BRCT domain which consists of three  $\alpha$ -helices and four  $\beta$ -strands. The p140(375-480) protein fragment has an absolute requirement for a free 5' phosphorylated terminus, and optimal binding requires 10 base pairs of fully duplex DNA. The binding stoichiometry of the DNA-protein complex is 1: 1, which reflects its specific binding to the single available 5'-phosphate on the hairpin oligonucleotide. However, there appears to be no restriction on the sequences that form the duplex region. Thus it is likely that the majority of protein contacts are made to the backbone and 5' phosphate of the DNA. Whether or not the BRCT domain actually contacts all ten base pairs of DNA awaits the elucidation of the structure of the complex. Alternatively, it is possible that stability of the duplex became limiting in our gel-shift experiments as the number of base pairs was reduced. This does not seem

likely however, since the entire assay was performed at 4 °C, a temperature at which even the short 6 bases duplex should be stable in the context of a hairpin oligonucleotide.

#### *DNA binding model of p140(375-545)*

Expression of a larger fragment of RFC p140 encompassing residues (375-545) resulted in a protein with different structural and DNA binding properties. RFC p140(375-545) forms two different complexes with oligo 1 that are resolvable in the gel retardation assay however with a much reduced DNA binding compared to that of the p140(375-480). The two complexes can be competed by DNA containing either a non-natural biotin or a hydroxyl group at the 5' position and therefore likely result from non-specific, dsDNA binding. This non-specific DNA binding of p140(375-545) is not due to the independent activity of the C-terminal aa's 480-545 because the isolated p140(480-545) is inactive in DNA binding. Instead the C-terminal aa's 480-545, may sterically hinder access to the putative 5' phosphate binding motif present in the residues 375-480 resulting in the reduced 5' phosphate-dependent DNA binding. Meanwhile the non-specific dsDNA binding is likely mediated by the aa's in the N-terminus 375-403 and the C-terminus 480-545, since the preferential non-specific dsDNA binding is present only in p140(375-545) and neither in p140(403-545) or p140(375-480) (Figure 2.8). However this model does not explain the observed two distinct DNA-p140(375-545) complexes. Further study is therefore necessary to understand in how the non-specific dsDNA binding and two distinct complexes are formed by p140(375-545).

While the p140(480-545) does not have a fold by itself, it is likely to take a part in the overall structure of p140(375-545), which is structurally better defined than p140(375-480), as judged by the good peak dispersion and relatively homogenous linewidth distribution exhibited in the [<sup>15</sup>N,<sup>1</sup>H] HSQC spectrum. Consequently the participation of aa's 408-545 in the structure of p140(375-545) may result in a preformed non-specific dsDNA binding site, which lacks in the p140(375-480). In contrast, a significant fraction of the shorter p140(375-480) peptide is likely in conformational exchange in the absence of its ligand but becomes more rigid upon binding DNA. Therefore it seems that the complete structural unit in the intact protein comprises residues 375-545.

The dual modes of DNA binding helps to explain observations that made by others working on RFC. The groups of Burgers and Hurwitz observed that 5 subunit complexes of both yeast (22) and human (23) RFC formed with an N-terminally truncated



p140 subunit (trRFC), in which the BRCT region was deleted, supported PCNA dependent DNA replication by pol $\delta$  more efficiently than complexes with full-length p140. The difference in DNA synthesis activity between the native RFC complex and the trRFC was greater for the synthetic substrate poly dA-oligo dT than for singly primed M13 DNA. Since neither substrate provides a ds, 5' PO<sub>4</sub> terminus it is likely that the non-specific dsDNA binding activity of the amino acid region 375-545 “trapped” RFC away from the 3' terminus where it is required for PCNA loading. This was further supported by the observation that trRFC exhibited a significantly lower non-specific dsDNA binding activity than the native RFC (33). Since poly dA-oligo dT has considerably more dsDNA than singly primed M13, non-productive binding would be expected to be greater on this substrate.

It was recently shown that hRFC p140 encompassing residues 361-545, which includes the sequences used in this study, can physically and functionally interact with the bZIP domain of transcription factors, C/EBP $\alpha$ , C/EBP $\beta$ , CREB, Jun and ATF2 (28). We have been unable to detect any evidence of direct protein-protein interaction between the BRCT region of the RFC p140 and the bZIP domain of C/EBP $\alpha$  using essentially the same proteins and solution conditions described in this work (28). We have however, demonstrated simultaneous binding of both proteins to short oligonucleotides bearing the C/EBP $\alpha$  recognition sequence and a 5' PO<sub>4</sub> (Supplementary materials S1) thereby providing an alternate explanation for the results described earlier (28).

The function of the entire N-terminal half of the p140 subunit of RFC, including the DNA binding domain discussed here, remains unclear. Deletion of this region in yeast did not result in an observable phenotype (22). *In vitro*, the DNA binding region of RFC p140 is not required for loading PCNA onto DNA nor for the elongation of primed DNA template (23;34). In light of recent observations indicating that many or possibly all, BRCT domains are phosphoserine binding modules, it is fair to ask the question as to whether the observed DNA binding activity is not an artifact of a binding site designed to interact with a phosphate group. While it is not yet possible to give a definitive answer to this question, the weight of evidence does not support it. DNA binding is tight (K<sub>D</sub> ~10 nM) and specific for 5' phosphorylated dsDNA rather than the single stranded DNA. Both 5' PO<sub>4</sub> specific and non-specific, dsDNA binding modes are observed in the full-length protein suggesting that they are not an artifact of the truncated proteins used in this study. It is difficult to

rationalize the non-specific dsDNA binding as a byproduct of the phosphate specificity. Further, the 5' PO<sub>4</sub> specific mode of binding could be biologically relevant since a substrate would be available on the lagging strand upon encountering a previously completed Okazaki fragment. Interestingly, recent data in prokaryotes suggests that stalled replication forks reinitiate daughter strand synthesis downstream of blockage sites thereby providing another potential source of 5' phosphorylated DNA termini (35;36). Of course, what the function of a 5' PO<sub>4</sub> bound RFC complex would be remains pure conjecture and therefore the focus of ongoing studies.

---

### Acknowledgement

---

The authors would like to thank Professor Ulrich Hübscher of the Universität Zürich for the kind gift of the original expression plasmid for human RFC p140(369-480) and Prof Bruce Stillman of the Cold Spring Harbor Laboratory for providing the full-length cDNA clone of RFC p140.

---

### Supplementary materials

---

S1.

#### *Interaction with the DBD (DNA binding domain) of C/EBP $\alpha$*

It has been reported that RFC can modulate the activity of transcription factors of the C/EBP family (28). In this report, the authors suggest a direct protein-protein interaction between the BRCT region of RFC and the bZIP DNA binding domain of C/EBP $\alpha$ . In order to investigate this interaction further we cloned the bZIP domain of C/EBP $\alpha$  and expressed it in *E. coli* with a His6 tag. We used 5' biotinylated oligonucleotides containing the C/EBP consensus binding sequence in a gel shift assay to detect binding to the purified bZIP protein (See Methods). As expected, a unique bZIP-DNA complex was observed (Figure S1A lane 3). 5' biotinylation was used to prevent interaction with RFC p140 (375-480) *via* 5' phosphate specific binding. As expected from the data in Figure 2.4, RFC p140(375-545) is also capable of binding the 5' biotinylated oligonucleotide (Figure S1A, lane 1) while RFC p140(375-480) did not (data not shown). When both the bZIP DBD of C/EBP $\alpha$  and RFC p140(375-545) were added, no additional, slowly migrating band was observed. Instead two distinct bands migrating at the same rate

as when the two proteins were added individually (Figure S1A, lane 5) were observed, suggesting that no ternary complex is formed between DNA, bZIP DBD and p140(375-545).

However, when the 5' biotin was replaced with a 5' phosphate while keeping the C/EBP $\alpha$  recognition sequence, very different results were obtained (Figure S1B). In this case all three proteins bound the CREB oligonucleotide (RFC p140(375-480), p140(375-545) and C/EBP $\alpha$  DBD) and migrated at different rates (lane 2, 3 and 4 from the left). When the C/EBP $\alpha$  DBD is combined with either RFC p140(375-480) or RFC p140(375-545), multiple complexes were observed. First, a band corresponding to DNA-DBD and DNA-RFCp140 complexes can be discerned individually. Second, a more slowly migrating complex is observed. The amount of the slow migrating complex increases with the amount of RFC p140 added. This slowly migrating complex likely corresponds to one RFC p140 and one C/EBP $\alpha$  DBD bound to the same oligonucleotide. Taken together with the observation made in Figure S1A, the slowly migrating complex appears to be the result of two separate events: one, that the C/EBP $\alpha$  DBD binds its cognate DNA sequence while two, RFC p140(375-480) or p140(375-545) binds the 5' phosphorylated terminus. The data clearly indicate that the slowly migrating band is not the result of a physical interaction between the two different proteins. At the bottom of the gel, the free DNA migrates as two distinct complexes. These complexes correspond to a majority of intermolecular (upper band) and a minority of intramolecular (lower band) annealed strands deriving from the self-complementarity of the oligonucleotide.

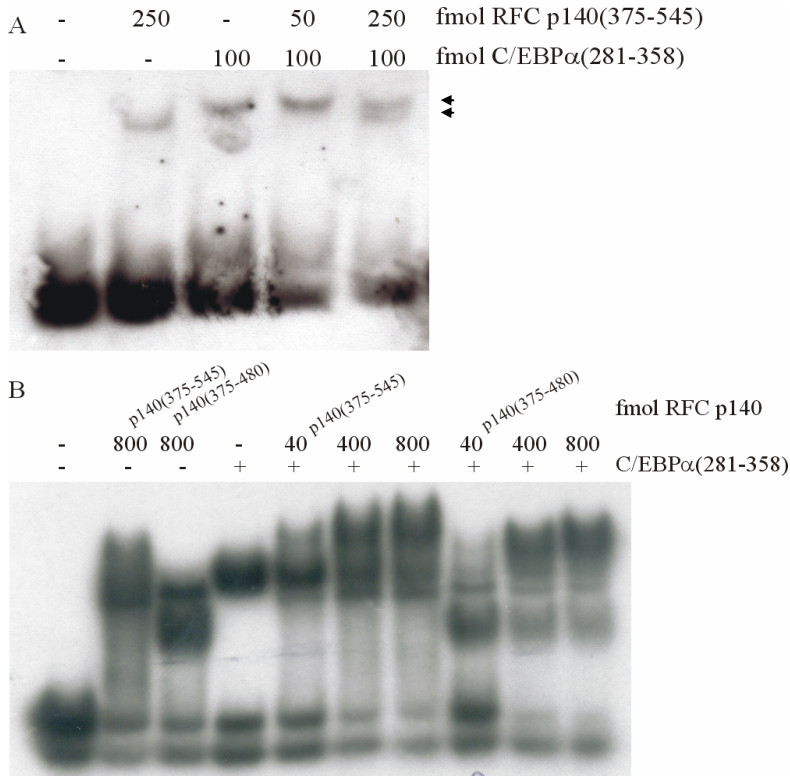


Figure S1. RFCp140 (375-545) does not interact with bZIP domain of C/EBP $\alpha$ .

**(A).** A gel retardation assay was used to investigate the interaction between RFCp140 (375-545) and the bZIP domain (281-358) of c/EBP $\alpha$ . 100fmoles of a 5' biotinylated oligonucleotide (5'B in Table I) containing the C/EBP binding sequence was mixed with the indicated amounts of either RFCp140 (375-545), C/EBP $\alpha$  (281-358) or both proteins together (black arrows). A complex containing both RFC and C/EBP $\alpha$  is not observed. **(B).** For comparison a similar experiment was performed using the same oligonucleotide as in A that was 5' labeled. The quantity of the C/EBP $\alpha$  bZIP domain (790 fmoles) and DNA (40 fmoles) was held constant, while the indicated amount of either of the two RFC p140 constructs was titrated in. The free DNA migrates as two bands indicating both inter- and intra-molecular hybridization of the self-complementary sequence. When a sufficient quantity of either of the two RFC p140 constructs is added, a band corresponding to just RFC bound DNA as well as a super-shifted band that appears to contain both RFC and C/EBP $\alpha$  can be observed.

---

## Reference list

---

1. Bork,P., Hofmann,K., Bucher,P., Neuwald,A.F., Altschul,S.F. and Koonin,E.V. (1997) A superfamily of conserved domains in DNA damage responsive cell cycle checkpoint proteins. *FASEB J.*, **11**, 68-76.
2. Callebaut,I. and Mornon,J.P. (1997) From BRCA1 to RAP1: A widespread BRCT module closely associated with DNA repair. *FEBS Lett.*, **400**, 25-30.
3. Nash,R.A., Caldecott,K.W., Barnes,D.E. and Lindahl,T. (1997) XRCC1 protein interacts with one of two distinct forms of DNA ligase III. *Biochemistry*, **36**, 5207-5211.
4. Masson,M., Niedergang,C., Schreiber,V., Muller,S., Menissier-de Murcia,J. and de Murcia,G. (1998) XRCC1 is specifically associated with poly(ADP-ribose) polymerase and negatively regulates its activity following DNA damage. *Mol.Cell.Biol.*, **18**, 3563-3571.
5. Soulier,J. and Lowndes,N.F. (1999) The BRCT domain of the *S-cerevisiae* checkpoint protein Rad9 mediates a Rad9-Rad9 interaction after DNA damage. *Curr.Biol.*, **9**, 551-554.
6. Liu,K., Lin,F.T., Ruppert,J.M. and Lin,W.C. (2003) Regulation of E2F1 by BRCT domain-containing protein TopBP1. *Mol.Cell.Biol.*, **23**, 3287-3304.
7. Manke,I.A., Lowery,D.M., Nguyen,A. and Yaffe,M.B. (2003) BRCT Repeats As Phosphopeptide-Binding Modules Involved in Protein Targeting. *Science*, **302**, 636-639.
8. Yu,X., Chini,C.C.S., He,M., Mer,G. and Chen,J.J. (2003) The BRCT Domain Is a Phospho-Protein Binding Domain. *Science*, **302**, 639-642.
9. Zhang,X.D., Morera,S., Bates,P.A., Whitehead,P.C., Coffey,A.I., Hainbuecher,K., Nash,R.A., Sternberg,M.J.E., Lindahl,T. and Freemont,P.S. (1998) Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. *EMBO J.*, **17**, 6404-6411.
10. Krishnan,V.V., Thornton,K.H., Thelen,M.P. and Cosman,M. (2001) Solution structure and backbone dynamics of the human DNA ligase III alpha BRCT domain. *Biochemistry*, **40**, 13158-13166.
11. Williams,R.S., Green,R. and Glover,J.N.M. (2001) Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1. *Nat.Struct.Biol.*, **8**, 838-842.
12. Derbyshire,D.J., Basu,B.P., Serpell,L.C., Joo,W.S., Date,T., Iwabuchi,K. and Doherty,A.J. (2002) Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor. *EMBO J.*, **21**, 3863-3872.
13. Lee,J.Y., Chang,C., Song,H.K., Moon,J., Yang,J.K., Kim,H.K., Kwon,S.T. and Suh,S.W. (2000) Crystal structure of NAD(+)-dependent DNA ligase: modular architecture and functional implications. *EMBO J.*, **19**, 1119-1129.
14. Waga,S. and Stillman,B. (1998) The DNA replication fork in eukaryotic cells. *Annu.Rev.Biochem.*, **67**, 721-751.
15. Mossi,R. and Hubscher,U. (1998) Clamping down on clamps and clamp loaders - The eukaryotic replication factor C. *Eur.J.Biochem.*, **254**, 209-216.
16. Reynolds,N., Fantes,P.A. and MacNeill,S.A. (1999) A key role for replication factor C in DNA replication checkpoint function in fission yeast. *Nucleic Acids Res.*, **27**, 462-469.

17. Fotedar,R., Mossi,R., Fitzgerald,P., Rousselle,T., Maga,G., Brickner,H., Messier,H., Kasibhatla,S., Hubscher,U. and Fotedar,A. (1996) A conserved domain of the large subunit of replication factor C binds PCNA and acts like a dominant negative inhibitor of DNA replication in mammalian cells. *EMBO J.*, **15**, 4423-4433.
18. Allen,B.L., Uhlmann,F., Gaur,L.K., Mulder,B.A., Posey,K.L., Jones,L.B. and Hardin,S.H. (1998) DNA recognition properties of the N-terminal DNA binding domain within the large subunit of Replication Factor C. *Nucleic Acids Res.*, **26**, 3877-3882.
19. Burbelo,P.D., Utani,A., Pan,Z.Q. and Yamada,Y. (1993) Cloning of the Large Subunit of Activator-1 (Replication Factor-C) Reveals Homology with Bacterial-Dna Ligases. *Proc.Natl.Acad.Sci.U.S.A.*, **90**, 11543-11547.
20. Tsurimoto,T. and Stillman,B. (1991) Replication Factors Required for Sv40 Dna-Replication In vitro .1. Dna Structure-Specific Recognition of A Primer-Template Junction by Eukaryotic Dna-Polymerases and Their Accessory Proteins. *J.Biol.Chem.*, **266**, 1950-1960.
21. Halligan,B.D., Teng,M., Guilliams,T.G., Nauert,J.B. and Halligan,N.L.N. (1995) Cloning of the Murine Cdna-Encoding Vdjp, A Protein Homologous to the Large Subunit of Replication Factor-C and Bacterial-Dna Ligases. *Gene*, **161**, 217-222.
22. Gomes,X.V., Gary,S.L. and Burgers,P.M.J. (2000) Overproduction in Escherichia coli and characterization of yeast replication factor C lacking the ligase homology domain. *J.Biol.Chem.*, **275**, 14541-14549.
23. Uhlmann,F., Cai,J.S., Gibbs,E., ODonnell,M. and Hurwitz,J. (1997) Deletion analysis of the large subunit p140 in human replication factor C reveals regions required for complex formation and replication activities. *J.Biol.Chem.*, **272**, 10058-10064.
24. Piotto,M., Saudek,V. and Sklenar,V. (1992) GRADIENT-TAILORED EXCITATION FOR SINGLE-QUANTUM NMR- SPECTROSCOPY OF AQUEOUS-SOLUTIONS. *J.Biomol.NMR*, **2**, 661-665.
25. Wishart,D.S. and SYKES,B.D. (1994) The C-13 Chemical-Shift Index - A Simple Method for the Identification of Protein Secondary Structure Using C-13 Chemical-Shift Data. *J.Biomol.NMR*, **4**, 171-180.
26. Andrade,M.A., Chacon,P., Merelo,J.J. and Moran,F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng*, **6**, 383-390.
27. Zhang,X., Morera,S., Bates,P.A., Whitehead,P.C., Coffey,A.I., Hainbuecher,K., Nash,R.A., Sternberg,M.J., Lindahl,T. and Freemont,P.S. (1998) Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. *EMBO J.*, **17**, 6404-6411.
28. Hong,S.H., Park,S.J., Kong,H.J., Shuman,J.D. and Cheong,J.H. (2001) Functional interaction of bZIP proteins and the large subunit of replication factor C in liver and adipose cells. *J.Biol.Chem.*, **276**, 28098-28105.
29. McGuffin,L.J., Bryson,K. and Jones,D.T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16**, 404-405.
30. Wishart,D.S. and SYKES,B.D. (1994) CHEMICAL-SHIFTS AS A TOOL FOR STRUCTURE DETERMINATION. *Methods in Enzymology*. Vol. 239, pp. 363-392.
31. Kobayashi,M. and Siegal,G. (2005) 1H, 15N and 13C resonance assignments of the BRCT region of the large subunit of human Replication Factor C. *J.Biomol.NMR*, **31**, 183-184.

32. Nelson,J.R., Gibbs,P.E., Nowicka,A.M., Hinkle,D.C. and Lawrence,C.W. (2000) Evidence for a second function for *Saccharomyces cerevisiae* Rev1p. *Mol.Microbiol.*, **37**, 549-554.
33. Hingorani,M.M. and Coman,M.M. (2002) On the specificity of interaction between the *Saccharomyces cerevisiae* clamp loader replication factor C and primed DNA templates during DNA replication. *J.Biol.Chem.*, **277**, 47213-47224.
34. Podust,V.N., Tiwari,N., Stephan,S. and Fanning,E. (1998) Replication factor C disengages from proliferating cell nuclear antigen (PCNA) upon sliding clamp formation, and PCNA itself tethers DNA polymerase delta to DNA. *J.Biol.Chem.*, **273**, 31992-31999.
35. Higuchi,K., Katayama,T., Iwai,S., Hidaka,M., Horiuchi,T. and Maki,H. (2003) Fate of DNA replication fork encountering a single DNA lesion during oriC plasmid DNA replication in vitro. *Genes Cells*, **8**, 437-449.
36. McNerney,P. and O'Donnell,M. (2004) Functional uncoupling of twin polymerases - Mechanism of polymerase dissociation from a lagging-strand block. *J.Biol.Chem.*, **279**, 21543-21551.

## Chapter 3

# Amino acid determinants for DNA binding by the BRCT region of human RFC p140 subunit

---

### Abstract

---

Sequence alignment of the BRCT region in the p140 subunit of Replication Factor C from various eukaryotes has identified very few conserved amino acid residues within the core BRCT domain, while none were found in sequences immediately N-terminal to the BRCT domain. However, mapping of the limited number of conserved residues that were found, onto a homology model of the BRCT domain revealed a cluster of conserved residues on one side of the molecular surface. Site-directed mutagenesis was performed on these residues as well as some lesser conserved amino acids of the N-terminal motif of the BRCT region of human p140. The structural changes due to the introduced mutations were rigorously checked using 1D  $^1\text{H}$ -NMR and CD spectroscopy of the purified proteins. Mutation of weakly conserved residues in the region N-terminal to the BRCT domain generally resulted in reduced DNA binding. Mutation of any of the conserved residues within the BRCT domain negatively affected DNA binding. This latter group of residues is conserved amongst the BRCT domains from the bacterial DNA ligases, which recently have been shown to bind DNA. Furthermore, these conserved residues are positionally equivalent to those required for phosphate recognition by the tandem BRCT domains from BRCA1 suggesting that they may be involved in binding the 5' phosphate of dsDNA.

M, Kobayashi, F, Figaroa and G. Siegal (Parts of this chapter and of chapter 2 will be submitted for publication)



---

Introduction

---

The BRCT domain (**BRCA1 C-Terminus**) was first identified as a tandem repeat of roughly 90 amino acids at the C- terminus of the BRCA1 (Breast Cancer susceptibility 1) protein. Extensive amino acid sequence profiling led to the discovery of a vast number of proteins (current entry 915 open reading frames in Pfam) carrying the BRCT domain and, strikingly, most of those characterized are either directly or indirectly associated with various aspects of DNA metabolism including: DNA repair, DNA replication or cell cycle-checkpoint (1). Despite the low sequence similarity, all the BRCTs with known structures share the same overall protein fold (2-4). BRCT domains have been reported to be responsible for homotypic (BRCT-BRCT) or heterotypic (BRCT-non BRCT) protein-protein interactions. For example, in the crystal structure of the BRCT domains of XRCC1 (5), there is evidence that a homo-dimer is formed through both hydrophobic and salt-bridge interactions, the same that mediate hetero dimer formation between the BRCT domains of XRCC1 and DNA ligase III (6). Additionally, recent studies have revealed that some BRCT domains bind specifically to a phosphoserine, pointing towards a role in signal transduction events that regulate the various DNA repair pathways (7-9). The subsequent structure determination of a complex between the tandem BRCT repeat of BRCA1 and a phosphopeptide derived from BACH1 revealed the detailed molecular basis of how the phosphopeptide is specifically recognized (10-12).

Replication factor C (RFC) is a five subunit complex that plays an important role in efficient loading of PCNA onto primer-template DNA during the synthesis of the daughter strands in DNA replication. RFC consists of four subunits of 35-40 kDa and a fifth large subunit (p140) of 140 kDa. The C-terminus of p140 shares homology with the four small subunits, while the unique N-terminal sequence (roughly 50 kDa), which contains a single BRCT domain, is dispensable for PCNA loading. This N-terminal sequence has been shown to bind dsDNA at the 5'-phosphorylated terminus. The region of the sequence responsible for this DNA binding activity was mapped and it includes the single BRCT domain (13). Truncation studies indicated that the BRCT domain from human RFC p140 cannot bind to DNA by itself but requires an extra 28 amino acids at its N-terminus (Figure 2.1). Similar requirements have also been reported for *Drosophila* RFC p140 (13), therefore extending the 5'-phosphate binding activity to what here is referred to as the "BRCT region".

While the majority of BRCT domains are, or are thought to be, bona fide protein-protein interaction modules, only a limited number appear to bind DNA. The BRCT region of RFC p140 is a member of this small class of DNA binding BRCT domains. In order to further characterize its DNA binding properties, site-directed mutagenesis has been performed on the BRCT region of RFC p140 to delineate potential residues involved. Sequence alignment and homology modeling of the human BRCT region revealed an interesting cluster of conserved residues on one side of the molecular surface. Amino acid substitution of any of these residues resulted in mutants with low or undetectable DNA binding activity. Despite the lack of absolute amino acid conservation, mutations of the moderately conserved amino acids within the N-terminal sequence of the BRCT region also diminished DNA binding activity, further supporting the direct involvement of these sequences in DNA binding.

---

## Materials and Methods

---

### *Site directed mutagenesis*

The procedure for generating point mutations was adopted from the QuickChange method (Stratagene) with some modifications. 125 ng of each primer (table 2) and 25 ng of template DNA, pET20b RFCp140 375-480, were used in the reaction. The PCR program used consisted of the following: 1 cycle of 95°C for 30 seconds and 18 cycles of steps consisting of 30 sec 95°C, 2 min 55°C, 8 min 68°C. After PCR, 10 U of DpnI was added to the reactions and incubated 1h at 37°C. Subsequently 10 U of Pfu polymerase was added to the reaction and incubated 30 min at 72°C, which was followed by addition of 10 U of DpnI and further incubation for 1h at 37°C. The reaction mix was then finally transformed into *E.coli* DH5 $\alpha$  supercompetent cells. The plasmids carrying the desired mutations were identified by the unique silent restriction site introduced via the mutagenesis primers and were further verified by DNA sequencing. The (R388A) mutation, was generated using SOE (Splice Overlap Extension)(14).

### *Protein expression and purifications*

*E.coli* strain BL21 (DE3) pLysS was transformed with pET20b RFC p140 (375-480) bearing one of the 9 site-directed mutations and grown in LB medium containing 4  $\mu$ g/ml Ampicilin and 34  $\mu$ g/ml Chloramphenicol. Protein expression was induced with 1mM IPTG and the cells were harvested after 3 hours. The mutant proteins were purified

using the identical protocol used for the wildtype protein described in Chapter 2. The concentration of mutants was measure using UV absorbance at 277 nm ( $\epsilon_{277\text{nm}} = 7.32 \text{ mM}^{-1} \text{ cm}^{-1}$ ). For the mutants Y382A and Y385E, extinction coefficients were calculated according to the amino acid contents using ProtParm (<http://au.expasy.org/tools/protparam.html>). The extinction coefficient of the native mutants was obtained by comparing the absorbance at 277 nm measured at native and unfolded mutants in 8M urea.

#### *Measurement of 1D $^1\text{H}$ -NMR and circular dichroism spectra*

Each mutated protein was exchanged into a buffer consisting of 25mM Tris-HCl pH 7.5, 50mM NaCl and 1mM DTT. Protein concentration varied from 0.05 to 0.1mM as determined by UV absorbance at 277 nm. 1D  $^1\text{H}$ -NMR spectra of each protein were recorded on a 500 MHz Bruker AVANCE instrument at 298 K (25 °C ) and processed with TOPSPIN 1.3 (Bruker). For the circular dichroism measurements, the purified mutants were exchanged into 10 mM Potassium Phosphate buffer (pH 7.7) using PD10 columns (Amersham Bioscience). Protein concentration was adjusted to 10  $\mu\text{M}$  (K397E, T415A, R423A, K458E and wildtype), and circular dichroism (CD) spectra were measured at 21 °C with scanning 1 nm increment between 195-260 nm in 0.1 mm path-length cuvette using a Jobin Yvon CD6 instrument. For each protein, five spectra were recorded to obtain an average spectrum from which the background spectrum of the buffer was subtracted.

#### *Detection of DNA-protein complex*

The sequence of the DNA used in the binding assay was designed to form a hairpin with 10 base pairs (5'pCTCGAGGTCG**TCATCGACCTCGAGATCA** 3', the letters in bold will form the loop). The 5' was radioactively labeled by exchange using T4 polynucleotide kinase and [ $\gamma$ - $^{32}\text{P}$ ] ATP as a substrate and purified as described in Methods Chapter 2. The mutant proteins were diluted in a buffer of 10 mM HEPES pH 7.8, 2 mM  $\text{MgCl}_2$ , 0.1 mM EDTA, 100  $\mu\text{g/ml}$  BSA, 15 % glycerol, 0.8  $\mu\text{g/ml}$  poly (dI-dC) and 2 mM DTT. Detection of DNA-protein complex was performed using the identical gel retardation assay as described in the Methods of Chapter 2. Radioactively labeled DNA was detected using film (Kodak Xomat XAR) or a phosphor-imager (BioRad) for quantification. The fraction of DNA in the DNA-protein complex was calculated as; Percentage of total shift (%) = [complex – background] / [total activity per lane - background] x 100, where total activity = complex + free DNA.

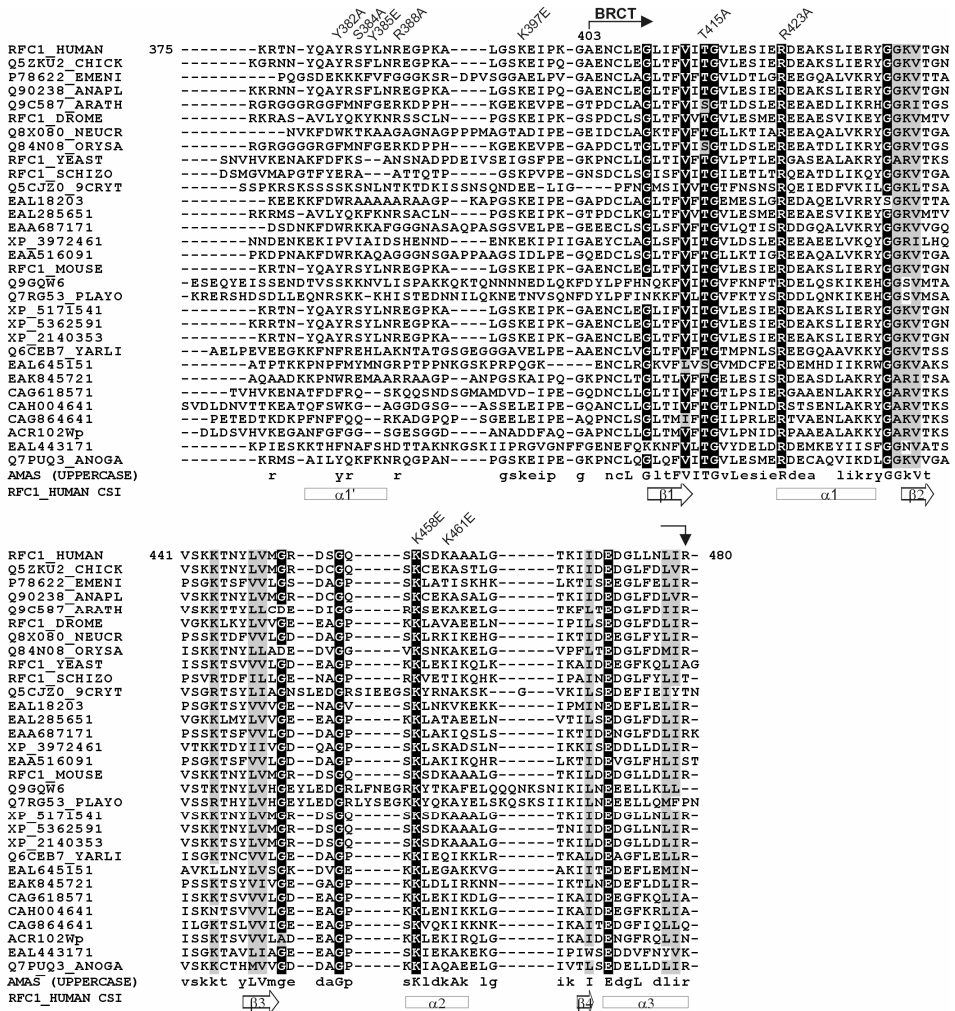
---

## Results

---

### *Design of mutations via amino acid sequence alignment and homology modeling of RFC p140 (375-480)*

The BRCT region is defined the sequences between aa 375 and 480 of the human RFC (or “p140(375-80)”) whereas the region between 403 and 480 (or “p140(403-480)”) is referred to as the BRCT domain. The BRCT region binds dsDNA with a 5' phosphorylated terminus (Figure 2.2). In order to find potential residues involved in the recognition of DNA, a sequence alignment of the BRCT region in the p140 subunit of 31 eukaryotic RFC complexes was analyzed, allowing the identification of a few highly conserved residues (Figure 3.1). The sequence alignment was generated using Clustal W (15;16) and the residue conservation in the alignment was further analyzed using AMAS (Analysis of Multiply Aligned Sequence)(17), which determines the degree of conservation for each position of the consensus on the basis of the amino acid properties. A limited number of conserved residues were identified within the BRCT domain while the N-terminal sequence of the BRCT region lacked any conservation among all 31 species (Figure 3.1) despite its indispensable nature in DNA binding (Figure 2.2). The amino acid backbone carbon and proton chemical shifts of p140(375-480) bound to dsDNA have been determined by NMR and with such values, one can reliably predict the secondary structure content in the protein by comparison with the chemical shift in the random coil state (Chapter 4, supplementary materials Figure 1S). This prediction suggested that the non-conserved N-terminal sequence contains a helix ( $\alpha 1'$ ) which is followed by a long loop that connects it to the BRCT domain (Figure 3.1 bottom). The BRCT domain itself consists of the consensus secondary structure ( $\beta 1$ - $\alpha 1$ - $\beta 2$ - $\beta 3$ - $\alpha 2$ - $\beta 4$ - $\alpha 3$ ) commonly found in BRCT folds (2;5;18) (Figure 3.1 bottom).

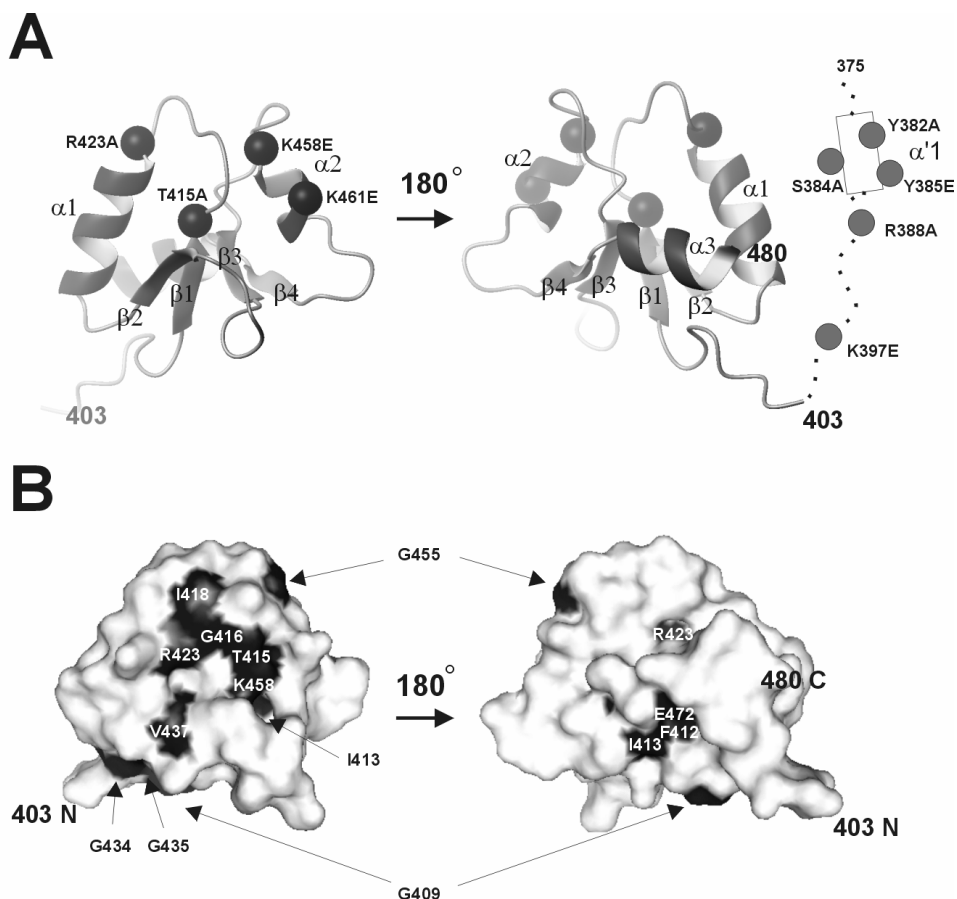


(Figure 3.1) Amino acid sequence alignment of the BRCT region of the large subunit of eukaryotic RFC. Amino acid sequences were retrieved from GenBank under the accession codes to the left of the sequence. The first raw contains the amino acid sequence of the BRCT region from human RFC p140 (RFC1\_HUMAN). The alignment was generated by ClustalW (16). Residue identity above 90 % is shaded black and residue similarity above 90% is shaded grey. Analysis of the alignment by AMAS identified the conserved residues (AMAS threshold above 7), which were presented in uppercase letters at the bottom of the sequence. The lowercase letter represents an amino acid identity above 50 %. The last row indicates the secondary structures defined by the NMR data of the human p140 (375-480) (box for alpha helix and arrow for beta strands) based on secondary structure determined by the program CSI 1.0 using chemical shifts of Ca, Cβ and Hα as input (Chapter 4 supplementary materials Figure 1S).

A homology model of the RFC BRCT domain (403-480) (Figure 3.2) was generated using 3D-JIGSAW using the available structure of the BRCT domain from the bacterial NAD<sup>+</sup> dependent DNA ligase, which has an amino acid sequence identity of 33 % (Figure 4.3), (19). The resulting model consists of four parallel β-strands (β1-β4) forming the core, which are surrounded by three α-helices (α1-α3) (Figure 3.2). The N-

terminal sequence between residues 375 and 402 is not included in the model due to the absence of sequence homology, hence the loop (387-402) and the predicted helix (379-386) are schematically presented as a dotted line and a rectangle ( $\alpha 1'$ ) respectively (Figure 3.2A right). Most of the conserved hydrophobic residues are involved in stabilizing the core created by residues within the parallel  $\beta$ -strands and the  $\alpha$ -helices packed against it (3-5). The structural importance of these conserved hydrophobic residues in the BRCT domain has been demonstrated by reduced stability of the folded protein when mutated (6). Mapping of the conserved residues onto the model structure reveals an interesting pattern where the majority of the conserved residues that are solvent-accessible cluster on either side of the molecular surfaces (Figure 3.2B).

Based on the sequence alignment and the structural model, amino acid residues identified as conserved or with an identity of at least 50 % were selected for mutagenesis. Conserved hydrophobic residues contributing purely to the stability of folding were excluded from mutagenesis. The surface exposed GG sequence (434, 435) (Figure 3.2B), highly conserved throughout the BRCT family, forms a tight turn between  $\alpha 1$  and  $\beta 2$ , which can be disrupted by substitution with bulky residues (Figure 2.2). Therefore this sequence was also excluded from mutagenesis. The assumption was made that most residues involved in DNA binding would do so via salt bridge or hydrogen bond interactions. Accordingly we selected basic and polar residues for mutagenesis. The final list includes three residues (Y382, S384, Y385) from the predicted  $\alpha 1'$  helix, two residues within the connecting loop (R388, K397), and four residues (T415, R423, K458 and K461) within the BRCT domain itself (Figure 3.1 and 2 A). These residues were preferably substituted with the negatively charged glutamic acid, if the substitution was not expected to disturb the secondary structure predicted by PSIPred (20). If the X to E substitution appeared to have deleterious structural consequences the residue X was substituted by an alanine. Therefore, four residues, Y385, K397, K458 and K461 were substituted with glutamic acid, and five residues, Y382, S384, R388, T415 and R423 were substituted with alanine. Mutated residues in this study are indicated as blue balls on the structure model in Figure 3.2A.



(Figure 3.2) Homology model of the BRCT domain from RFC p140 (403-480).

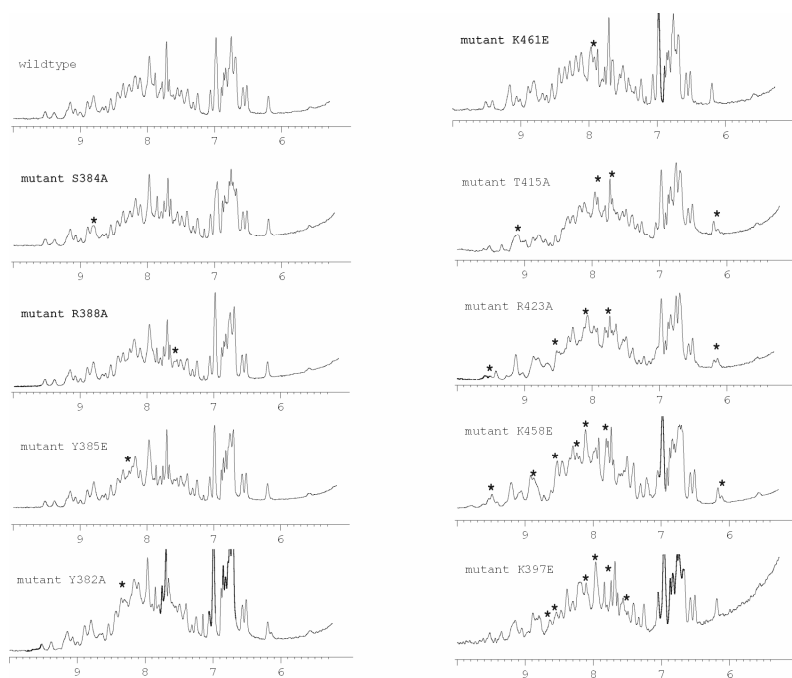
(A). Ribbon presentation of the model BRCT domain. The model was based on the NMR structure of the BRCT domain (PDB code: 1IB7) from the NAD<sup>+</sup> dependent DNA ligase which has 33 % amino acids identity with RFC p140 and was generated by 3D-JIGSAW (19). The residues where structural homology was not available are schematically indicated in the figure. The predicted helix between residues 379 and 386 is presented as a box and the loop (387-402) is presented as a dotted line. The blue balls represent the C $\alpha$  locations of the point mutations created in this study. (B). Surface presentation of the model BRCT domain. The blue-shaded areas represent the accessible surface occupied by the conserved amino acids defined from the amino acid sequence alignment (Figure 3.1) using AMAS (17). The total accessible surface area (ASA) were calculated using MOLMOL (21).

### Mutagenesis and purification of proteins

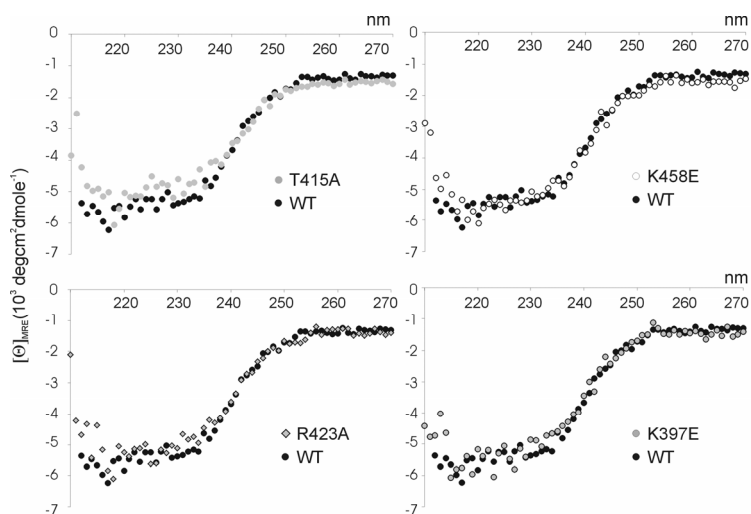
Mutagenesis was performed using the whole plasmid PCR (WHOP) (see Materials and Methods) or the splicing overlap extension (SOE) protocols (14). The resulting mutants were sequenced in the expression vector and the proteins were expressed in the BL21 (DE3) strain of *E. coli*. All nine mutants contained a 6 histidine tag at the C-terminus and were purified using the identical protocol as for the wildtype protein (this thesis Chapter 2). All mutants eluted at the same position as the wt protein during gel

filtration (superpose 12) indicating that the mutants have a similar radius of gyration as the wildtype. Presumably this similarity is due to preservation of the wildtype fold and multimeric state. Proteins were analyzed using SDS-PAGE to determine purity and quantified by UV-absorbance (Materials and Methods). The existence of natively folded protein in the purified fraction was further confirmed by 1D  $^1\text{H}$  NMR spectroscopy comparing the spectral region containing resonances from the backbone amide  $^1\text{H}$ 's (6 ~ 10 ppm) and aliphatic  $^1\text{H}$ 's (0 ~ 4 ppm). In Figure 3 the spectrum of each mutant protein is presented in order of decreasing similarity to the wt protein based on the number of missing and shifted peaks identified by visual inspection of each spectrum. Asterisks in the mutant spectra indicate the resonances missing or shifted from the  $^1\text{H}$  amide backbone region of the wt spectrum (Figure 3.3). Mutants (T415A and R423A) exhibit increased linewidth in comparison to the wt, indicating possible aggregation in solution or structural dynamics displayed by the mutants, which are different from that of wildtype. There were no significant differences in the resonances arising from  $^1\text{H}$  aliphatic carbons (results not shown). The mutants (T415, R423, K397E and K458E) that displayed spectra with some deviations from the wt protein were further checked using CD spectroscopy for secondary structure content (Figure 3.4). Due to the quality of the CD spectra, it is not possible to deduce the exact content of secondary structures in the proteins. However there are no noticeable major difference between the CD spectra of the wildtype and the mutants (Figure 3.4). The overall similarity of the  $^1\text{H}$  NMR spectra and the qualitative measure indicating existence of secondary structures in the CD spectra suggested that all the mutants were folded with only minor structural changes.





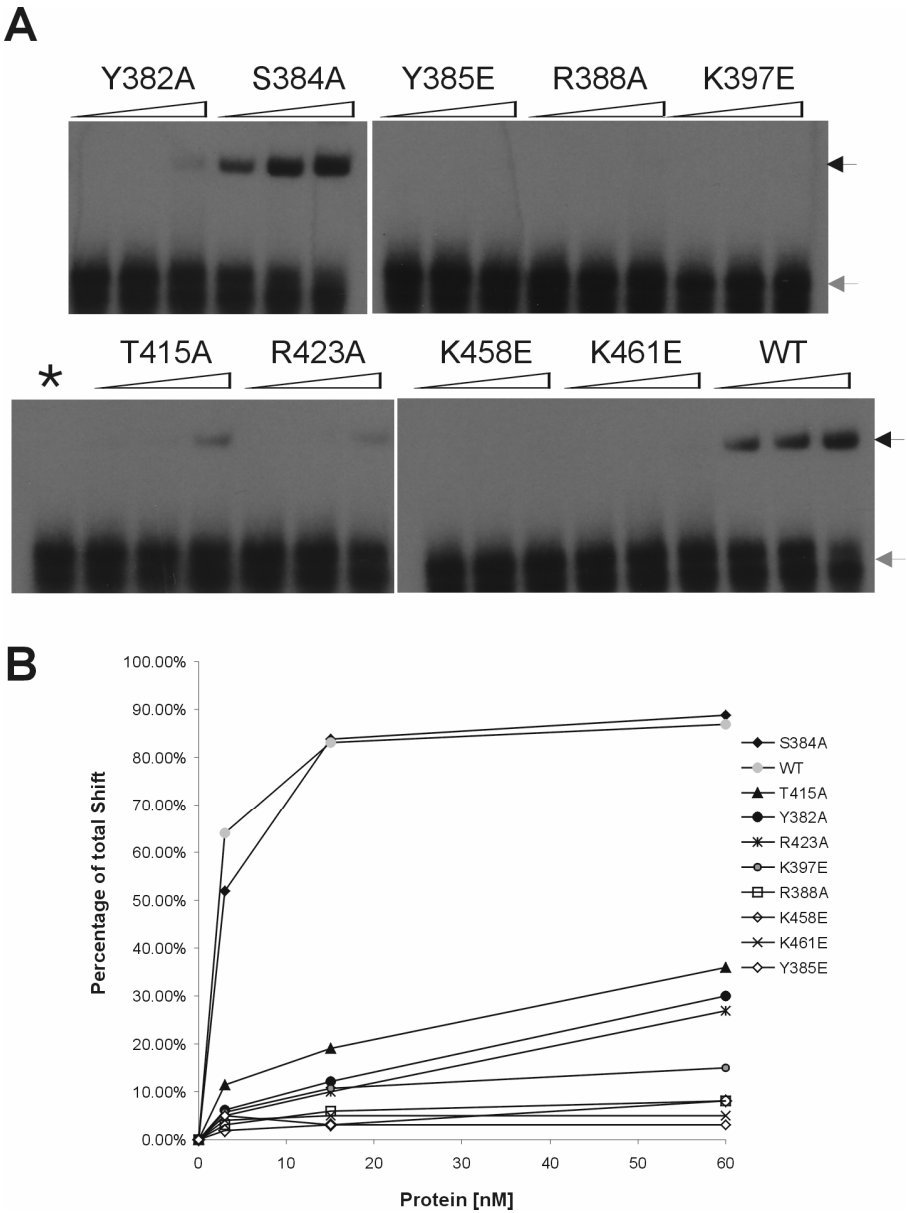
(Figure 3.3) 1D  $^1\text{H}$ -NMR spectra of the purified p140(375-480) mutants. 1D  $^1\text{H}$ -NMR spectra were recorded of each purified mutant protein described in the text. The good dispersion of resonances in the spectra region  $> 6.5$  ppm, representing primarily amide protons, is indicative of a folded protein. The similarity of each spectrum to that of the wt protein (the top left lane) indicates a nearly native fold. Resonances that either disappear or are different from the wt spectrum are indicated with asterisks. The spectra of the mutants (indicated on top left corner of the each spectra) are ordered in decreasing similarity to the wt protein (descending order of spectra from the top left to the bottom left and then from top right to the bottom right).



(Figure 3.4) CD spectra of the mutants T415A, R423A, K458E and K397E in comparison to that of wildtype (WT) protein. No significant differences in the secondary structure content between the mutants and wildtype can be seen.

*Effect of mutations on DNA-protein complex formation*

The effect of each mutation on dsDNA binding was assessed using the gel shift assay. The previously determined binding constant ( $K_D$ ) of p140 (375-480) for 5' phosphorylated, dsDNA is approximately 10 nM. Under identical conditions, a constant amount of 5' [ $^{32}$ P]-labeled ds oligonucleotide (3 nM) was titrated with an increasing amount of each mutant protein (Figure 3.5 A)(3 nM, 16 nM and 66 nM). The data was analyzed with a phosphor-imager and the fraction of DNA bound protein was calculated (Materials and Methods) for each mutant (Figure 3.5B). Only the S384A mutant exhibited DNA binding equivalent to wt (Figure 3.5B). The mutations, Y385E, R388A, K458E and K461E, essentially abrogated DNA binding activity (Figure 3.5). Significant reductions in DNA binding were also observed for the mutations Y382A, T415A and R423A (Figure 3.5). As might be expected, the substitution of residues with the negatively charged glutamic acid sidechain affected DNA binding more drastically than substitution by alanine (Figure 3.5B). This observation likely reflects the greater change in surface electrostatic potential caused by incorporation of a charged residue than a neutral residue.



(Figure 3.5) **The effect of various point mutations on the DNA binding activity of RFC p140 (375-480).** (A) Titration of DNA-protein complex was performed by adding an increasing amount of each protein (3, 15, 60 nM) to a constant amount (3 nM) of the [ $^{32}$ P] labeled hairpin oligonucleotide. The resulting DNA-protein complex (black arrow) was separated from the free DNA (grey arrow) in a gel retardation assay. The amount of the complex formed by wildtype (WT) and mutants proteins (labeled on top of the lanes) were compared. The lanes with an asterisk contained only the labeled oligonucleotides as a negative control. (B) Quantification of DNA-protein complex formation. The intensity of the bands in the A were determined using a Phosphor-imager. The percentage of the total shift (DNA-protein complex, black arrow in A) was calculated as described in Materials and Methods.

---

Discussion

---

*Structural consequences of mutagenesis*

We used site-directed mutagenesis in order to delineate a potential DNA binding site within the BRCT region of RFC p140. This region contains 28 residues N-terminal to the consensus BRCT sequence (Figure 2.2), and the NMR data suggests that this N-terminal sequence forms an  $\alpha$ -helix (called  $\alpha 1'$ ). Residues targeted for mutagenesis were identified by sequence homology, and the resulting protein was purified to apparent homogeneity. Since structural changes caused by a particular mutation could potentially indirectly affect the activity of the protein, two qualitative measurements of a different nature,  $^1\text{H}$ -NMR and CD spectroscopy, were employed to inspect structural differences between the wildtype and the mutant proteins. Slight changes in 1D NMR spectra were observed for those mutations within the BRCT domain (T415A, R423A and K458E) and within the loop (K397E), but no major differences in the secondary structure contents were seen between the wildtype and those mutants judging from their CD spectra hence indicating that the structural changes in the mutants are minor.

*Effects of Mutagenesis on DNA binding*

The sequence alignment of the BRCT region from 31 different species revealed a striking lack of amino acid conservation in the N-terminal sequence of the BRCT region. However, mutation of the weakly conserved aromatic groups Y382 and Y385 within the  $\alpha 1'$  helix and the positively charged residues R388 and K397 in the adjacent loop resulted in a significant reduction in DNA binding. The interaction site on the  $\alpha 1'$  helix was further localized by the observation that the S384A mutation had no effect on DNA binding suggesting that a specific side of the helix makes contacts to the DNA. These results support our earlier observation that deletion of the N-terminal sequence of the BRCT region abrogated DNA binding. Interestingly, both aromatic residues in the  $\alpha 1'$  helix and the arginine and lysine in the adjacent loop that were mutated in this study are conserved in *Drosophila* and human p140.

The absence of conserved residues in the N-terminal sequence of the BRCT region in combination with the presence of a limited number of highly conserved residues within the BRCT domain suggests that 5' phosphate recognition may be predominantly modulated by the BRCT domain. The reduced DNA binding observed for mutations at the

conserved T415, R423 and K458 and weakly conserved K461, which is located close to the first three, indicates a potential surface for DNA binding. Remarkably T415, R423, and K458 are conserved widely throughout the bacterial NAD<sup>+</sup> dependent DNA ligases (Figure 3.6), which together with the RFC p140, form a distinct subclass of the BRCT family (1). A site directed mutagenesis study of the BRCT domain of the NAD<sup>+</sup> dependent DNA ligase from *Thermus* species AK16D (DNLJ\_THESK) indicated that the conserved residues including T599, R606 and G638 (T415, R423 and G455 respectively in p140) are important for DNA binding activity associated with the subsequent adenyl-moiety transfer to the 5' phosphate of the dsDNA nick (22). Deletion of the BRCT domain from the NAD<sup>+</sup> dependent DNA ligases of several bacterial species has indicated its crucial role in DNA binding (22-25), and recently DNA binding by the isolated BRCT domain from *E.coli* LigA has been confirmed (25). Hence this particular class of the BRCT family, with the conserved amino acids described above, may form a group of DNA binding BRCTs.



(Figure 3.6) The positional conservation of amino acids involved in phospho-moiety binding by the tandem BRCT domains from BRCA1 and the conserved residues of the distinct subclass of BRCT superfamily. The alignment was prepared using ClustalW (16) and gaps were introduced to the alignment based on the secondary structure of BRCA1 BRCT (PDB entry: 1T29) and human RFC p140. The asterisks indicate the phospho-moiety binding residues by the N-terminal domain of the tandem BRCT repeat from BRCA1. The representatives of the distinct subclass of BRCT superfamily used for the alignment are RFC p140 subunit of RFC1\_HUMAN (human) and RFC1\_DROME (*Drosophila melanogaster*), and NAD<sup>+</sup> dependent ligases from DNLJ\_THEFI (*Thermus filiformis*), DNLJ\_THESK (*Thermus sp AK16D*), DNLJ\_ECOLI (*Escherichia coli*) and DNLJ\_CHLTR (*Chlamydia trachomatis*). BRCA1\_HUMAN sequence is from the N-terminal BRCT of the tandem BRCTs. The secondary structure of the BRCT domain from RFC p140 as determined by NMR (this thesis) is indicated above the alignment. Arrows indicate  $\beta$ -sheet and square indicate  $\alpha$ -helix.

### Potential 5' phosphate binding residues

The tandem BRCT domains from BRCA1 bind a short peptide (ISRSTpSPTFNKQT) from the protein BACH1 (BRCA1 associated Carboxyl-terminal Helicase) in a phosphoserine dependent manner (7-9). The crystal structure of the complex formed between the BRCA1 tandem BRCT domains and the phosphoserine peptide revealed that specific interaction is achieved via a network of hydrogen bonds formed by T1655, G1656 and K1702 to the phosphate oxygens. These residues belong to the first domain of the tandem BRCT repeat (10-12) and are also conserved among other tandem BRCT domains with phospho-peptide binding activity (8). Interestingly, despite the

overall lack of sequence homology between the BRCT domains of BRCA1 and RFC p140, sequence alignment based on the secondary structure elements shows a positional conservation of three key residues of BRCA1 BRCTs that are involved in phospho-moiety recognition. These residues are equivalent to the conserved T415, G416 and K458 on the surface of p140 (375-480) (Figure 3.6). The importance of T415 and K458 for DNA binding is shown here. A possible conclusion is that these residues form a patch on the surface of RFC p140 that is involved in 5' phosphate binding. Structure elucidation of p140 (375-480) will likely confirm this conclusion.

It has to be noted that the DNA binding is not restricted to the distinct class of the BRCT superfamily formed by RFC p140 and NAD<sup>+</sup> dependent DNA ligases. A recent report suggests that polymerase X-family members TdT, pol mu and pol lambda, all of which bear an N-terminal BRCT domain, have DNA binding activity, albeit with weak apparent affinity (26). The BRCT domains from these polymerases have little amino acid similarity to that of RFC p140 (375-480). Similarly the BRCT domain from mouse Rev1 has been shown to bind DNA stably enough to be observed on a gel retardation assay (unpublished data provided by G. Siegal). It will be interesting to see if DNA binding by the BRCTs from polX family or Rev1 has similar biochemical properties to the BRCT region of p140(375-480).

Although we could not identify which residues were directly interacting with the 5' phosphate, conserved residues which directly influence DNA binding were identified. The data presented here could potentially be useful for modeling the DNA-protein complex formed by the BRCT domain.

---

## Reference list

---

1. Bork, P., Hofmann, K., Bucher, P., Neuwald, A. F., Altschul, S. F., and Koonin, E. V. (1997) *FASEB J.* **11**, 68-76
2. Gaiser, O. J., Ball, L. J., Schmieder, P., Leitner, D., Strauss, H., Wahl, M., Kuhne, R., Oschkinat, H., and Heinemann, U. (2004) *Biochemistry* **43**, 15983-15995
3. Krishnan, V. V., Thornton, K. H., Thelen, M. P., and Cosman, M. (2001) *Biochemistry* **40**, 13158-13166
4. Williams, R. S., Green, R., and Glover, J. N. (2001) *Nat.Struct.Biol.* **8**, 838-842

5. Zhang, X., Morera, S., Bates, P. A., Whitehead, P. C., Coffey, A. I., Hainbucher, K., Nash, R. A., Sternberg, M. J., Lindahl, T., and Freemont, P. S. (1998) *EMBO J.* **17**, 6404-6411
6. Dulic, A., Bates, P. A., Zhang, X., Martin, S. R., Freemont, P. S., Lindahl, T., and Barnes, D. E. (2001) *Biochemistry* **40**, 5906-5913
7. Manke, I. A., Lowery, D. M., Nguyen, A., and Yaffe, M. B. (2003) *Science* **302**, 636-639
8. Rodriguez, M., Yu, X., Chen, J., and Songyang, Z. (2003) *J.Biol. Chem.* **278**, 52914-52918
9. Yu, X., Chini, C. C., He, M., Mer, G., and Chen, J. (2003) *Science* **302**, 639-642
10. Clapperton, J. A., Manke, I. A., Lowery, D. M., Ho, T., Haire, L. F., Yaffe, M. B., and Smerdon, S. J. (2004) *Nat.Struct.Mol.Biol.* **11**, 512-518
11. Shiozaki, E. N., Gu, L., Yan, N., and Shi, Y. (2004) *Mol.Cell* **14**, 405-412
12. Williams, R. S., Lee, M. S., Hau, D. D., and Glover, J. N. (2004) *Nat.Struct.Mol.Biol.* **11**, 519-525
13. 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
14. Horton, R. M., Cai, Z. L., Ho, S. N., and Pease, L. R. (1990) *Biotechniques* **8**, 528-535
15. Higgins, D. G., Thompson, J. D., and Gibson, T. J. (1996) *Methods Enzymol.* **266**, 383-402
16. Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. (2000) *Trends Biochem.Sci.* **25**, 147-150
17. Livingstone, C. D. and Barton, G. J. (1993) *Comput.Appl.Biosci.* **9**, 745-756
18. Thornton, K. H., Krishnan, V. V., West, M. G., Popham, J., Ramirez, M., Thelen, M. P., and Cosman, M. (2001) *Protein Expr.Purif.* **21**, 401-411
19. Bates, P. A., Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. (2001) *Proteins Suppl* **5**, 39-46
20. Jones, D. T. (1999) *J.Mol.Biol.* **292**, 195-202
21. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J.Mol.Graph.* **14**, 51-32
22. Feng, H., Parker, J. M., Lu, J., and Cao, W. (2004) *Biochemistry* **43**, 12648-12659
23. Jeon, H. J., Shin, H. J., Choi, J. J., Hoe, H. S., Kim, H. K., Suh, S. W., and Kwon, S. T. (2004) *FEMS Microbiol.Lett.* **237**, 111-118
24. Lim, J. H., Choi, J., Kim, W., Ahn, B. Y., and Han, Y. S. (2001) *Arch.Biochem.Biophys.* **388**, 253-260
25. 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
26. Ma, Y., Lu, H., Tippin, B., Goodman, M. F., Shimazaki, N., Koiwai, O., Hsieh, C. L., Schwarz, K., and Lieber, M. R. (2004) *Mol.Cell* **16**, 701-713

## Chapter 4

# $^1\text{H}$ , $^{15}\text{N}$ and $^{13}\text{C}$ resonance assignments and secondary structure determination of the BRCT Region of the large subunit of human Replication Factor C, bound to DNA

---

### Abstract

---

Essentially complete  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  resonances of the protein moiety of the 19 kDa p140(375-480)-DNA complex are presented. Secondary structure prediction based on the chemical shifts of  $\text{C}\alpha$ ,  $\text{C}\beta$  and  $\text{H}\alpha$  and on the pattern of backbone NOEs indicate the presence of a consensus BRCT domain with an extra alpha helix in sequences N-terminal to the BRCT domain.

Parts of this chapter have been published as:

Kobayashi, M. and Siegal, G. (2005) *J.Biomol.NMR* **31**, 183-184



## Introduction

---

Replication Factor C (RFC) is a complex of five proteins required for replication and repair of chromosomal DNA (1). The primary function of RFC appears to be to open the toroidally shaped, “sliding clamp” protein PCNA and “load” it onto DNA where it serves as a binding platform for a multitude of enzymes and regulatory proteins involved in the replication and repair of DNA. RFC consists of five subunits, four homologous proteins with molecular mass between 35 and 40 kDa, and a fifth, which has a molecular mass of 140 kDa in mammals (referred to as p140). The N-terminal half of RFC p140 contains sequences unique to RFC, including a region shown to have DNA binding activity(2) (3-6), but that is not required for the clamp loading activity (7;8).

The DNA binding region, which is homologous to bacterial DNA ligases (5), has an unusual specificity for the 5' phosphorylated terminus of dsDNA (4). Interestingly, binding is independent of the sequence of the double-stranded region, so long as there are at least 10 fully paired bases (Figure 2.3). Since the ligase homology was initially noted, it has been recognized that sequences between residues 403 and 486 of human RFC p140 form part of a distinct class of BRCT domains (9;10). However, it has been shown that residues outside of this region are additionally required for DNA binding ((4) and Figure 2.2) thus bringing into question whether residues 403-486 actually fold similarly to known BRCT domains. The DNA binding region of human p140 subunit has been mapped between 375 and 480, which consists of an N-terminus of 28 amino acids followed by a BRCT domain (403-480), and hereafter referred to as the “BRCT region”. As part of a project to determine the biological role of the N-terminal half of RFC p140 and to further understand the molecular mechanism of the unusual DNA recognition, we have determined the solution structure of a protein-DNA complex that consists of residues 375-480 and a 10 bp, hairpin oligonucleotide with a recessed, 5' phosphorylated terminus using NMR spectroscopy. Presumably due to dynamics that occur intermediate on the NMR timescale, the spectra of some residues of the protein are of poor quality. Despite this fact, we report here essentially complete  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  resonances of the protein moiety of the 19 kDa protein-DNA complex.

---

Methods and experiments

---

The gene coding for human RFC p140 residues 375-480 was cloned into pET-20b (Novagen) to allow its expression in fusion with an C-terminal His<sub>6</sub>-tag. The recombinant gene codes for Met and Asn prior to residue 375 and Asn, Leu and Glu before the C-terminal His<sub>6</sub>-tag. Recombinant protein was produced in *E. coli* BL21(DE3), purified by immobilized metal affinity chromatography on HisBind resin (Novagen) charged with Ni<sup>2+</sup> ions and subsequently gel filtration using Superose 12 resin (1.6 x 75 cm, Amersham Biosciences). Isotopically labeled proteins were prepared from cells grown in M9-based minimal medium supplemented with <sup>15</sup>NH<sub>4</sub>Cl as the sole nitrogen source, and either <sup>13</sup>C<sub>6</sub>-glucose or unlabeled glucose. To form the complex, RFC p140(375-480) was diluted to 10 μM in 25 mM Tris-HCl pH 7.5, 5 mM NaCl, 1 mM DTT and 1.2 equivalents of the oligonucleotide (pCTCGAGGTCGTCATCGACCTCGAGATCA) were added. The complex was concentrated to 0.5 mM using vacuum dialysis (Spectrum Labs) and the buffer was exchanged to 25 mM D<sub>11</sub>-Tris-HCl pH 7.5, 5mM NaCl in 95/5 H<sub>2</sub>O/D<sub>2</sub>O.

All NMR data were acquired at 25 °C on a Bruker DMX600 spectrometer. Most of the sequential assignments for the backbone were obtained using 3D HNCACB, CBCA(CO)NH and HBHA(CO)NH spectra. Aliphatic side-chain resonances were derived from 3D HCCH-TOCSY and CCH-TOCSY spectra. Additional data provided by 2D [<sup>1</sup>H, <sup>1</sup>H] NOESY, 3D [<sup>15</sup>N, <sup>1</sup>H] NOESY-HSQC and [<sup>13</sup>C, <sup>1</sup>H] NOESY-HSQC experiments were used for further assignment as well as confirmation of the through-bond data.

---

Results and Discussion

---

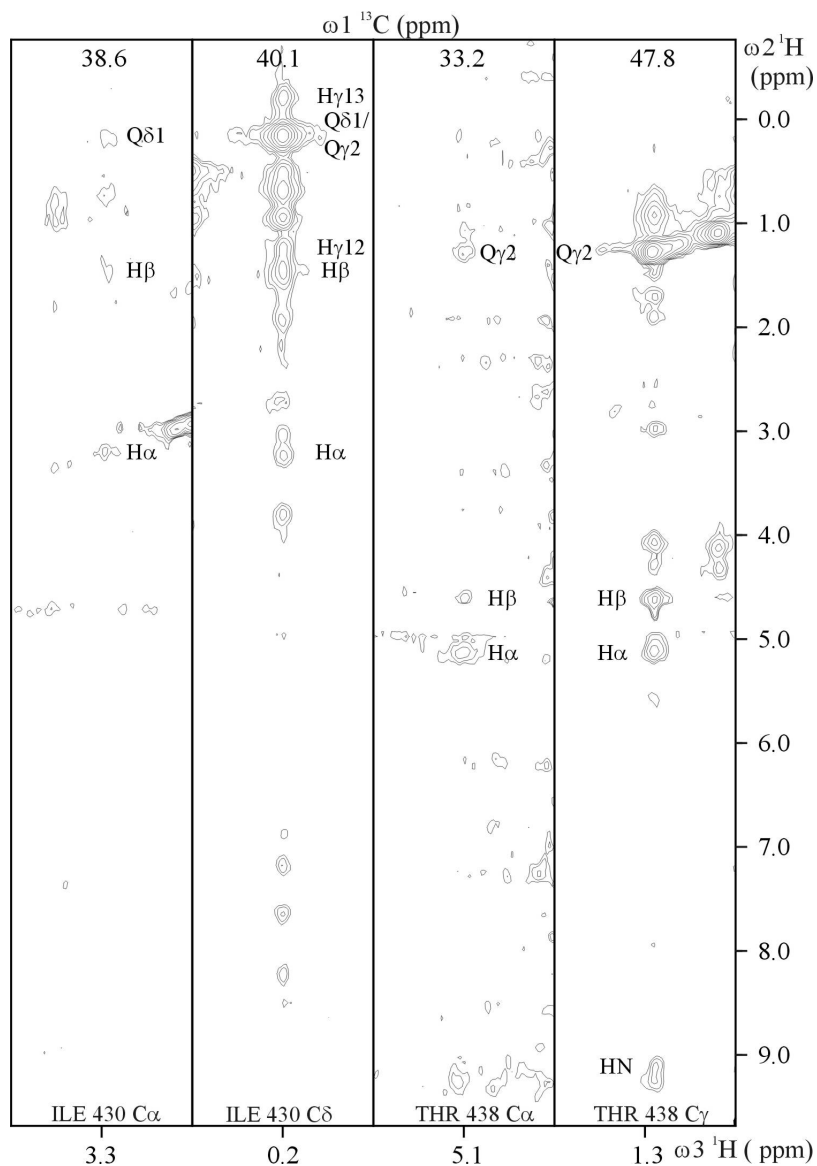
In general, the spectra of the protein-DNA complex are of moderate quality (Figure 4.1) displaying linewidths that are broader than expected for a 19 kDa complex and with missing correlations. However, we have been able to assign over 95 % of the non-labile aliphatic <sup>1</sup>H's and 96 % of the protonated aliphatic <sup>13</sup>C resonances. More than 99 % of the amide <sup>1</sup>H and <sup>15</sup>N resonances of the backbone have been assigned with chemical shifts missing only for Tyr 379. <sup>1</sup>H-<sup>15</sup>N correlations have been found for all glutamine and asparagine side chains. RFC p140(375-480) has a rather low aromatic content (5 Tyr and 1 Phe). Of these, we have unambiguously assigned the <sup>1</sup>H aromatic resonances of Tyr 382, 385 433 and 447, and Phe 412 using a combination of homonuclear and heteronuclear

edited NOESY spectra. The chemical shift values (Supplementary materials Table 1S) have been deposited in the BioMagResBank database under the accession number 6353.

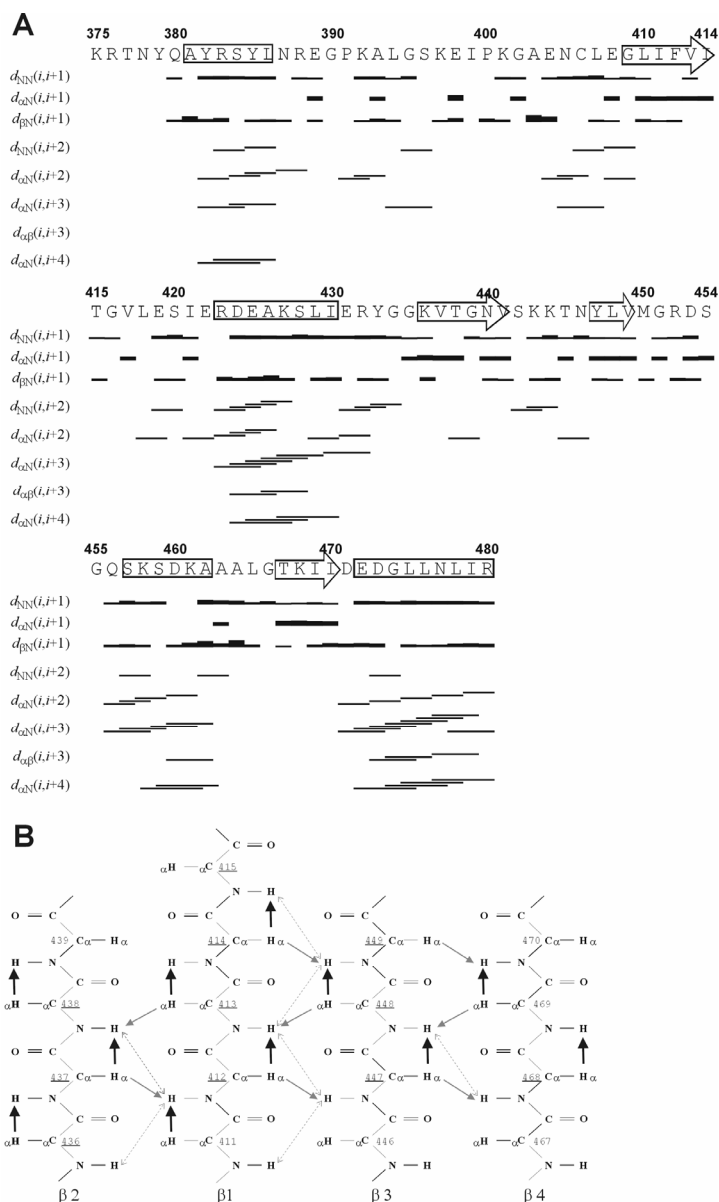
#### *Secondary structure analysis*

Regular secondary structures in a polypeptide can be distinguished by the presence of medium and long-range  $H^1-H^1$  distances that are readily observed by NOE's (11). Using the essentially complete chemical shift assignment, the presence of such NOE patterns in the 3D [ $^{15}N, ^1H$ ] NOESY-HSQC spectrum was analyzed by manual assignment of NOESY crosspeaks. The resulting sequential and medium range  $H^1-H^1$  NOE's observed are summarized in Figure 4.2A, which indicates that the BRCT region, p140(375-480) consists of four  $\alpha$  helices spanning residues 381-386, 423-431, 458-462 and 471-480 and four  $\beta$  - strands spanning residues 410-414, 436-438, 447-449 and 467-470.

In regular  $\beta$  – sheets, the long range  $H^1-H^1$  distances involving the polypeptide backbone are sufficiently short to be observable by NOEs. The pattern of such NOEs is unique for anti-parallel and parallel  $\beta$  – sheets (11). Analysis of the 3D [ $^{15}N, ^1H$ ] NOESY-HSQC spectrum reveals a pattern of cross-strand NOE's that includes medium  $d\alpha_N(i,j)$  weak  $d_{NN}(i,j)$  and strong sequential  $d\alpha_N(i, i+1)$  NOE's, the signature of a parallel stranded  $\beta$  -sheet (Figure 4.2B) (11). Due to the absence of  $H\alpha-C\alpha$  correlations in the 3D [ $^{13}C, ^1H$ ] NOESY-HSQC,  $d\alpha\alpha(i,j)$  NOE's could not be found. The pattern of secondary structure is in good agreement with all known BRCT domains which have a four stranded parallel  $\beta$  – sheet that comprises the core of the domain (12;13). In general the length of  $\beta$  - strands in BRCT domains (12;13) are rather short, typically between 2 and 4 residues, as also observed here.



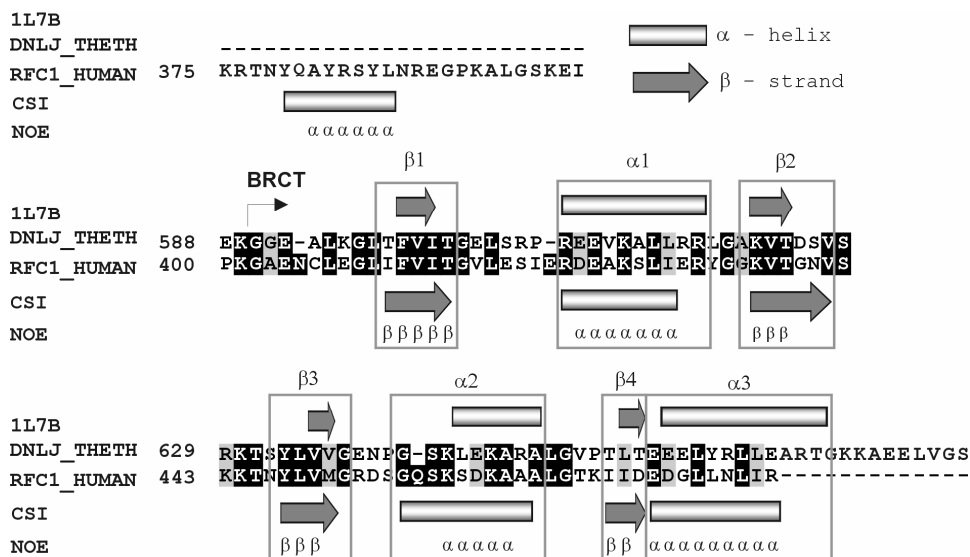
(Figure 4.1) Selected strips from the 3D  $^{13}\text{C}, ^1\text{H}$ -NOESY HSQC of the RFC p140(375-480)-DNA complex. Strips from the  $\text{H}\alpha$ - $\text{C}\alpha$  and  $\text{Q}\delta 1$ - $\text{C}\delta 1$  correlation of Ile 430 and the  $\text{H}\alpha$ - $\text{C}\alpha$  and  $\text{Q}\gamma 2$ - $\text{C}\gamma 2$  correlation of Thr 438 are shown. The  $^{13}\text{C}$  chemical shift is shown above each strip.  $^1\text{H}$ - $^{13}\text{C}$  correlations are significantly weaker in the backbone than in the sidechain. This pattern is observed 83% of the residues in the protein.



(Figure 4.2) The secondary structure prediction generated from NOE connectivity.

**(A)** Short and medium range upper-distance limits were identified by manual assignments of NOEY crosspeaks in  $^{15}\text{N}$ ,  $^1\text{H}$  NOESY-HSQC spectrum, and plotted against the sequence using the DYANA (14). The resulting secondary structures determined by the NOE patterns were drawn on the sequence where the arrow and the square indicate the presence of a  $\beta$  strand and  $\alpha$  helix respectively. **(B)** The backbone amide representing the parallel  $\beta$ -strands. The cross-strand NOE's particular to the parallel  $\beta$ -strands observed in  $^{15}\text{N}$ ,  $^1\text{H}$  NOESY-HSQC spectrum are indicated with arrows. The arrows represent the observed strong sequential  $d_{\alpha N}(i, i+1)$  (in black), the medium cross-strand NOE's  $d_{\alpha N}(i, j)$  (in gray), and weak  $d_{NN}(i, j)$  (dotted) correlations. The underlined residue numbers are conserved amino acid residues, which are also found to participate in forming the parallel  $\beta$ -strands of the homologous BRCT domain from the bacterial NAD $^{+}$  dependent DNA ligase (see Figure 4.3 for details).

The secondary structure determined from the NOE connectivity is further supported by prediction based on the chemical shift index (CSI) using the  $^{13}\text{C}\alpha$ ,  $^{13}\text{C}\beta$  and  $^1\text{H}\alpha$  resonance assignment of DNA bound p140(375-480) (15). The deviation of chemical shifts of those nuclei from their random coil values can be used to predict secondary structure in proteins. The CSI plot for each of the  $^{13}\text{C}\alpha$ ,  $^{13}\text{C}\beta$  and  $^1\text{H}\alpha$  nuclei can be found as a supplement data (Figure S1) at the end of this chapter. Comparison of the secondary structure derived from the CSI plot and the NOE pattern shows slight differences, especially in the size of strand  $\beta 2$  and helix  $\alpha 3$ , however in general they are in good agreement with each other (Figure 4.3). Comparison with one member of the distinct class of BRCT superfamily indicates that the secondary structure of the BRCT domain (403-480) is indeed consistent with the known structures of BRCT domains from the NAD $^{+}$  dependent DNA ligase (Figure 4.3). In addition, the BRCT region includes an extra  $\alpha$ -helix (379-386) at the N-terminus which is followed by a long loop (387-410) before the BRCT domain (Figure 4.3).



(Figure 4.3) Amino acid sequence and secondary structure determination of the DNA bound RFC p140(375-480). The secondary structure as determined by CSI analysis (in Supplementary materials) or from the pattern of NOE's is indicated schematically below the sequence. The start of the conserved BRCT domain is indicated. For comparison, the secondary structure of the BRCT domain from the NAD $^{+}$  dependent ligase from *Th. filiformis* determined by NMR, is also shown.

The complete chemical shift assignment of the free protein is not currently available. A considerable number of the expected peaks were missing from the HNCACB spectrum and 40 % of the backbone amide correlations were missing from the [ $^{15}\text{N}$ ,  $^1\text{H}$ ] NOESY-HSQC spectrum. Presumably the amide protons are in rapid exchange with those of water. Accordingly poor dispersion and non-uniform linewidth of resonances were observed in the [ $^{15}\text{N}$ ,  $^1\text{H}$ ]-HSQC spectrum of the free p140(375-480) indicating that the protein may undergo conformational exchange intermediate on the NMR time scale. However partial backbone assignments and circular dichroism spectroscopic data (Figure 2.1) suggest the existence of secondary structure in the free protein. Upon DNA binding, the conformational dynamics of the p140(375-480) apparently become restricted resulting in a more defined conformation. The similarity of the secondary structure content to the BRCT domains of known structure suggests that the BRCT domain of the p140(375-480) folds similarly to that of the NAD<sup>+</sup> dependent DNA ligase upon DNA binding. The position of the  $\alpha$ -helix and the loop at the N-terminus relative to the BRCT domain awaits elucidation of the three-dimensional structure of p140(375-480).

---

## Reference list

---

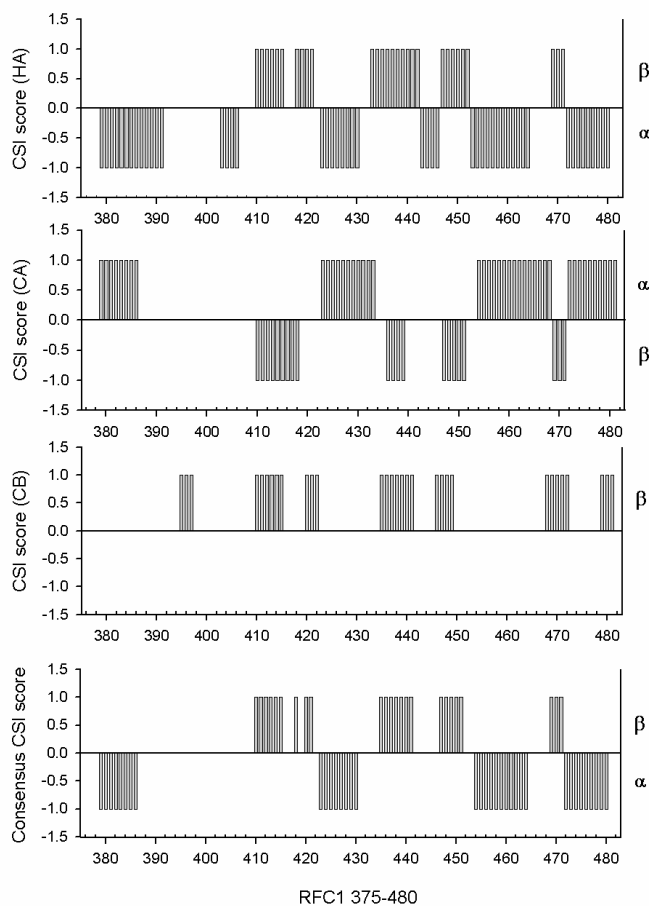
1. Waga, S. and Stillman, B. (1998) *Annual Review of Biochemistry* **67**, 721-751
2. Allen, B. L., Uhlmann, F., Gaur, L. K., Mulder, B. A., Posey, K. L., Jones, L. B., and Hardin, S. H. (1998) *Nucleic Acids Res.* **26**, 3877-3882
3. Fotedar, R., Mossi, R., Fitzgerald, P., Rousselle, T., Maga, G., Brickner, H., Messier, H., Kasibhatla, S., Hubscher, U., and Fotedar, A. (1996) *EMBO Journal* **15**, 4423-4433
4. Allen, B. L., Uhlmann, F., Gaur, L. K., Mulder, B. A., Posey, K. L., Jones, L. B., and Hardin, S. H. (1998) *Nucleic Acids Research* **26**, 3877-3882
5. Burbelo, P. D., Utani, A., Pan, Z. Q., and Yamada, Y. (1993) *Proceedings Of The National Academy Of Sciences Of The United States Of America* **90**, 11543-11547
6. Tsurimoto, T. and Stillman, B. (1991) *Journal of Biological Chemistry* **266**, 1950-1960
7. Gomes, X. V., Gary, S. L., and Burgers, P. M. J. (2000) *Journal of Biological Chemistry* **275**, 14541-14549
8. Uhlmann, F., Cai, J. S., Gibbs, E., O'Donnell, M., and Hurwitz, J. (1997) *Journal of Biological Chemistry* **272**, 10058-10064
9. Bork, P., Hofmann, K., Bucher, P., Neuwald, A. F., Altschul, S. F., and Koonin, E. V. (1997) *FASEB Journal* **11**, 68-76

10. Callebaut, I. and Mornon, J. P. (1997) *FEBS Letters* **400**, 25-30
11. Wuthrich, K. (1986) *NMR of Proteins and Nucleic acids*, Wiley, New York,
12. Krishnan, V. V., Thornton, K. H., Thelen, M. P., and Cosman, M. (2001) *Biochemistry* **40**, 13158-13166
13. Zhang, X., Morera, S., Bates, P. A., Whitehead, P. C., Coffey, A. I., Hainbucher, K., Nash, R. A., Sternberg, M. J., Lindahl, T., and Freemont, P. S. (1998) *EMBO J.* **17**, 6404-6411
14. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) *J.Mol.Biol.* **273**, 283-298
15. Wishart, D. S. and Sykes, B. D. (1994) *Journal of Biomolecular NMR* **4**, 171-180
16. Lee, J. Y., Chang, C., Song, H. K., Moon, J., Yang, J. K., Kim, H. K., Kwon, S. T., and Suh, S. W. (2000) *EMBO J.* **19**, 1119-1129



## Supplementary materials

Figure S1. CSI calculation.



**CSI scoring and secondary structures generated based on the CSI consensus.** Chemical shift data of residue number 375-480 were used as input for CSI calculation. The secondary structures were assigned to the residues based on their CSI score (left side the plots). The consensus score was derived from the CSI score of the nuclei CA, CB and HA and was used to generate the final secondary structures for the Figure 3.

Table 1S : RFC p140 375-480 BRCT region : assignments; pH 7.5, 298 ° K (proton chemical shifts are given in parentheses).

residue	N	C <sup>α</sup>	C <sup>β</sup>	other
K375	122.3 (8.35)	56.8 (4.33)	33.2 (1.79, 1.86)	C <sup>γ</sup> , 25.0 (1.46, 1.46); C <sup>δ</sup> , 29.1 (1.71, 1.71); C <sup>ε</sup> , 30.7 (3.05, 3.05)
R376	123.4 (8.60)	56.6 (4.38)	30.8 (1.80, 1.87)	C <sup>γ</sup> , 27.2 (1.68, 1.52); C <sup>δ</sup> , 43.6 (3.17, 3.17)
T377	115.7 (8.28)	62.2 (4.32)	70.1 (4.13)	C <sup>γ2</sup> , 21.6 (1.11)
N378	122.4 (8.57)	53.0 (4.70)	38.2 (2.99, 2.77)	N <sup>δ2</sup> , 112.9 (7.58, 6.89)
Y379		59.4 (4.46)	38.4 (3.18, 3.00)	C <sup>δ1</sup> , * (7.00); C <sup>δ2</sup> , * (7.00); C <sup>ε1</sup> , * (6.77); C <sup>ε2</sup> , * (6.77)
Q380	119.4 (8.46)	59.2 (3.92)	28.1 (2.10, 2.16)	C <sup>γ</sup> , 33.9 (2.44, 2.44); N <sup>δ2</sup> , 112.3 (6.94, 7.54)
A381	122.0 (8.21)	54.4 (4.16)	18.3 (1.53)	
Y382	121.0 (7.94)	58.8 (4.42)	39.4 (2.75, 3.14)	C <sup>δ1</sup> , * (6.53); C <sup>δ2</sup> , * (6.53); C <sup>ε1</sup> , * (5.89); C <sup>ε2</sup> , * (5.89)
R383	117.9 (8.16)	59.3 (3.33)	29.2 (1.64, 1.64)	C <sup>γ</sup> , 27.0 (1.38, 1.56); C <sup>δ</sup> , 43.1 (3.04, 3.04)
S384	113.5 (7.94)	61.6 (3.94)	63.1 (4.09, 4.09)	
Y385	124.0 (7.52)	60.8 (4.06)	37.3 (3.28, 3.03)	C <sup>δ1</sup> , * (6.66); C <sup>δ2</sup> , * (6.66)
L386	119.3 (8.00)	56.7 (3.33)	42.1 (1.31, 0.91)	C <sup>γ</sup> , 25.8 (0.82); C <sup>δ1</sup> , 24.3 (-0.24); C <sup>δ2</sup> , 21.5 (0.41)
N387	113.8 (7.22)	52.8 (4.61)	39.8 (2.81, 2.54)	N <sup>δ2</sup> , 112.5 (6.89, 7.43)
R388	122.1 (7.22)	56.7 (4.24)	30.8 (1.88, 1.88)	C <sup>γ</sup> , * (1.37, 1.37); C <sup>δ</sup> , * (3.33, 3.33)
E389	125.8 (9.00)	56.6 (4.23)	30.8 (2.11, 1.95)	C <sup>γ</sup> , 36.9 (2.38, 2.38)
G390	109.8 (8.55)	45.3 (3.85, 4.25)		
P391		64.4 (4.09)	32.8 (1.56, 2.49)	C <sup>γ</sup> , 28.0 (1.80, 2.15); C <sup>δ</sup> , 49.6 (3.62, 3.62)
K392	123.0 (9.22)	57.5 (4.31)	33.8 (1.99, 1.53)	C <sup>γ</sup> , 25.5 (1.66, 1.66); C <sup>δ</sup> , 29.3 (1.63, 1.63); C <sup>ε</sup> , 42.3 (2.91, 2.91)
A393	127.3 (9.00)	50.4 (4.80)	19.0 (1.08)	
L394	123.2 (8.46)	56.9 (3.99)	41.4 (1.68, 1.46)	C <sup>γ</sup> , * (1.55); C <sup>δ1</sup> , 23.8 (0.77); C <sup>δ2</sup> , 25.4 (0.94)
G395	111.8 (8.74)	46.3 (4.16, 4.00)		
S396	113.3 (7.95)	60.0 (4.19)	63.7 (3.96, 3.80)	
K397	121.6 (7.89)	55.3 (4.62)	35.1 (1.47, 1.60)	C <sup>γ</sup> , 25.6 (1.39, 1.39); C <sup>δ</sup> , 29.4 (1.43, 1.43)
E398	124.3 (8.52)	56.3 (4.04)	30.2 (1.89, 1.83)	C <sup>γ</sup> , 36.2 (1.97, 2.23)
I399	128.2 (8.63)	56.7 (4.40)	36.6 (2.00)	C <sup>γ1</sup> , 26.7 (1.52, 1.45); C <sup>γ2</sup> , 17.5 (0.96); C <sup>δ1</sup> , 10.4 (0.73)
P400		68.2 (4.42)	32.7 (1.63, 2.12)	C <sup>γ</sup> , 26.6 (1.91, 2.02); C <sup>δ</sup> , 51.6 (4.01, 3.56)
K401	120.1 (8.26)	55.7 (4.35)	32.7 (1.85, 1.74)	C <sup>γ</sup> , 25.0 (1.53, 1.46); C <sup>δ</sup> , 29.1 (1.73, 1.73); C <sup>ε</sup> , 42.3 (3.06, 3.06)
G402	110.9 (7.42)	44.6 (3.07, 4.34)		
A403	123.4 (8.58)	51.7 (4.35)	19.2 (1.38)	
E404	120.7 (8.55)	58.0 (4.15)	29.6 (2.02, 2.02)	C <sup>γ</sup> , 36.2 (2.38, 2.38)
N405	117.7 (9.35)	54.3 (4.52)	38.1 (3.17, 2.84)	N <sup>δ2</sup> , 112.5 (7.58, 6.87)
C406	114.6 (7.93)	60.8 (4.35)	27.9 (3.05, 3.05)	
L407	118.5 (8.58)	53.4 (4.77)	40.3 (1.85, 1.58)	C <sup>γ</sup> , 26.2 (0.82); C <sup>δ1</sup> , 22.6 (0.72); C <sup>δ2</sup> , * (0.72)
E408	119.0 (7.36)	58.5 (3.99)	29.4 (2.19, 1.98)	C <sup>γ</sup> , * (2.31, 2.31)
G409	113.9 (8.28)	45.5 (3.72, 4.26)		
L410	120.7 (8.10)	54.1 (4.71)	45.4 (2.38, 2.38)	C <sup>γ</sup> , * (1.54); C <sup>δ1</sup> , 22.9 (0.80); C <sup>δ2</sup> , 26.3 (1.04)
I411	123.3 (9.35)	61.0 (5.02)	39.3 (1.90)	C <sup>γ1</sup> , 28.2 (1.05, 1.70); C <sup>γ2</sup> , 19.3 (0.94); C <sup>δ1</sup> , 13.5 (0.90)
F412	127.8 (9.77)	56.4 (5.71)	43.7 (3.08, 2.75)	C <sup>δ1</sup> , * (7.18); C <sup>δ2</sup> , * (7.18); C <sup>ε1</sup> , * (6.89); C <sup>ε2</sup> , * (6.89)
V413	119.6 (8.32)	61.3 (4.60)	35.6 (1.64)	C <sup>γ1</sup> , 21.2 (0.76); C <sup>γ2</sup> , 22.7 (0.80)
I414	127.7 (9.18)	60.7 (4.95)	40.2 (1.55)	C <sup>γ1</sup> , 27.1 (1.60, 1.60); C <sup>γ2</sup> , 14.3 (0.77); C <sup>δ1</sup> , 18.2 (0.84)
T415	120.7 (9.04)	59.4 (5.17)	70.5 (3.88)	C <sup>γ2</sup> , 20.0 (1.22)
G416	114.6 (9.94)	44.0 (3.81, 3.65)		
V417	120.0 (9.59)	63.2 (3.84)	34.5 (1.74)	C <sup>γ1</sup> , 20.3 (0.88); C <sup>γ2</sup> , 21.1 (0.83)
L418	131.1 (8.62)	53.8 (4.53)	37.9 (2.19, 1.77)	C <sup>γ</sup> , 26.9 (1.70); C <sup>δ1</sup> , 21.8 (0.60); C <sup>δ2</sup> , 25.1 (0.67)
E419	118.6 (8.44)	57.5 (4.57)	29.9 (2.00, 2.26)	C <sup>γ</sup> , 34.8 (2.48, 2.38)
S420	111.6 (10.16)	57.4 (4.94)	65.3 (3.80, 3.69)	
I421	119.8 (7.17)	60.1 (4.41)	42.3 (1.66)	C <sup>γ1</sup> , 26.6 (1.17, 1.58); C <sup>γ2</sup> , 16.8 (0.88); C <sup>δ1</sup> , 13.6 (0.66)
E422	123.8 (8.88)	56.7 (4.23)	30.8 (1.94, 1.94)	C <sup>γ</sup> , 36.8 (2.56, 2.35)
R423	124.0 (8.92)	60.1 (3.84)	(1.93, 1.93)	C <sup>γ</sup> , * (1.53, 1.53); C <sup>δ</sup> , * (3.26, 3.26); N <sup>ε</sup> , 81.2 (7.81)
D424	116.3 (8.98)	57.6 (4.37)	39.8 (2.68, 2.68)	
E425	121.0 (7.05)	51.1 (4.18)	29.7 (2.08, 2.24)	C <sup>γ</sup> , 36.3 (2.39, 2.25)
A426	124.3 (8.55)	55.4 (3.86)	17.3 (1.28)	
K427	117.2 (8.23)	60.2 (3.85)	32.7 (1.99, 1.80)	C <sup>γ</sup> , 24.9 (1.53, 1.53); C <sup>δ</sup> , 30.0 (1.75, 1.75); C <sup>ε</sup> , 42.0 (3.00, 3.00)

# Chapter 4: Chemical shift assignments of the BRCT region

S428	113.0 (7.90)	61.6 (4.23)	62.8 (3.98, 3.92)	
L429	124.7 (8.13)	58.3 (4.00)	42.2 (2.00, 1.62)	C <sup>γ</sup> , 26.9 (1.49); C <sup>δ1</sup> , 25.5 (0.74); C <sup>δ2</sup> , 24.3 (1.01)
I430	117.2 (7.62)	65.1 (3.27)	38.2 (1.55)	C <sup>γ1</sup> , 29.9 (1.46, -0.13); C <sup>γ2</sup> , 19.1 (0.30); C <sup>δ1</sup> , 13.6 (0.21)
E431	118.4 (8.21)	52.1 (4.36)	31.2 (2.05, 2.05)	C <sup>γ</sup> , 36.8 (2.35, 2.56)
R432	123.2 (8.49)	58.7 (4.03)	29.1 (1.96, 1.83)	C <sup>γ</sup> , 26.7 (1.22, 1.22); C <sup>δ</sup> , 43.7 (2.92, 2.86)
Y433	116.0 (7.13)	58.9 (4.76)	38.7 (3.94, 2.50)	C <sup>δ1</sup> , * (7.39); C <sup>δ2</sup> , * (7.39); C <sup>ε1</sup> , * (6.69); C <sup>ε2</sup> , * (6.69)
G434	107.0 (7.91)	45.9 (3.85, 4.31)		
G435	109.8 (8.24)	44.9 (4.37, 3.34)		
K436	119.5 (8.53)	54.6 (4.98)	36.1 (1.87, 1.87)	C <sup>γ</sup> , 25.2 (1.51, 1.41); C <sup>δ</sup> , 29.3 (1.73, 1.73); C <sup>ε</sup> , 42.4 (2.99, 2.99)
V437	125.6 (9.27)	60.4 (5.56)	33.5 (1.94)	C <sup>γ1</sup> , 23.4 (0.98); C <sup>γ2</sup> , 22.2 (1.03)
T438	120.1 (9.10)	59.7 (5.11)	71.4 (4.61)	C <sup>γ2</sup> , 21.3 (1.29)
G439	106.1 (9.23)	45.2 (3.81, 4.62)		
N440	118.2 (7.91)	52.2 (4.94)	42.6 (2.80, 2.18)	N <sup>δ2</sup> , 117.1 (7.42, 8.71)
V441	125.3 (8.69)	63.2 (4.07)	31.6 (1.92)	C <sup>γ1</sup> , 22.8 (1.01); C <sup>γ2</sup> , 22.0 (0.82)
S442	125.6 (9.37)	57.3 (4.67)	66.6 (5.01, 5.01)	
K443	120.9 (9.22)	59.5 (4.05)	32.1 (2.08, 1.97)	C <sup>γ</sup> , 25.8 (1.64, 1.64); C <sup>δ</sup> , 29.4 (1.80, 1.80); C <sup>ε</sup> , 41.9 (3.06, 3.06)
K444	116.2 (8.09)	56.2 (4.09)	32.8 (1.84, 1.69)	C <sup>γ</sup> , 24.8 (1.46, 1.46); C <sup>δ</sup> , 29.2 (1.70, 1.70); C <sup>ε</sup> , 42.3 (3.00, 3.00)
T445	114.6 (7.40)	65.1 (3.40)	68.9 (3.92)	C <sup>γ2</sup> , 25.3 (1.07)
N446	126.0 (8.99)	57.4 (4.56)	41.4 (2.45, 2.37)	N <sup>δ2</sup> , 109.0 (6.87, 7.67)
Y447	114.1 (7.42)	57.0 (5.22)	44.0 (2.45, 2.26)	C <sup>δ1</sup> , * (6.63); C <sup>δ2</sup> , * (6.63); C <sup>ε1</sup> , * (6.29); C <sup>ε2</sup> , * (6.29)
L448	124.5 (9.03)	53.1 (5.15)	44.6 (2.02, 0.99)	C <sup>γ</sup> , 27.5 (1.27); C <sup>δ1</sup> , 24.8 (0.80); C <sup>δ2</sup> , 26.7 (0.76)
V449	128.6 (9.61)	61.2 (4.55)	31.3 (2.29)	C <sup>γ1</sup> , 21.4 (1.03); C <sup>γ2</sup> , 21.4 (0.76)
M450	127.4 (9.15)	55.1 (4.98)	35.9 (2.07, 1.62)	C <sup>γ</sup> , 32.2 (2.39, 2.39)
G451	114.4 (8.96)	43.6 (4.50, 3.45)		
R452	118.4 (8.05)	55.7 (4.51)	29.8 (1.76, 1.76)	C <sup>γ</sup> , 25.9 (1.65, 1.65); C <sup>δ</sup> , * (2.98, 2.98); N <sup>ε</sup> , 84.9 (9.30)
D453	119.8 (9.21)	56.7 (4.41)	40.1 (2.84, 2.84)	
S454	110.6 (7.70)	58.2 (3.95)	65.3 (3.94, 3.84)	
G455	108.5 (9.10)	47.9 (3.81, 4.09)		
Q456	128.6 (9.13)	58.3 (4.01)	29.4 (2.16, 2.07)	C <sup>γ</sup> , 33.9 (2.44, 2.44); N <sup>ε2</sup> , 114.8 (7.78, 6.84)
S457	115.7 (8.86)	61.8 (4.11)	62.5 (3.96, 3.96)	
K458	121.4 (8.23)	61.6 (4.00)	33.8 (2.15, 2.15)	C <sup>γ</sup> , * (1.51, 1.36); C <sup>δ</sup> , * (1.93, 1.93); C <sup>ε</sup> , * (2.94, 2.94)
S459	112.5 (7.91)	62.4 (3.90)	68.0 (3.86, 3.86)	
D460	123.8 (8.40)	57.5 (4.45)	40.2 (2.95, 2.69)	
K461	122.5 (8.39)	59.5 (4.08)	32.5 (1.93, 1.71)	C <sup>γ</sup> , 25.5 (1.46, 1.46); C <sup>δ</sup> , 29.4 (1.65, 1.65); C <sup>ε</sup> , 42.3 (2.92, 2.92)
A462	120.4 (8.28)	54.3 (4.21)	19.4 (1.43)	
A463	120.1 (8.03)	54.8 (4.13)	18.0 (1.56)	
A464	120.7 (7.76)	54.7 (4.15)	18.4 (1.56)	
L465	115.7 (8.14)	55.4 (4.25)	42.9 (1.31, 1.88)	C <sup>γ</sup> , 26.4 (1.82); C <sup>δ1</sup> , 22.7 (0.82); C <sup>δ2</sup> , 25.8 (0.56)
G466	108.7 (7.92)	45.9 (4.11, 3.83)		
T467	120.1 (8.11)	64.0 (3.66)	69.4 (3.39)	C <sup>γ2</sup> , 22.9 (1.00)
K468	126.7 (7.56)	57.8 (4.14)	32.3 (1.77, 1.77)	C <sup>γ</sup> , 25.3 (1.49, 1.49); C <sup>δ</sup> , 29.1 (1.59, 1.59); C <sup>ε</sup> , 41.8 (2.96, 2.96)
I469	125.4 (8.30)	60.5 (5.13)	39.4 (1.76)	C <sup>γ1</sup> , 28.0 (1.03, 1.03); C <sup>γ2</sup> , 18.6 (0.96); C <sup>δ1</sup> , 13.5 (0.87)
I470	122.5 (9.12)	59.0 (5.04)	43.2 (2.01)	C <sup>γ1</sup> , 25.3 (1.56, 1.56); C <sup>γ2</sup> , 18.2 (1.00); C <sup>δ1</sup> , 13.5 (0.54)
D471	120.0 (7.87)	50.9 (5.04)	41.8 (2.84, 3.46)	
E472	119.0 (8.68)	60.8 (3.64)	31.1 (2.14, 2.14)	C <sup>γ</sup> , 36.7 (2.01, 2.01)
D473	117.0 (7.97)	57.6 (4.39)	39.7 (2.65, 2.65)	
G474	110.5 (8.70)	47.3 (3.87, 3.87)		
L475	124.6 (8.47)	58.4 (3.95)	41.0 (2.17, 2.17)	C <sup>γ</sup> , 27.5 (1.28); C <sup>δ1</sup> , 22.2 (0.56); C <sup>δ2</sup> , 24.6 (0.69)
L476	117.0 (7.85)	58.2 (3.82)	39.9 (1.28, 2.00)	C <sup>γ</sup> , 26.5 (2.04); C <sup>δ1</sup> , 21.9 (0.72); C <sup>δ2</sup> , 25.1 (0.86)
N477	116.2 (8.57)	55.5 (4.45)	38.2 (2.75, 2.87)	N <sup>δ2</sup> , 113.1 (7.64, 7.00)
L478	122.3 (7.73)	57.9 (4.06)	42.3 (1.85, 1.57)	C <sup>γ</sup> , 25.8 (1.65); C <sup>δ1</sup> , 25.3 (0.54); C <sup>δ2</sup> , 23.9 (0.61)
I479	116.6 (7.03)	64.0 (3.48)	38.3 (1.59)	C <sup>γ1</sup> , 29.5 (0.59, 1.57); C <sup>γ2</sup> , 17.2 (0.17); C <sup>δ1</sup> , 13.3 (0.60)
R480	116.3 (7.64)	58.6 (3.91)	31.3 (1.80, 1.80)	C <sup>γ</sup> , 28.0 (1.68, 1.60); C <sup>δ</sup> , 43.5 (3.20, 3.20)

## Chapter 5

# Structure of the BRCT domain from RFC p140: A model Protein-DNA complex determined by NMR and mutagenesis data<sup>1</sup>

---

### Abstract

---

Based on the chemical shift assignment of human RFC p140(375-480) in complex with double stranded DNA containing a 5'-recessed phosphate, 3D [<sup>15</sup>N, <sup>1</sup>H] NOESY-HSQC and [<sup>13</sup>C, <sup>1</sup>H] NOESY-HSQC spectra were assigned using the automated protocol CANDID. The resulting distance restraints and predicted dihedral angle restraints were used as input to CYANA 2.0, to produce an ensemble of 20 structures of the protein moiety, p140(375-480), bound to DNA. The protein consists of a well-defined core, corresponding to a consensus BRCT domain, and an N-terminal  $\alpha$  helix, whose spatial orientation with respect to the core of the protein is less well defined. Due to a lack of sequence specific resonance assignments, the NMR data was insufficient to determine the structure of the DNA moiety of the complex. Therefore we used the HADDOCK protocol to dock the protein onto the DNA using ambiguous restraints derived from mutagenesis, amino acid residue conservation and ambiguously assigned intermolecular NOEs. In this model the 5' phosphate interacts with the positively charged surface of the BRCT domain while the N-terminal helix lies in the major groove of the DNA. The model supports previous observations that both the N-terminal sequence and the BRCT domain directly interact with the DNA and explains why no NOEs were observed between these two regions of the protein. Comparison of the 3D structure of the RFC BRCT domain with that of BRCA1 and the NAD<sup>+</sup> dependent DNA ligases reveals a remarkable structural conservation of the phospho-moiety binding residues. The shallow phospho-moiety binding surface may explain why the additional interactions with DNA made by the N-terminal  $\alpha$  helix are essential for the stability of the complex.

---

<sup>1</sup> Parts of this chapter will be used to prepare a manuscript by M. Kobayashi, E. AB, A. M. J. J. Bonvin, and G. Siegal.

---

## Introduction

---

Replication factor C (RFC) is a five subunit complex, which plays an important role in efficiently loading PCNA onto primer-template DNA during synthesis of the daughter strand in DNA replication (1). Human RFC consists of four subunits of 35-40 kDa and a fifth large subunit (p140) of 140 kDa. The C-terminus of p140 shares homology with the four small subunits, while the unique N-terminal sequence contains a single BRCT domain that is dispensable for its function in PCNA loading (2). The crystal structure of yeast RFC carrying a BRCT-truncated p140 indicated that the five subunits form a spiral complex that precisely matches that of B form DNA (3). Despite the lack of a role in DNA replication, the region including the BRCT domain (subsequently referred to as the BRCT region, residues 375-480) was shown to have binding activity specific for 5' phosphorylated dsDNA (4). Currently there is no structural information available regarding this type of structure-specific DNA recognition.

BRCT domains are small, consisting of roughly 90 amino acids, and are found in more than 900 proteins from all biological kingdoms (5). These proteins, which may bear more than a single copy of the BRCT domain, exhibit a broad range of functional activities in DNA replication, DNA repair and cell-cycle checkpoint regulation (5-7). Structural information is available for the BRCT domains of XRCC1 (8), BRCA1 (9), 53BP1 (10;11), DNA ligase III (12) and the bacterial NAD<sup>+</sup> dependent DNA ligase (13), all displaying a conserved fold. XRCC1 contains two copies of the BRCT domain, of which the C-terminal one forms a hetero-dimer with the BRCT domain of DNA ligase III through residues conserved between the two domains (14). BRCA1 also contains two BRCT domains, eliciting a very different function: they form an obligate paired structural unit that specifically binds to a phospho-serine containing sequence in the protein BACH1 (15) and CtIP (16). Hence, despite conservation of the three dimensional structure of each domain, the mechanism by which BRCT domains execute their function differs significantly within the BRCT superfamily.

Although a limited number of BRCT-DNA interactions are known or have been implied from biochemical data, there is at present no structure of a BRCT-DNA complex. Deletion and mutagenesis data (Figures 2.2 and 3.5) suggest that the region spanning residues 375-480 in RFC p140 (hereafter called p140 (375-480)) is important for DNA binding. In order to investigate the structural basis of this recognition, we employed NMR

methods to determine the structure of p140 (375-480) bound to dsDNA. Despite numerous efforts, the data obtained were not sufficient to determine the solution structure of the DNA portion of the complex, therefore only the structure of the protein in the complex was determined from experimentally derived restraints. The resulting structure of p140(375-480) consists of a consensus BRCT fold preceded by an  $\alpha$ -helix that connects to the core domain by a long loop. A model of the protein-DNA complex was generated using HADDOCK (17), an algorithm that docks two molecules using ambiguous interaction restraints based on a variety of experimental data including mutagenesis, ambiguously assigned intermolecular NOEs and amino acid conservation.

---

## Materials and Methods

---

### *Expression and purification of RFC p140 (375-480)*

The expression and purification of RFC p140 (375-480) were performed as described in the Materials and Methods in Chapter 4.

### *Preparation of DNA*

The oligonucleotide of sequence pCTCGAGGTCGTCATCGACCTCGAGATCA was produced by standard solid state synthesis. The synthesized DNA was dissolved in 0.1 M NaOH and applied to a Q-sepharose column (Amersham Bioscience), also pre-equilibrated with 0.1 M NaOH. The DNA was eluted by increasing concentrations of NaCl in the same buffer. The volume of the collected peak fraction was reduced by rotary evaporator and the buffer was exchanged to 25 mM Tris-HCl pH 7.5, 50mM NaCl by PD10 desalting column (Amersham Bioscience). The purity of the DNA was analyzed by MS.

### *Protein-DNA complex preparations*

Both the protein and the DNA solutions were diluted in 25mM Tris-HCl pH7.5, 50mM NaCl and 1mM DTT to 10  $\mu$ M in order to prevent aggregation, mixed in the molar ratio of 1 to 1.2 and concentrated to approximately 0.5 mM by vacuum dialysis (Spectrum Labs) using a 10 kDa cut-off membrane. Subsequently, the buffer was exchanged to 25 mM D<sub>11</sub>-Tris-HCl pH 7.5, 5mM NaCl in 95/5 H<sub>2</sub>O/D<sub>2</sub>O.

### *NMR spectroscopy*

All NMR data were acquired at 298 K on a Bruker DMX600 spectrometer. Most of the sequential assignments for the backbone were obtained using 3D HNCACB,

CBCA(CO)NH and HBHA(CO)NH spectra. Aliphatic side-chain resonances were derived from 3D HCCH-TOCSY and CCH-TOCSY spectra. Additional data provided by 2D [ $^1\text{H}, ^1\text{H}$ ] NOESY (150 ms), 3D [ $^{15}\text{N}, ^1\text{H}$ ] NOESY-HSQC (150 ms) and [ $^{13}\text{C}, ^1\text{H}$ ] NOESY-HSQC experiments (150 ms) were used for further assignment of aromatic side-chain resonances as well as confirmation of the through-bond data. An additional 3D [ $^{15}\text{N}, ^1\text{H}$ ] NOESY-HSQC was recorded at 310 K for structure calculation. Spectral data were processed using NMRPipe (18).

The following half- and double- filtered experiments were acquired: a 2D NOESY ( $\tau_m = 150$  ms) recorded at 900 MHz with HMQC purge set to reject  $^{13}\text{C}$ - and  $^{15}\text{N}$ - coupled protons during  $t_1$  and to accept  $^{13}\text{C}$ - and  $^{15}\text{N}$ - coupled protons during  $t_2$ , and a 2D NOESY ( $\tau_m = 150$  ms) run at 900M Hz with HMQC purge set to reject  $^{13}\text{C}$ - and  $^{15}\text{N}$ - coupled protons during both  $t_1$  and  $t_2$ . Details of the pulse sequences used are given in the Supplementary Materials S2 and S3.

#### *Resonance assignment*

The assignment and the integration of NOE peaks was performed using the computer program CARA (19) available at <http://www.nmr.ch>) The majority of the chemical shift assignments of the protein bound to the DNA were obtained by comparing the data from the through-bond coupling experiment 3D CCH-TOCSY to the 3D [ $^{15}\text{N}, ^1\text{H}$ ] NOESY-HSQC. Approximately 83 % of the  $\text{C}\alpha$ - $\text{H}\alpha$  and  $\text{C}\beta$ -  $\text{H}\alpha$  correlations were missing in the [ $^{13}\text{C}, ^1\text{H}$ ] NOESY-HSQC. The chemical shift assignments of the protein bound to DNA have been reported (20) and deposited (BMRB accession number 6353).

Spectra from through-bond coupling experiments contained substantially fewer peaks in the case of free protein than in that of the complex, at the same time approximately 43 % of the amide backbone correlations were missing in the [ $^{15}\text{N}, ^1\text{H}$ ] NOESY-HSQC in the free protein, presumably due to rapid exchange with water. Therefore no reliable chemical shift assignments were obtained for the free protein.

#### *Structure Calculations*

Distance restraints were derived from the automated NOE crosspeak assignment of the 3D [ $^{15}\text{N}, ^1\text{H}$ ] NOESY-HSQ (recorded at 310 K) and the [ $^{13}\text{C}, ^1\text{H}$ ] NOESY-HSQC (recorded at 298 k) using the algorithm CANDID implemented in the computer program CYANA 2.0 (21). The chemical shift tolerances used in the automated assignment were 0.02 ppm for protons and 0.1 ppm for heavy atoms. Structure calculations were performed

with CYANA 2.0 (22). The structures were calculated using the NOE derived distance restraints and the dihedral angle restraints calculated from the chemical shift values of C $\alpha$  and C $\beta$  by TALOS (23). One hundred structures were calculated starting from conformers with random dihedral angles and using simulated annealing and torsion angle dynamics (TAD) as implemented in CYANA 2.0. The distance restraints which resulted in the 20 structures with the lowest CYANA 2.0 target function were converted to CNS format. Since a water refinement protocol is not available for CYANA, the water-refinement protocol integrated in CNS 1.0 was used. The structures were then recalculated with the distance restraints and dihedral angle restraints using the standard simulated annealing protocol in the computer program CNS 1.0 (24). The 20 lowest-energy structures with no distance violations greater than 0.3 Å and no angle violations greater than 5° were subjected to water-refinement following the scheme described (25). The 20-lowest energy structures with no NOE violation greater than 0.3 Å and no angle violations greater than 5° were accepted as the final structures representing the solution conformation. The quality of the structures was assessed using the program PROCHECK (26).

#### *Docking protocol*

The docking was performed with HADDOCK (17) and the average relative solvent accessible surface for each residue was calculated using NACCESS (<http://wolf.bms.umist.ac.uk/naccess>). In HADDOCK, the docking is driven by ambiguous intermolecular restraints (AIRs) which are defined according to established criteria of the program (17). The "active" residues are either those that have been shown by mutations to abolish or significantly perturb complex formation with average solvent accessible surface above 50 %, or those with observable intermolecular NOE to the dsDNA. In this case, the counterpart "passive residues" correspond to the nucleotide residues of the dsDNA. As a result, 8 amino acids of p140(375-480) (Y382, Y385, R388, T415, G415, R423, K458, and K461) with > 50% solvent accessibility were defined as "active" residues. Ambiguous interaction restraints (AIRs) were then generated between the active residues of p140(375-480) and the passive residues of dsDNA (Table 5.4). A more specific AIR restraint was introduced between T415 and the 5' phosphate group of dsDNA based on its conservation in the N-terminal BRCT domain of BRCA1 which is known to form hydrogen bonds to a phosphate. Side chain and backbone flexibility that allow local rearrangement during the semi-flexible simulated annealing step of HADDOCK, are confined to the segments around



the active and passive residues, that is residues 377-392 and 414-462 in p140(375-480) and the entire DNA molecule (Table 5.4).

The starting structures for docking were the 20 NMR structures of p140(375-480) and 3 models of dsDNA. Since no structure of the DNA portion of the complex is available, a model structure of 5' phosphorylated dsDNA with a 3' single stranded overhang in the standard B-form DNA with 3 conformations was generated, using the sequence of an oligonucleotide identical to that used for the NMR studies except for the fact that it did not contain a hairpin (5'pCTCGAGGTCG3'/5'CGACCTCGAGATCA3'). Docking of the p140(375-480)-dsDNA complex was performed following the protocol of HADDOCK1.3 (17). Inter- and intramolecular energies are evaluated using full electrostatic and van der Waal's energy terms with a distance cutoff using OPLS nonbonded parameters as defined in the default protocol (17). During the rigid body energy minimization, 2400 docking structures were generated (4 cycles of orientational optimization for each combination of starting structures were repeated 10 times). The best 200 structures in terms of intermolecular energies were then used for the semi-flexible simulated annealing, followed by explicit water-refinement. Finally the structures were clustered using a 5 Å RMSD as a cut-off based on the pairwise backbone RMSD.

#### *Analysis of intermolecular contacts*

Intermolecular contacts (hydrogen bonds and nonbonded contacts) in the ensemble of five best complex structures from the clusters with the lowest HADDOCK score were analyzed with the NBPLUS which is the part of NUCPLOT software (27). Used settings for hydrogen bonds were the distance cut-offs of < 2.7 Å and < 3.35 Å respectively for proton -acceptor (H-A) and proton donor -acceptor(D-A) provided that the D-H-A angle and H-A-AA are > 90°, where AA is the atom attached to the acceptor (for more details see (27)). 3.9 Å was the distance cut-off for nonbonded contacts.

---

## Results

---

### *Preparation of the protein-DNA complex*

The oligonucleotide used to form the protein-DNA complex was a 5' phosphorylated, 28-mer containing two self-annealing 10 nucleotide sequences separated by a 4 base hairpin turn. Upon annealing the oligonucleotide contains a single binding site for RFC p140 (375-480) with a 4 nucleotide long, 3' single stranded tail. This

oligonucleotide, has been shown by electrophoresis mobility shift assays (Figure 2.3) to bind RFC p140 (375-480) with  $K_D \sim 10$  nM. Since mixing of protein and DNA at high concentrations ( $> 0.1$  mM) resulted in severe protein precipitation, complex formation was performed under dilute conditions (range 5-10  $\mu$ M of each constituent). The complex was then concentrated to 0.5 mM by vacuum dialysis using a 10 kDa cut-off membrane. The starting protein/DNA ratio was 1:1.2 to ensure the formation of a full complex with a 1:1 stoichiometry (Figure 2.2). Excess DNA eluted through the dialysis membrane. No signals from unbound protein could be detected in the NMR spectra.

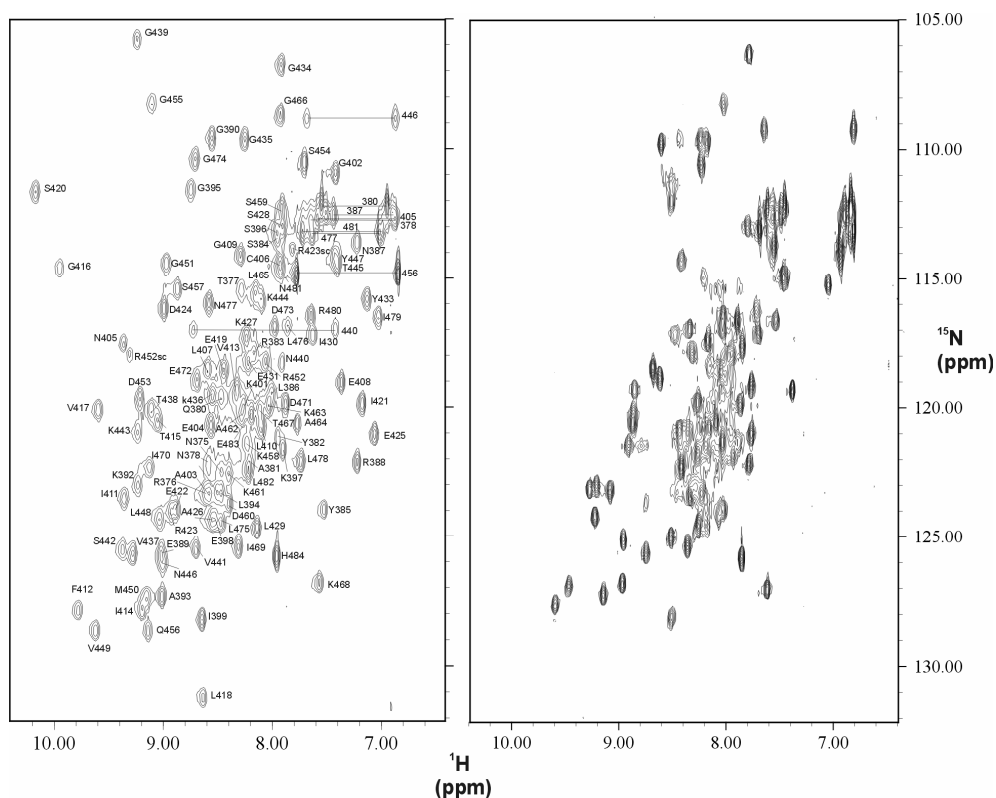
### *NMR spectroscopy and Resonance Assignment*

Despite the moderate quality of the NMR data (Figure 4.1), we obtained over 90% of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shift assignments for the observable resonances of RFC p140(375-480) bound to the oligonucleotide (20). Approximately 83% of the  $\text{C}\alpha$ - $\text{H}\alpha$  and  $\beta$ - $\text{H}\alpha$  correlations were missing from the [ $^{13}\text{C}$ ,  $^1\text{H}$ ] NOESY-HSQC. A likely explanation is highly efficient relaxation due to dynamic behavior intermediate on the NMR time scale within the complex. The majority of the chemical shift assignments were obtained by analysis of data from the through-bond experiments HNCABC, CACBCONH, HCCH/CCH-TOCSY while the 3D [ $^{15}\text{N}$ ,  $^1\text{H}$ ] NOESY-HSQC supplied confirmatory and supplemental correlations. 99% of the backbone assignments were determined, with the only missing residues being P391, P400 and Y379. As typically is the case in NMR, the N-terminal residue, here methionine, was also not observed. The sequential assignment is available from the BMRB (<http://www.bmrb.wisc.edu/>) under the accession number 6353.

RFC p140(375-480) is rich in basic residues (11 lysines and 7 arginines). Many of the side chain carbon and proton resonances from the basic residues suffered from severe overlap in the 3D HCCH/CCH-TOCSY and 3D [ $^{13}\text{C}$ ,  $^1\text{H}$ ] NOESY-HSQC spectra. The side chain  $\text{C}\epsilon$ / $\text{H}\epsilon$  correlation of 9 lysines (residue numbers 375, 392, 401, 443, 444, 427, 436, 461 and 468) resonated very close to the random coil values, suggesting high mobility. The resonances of the side chains of K397 and K458 were mostly absent from the CCH-TOCSY and [ $^{13}\text{C}$ ,  $^1\text{H}$ ] NOESY-HSQC and only some of the resonances were assignable using the 3D [ $^{15}\text{N}$ ,  $^1\text{H}$ ] NOESY-HSQC and 2D NOESY spectra. Similarly, the side chain  $\text{C}\delta$ / $\text{H}\delta$  correlation of R376 and R480 resonated close to their random coil values, while those of R383 and R432 had unique chemical shift values for  $\text{H}\delta$ . Much of the resonances

of R388, R423 and R452 were missing from both CCH-TOCSY and [ $^{13}\text{C}$ ,  $^1\text{H}$ ] NOESY-HSQC and were therefore assigned using [ $^{15}\text{N}$ ,  $^1\text{H}$ ] NOESY-HSQC and 2D NOESY spectra. The  $^1\text{H}$  and  $^{15}\text{N}$  resonances of the side chain HεNε correlation of R423 and R452 were assigned using the [ $^{15}\text{N}$ ,  $^1\text{H}$ ] NOESY-HSQC spectrum.

The [ $^{15}\text{N}$ ,  $^1\text{H}$ ] HSQC spectrum of the free protein was poorly dispersed (Figure 5.1) and exhibited heterogeneous linewidth and intensity, all characteristics of a flexible and dynamic molecule. A considerable number of the expected peaks were missing from the HNCACB spectrum and 40% of the backbone amide correlations were missing from the [ $^{15}\text{N}$ ,  $^1\text{H}$ ] NOESY-HSQC spectrum. Presumably the amide protons are in rapid exchange with those of water. The completed chemical shift assignment of the free protein is therefore not currently available. However, analysis of the partial sequential assignment and circular dichroism spectroscopy data (Figure 2.1) suggested the existence of secondary structure elements. The spectrum of the DNA-bound protein was clearly better than that of the free form (Figure 5.1). The spectrum of the complex contained 105 of the 106 expected amide correlations and exhibited good dispersion with more homogeneous linewidths. The observation of two distinct sets of resonances for the bound and the free protein is characteristic of slow dissociation of the Protein-DNA complex on the NMR timescale. As a consequence, it was not possible to deduce the DNA binding site on the protein by chemical shift perturbation analysis.



(Figure 5.1)  $^{15}\text{N}, ^1\text{H}$  HSQC spectra of the protein-DNA complex (left) and of the free protein (right). The amide backbone assignments are plotted on the  $^{15}\text{N}, ^1\text{H}$  HSQC spectra of the protein-DNA complex. Only 5% of the total amide correlations overlap in the complex and free-protein spectra. All the side chain amides Gln and Asn were identified and are connected by lines. The side chain guanidinium resonances (folded) are indicated as “R423sc” and “R452sc”

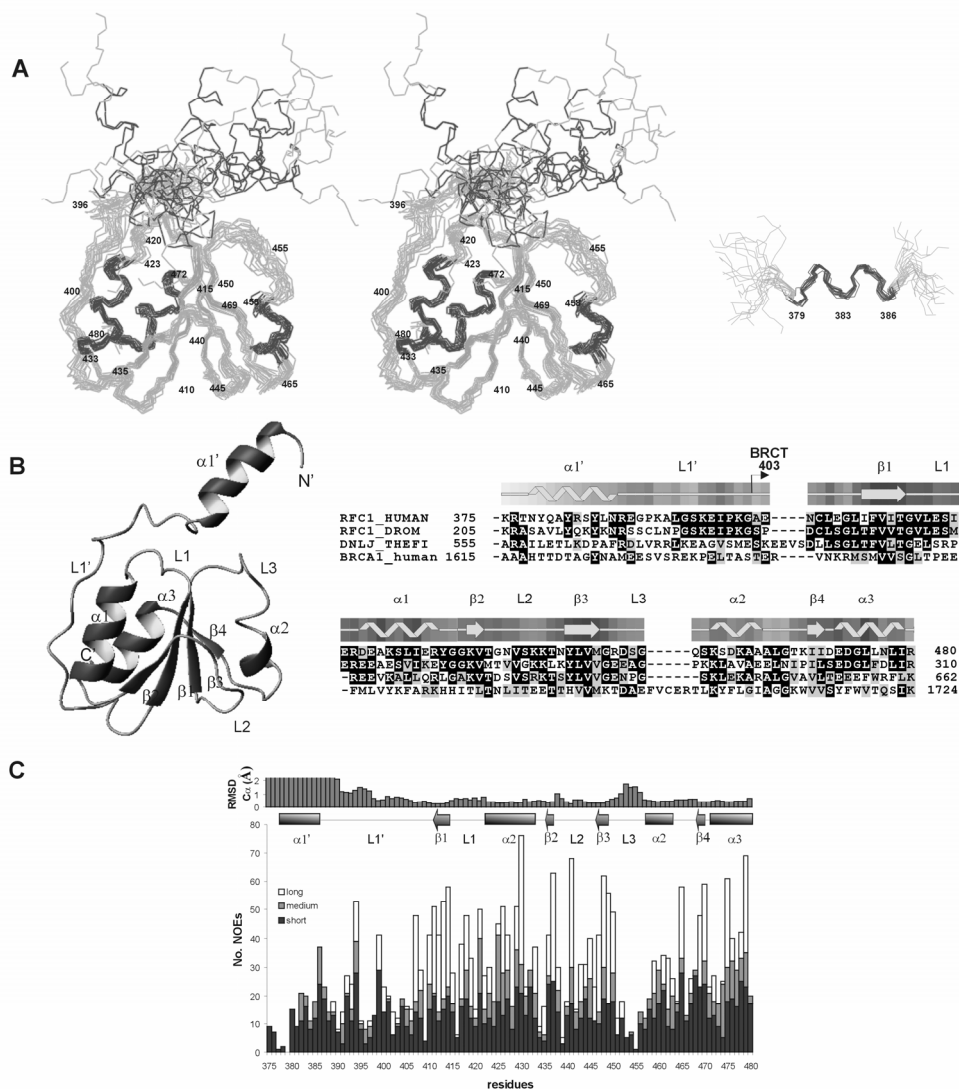
#### *Structure calculation of p140(375-480) when bound to DNA*

The structure of the protein moiety of the complex was determined primarily from distance restraints derived from NOEs in the 3D  $^{15}\text{N}, ^1\text{H}$  NOESY-HSQC and 3D  $^{13}\text{C}, ^1\text{H}$  NOESY-HSQC spectra as described in Materials and Methods. The best-fit superposition of the 20-lowest energy conformers is depicted in Figure 5.2A and the quality statistics of the structures are summarized in Table 5.1. With the exception of the N-terminal 16 residues (375-389), which were disordered with respect to the rest of the protein, and residues 451-457 in the loop 3 (L3), the backbone of p140 (375-480) is moderately well defined, with an average RMSD of  $1.07 (\pm 0.17)$  Å for backbone atoms and  $1.67 (\pm 0.23)$  Å for all heavy atoms (Table 5.1). The moderate definition of the structures reflects the low

number of long range distance restraints used in the structure calculation, which amounts to less than 5 restraints per residue (Table 5.1). None of the distances or dihedral angle violations was greater than 0.3 Å or 5.0° respectively in the final 20 structures. Analysis of the Ramachandran plot for all residues, using the computer program PROCHECK (26), showed that 83% of  $\phi$  and  $\psi$  angles lie within the most favored and 14.8% lie in the additionally allowed regions, while only 2.4 % are in the generously allowed or disallowed regions (Table 5.1). The residues that fall into the latter regions are mostly found in loops. The 3D structures of a number of other BRCT domains have been published (8-13) and the fold is highly conserved and consists of the following secondary structure elements:  $\beta$ 1- $\alpha$ 1- $\beta$ 2- $\beta$ 3- $\alpha$ 2- $\beta$ 4- $\alpha$ 3. The various structural elements of p140(375-480) have been labeled as  $\alpha$ (helix),  $\beta$ (sheet) and L(loop) for consistency with the reported structures of other BRCT domains and are depicted in Figure 5.2B.

(Table 5.1) Summary of restraints and structural statistics for the final RFCp140 (375-480) ensemble		
Restraints used in the calculation		
total number of NOE upper distance limits :		1782
intraresidue and sequential ( $ i-j  \leq 1$ )		1003
medium-range ( $1 <  i-j  < 5$ )		283
long-range ( $ i-j  \geq 5$ )		496
total number of dihedral angle restraints		
predicted (TALOS)		36
CYANA 2.0 outputs (20 structures)		
target function value ( $\text{\AA}^2$ )		0.85
number of distance restrain violations ( $> 0.2 \text{ \AA}$ )		1
number of dihedral angle constrain violations ( $> 5^\circ$ )		0
CNS 1.0/water-refinement (20 structures)		
number of distance restrain violations ( $> 0.3 \text{ \AA}$ )		0
number of dihedral angle constrain violations ( $> 5^\circ$ )		0
Structure statistics		
Final energies	(kcal/mol)	
total	(kcal/mol)	-3830
bonds	(kcal/mol)	42.5
angles	(kcal/mol)	151
improper	(kcal/mol)	54.4
dihedral	(kcal/mol)	485
van der Waals	(kcal/mol)	-430
electrostatic	(kcal/mol)	-4130
NOE	(kcal/mol)	0.351
RMS deviation from ideal values		
Bonds	( $\text{\AA}$ )	0.010
angles	( $^\circ$ )	1.25
PROCHECK Ramachandran plot analysis (375-480)		
Residues in most favored region (%)		83.0
Residues in additionally allowed region (%)		14.8
Residues in generously allowed region (%)		1.3
Residues in disallowed region (%)		0.9
RMSD to the averaged coordinates*		
		391-480
Backbone atoms		$0.91 \pm 0.16$
Heavy atoms		$1.44 \pm 0.20$

\* RMSD to the average coordinates of the 20 structures calculated excluding residues 375-389 and 451-457.



(Figure 5.2) Structure of RFC p140(375-480) when bound to dsDNA. **(A)** Left; Stereoview of a superposition of the backbone (N, C $\alpha$  and C') atoms for the 20 lowest-energy structures of RFC p140(375-480).  $\alpha$ -helices and  $\beta$ -strands are colored in red and cyan respectively. Right; superposition of the backbone of residues 379-386 demonstrating the well defined helix  $\alpha$ 1' the N-terminus. **(B)** On the left, sequence alignment of the p140(375-480) with the homologous region of RFC p140 from *Drosophila melanogaster* (RFC1\_DROM), NAD<sup>+</sup> dependent ligase of *Thermus filiformis* (DNLJ\_THEFI) and human BRCA1 BRCT-n. The secondary structure of human RFC p140(375-480) was determined by PROCHECK-NMR(26) and the level of solvent accessibility of the residues is colored from light (accessible) to dark blue (not accessible). The loop regions are labeled as L1', L1, L2 and L3. On the left a ribbon diagram of the p140(375-480) structure with the lowest-energy is presented. The orientation is equivalent to (A). **(C)** The number of short, medium and long range distance restraints per residue are plotted along with the RMSD (root mean square deviation) values of the C $\alpha$  atoms. The RMSD values were determined by superimposing the backbone atoms of each member of the ensemble of 20 conformers that form the BRCT domain itself (residues 403-480). Distant restraints are intra-residue NOEs ( $|i - j| < 1$ ), inter-residue medium NOEs ( $1 < |i - j| < 5$ ) and long-range NOEs ( $|i - j| > 5$ ). Figures (A) and (B) (right) were generated using MOLMOL (28)

### Structure description

Residues 403-480 of RFC p140 (375-480), which form the conserved BRCT domain, fold into a compact unit which consists of four parallel  $\beta$ -strands surrounded by helices  $\alpha 1$  and  $\alpha 3$  on one side and by helix  $\alpha 2$  on the other (Figure 5.2A). With the  $\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4\alpha 3$  topology, the 3D fold of residues 403-480 (Figure 5.2B) closely resembles that of BRCT domains whose 3D structure has previously been determined. The packing of the parallel  $\beta$ -strands forming the central core of the BRCT domain is stabilized mainly by hydrophobic interactions between I411, F412, V413, I414 ( $\beta 1$ ), and Y447, L448, V449 ( $\beta 3$ ). These strands are further surrounded by V437 of  $\beta 2$  and I469-470 of  $\beta 4$ . The packing of helices  $\alpha 1$  and  $\alpha 3$  against the central  $\beta$ -sheet occurs via hydrophobic interactions of A426, L430 (helix  $\alpha 1$ ) and L475, L478 (helix  $\alpha 3$ ) against the hydrophobic core. The BRCT domain of p140(375-480) belongs to a distinct subclass of the BRCT superfamily due to its amino acid sequence divergence from the rest of the family (5). One such difference from the rest of the superfamily is the substitution of the conserved tryptophan ( $\alpha 3$ ) by G474 in RFC p140(375-480), even though the amino acid substitution of W74 ( $\alpha 3$ ) to L in the XRCC1 BRCT revealed its significant contribution to the stability of the BRCT fold (14). The amino acid sequence identity among the BRCT homologs is very low (~14 %) (5), however the residues involved in hydrophobic interactions that make up the core of the protein are well conserved since substitution, when it occurs, is by a similar residue (Figure 3.1) (5). This pattern of conservation suggests that BRCTs from different species all adopt a similar, canonical three dimensional structure. Two of the most conserved residues in the BRCT superfamily are G434 and G435, which form a tight turn between  $\alpha 1$  and  $\beta 2$  (5). The substitution of any of these glycines by a larger residue could potentially destabilize the 3D structure. In the case of BRCA1 for instance, the G1788V mutation renders the tandem BRCT repeat more sensitive to proteolytic digestion (29). Likewise, the G617I mutation reduces both DNA binding and nick-adenylation activity of the bacterial NAD<sup>+</sup> dependent DNA ligase(30). Interestingly, the G193R mutation in the BRCT domain of Rev1 has been shown to interfere with the *in vivo* translesion bypass activity of REV1 in *S. cerevisiae* (31). This structural analysis suggests that the Rev1 G193R mutant results in decreased stability of the fold. The analogous G435R mutation in human RFC p140(375-480) resulted in a mutant prone to precipitation and with reduced DNA binding activity (Figure 2.2), both characteristics suggestive of a decrease in  $\Delta G_{\text{fold}}$ .



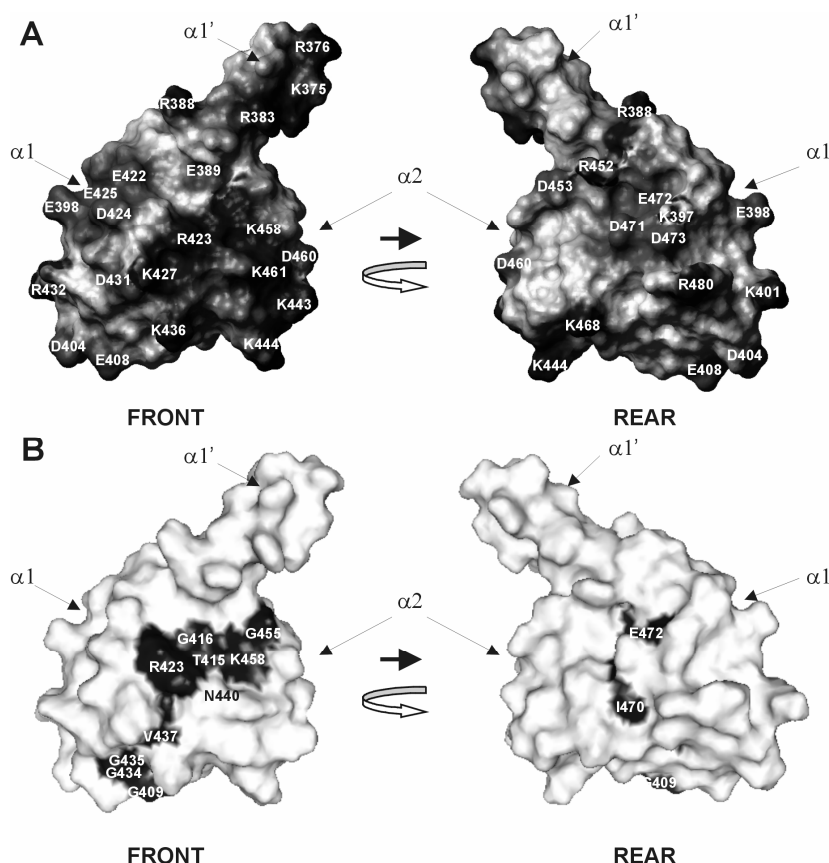
Helix  $\alpha 2$  is formed by residues K458-A463 (Figure 5.2B), which together with the preceding loop L3 are the most variable in size and sequence in the BRCT family. For instance helix  $\alpha 2$  is completely replaced by an extended L3 loop in the BRCT domain of human DNA ligase III (12). In the present structure, the L3 loop displays a high degree of disorder (Figure 5.1A) due to the limited number of restraints available within this region. It is not yet clear whether the disorder reflects actual dynamic motions within the L3 loop or simply a paucity of structural restraints (Figure 5.2C). To a lesser extent, loops L1, L2 and L4 display some conformational variation in the ensemble (Figure 5.2A and C). In most BRCT domains loop L1 appears to be more or less flexible as reflected by the high B-factors in X-ray crystal structures and poor definition in NMR structures (8;9;12;32). In relation to these other structures, the L1 loop is better defined and buried under loop L1' in the structure of RFC p140 (375-480) (Figure 5.2A and C).

The N-terminus of p140 (375-480), residues 375-403, forms an  $\alpha$ -helix ( $\alpha 1'$ ) which is separated by a loop (L1') from the core of the protein. Loop L1' packs against helices  $\alpha 1$  and  $\alpha 3$  of the conserved BRCT domain. Helix  $\alpha 1'$  (residues 379-386) appears consistently in all 20 structures (Figure 2A, right), however it is poorly defined with respect to the rest of the protein (Figure 2A). This lack of definition certainly reflects the absence of observable long range NOEs between helix  $\alpha 1'$  and the core of the protein (Figure 5.2C). The loop L1' is anchored to helices  $\alpha 1$  and  $\alpha 2$  through burying of the side chains of residues L407, P400 and L399 between the two helices, and through potential salt-bridging between the side chains of K397 (L1') and E472 or D473 ( $\alpha 3$ ), and of K392 (L1') and E419 (L1).

#### *A potential DNA interaction site*

The electrostatic potential at the surface of RFC p140(375-480) was calculated using the computer program MolMol (28) and is presented in Figure 5.3A. Casual inspection of Figure 5.3A reveals a large basic patch, composed of R423, K427, K436, K443, K444, K458 and K461 ("front" in Figure 5.3A), which might be a binding site for the negatively charged phosphate backbone of the DNA. Based on the amino acid sequence alignment (Figure 3.1), the surface exposed conserved residues of a number of orthologs of RFC p140 are mapped onto the surface of the p140(375-480) structure (Figure 5.3 B). The conserved residues are distributed mainly on the basic patch ("front" in Figure

5.3) of p140 (375-480) and localized within the BRCT domain rather than within the loop L1' or helix  $\alpha 1'$ . Comparison of Figures 3A and B indicates that the highly conserved residues R423 and K458 form part of the basic patch. Furthermore in the vicinity of this region, the highly conserved residues T415, G416 and G455 can also be found. Negatively charged surfaces, on the other hand, extend from the front to the “rear” of the molecule (Figure 5.3). A large patch of negative charge is located along helix  $\alpha 1$  (423-432) on the front side while the conserved E472 is found in the negative charged patch on the rear. It is important to note that the location of helix  $\alpha 1'$  relative to the core of the protein in Figure 5.3A is arbitrary.



(Figure 5.3) Surface potential of p140(375-480). (A) Electrostatic potential of the accessible surface of p140(375-480) is shown. Negative potential is colored in red and positive potential in blue. The residues contributing to the surface charge are indicated on the surface. (B) Surface accessible residues that are conserved (defined in Figure 3.1) are colored with blue onto the structure of p140 (375-480). N440 is not a conserved residue. An intermolecular NOE transfer between N440 and unassigned DNA proton was observed (Table 5.3). The “front” and “rear” images were generated by 180° rotation around the Z-axis. The electrostatic potential surface was calculated using MolMol (28) and the residue conservation map (B) was created with Pymol (33).

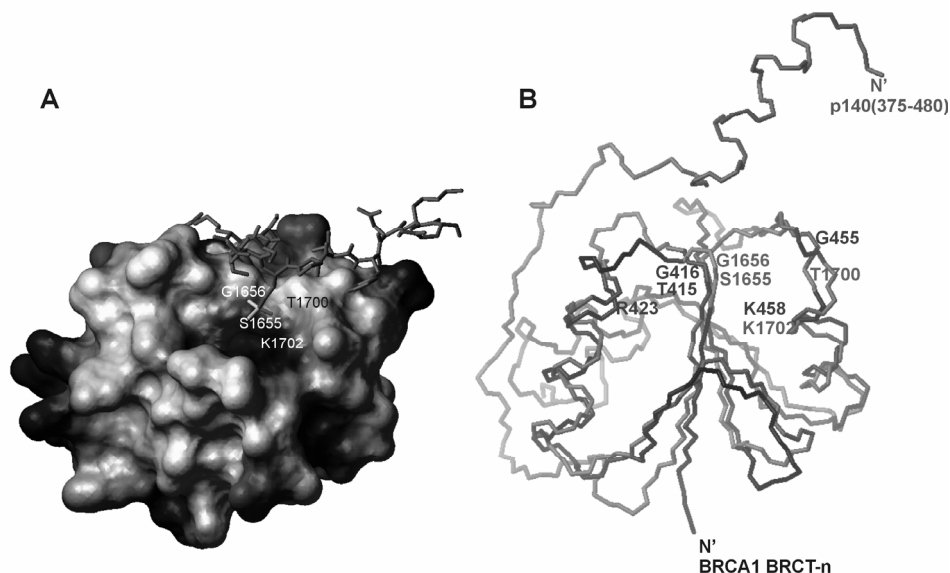
(Table 5.2) Comparison of the backbone fold of RFC p140(375-480) and various BRCT domains.

PDB	Description (reference)	Methods	RMSD* (Å)	Ref
1T29	BRCA1 BRCT-N	X-ray	2.5	(34)
1T29	BRCA1 BRCT-C	X-ray	3.0	(34)
1cdz	XRCC1 BRCT-N	X-ray	3.0	(8)
1l7b	NAD <sup>+</sup> dependent ligase BRCT	NMR	3.2	N/P
1gzh	p53BP BRCT-N	X-ray	1.9	(10)
1gzh	p53BP BRCT-C	X-ray	1.9	(10)
1wf6	RAD4+/CUT5+ PRODUCT	NMR	2.3	N/P
1in1	DNA ligase III $\alpha$ BRCT	NMR	2.7	(12)

N/P = Not published. \*The backbone RMSD (Å) between the p140(375-480) and the PDB structures. The segments required for optimal structural alignment between the two structures were identified using DALI server (35).

### *Comparison with the structure of other BRCT domains*

In order to compare the fold of RFC p140(375-480) to that of other BRCT domains, the backbone C $\alpha$  atoms were superimposed (Table 5.2). Recently, the crystal structure of a complex of the BRCA1 tandem BRCT repeats bound to a phosphoserine peptide has been elucidated (34;36;37). Structural comparison of p140(375-480) with the N-terminal domain of the BRCA1 tandem BRCTs revealed a striking similarity between the binding site for the phosphate moiety of the phosphoserine on BRCA1 (Figure 5.4A) and the conserved, basic patch of RFC p140(375-480) (Figure 5.4B). Residues of BRCA1 within direct hydrogen-bonding distance to the phosphate-moiety are identified as S1655 ( $\beta$ 1), G1656 (L2) and K1702 ( $\alpha$ 2) (Figure 5.4A)(36-38). Despite the low amino acid conservation, the position of the phosphate binding residues, corresponds well with conserved residues T415, G416 and K458 respectively of p140(375-480) (Figure 5.2B and 4). The hydrogen bond donor S1655 of BRCA1 is conservatively substituted by T1898 in the related protein MDC1. The recent crystal structure of the tandem BRCT domains of MDC1 bound to a phosphoserine peptide (39) suggests that the phosphate binding pocket is highly conserved. Analogously, the positive patch created by the conserved residues of p140(375-480) is likely to be important for the interaction with the 5' phosphate of dsDNA.



(Figure 5.4) Structure comparison between the BRCT domains

**(A).** Electrostatic surface presentation of the N-terminal BRCT (BRCT-n) of BRCA1 (PDB:1T29) in complex with a phosphoserine peptide (in magenta). The C-terminal BRCT domain is not directly involved in the phosphate binding and therefore has been deleted from this figure for clarity. Positive potential is shown in blue and negative potential in red. The amino acid residues forming the pocket that binds the phosphate moiety (in yellow) of phosphoserine are indicated on the surface. Phosphate is directly hydrogen bonded by S1655, G1656 and K1702. T1700 is a conserved residue which forms a hydrogen bond to the side chain oxygen of S1655 stabilizing the S1655 side chain conformation. **(B)** Backbone superposition of p140 (375-480) (red) and the BRCT-n from BRCA1 (black). The orientation of the BRCT-n is identical to that of A. The backbone C $\alpha$ , N and C' of the proteins are presented in red for p140(375-480) and in black for the BRCA1 BRCT-n. The conserved residues of p140(375-480) are presented in blue and the residues essential for the phosphate-moiety recognition of BRCA1 BRCT-n are presented in magenta (36-38). The overlay regions of p140 and (BRCA1) are 409-417(1649-1657), 421-452(1660-1691) and 453-479(1697-1723).

### Detection of intermolecular NOEs

The conventional approach to structure elucidation of molecular complexes using NMR is based upon the assignment of intermolecular NOEs that can be used as structure restraints in the calculations. The combination of  $^{13}\text{C}/^{15}\text{N}$  isotope labeling of RFC p140(375-480) and use of  $^{13}\text{C}/^{15}\text{N}$  isotope editing NMR techniques enables selective observation of the components, isotope attached, non-attached or both of a complex (reviewed in ref (40)). The technique has been developed to distinguish between intra and intermolecular NOEs of a complex in which one component is uniformly isotope-labeled. In order to obtain intermolecular NOEs and structural information on p140(375-480) bound to DNA, an  $F_1F_2$ -double-half filtered 2D NOESY spectrum was recorded. Although this experiment has been successfully used to selectively observe intra- and intermolecular NOEs in the dimeric *arc* repressor (41), it did not yield high quality data on the RFC

p140(375-480)-dsDNA complex. The failure of the isotope-filtered experiment may possibly be due to magnetization loss due to  $T_2$  relaxation during the long delay imposed by the refocused half-filter. We therefore tried an alternative approach based on purge pulses (42) which proved to be moderately successful. Two NOESY spectra were obtained by simultaneous suppression of  $^{13}\text{C}/^{15}\text{N}$  attached protons in both F1 and F2, or only in F1. The resulting  $F_1F_2$  double-filtered spectrum, which contains exclusively resonances from the unlabeled DNA, was different from the free oligonucleotide but was not sufficiently well resolved to perform a reliable sequential assignment. However, detailed comparison of the NOESY spectra listed in Table 5.3 allowed us to identify a few peaks arising from intermolecular magnetization transfer from DNA to protein (Figure 1S of supplementary materials). Due to the lack of sequence specific resonance assignments for the DNA, the identity of the source proton could not be ascertained.

(Table 5.3) Intermolecular NOEs observed between the p140(375-480) and DNA.

RFCp140(375-480)	DNA (possibilities) Ambiguously assigned	Intermolecular NOE (ppm)	Experiments
Y385 QD	CYT H5 or THY H1'	5.51	A,B
	CYT H6 or THY H6	7.72	
N440 HD21	CYT H5 or THY H1'	5.37	A,B,C
HB3	TCH <sub>3</sub>	1.5	A,B
G416 HN	CYT H6, THY H6	7.7	C
	ADE H2, H8 or NH <sub>2</sub> of ADE, CYT and GUA		
R423 HE	Ribose H2'' or H2'	2.19	B,C
	Ribose H4', H5'' or H5'	3.88	
G455 HN	Ribose H4', H5'' or H5'	3.94	A,B,C

A. 2D  $F_1$ , [ $^{13}\text{C}/^{15}\text{N}$ ]-filtered NOESY, B. 2D [ $^1\text{H}$ , $^1\text{H}$ ]-NOESY, C. 3D [ $^{13}\text{N}$ , $^1\text{H}$ ]-NOESY HSQC

As can be seen in Table 5.3, intermolecular NOEs between residue Y385 and the DNA were observed. The importance of this interaction was also demonstrated by mutagenesis (Figure 3.5). This observation supports the earlier speculation that the N-terminal helix ( $\alpha 1'$ ) is directly in contact with the DNA. The non-assignable NOE crosspeak correlations to the amide backbone of G416 and to the side chain guanidinium of R423 originate from DNA as well. The intermolecular NOEs to the amide side chain of N440 and the backbone of G455 were identified on the basis of the model p140(375-480)-dsDNA complexes generated by docking (described below). The majority of intermolecular NOEs that have been identified involve residues on one side of the protein surface within or in the vicinity of the conserved basic patch (Figure 5.3, Front side). However, the number of identifiable intermolecular NOEs is very small. This likely reflects

the mode of interaction between basic side chains of the protein and backbone phosphates of the DNA (although it is not possible to entirely rule out experimental limitations).

*Protein-DNA docking by HADDOCK using mutagenesis and ambiguous intermolecular NOE data*

Due to the lack of sequence specific resonance assignments for the DNA, it was not possible to calculate the structure of the bound dsDNA. In order to generate a model of the p140(375-480)-DNA complex despite the limited number of intermolecular structure restraints, an alternative approach using the docking program HADDOCK (17) was employed. HADDOCK can make use of a broader array of restraints including those derived from biochemical and biophysical data. This data was introduced as so-called Ambiguous Interacting Restraints (AIR) to drive the docking process. The mutagenesis (Figure 3.5), the intermolecular NOEs (Table 5.3) and the structural conservation (Figure 5.4) clearly indicate at least some of the residues that interact with the DNA. In the docking procedure, AIRs are therefore defined as ambiguous distances between these residues (called “active”) and the 5' PO<sub>4</sub> or any/specific (if known) nucleotides on the DNA (called “passive”) (Table 5.4). Although mutation of K379 resulted in the loss of DNA binding (Figure 3.5), it was not included as an AIR because its average solvent accessible surface is below 50% (see Materials and Methods).

An AIR was introduced between T415 and any nucleotide in the DNA and an additional AIR was specifically generated to the 5' phosphate of the DNA on the basis of the following three observations. (Table 5.4). 1) The resonance of the  $\gamma$  <sup>1</sup>H of T415 has been tentatively assigned indicating that it is in slow exchange with the solvent. Reduced exchange is a strong indication of involvement in hydrogen bonding while inspection of the protein structure indicates that there are no neighboring residues within sufficiently close distance to form a hydrogen bond. 2) Residue T415 is structurally conserved with S1655 of the BRCA1 and T1898 of the MDC1, which are crucial for binding the phosphate moiety of phosphoserine (Figure 5.3). 3) The T415A mutant (Figure 3.5) exhibited reduced DNA binding.

(Table 5.4) Active and passive residues used in the definition of the ambiguous distance restraints (AIRs) and the flexible fragments used in the HADDOCK of p140(375-480) and DNA.

<i>Active residues of p140(375-480)</i>	<i>Method determined</i>	<i>Passive residues of dsDNA</i>
Y382, Y385, R388, T415, R423	Mutagenesis data (Figure 3.5)	Any DNA nucleotide (5'pCTCGAGGTCG3'/
K458, K461	Residue conservation	5'CGACCTCGAGATCA3')
G416		
Y385 H $\delta$	Intermolecular NOEs (Table 5.3)	Any H2, H6, H8
T415 H $\gamma$		O1P, O2P, O3P of 5'pC19
G416 HN		Any H2, H6, H8
R423 H $\epsilon$		Any H4', H5', H5''
	<i>Residue numbers of p140(375-480)</i>	<i>dsDNA</i>
Flexible fragments	377-392, 414-464	All nucleotides

In HADDOCK, the docking protocol consists of three stages. Initially a protein and a DNA molecule are randomly rotated around the center of mass. Subsequently the orientation of each molecule is optimized in order to minimize the intermolecular energy functions, and then the two are docked using a rigid body energy minimization (EM). During this initial docking, proteins are treated as rigid bodies forbidding movement of side-chains or flexible loops. However, this lack of flexibility could be detrimental to determining the correct initial orientation of the interacting surfaces. In order to overcome this problem, the HADDOCK protocol was applied to each member of the ensemble of 20 NMR structures of p140(375-480). Since different side chain and/or backbone orientations allow better sampling of conformational possibilities, this method has been shown to produce better docking solutions that are closer to known complexes than use of a single structure (43). As no structure of the DNA portion of the complex is available, a model structure of 5' phosphorylated, dsDNA with a 3' single stranded overhang in the standard B-form conformation was generated using the sequence of an oligonucleotide used in the NMR studies but lacking the hairpin since it does not interact with p140(375-480). In the second stage of the docking protocol, HADDOCK accommodates conformational rearrangements of side chain and backbone of defined segments (Table 5.4) to optimize interaction at the protein-DNA interface. The N-terminal amino acid segment residues 377-392 of p140(375-480), which was less defined in the ensemble of structures, was allowed

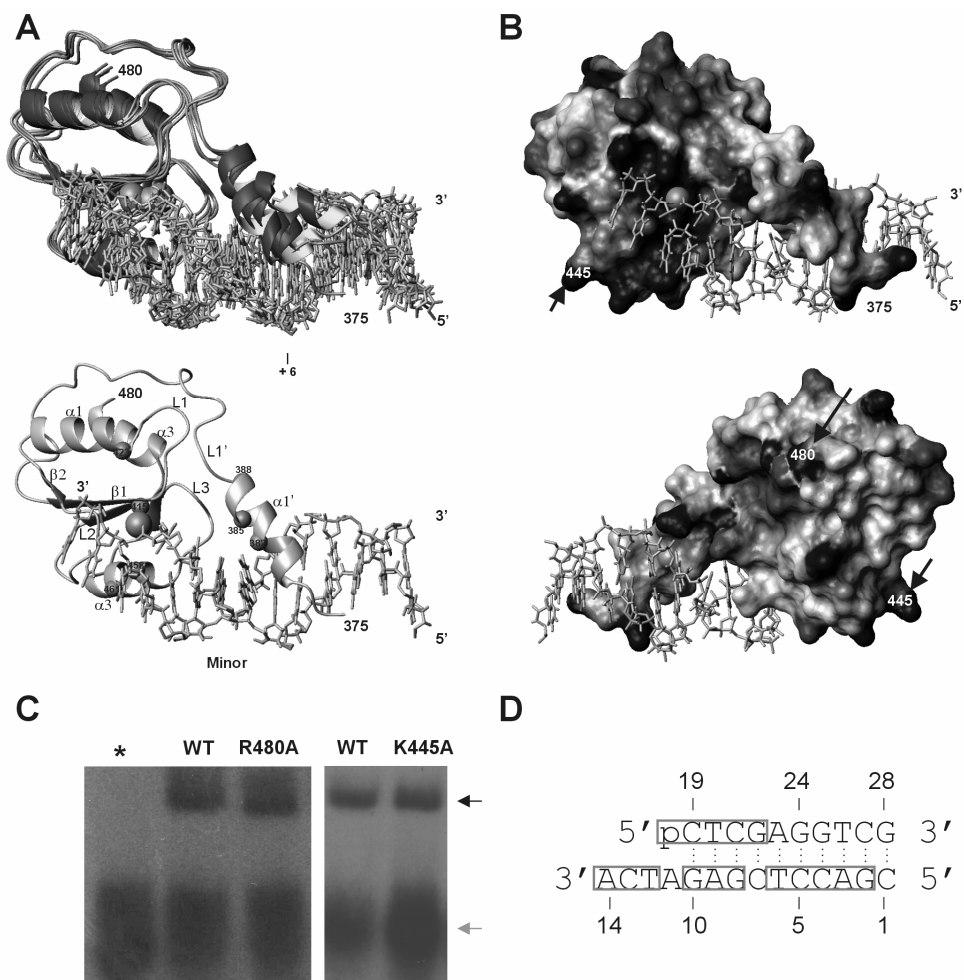
to move freely. In the last stage of docking, the protein-DNA complexes are refined in an explicit water bath. As a result, the docking process generated 200 solutions that were sorted into clusters using a pairwise backbone RMSD of 5Å as a cutoff criterion (Table 5.5). This procedure resulted in 10 clusters, which were then ranked according to their HADDOCK scores calculated on the basis of the intermolecular energy (the sum of electrostatic, van der Waal's, and AIR energy terms) and their average buried surface area (Table 5.5). The five best structures from cluster 7, which had the lowest HADDOCK score of any cluster, were accepted as the best representative model of the complex (Figure 5.5A).

Table 5.5. Statistics of the five structures from each 8 clusters, which are ranked on the basis of their HADDOCK scores.

#cluster	RMSD <sup>1</sup>	Sd <sup>2</sup>	HADDOCK score <sup>3</sup>	Sd <sup>4</sup>	BSA <sup>5</sup>
clust7	1.3	0.8	-63	55	2081
clust1	4.8	1.9	-15.9	11.6	1924
clust4	3.3	1.5	17.4	32.5	2039
clust2	4.7	1.9	64.4	16.2	1419
clust6	5	2.4	117.9	43.9	1483
clust5	5	1.5	114.7	23.5	1346
clust3	4	1	135.2	23.3	1660
clust8	4.3	1.3	151.4	48.4	1502

<sup>1</sup> RMSD from the lowest energy structure in the cluster; <sup>2</sup> standard deviation of RMSD; HADDOCK score calculated in the basis of the intermolecular in interaction energy which is the sum of electrostatic, van der Waals and AIR energies at the interface. <sup>4</sup>Standard deviation of HADDOCK score. <sup>5</sup> BSA is a "buried surface area" in the complex (Å<sup>2</sup>).





(Figure 5.5) Docking model of the p140 (375-480)-dsDNA complex generated by HADDOCK. **(A)** Ribbon representation of ensemble of the five lowest energy structures from the cluster 7. The 5'phosphate is indicated by a magenta sphere. One of the ensemble structures in which the active residues defined in HADDOCK are presented in yellow spheres (bottom). **(B)** Electrostatic surface potential presentation of p140 (375-480) bound to dsDNA. The upper figure has the same orientation as in A while the lower has been rotated 180° around the Z-axis (bottom). The positive and negative charged surfaces are colored in blue and red respectively. Exposed hydrophobic residues are white or slightly colored. **(C)** DNA binding activity of R480A and K445A. The mutations were designed as tests of the model of the Protein-dna complex. The wildtype DNA binding activity of the R480A and K445A mutants suggests that the “rear” side and the bottom of p140(375-480) (indicated with an arrow in **(B)**) do indeed not contact the DNA. **(D)** The sequence of the dsDNA used in the HADDOCK docking. The nucleotides with phosphates and bases in contact with the protein are boxed.

(Table 5.6) Intermolecular contacts identified over the ensemble of five best structure of cluster 7 representing p140(375-480)-dsDNA complex.

Residues	Secondary structure	DNA*	Hydrogen bonds				Nonbonded contacts		
			B-P	B-b	S-P	S-b	B-b	S-P	S-b
K375	$\alpha 1'$	G2				2			
		A3				1			
		C4				1			
Y382	$\alpha 1'$	C21				1		5	1
		C20				1			
		G22				1			
Y385	$\alpha 1'$	C21							3
		T6				1			1
		G22				1			
		C5							1
		T20							3
R388	L1'	T6				3		1	
		C5				1			
T415	$\beta 1$	C19				4		3	3
G416	L1	C19	4					2	
R423	$\alpha 1$	C19				2		1	
		A14				2			
T438	L2	A14					3		
N440	L2	C13				3			2
R452	L3	C19				1		2	
		T20				1			
S454	L3	C19					4		
		T20					3		1
G455		C19					3		
Q456	L3	G8				2		1	1
		A9					1		
S457	L3	G10				3			4
		A9							4
K458	$\alpha 2$	C19				2	1	3	1
		T12				1			

\* The numbering of the DNA refers to the Figure 5.5D.

Intermolecular contacts were analyzed by NUCPLOT (see Materials and Methods) and listed only when the interacting residues were present in three out of five structures. The identified hydrogen bonds and nonbonded contacts are classified into protein Backbone-DNA phosphate (B-P), Backbone-DNA base (B-b), side chain – DNA phosphate (S-P), side chain-DNA Base (S-b).

The residues interacting with the DNA in the five model complexes were identified using the HBPLUS (44) in the NUCPLOT (45) package. HBPLUS identified hydrogen bonds and non-bonded contacts made by pairs of atoms (Materials and Methods for the details), which were sorted according to the amino acid residue numbers. Residues which were identified as interacting with DNA in three out of five model complexes are listed in Table 5.6. In the model of the p140(375-480)-dsDNA complex, the 5' phosphate (C19, Figure 5.5D) of the dsDNA is accommodated by the positively charged surface (Figure 5.5B) mainly formed by the conserved residues T415, G416, R423 and K458 (Figure 5.3). The side chains of T415, R423 and K458, and the amide backbone of G416 are within the hydrogen bonding or salt-bridging distance to the oxygens of the 5' phosphate (Table 5.6). This is due to the AIR restraint introduced between T415 and the 5' phosphate (C19), while in the absence of this restraint; we did not obtain docking solutions in which the 5' phosphate is bound by the protein. Both mutants (R423A and T415A) exhibited reduced DNA binding but were not as severely affected as the K458E mutation (Figure 3.5). It is possible that the network of hydrogen bonds to the 5' phosphate by the remaining residues is sufficient to partially compensate for the loss of one hydrogen bond donor resulting from the electrically neutral alanine mutation. On the other hand, the introduction of negative charge in the K458E mutation, as might be expected, has a more drastic effect on DNA binding. Other than the conserved residues, R452 also interacts with the 5' phosphate via dominantly electrostatic interaction. Although no intermolecular NOE was observed between R452 and the DNA, the H $\epsilon$ N $\epsilon$  resonance of R452 is folded in [ $^1\text{H}$ ,  $^{15}\text{N}$ ] HSQC spectrum of the p140(375-480)-dsDNA complex, which suggests stabilization of the side chain conformation upon DNA binding often through a charge mediated, nonbonded interaction. In this model, K461 approaches the phosphates within coulombic distance ( $\sim 6$  Å) providing additional charge interaction. However in contrast to the K461E mutation data, which resulted in dramatic reduction of DNA binding (Figure 3.5), K461 does not seem to play a major role in DNA binding in the model structure.

In the model presented in Figure 5.5, the orientation of the helix  $\alpha 1'$  is better defined than in the free protein, and lies in the major groove (Figure 5.5A) clearly separated from the core of the protein, which explains the lack of long range NOEs in this region. The amino group of K375 ( $\alpha 1'$ ), the hydroxyl group of Y382 ( $\alpha 1'$ ) and the guanidinium group of R388 ( $L1'$ ) participate in hydrogen bonds or electrostatic

interactions with the phosphate backbone of the DNA (Table 5.6). In addition, numerous interactions including hydrogen bonds and van der Waal's contacts occur between the aromatic side chain of Y385 ( $\alpha 1'$ ) and nucleotide bases (Table 5.6) of the major groove. Reduced DNA binding of p140(375-480) was observed when the size of the DNA duplex becomes less than 7 nucleotides long or when the +6 nucleotide position (G24) from the 5' phosphate end contains a non Watson-Crick basepair (T24:C5) (Figure 2.3). Accordingly the model complexes suggest that the shortening of the duplex and the non Watson-Crick base pair may interfere with the interaction of Y385 as well as with the interaction of R388 and K375 with dsDNA. No interaction between S384 and dsDNA was found in the model complexes, a result that is in concert with the mutagenesis study (Figure 3.5).

The model of the protein-DNA complex suggests that the 3' single stranded DNA tail (nucleotides C13 and A14) interacts via the backbone phosphates with the side chain of R423 and via the bases with the side chains of T438 and N440. Importantly, there are no restraints in the calculations that explicitly define this interaction. Therefore, the model explains the earlier observation that p140 (375-480) binds a 5' recessed dsDNA with higher affinity than blunt ended DNA (Figure 2.3). Based on the structure, we identified an intermolecular NOE between the side chain amide of N440 and the DNA (Table 5.3). The side chain amide resonance of N440 is shifted away from the random coil value (Figure 5.1) suggesting involvement in hydrogen bonding. Since there are no hydrogen bond acceptors on the protein in close vicinity of N440, the likely acceptor is the DNA. However, the present model does not include this restraint.

The backbone atoms of the S454, G455, Q456 and S457 in Loop 3 (L3) make a number of van der Waal's contacts with the bases of the DNA (Table 5.6). Although Loop 3 (L3) is disordered in the ensemble of 20 structures of the protein without DNA (Figure 5.2A), the loop orientation is better defined in the protein-DNA complexes (RMSD of the L3(451-457) is  $1.36 \pm 0.4$  Å in the free protein in comparison to  $0.39 \pm 0.14$  Å in the model complex). Similarly to N440, an intermolecular NOE between the backbone HN of G455 and an unassigned DNA proton was identified based on this model. Since the backbone HN of G455 resonates at the random coil position, it apparently does not take part in a hydrogen bond interaction with the DNA or protein, suggesting that it serves a structural role in the loop instead.

In order to test the veracity of the proposed model we sought to make further site directed mutations. Visual inspection of the structures presented in Figure 5.5 indicates that

residues R480 and K445, both of which have  $\geq 60\%$  amino acid identity among the 31 species analyzed (Figure 3.1), lie on the face of p140(375-480) that does not contact the DNA (Figure 5.5B arrows). We therefore generated R480A and K445A mutations and tested their effect on DNA binding. As shown in Figure 5.5C, neither of the mutations disrupts DNA binding, as was expected. This experiment serves as a further control that disruption of DNA binding observed in the earlier mutagenesis studies was not the result of a subtle global change in the protein fold that could not be detected in the 1D NMR spectra.

---

## Discussion

---

### *Structure of p140(375-480) in the presence of DNA*

Here the first structure of a BRCT domain bound to DNA is presented. Preliminary analysis of the NMR data suggests that the free protein is flexible in solution but becomes more rigid upon binding to DNA. Even in the presence of DNA, p140(375-480) still exhibits dynamic behaviour which caused substantial loss of NMR signals. The moderate resolution of the structures based on NOE derived restraints is a direct consequence of the poor quality NMR data.

### *The p140(375-480)-dsDNA complex generated by HADDOCK*

The model complex of p140(375-480) and dsDNA was generated using the program HADDOCK which uses Ambiguous Interaction Restraints (AIRs) derived from mutagenesis, intermolecular NOEs and residue conservation to drive intermolecular docking. Although the model complex is in good agreement with all of the DNA binding properties characterized by biochemical analysis, it should nonetheless be considered as preliminary pending further experimental analysis.

Despite the fact that no specific restraints were included in the calculations, helix  $\alpha 1'$  lies in the major groove of the DNA helix in the best structures generated by HADDOCK. The  $\alpha$ -helix is a commonly used structural element for recognition of bases as well as backbone phosphates in sequence specific (reviewed in (46)) and non-sequence specific DNA binding (47). In the non-specific complex of DNA *-lac* headpiece-62, many of the side chains (Y7, Y17, Q18 and R22) that confer direct interactions with the base pairs in the major groove of the sequence specific complex, shift and participate in hydrogen bonds or electrostatic interactions with the backbone phosphates (47). Similarly in the p140(375-480)-dsDNA models, non-sequence specific interaction is mediated by

hydrogen bonds or salt bridges between K375, Y382 ( $\alpha 1'$ ) and R388 (L1') and backbone phosphates. In contrast to the sequence specific *lac* headpiece-DNA complex in which base specific contacts are made by several residues, in p140(375-480)-dsDNA Y385 is the only residue in helix  $\alpha 1'$  that points into the major groove. Amongst the five different structures in the cluster, Y385 contacts five different base pairs. This type of behavior is consistent with a dynamic, non-sequence specific complex.

Based upon analysis of the structure of the complex, 36 intermolecular NOEs would be expected. However, to date only intermolecular NOEs involving G455 and N440 have been definitively identified. This discrepancy can be explained in three ways. Clearly the model of the complex may be incorrect. However, since the model explains a considerable amount of biochemical data that was not used to generate it, this seems unlikely. Alternatively, the expected NOE correlations may be missing or overlapped in the NMR spectra. It is likely that at least some of the missing correlations are due to overlap in crowded regions of the spectra or to experimental limitations (Figure 1S). NMR methods commonly used to detect intermolecular NOE correlations failed on the p140-DNA complex despite the fact that the complex is not unusually large. An experiment based on purge pulses (42), which reduces the amount of time required to perform the magnetization filter, yielded moderate results. The limited amounts of intermolecular NOEs that have been assigned were derived from this experiment. This observation, in conjunction with the previously reported missing correlations in the 3D [ $^{13}\text{C}$ , $^1\text{H}$ ]-NOESY HSQC spectrum (Figure 4.1) and the below normal number of long range NOEs, reflects the nature of the complex in which the residues making contact with DNA likely undergo intermediate time scale exchange between conformations. This intermediate exchange rate will lead to loss of resonance signals due to efficient relaxation of transverse magnetization. In the case of the non specific DNA-*lac* headpiece-62 complex, residues located in the protein-DNA interface were clearly shown to undergo exchange dynamics on the  $\mu\text{s}$ -ms timescale indicating that they sample different base pair environments (47). Water molecules play important roles in filling the gaps in protein-DNA interfaces, which are otherwise energetically unfavorable, as well as in water-mediated hydrogen bond with bases and phosphate backbone at the protein-DNA interface of both specific (48) and non-specific complexes (49). In the interface of a non-specific protein-DNA complex of BamHI, all specific interactions are replaced by non-specific interactions mediated by hydrogen bonds

to the DNA backbone via highly mobile water molecules (49). Although water molecules were not included during the docking, one can not rule out their potential participation at the p140(375-480)-dsDNA interface, which maybe also contribute to the limited number of observable intermolecular NOEs.

The model of the complex shown in Figure 5.5 depends on the inclusion of a specific AIR between T415 and the 5' phosphate of the dsDNA. Justification of this assumption was previously provided (*vide infra*). While this assumption clearly demands further experimental support ("the reason for referring to the structure of the complex as preliminary"), the resulting structure explains a number of observations that were not included in the calculation and is therefore likely correct. This specific AIR was introduced, in part, based on the structural conservation with the functional residues of BRCA1 protein. Many of the tumor associated missense mutations are known to occur in the BRCT repeats of BRCA1 (50-52), which bind to phosphoserine containing peptide sequences (15;53;54). Peptide recognition by the BRCA BRCTs mainly involves the phosphoserine and a phenylalanine, that is located 3 residues carboxy terminal to the phosphoserine (34;37;38). The residues that bind the phosphoserine are found exclusively in the N-terminal domain of the BRCA1 BRCT repeat, while the phenylalanine is accommodated by a hydrophobic cleft created by the interface of the two BRCT repeats (Figure 5.4). Superposition of the N-terminal domain of the BRCA1 BRCT repeat with p140(375-480) showed a remarkable similarity in the positions of conserved residues of p140(375-480) and those residues involved in phosphoserine binding by BRCA1 (34;36;37). If these conserved residues, T415, G416, and K458 of the p140(375-480) bind specifically to the 5' phosphate, it would provide a possible explanation for the necessity of the N-terminal extension for high affinity DNA binding. In contrast to SH2 domains, which bind phosphotyrosine in a deep, positively charged pocket, the BRCT domains interact weakly with phosphoserine by a shallow basic site. High affinity binding requires a second interaction, such as the hydrophobic interaction involving the peptide phenylalanine in the BRCA1 tandem BRCTs. Analogously, p140(375-480) presents a shallow, positively charged pocket for 5' phosphate binding but in order to achieve high affinity DNA binding, additional interactions by the N-terminal  $\alpha 1$  helix are necessary.

*Other DNA binding BRCT domains*

The BRCT domain of RFC p140 belongs to a distinct subgroup of the BRCT superfamily. Within the distinct subgroup, there is increasing evidence to suggest that the BRCT domain from the bacterial NAD<sup>+</sup> dependent ligase binds to DNA (30;55;56). This BRCT domain is located at the C-terminus of the multi-domain enzyme and recent studies (30) showed that the domain is responsible for stable association of protein and DNA (56). Amino acid sequence analysis of the distinct subgroup of BRCT domains indicates that the potential DNA-binding residues, including T415, G416, R423, G455 and K458, are absolutely conserved between the NAD<sup>+</sup> dependent DNA ligases and RFC p140 (Figure 3.6). Mutations in these residues severely affect the DNA binding or the adenylate-moiety transfer activities of this class of ligases (30). Similar effects on DNA binding observed upon mutations to the conserved residues shared between the bacterial ligases and RFC p140, implies that the 5' phosphate could also be the specific target for DNA binding by the BRCT domain of the DNA ligases. However, ligase activity of RFC p140 has not been reported and the cellular role of 5' phosphate binding remains unknown.

Acknowledgments

The author sincerely thanks Prof. R. Boelens for recording the isotope-filtered NOESY spectra, Marc van Dijk for initiating the early part of the HADDOCK calculation and Anneloes Blok for generation of the R480A and K445A mutants.

---

Reference list

---

1. Waga, S. and Stillman, B. (1994) *Nature* **369**, 207-212
2. Uhlmann, F., Gibbs, E., Cai, J., O'Donnell, M., and Hurwitz, J. (1997) *J.Biol.Chem.* **272**, 10065-10071
3. Bowman, G. D., O'Donnell, M., and Kuriyan, J. (2004) *Nature* **429**, 724-730
4. 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
5. Bork, P., Hofmann, K., Bucher, P., Neuwald, A. F., Altschul, S. F., and Koonin, E. V. (1997) *FASEB J.* **11**, 68-76
6. Caldecott, K. W. (2003) *Science* **302**, 579-580
7. Koonin, E. V., Altschul, S. F., and Bork, P. (1996) *Nat.Genet.* **13**, 266-268



8. Zhang, X., Morera, S., Bates, P. A., Whitehead, P. C., Coffey, A. I., Hainbucher, K., Nash, R. A., Sternberg, M. J., Lindahl, T., and Freemont, P. S. (1998) *EMBO J.* **17**, 6404-6411
9. Williams, R. S., Green, R., and Glover, J. N. (2001) *Nat.Struct.Biol.* **8**, 838-842
10. Derbyshire, D. J., Basu, B. P., Serpell, L. C., Joo, W. S., Date, T., Iwabuchi, K., and Doherty, A. J. (2002) *EMBO J.* **21**, 3863-3872
11. Joo, W. S., Jeffrey, P. D., Cantor, S. B., Finnin, M. S., Livingston, D. M., and Pavletich, N. P. (2002) *Genes Dev.* **16**, 583-593
12. Krishnan, V. V., Thornton, K. H., Thelen, M. P., and Cosman, M. (2001) *Biochemistry* **40**, 13158-13166
13. Lee, J. Y., Chang, C., Song, H. K., Moon, J., Yang, J. K., Kim, H. K., Kwon, S. T., and Suh, S. W. (2000) *EMBO J.* **19**, 1119-1129
14. Dulic, A., Bates, P. A., Zhang, X., Martin, S. R., Freemont, P. S., Lindahl, T., and Barnes, D. E. (2001) *Biochemistry* **40**, 5906-5913
15. Yu, X., Chini, C. C., He, M., Mer, G., and Chen, J. (2003) *Science* **302**, 639-642
16. Yu, X. and Chen, J. (2004) *Mol.Cell Biol.* **24**, 9478-9486
17. Dominguez, C., Boelens, R., and Bonvin, A. M. (2003) *J.Am.Chem.Soc.* **125**, 1731-1737
18. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J.Biomol.NMR* **6**, 277-293
19. Keller, R. (2004) *The Computer Aided Resonance Assignment Tutorial*, 1st Ed., CANTINA Verlag,
20. Kobayashi, M. and Siegal, G. (2005) *J.Biomol.NMR* **31**, 183-184
21. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *J.Mol.Biol.* **319**, 209-227
22. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) *J.Mol.Biol.* **273**, 283-298
23. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J.Biomol.NMR* **13**, 289-302
24. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr.D.Biol.Crystallogr.* **54** ( Pt 5), 905-921
25. Nederveen, A. J., Doreleijers, J. F., Vranken, W., Miller, Z., Spronk, C. A., Nabuurs, S. B., Guntert, P., Livny, M., Markley, J. L., Nilges, M., Ulrich, E. L., Kaptein, R., and Bonvin, A. M. (2005) *Proteins* **59**, 662-672
26. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J.Biomol.NMR* **8**, 477-486
27. Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (1997) *Nucleic Acids Res.* **25**, 4940-4945
28. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J.Mol.Graph.* **14**, 51-32

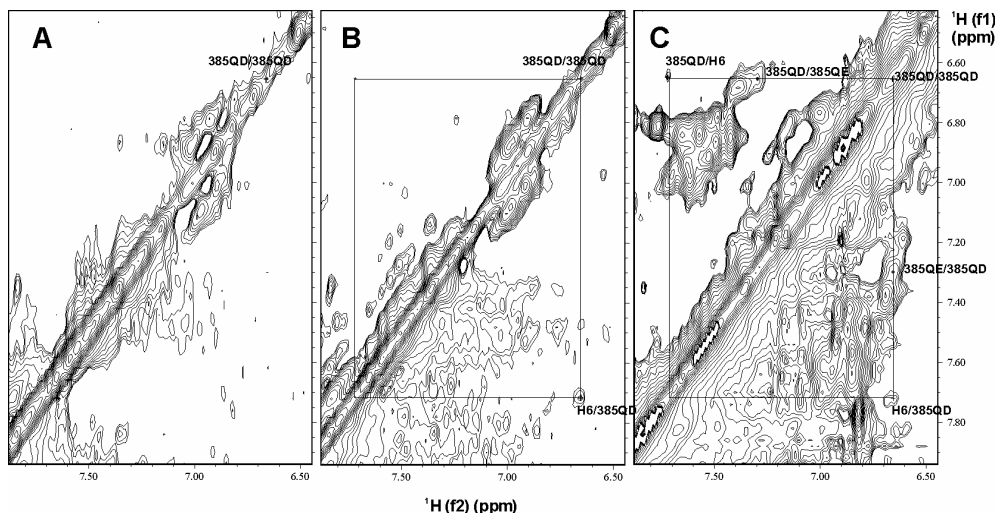
29. Ekblad, C. M., Wilkinson, H. R., Schymkowitz, J. W., Rousseau, F., Freund, S. M., and Itzhaki, L. S. (2002) *J.Mol.Biol.* **320**, 431-442
30. Feng, H., Parker, J. M., Lu, J., and Cao, W. (2004) *Biochemistry* **43**, 12648-12659
31. Nelson, J. R., Gibbs, P. E., Nowicka, A. M., Hinkle, D. C., and Lawrence, C. W. (2000) *Mol.Microbiol.* **37**, 549-554
32. Gaiser, O. J., Ball, L. J., Schmieder, P., Leitner, D., Strauss, H., Wahl, M., Kuhne, R., Oschkinat, H., and Heinemann, U. (2004) *Biochemistry* **43**, 15983-15995
33. Warren, G. L. (2005) *The PyMOL Molecular Graphics System*, Delno Scientific LLC, San carlos, CA, USA <http://www.pymol.org>.
34. Shiozaki, E. N., Gu, L., Yan, N., and Shi, Y. (2004) *Mol.Cell* **14**, 405-412
35. Holm, L. and Sander, C. (1993) *J.Mol.Biol.* **233**, 123-138
36. Clapperton, J. A., Manke, I. A., Lowery, D. M., Ho, T., Haire, L. F., Yaffe, M. B., and Smerdon, S. J. (2004) *Nat.Struct.Mol.Biol.* **11**, 512-518
37. Williams, R. S., Lee, M. S., Hau, D. D., and Glover, J. N. (2004) *Nat.Struct.Mol.Biol.* **11**, 519-525
38. Botuyan, M. V., Nomine, Y., Yu, X., Juranic, N., Macura, S., Chen, J., and Mer, G. (2004) *Structure.(Camb.)* **12**, 1137-1146
39. Lee, M. S., Edwards, R. A., Thede, G. L., and Glover, J. N. (2005) *J.Biol.Chem.*
40. Breeze, A. L. (2000) *Progress in Nuclear Magnetic Resonance Spectroscopy* **36**, 323-372
41. Burgering, M. J., Boelens, R., Caffrey, M., Breg, J. N., and Kaptein, R. (1993) *FEBS Lett.* **330**, 105-109
42. Ikura, M. and Bax, A. (2005) *J.Am.Chem.Soc.* **114**, 2433-2440
43. Dominguez, C., Bonvin, A. M., Winkler, G. S., van Schaik, F. M., Timmers, H. T., and Boelens, R. (2004) *Structure.(Camb.)* **12**, 633-644
44. McDonald, I. K. and Thornton, J. M. (1994) *J.Mol.Biol.* **238**, 777-793
45. Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (1997) *Nucleic Acids Res.* **25**, 4940-4945
46. Garvie, C. W. and Wolberger, C. (2001) *Mol.Cell* **8**, 937-946
47. Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R., and Kaptein, R. (2004) *Science* **305**, 386-389
48. Lawson, C. L. and Carey, J. (1993) *Nature* **366**, 178-182
49. Viadiu, H. and Aggarwal, A. K. (2000) *Mol.Cell* **5**, 889-895
50. Mirkovic, N., Marti-Renom, M. A., Weber, B. L., Sali, A., and Monteiro, A. N. (2004) *Cancer Res.* **64**, 3790-3797

51. Williams, R. S. and Glover, J. N. (2003) *J.Biol.Chem.* **278**, 2630-2635
52. Williams, R. S., Chasman, D. I., Hau, D. D., Hui, B., Lau, A. Y., and Glover, J. N. (2003) *J.Biol.Chem.* **278**, 53007-53016
53. Manke, I. A., Lowery, D. M., Nguyen, A., and Yaffe, M. B. (2003) *Science* **302**, 636-639
54. Rodriguez, M., Yu, X., Chen, J., and Songyang, Z. (2003) *J.Biol.Chem.* **278**, 52914-52918
55. Jeon, H. J., Shin, H. J., Choi, J. J., Hoe, H. S., Kim, H. K., Suh, S. W., and Kwon, S. T. (2004) *FEMS Microbiol.Lett.* **237**, 111-118
56. 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

---

## Supplementary materials

---



(Figure 1S) Identification of intermolecular NOE between DNA (H6) and Y385(QD) using spectra; **A**)  $F1, F2$ ,  $[^{15}\text{N}/^{13}\text{C}]$ -filtered  $[^1\text{H}, ^1\text{H}]$ -NOESY contains only intra NOE peaks between protons of DNA, **B**)  $F1$   $[^{15}\text{N}/^{13}\text{C}]$ -filtered  $[^1\text{H}, ^1\text{H}]$ -NOESY contains intra NOE peaks of DNA and intermolecular NOE peaks between DNA (f1) and protein (f2) and **C**) 2D  $[^1\text{H}, ^1\text{H}]$ -NOESY contains intramolecular NOE peaks of both protein and DNA, and intermolecular NOE peaks. In **B**), the magnetization of protons attached to the  $^{15}\text{N}/^{13}\text{C}$  isotopes (protein) were suppressed during the first  $t_1$  and the magnetization were allowed to transfer to protein from DNA during the mixing time, which results in the intermolecular NOE peaks observed only one side of the diagonal peaks (H6(f1)/385QD(F2) in **B**) in comparison to the symmetric peaks observed between intramolecular NOE transfers (Y385 QE in **C**).

S2. Bruker pulse program

2D  $F1, F2$ ,  $[^{13}\text{C}/^{15}\text{N}]$ -filtered NOESY (Isotope filter based on HMQC purging)

Written by Rolf Boelens (Department of NMR spectroscopy, Bijvoet Center for Biomolecular research, University of Utrecht)

```
# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2"
;qnoesyphpr.rb2
```

```

;wgMQ4N1d.rb
;avance-version (00/02/07)
;2D homonuclear correlation via dipolar coupling
;dipolar coupling may be due to noe or chemical exchange.
;phase sensitive
;with presaturation during relaxation delay and mixing time
;;invieaV1_22.rb
;

# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance2.incl
;   for 1
;
;avance-version (02/08/12)

;$Id: Avance2.incl,v 1.7.2.1 2002/08/12 13:19:57 ber Exp $
# 12 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2" 2

# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/Grad.incl" 1
;Grad2.incl - include file for Gradient Spectroscopy
;   for 1
;
;avance-version (02/05/31)

define list<gradient> EA=<EA>

;$Id: Grad2.incl,v 1.7 2002/06/12 09:04:22 ber Exp $
# 13 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2" 2

# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/Delay.incl" 1
;Delay.incl - include file for commonly used delays
;
;version 00/02/07

;general delays

define delay DELTA
define delay DELTA1
define delay DELTA2
define delay DELTA3
define delay DELTA4
define delay DELTA5
define delay DELTA6
define delay DELTA7
define delay DELTA8

define delay TAU
define delay TAU1
define delay TAU2
define delay TAU3
define delay TAU4
define delay TAU5

;delays for centering pulses

define delay CEN_HN1
define delay CEN_HN2
define delay CEN_HN3
define delay CEN_HC1
define delay CEN_HC2
define delay CEN_HC3
define delay CEN_HC4
define delay CEN_HP1
define delay CEN_HP2
define delay CEN_CN1
define delay CEN_CN2
define delay CEN_CN3
define delay CEN_CN4
define delay CEN_CP1

```

## Chapter 5: Structure of the BRCT domain

```
define delay CEN_CP2

;loop counters

define loopcounter COUNTER
define loopcounter SCALEF
define loopcounter FACTOR1
define loopcounter FACTOR2
define loopcounter FACTOR3

;$Id: Delay.incl,v 1.11 2002/06/12 09:04:22 ber Exp $
# 14 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2" 2

"p2=p1*2"
"p22=p21*2"

;"d24=2m" ;wurst
"d5=5.55m"
"d11=30m"
"d12=20u"
"d13=4u"
"d0=4u"
"d18=d8-p16-d16-d12*5+d13*2"

"DELTA2=d24-p24*0.5+p2*0.5-d12-p16"
"DELTA6=d24-p24*0.5+p2*0.5-d12-p16-p1*2/3.1416"
"d15=p16-p3*2-d13"
"DELTA3=d24-p24*0.5+p2*0.5-d12-p16*2-d16"
"d25=p16+d16-p21-d0-p1*2/3.1416"
"d26=p16+d16-p21-d13-d12*2"
"DELTA4=d5-d24*2-d24+p24*0.5-p2*0.5"
"DELTA5=p21-p3*2-d13"
"CEN_HC2=(p24-p2)/2"
"CEN_HN2=(p24-p22)/2"

# 1 "mc_line 38 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2 expanding
    definition part of mc command before ze"
define delay MCWRK
define delay MCREST
define loopcounter ST1CNT
"ST1CNT = td1 / (2)"
"MCWRK = 0.500000*d11"
"MCREST = d11 - d11"
# 38 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2"
1 ze
# 1 "mc_line 38 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2 expanding
    definition of mc command after ze"
# 39 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2"
3m pl12:f2
3m pl16:f3
# 1 "mc_line 41 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2 expanding start
    label for mc command"
2 MCWRK do:f2
LBLSTS1, MCWRK
LBLF1, MCREST
# 42 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2"
d12 do:f3
d12 setnmr3|0 setnmr0|34|32|33
p16:gp0
d16
d12 setnmr3^0 setnmr0^34^32^33
d12 pl9:f1
d1 cw:f1 ph29
d13 do:f1
d12 pl1:f1
3 (pl ph1)
d12
```

```

    p16:gp1
    DELTA6 p10:f2 p13:f3
(CEN_HC2 p2 ph10) (p24:sp24 ph6):f2 (CEN_HN2 p22 ph0):f3
    d12
    DELTA2
    d15 p12:f2
(p3 ph20 d13 p3 ph10):f2
;
    d12
    p16:gp1
    DELTA3
    p16:gp2
    d16 p10:f2
(CEN_HC2 p2 ph10) (p24:sp24 ph6):f2 (DELTA4 p21 ph6):f3 (CEN_HN2 p22 ph0):f3
    d12
    DELTA3
    p16:gp2
    d25 p12:f2
    (DELTA5 p3 ph16 d13 p3 ph10):f2 (p21 ph20):f3
;
    d0
    p1 ph2
    d13
    d12 p19:f1 setnmr3|0 setnmr0|34|32|33
    p16:gp3
    d16
    d12 setnmr3^0 setnmr0^34^32^33
    d18 cw:f1
    d13 do:f1
    d12 p11:f1
;
    (p1 ph3)
    d12
    p16:gp1
    DELTA6 p10:f2 p13:f3
(CEN_HC2 p2 ph11) (p24:sp24 ph6):f2 (CEN_HN2 p22 ph0):f3
    d12
    DELTA2
    d15 p12:f2
(p3 ph21 d13 p3 ph10):f2
;
    d12
    p16:gp1
    DELTA3
    p16:gp2
    d16 p10:f2
(CEN_HC2 p2 ph11) (p24:sp24 ph6):f2 (DELTA4 p21 ph6):f3 (CEN_HN2 p22 ph0):f3
    d12
    DELTA3
    p16:gp2
    d26 p12:f2
    (DELTA5 p3 ph17 d13 p3 ph10):f2 (p21 ph16):f3
    d13
    d12 p112:f2
    d12 p116:f3
;
    go=2 ph31 cpd2:f2 cpd3:f3
# 1 "mc_line 110 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2 expanding mc
    command in line"
    MCWRK do:f2 do:f3 wr #0 if #0 zd dp2
    lo to LBLSTS1 times 2
    MCWRK id0
    lo to LBLF1 times ST1CNT
# 111 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2"

exit

ph0=0
ph10=1

```

```

ph6=0 2
ph20=0 0 2 2
ph16=0 0 0 0 2 2 2 2
ph21=0 0 0 0 0 0 0 2 2 2 2 2 2 2
ph17=0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2 2 2 2 2 2 2
ph1=0 2
ph2=0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
      2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
ph3=0 0 0 0 2 2 2 2 1 1 1 1 3 3 3 3
ph11=1 1 1 1 3 3 3 3 2 2 2 2 0 0 0 0
ph29=0
ph31=0 2 0 2 2 0 2 0 1 3 1 3 3 1 3 1 0 2 0 2 2 0 2 0 1 3 1 3 3 1 3 1 2 0 2 0 0 2 0
      2 3 1 3 1 1 3 1 3 2 0 2 0 0 2 0 2 3 1 3 1 1 3 1 3
;pll : f1 channel - power level for pulse (default)
;pl9 : f1 channel - power level for presaturation
;pl : f1 channel - 90 degree high power pulse
;d0 : incremented delay (2D)
;d1 : relaxation delay; 1-5 * T1
;d8 : mixing time
;d11: delay for disk I/O [30 msec]
;d12: delay for power switching [20 usec]
;d13: short delay [4 usec]
;in0: 1/(1 * SW) = 2 * DW
;nd0: 1
;NS: 8 * n
;DS: 16
;td1: number of experiments
;FnMODE: States-TPPI, TPPI, States or QSEC

;Processing
;PHC0(F1): 90
;PHC1(F1): -180
;FCOR(F1): 1

;$Id: noesyphpr,v 1.4 2002/06/12 09:05:10 ber Exp $

S3. 2D [F1 filtered-13C/15N]] NOESY (Isotope filter based on HMQC purging)
Written by Rolf Boelens (Department of NMR spectroscopy, Bijvoet Center for biomolecular research, University of Utrecht

# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2"
;qnoesyphpr.rb1
;wgMQ4N1d.rb
;avance-version (00/02/07)
;2D homonuclear correlation via dipolar coupling
;dipolar coupling may be due to noe or chemical exchange.
;phase sensitive
;with presaturation during relaxation delay and mixing time
;;invieaV1_22.rb
;

# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance2.incl
; for 1
;
;avance-version (02/08/12)

;$Id: Avance2.incl,v 1.7.2.1 2002/08/12 13:19:57 ber Exp $
# 12 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2" 2

# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/Grad.incl" 1
;Grad2.incl - include file for Gradient Spectroscopy
; for 1

```

```

;
;avance-version (02/05/31)

define list<gradient> EA=<EA>

;$Id: Grad2.incl,v 1.7 2002/06/12 09:04:22 ber Exp $
# 13 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2" 2

# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/Delay.incl" 1
;Delay.incl - include file for commonly used delays
;
;version 00/02/07

;general delays

define delay DELTA
define delay DELTA1
define delay DELTA2
define delay DELTA3
define delay DELTA4
define delay DELTA5
define delay DELTA6
define delay DELTA7
define delay DELTA8

define delay TAU
define delay TAU1
define delay TAU2
define delay TAU3
define delay TAU4
define delay TAU5

;delays for centering pulses

define delay CEN_HN1
define delay CEN_HN2
define delay CEN_HN3
define delay CEN_HC1
define delay CEN_HC2
define delay CEN_HC3
define delay CEN_HC4
define delay CEN_HP1
define delay CEN_HP2
define delay CEN_CN1
define delay CEN_CN2
define delay CEN_CN3
define delay CEN_CN4
define delay CEN_CP1
define delay CEN_CP2

;loop counters

define loopcounter COUNTER
define loopcounter SCALEF
define loopcounter FACTOR1
define loopcounter FACTOR2
define loopcounter FACTOR3

;$Id: Delay.incl,v 1.11 2002/06/12 09:04:22 ber Exp $
# 14 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2" 2

"p2=p1*2"
"p22=p21*2"

```



## Chapter 5: Structure of the BRCT domain

```
; "d24=2m" ;wurst
"d5=5.55m"
"d11=30m"
"d12=20u"
"d13=4u"
"d0=4u"
"d18=d8-p16-d16-d12*5+d13*2"

"DELTA2=d24-p24*0.5+p2*0.5-d12-p16"
"DELTA6=d24-p24*0.5+p2*0.5-d12-p16-p1*2/3.1416"
"d15=p16-p3*2-d13"
"DELTA3=d24-p24*0.5+p2*0.5-d12-p16*2-d16"
"d25=p16+d16-p21-d0-p1*2/3.1416"
"DELTA4=d5-d24*2-d24+p24*0.5-p2*0.5"
"DELTA5=p21-p3*2-d13"
"CEN_HC2=(p24-p2)/2"
"CEN_HN2=(p24-p22)/2"

# 1 "mc_line 37 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2 expanding
definition part of mc command before ze"
define delay MCWRK
define delay MCREST
define loopcounter ST1CNT
"ST1CNT = tdl / (2)"
"MCWRK = 0.500000*d11"
"MCREST = d11 - d11"
# 37 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2"
1 ze
# 1 "mc_line 37 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2 expanding
definition of mc command after ze"
# 38 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2"
3m pl12:f2
3m pl16:f3
# 1 "mc_line 40 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2 expanding start
label for mc command"
2 MCWRK do:f2
LBLSTS1, MCWRK
LBLF1, MCREST
# 41 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2"
d12 do:f3
d12 setnmr3|0 setnmr0|34|32|33
p16:gp0
d16
d12 setnmr3^0 setnmr0^34^32^33
d12 p19:f1
d1 cw:f1 ph29
d13 do:f1
d12 p11:f1
3 (p1 ph1)
d12
p16:gp1
DELTA6 p10:f2 p13:f3
(CEN_HC2 p2 ph10) (p24:sp24 ph6):f2 (CEN_HN2 p22 ph0):f3
d12
DELTA2
d15 p12:f2
(p3 ph20 d13 p3 ph10):f2
;
d12
p16:gp1
DELTA3
p16:gp2
d16 p10:f2
(CEN_HC2 p2 ph10) (p24:sp24 ph6):f2 (DELTA4 p21 ph6):f3 (CEN_HN2 p22 ph0):f3
d12
DELTA3
p16:gp2
d25 p12:f2
(DELTA5 p3 ph16 d13 p3 ph10):f2 (p21 ph16):f3
;
```

```

d0
p1 ph2
d13
d12 p19:f1 setnmr3|0 setnmr0|34|32|33
p16:gp3
d16
d12 setnmr3^0 setnmr0^34^32^33
d18 cw:f1
d13 do:f1
d12 p11:f1
d12 p112:f2
d12 p116:f3
p1 ph3
go=2 ph31 cpd2:f2 cpd3:f3
# 1 "mc_line 86 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2 expanding mc
command in line"
MCWRK do:f2 do:f3 wr #0 if #0 zd dp2
lo to LBLSTS1 times 2
MCWRK id0
lo to LBLF1 times ST1CNT
# 87 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2"

exit

ph0=0
ph1=0 2
ph20=0 0 2 2
ph6=0 2
ph16=0 0 0 0 2 2 2 2
ph10=1

ph2=0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2 2 2 2 2 2 2
ph3=0 0 0 0 2 2 2 2 1 1 1 1 3 3 3 3
ph29=0
ph31=0 2 0 2 2 0 2 0 1 3 1 3 3 1 3 1 2 0 2 0 0 2 0 2 3 1 3 1 1 3 1 3
;pl1 : f1 channel - power level for pulse (default)
;p19 : f1 channel - power level for presaturation
;p1 : f1 channel - 90 degree high power pulse
;d0 : incremented delay (2D)
;d1 : relaxation delay; 1-5 * T1
;d8 : mixing time
;d11: delay for disk I/O [30 msec]
;d12: delay for power switching [20 usec]
;d13: short delay [4 usec]
;in0: 1/(1 * SW) = 2 * DW
;nd0: 1
;NS: 8 * n
;DS: 16
;td1: number of experiments
;FnMODE: States-TPPI, TPPI, States or QSEC

;Processing

;PHC0(F1): 90
;PHC1(F1): -180
;FCOR(F1): 1

;$Id: noesyphpr,v 1.4 2002/06/12 09:05:10 ber Exp $

```



## 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 indiscriminately 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  $\beta$ -sheet surrounded by three  $\alpha$ -helices. Helices  $\alpha 1$  and  $\alpha 3$  are more conserved in sequence and in length (4). Loop 3 and helix  $\alpha 2$  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 ( $\alpha 1'$ ) 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<sup>+</sup> 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 <sup>1</sup>H 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 [<sup>1</sup>H, <sup>15</sup>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.

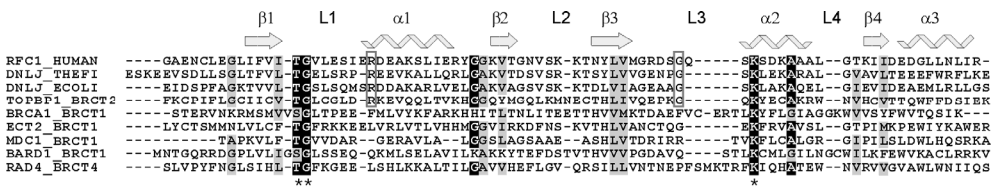
Because of the dynamic behaviour often observed in non-specific protein-DNA complexes, NMR analysis using short distance interactions such as the NOE ( $< 5\text{\AA}$ ) are not sufficient to completely characterize the behaviour of the complex. Long-range distance information (up to  $35\text{\AA}$ ) 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  $\text{Mn}^{+2}$  – chelating EDTA attached to dT, which is located at either side of dsDNA sequence allowed observation of PRE induced line broadening of the  $^1\text{H}$  amide in the  $[^1\text{H}, ^{15}\text{N}]$ -HSQC spectra of HMGB, which revealed HMGB binding non-specifically to multiple sites on dsDNA (13). Similarly, using fluorescence 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<sub>4</sub> 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'-phosphate 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  $\alpha 1$  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-A/G within helix  $\alpha 2$  (Figure 6.1). However a special attention must be paid in identifying the conserved K (in helix  $\alpha 2$ ), which can fall in either L3 or helix  $\alpha 2$  depending on the level of confidence in the structure prediction.

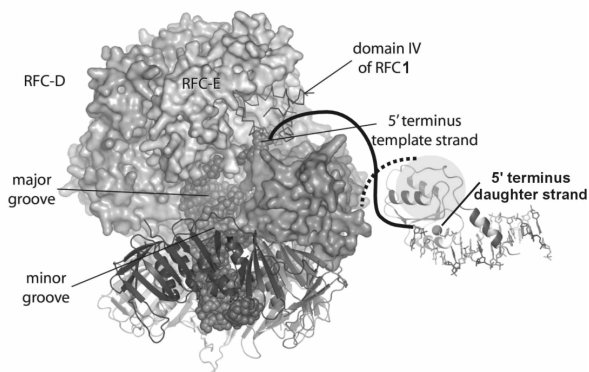


(Figure 6.1) Sequence alignment of BRCT domains with known DNA or pS binding activities. Sequence alignment was first performed using ClustalW, and then the alignment was further adjusted manually according to the alignment of the secondary structure. DNA binding BRCTs are RFC1\_BRCT (RFC p140), DNLJ\_THEFI and DNLJ\_ECOLI (NAD<sup>+</sup> 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  $\beta$ -strand and a ribbon for an  $\alpha$ -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



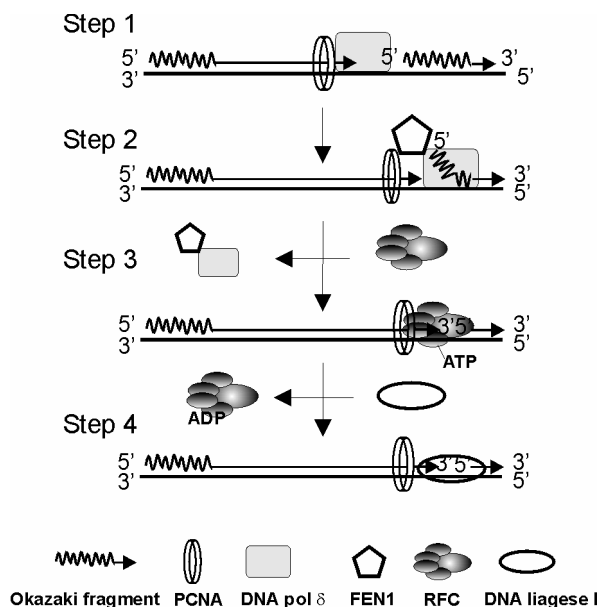
(Figure 6.2). Schematic model of the complete RFC-PCNA-DNA complex. The model is formed from our NMR structure of the RFC p140 BRCT region bound to 5' P dsDNA (right side ribbon view). The purple sphere behind it represents the 60 aa's C-terminal to the BRCT region we believe form a single domain. The 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.

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).

### 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 downstream Okazaki fragment or reinitiated DNA synthesis. Efficient completion of the Okazaki fragment maturation requires coordinated activities of DNA pol  $\delta$ , 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  $\delta$ /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  $\delta$ . 2) strand displacement of the Okazaki fragment and subsequent removal of RNA molecules by FEN1. 3) pol  $\delta$  and FEN1 has to be displaced. The displacement could be achieved via RFC binding at 3' and 5' ends of the ligatable nick releasing the pol  $\delta$ -FEN1 from PCNA or off-loading of PCNA-pol $\delta$ -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  $\delta$ ), 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<sub>4</sub> 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  $\epsilon$ ) 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  $\zeta$  and Pol  $\eta$ .

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. Knegt, R. M., Boelens, R., and Kaptein, R. (1994) *Protein Eng* **7**, 761-767
11. Chandrashekar, 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

# Summary

Every time the cell divides, DNA, which carries the genetic information, must be duplicated. Accurate duplication of the genetic material is carried out by many proteins, each of which participates in a specific step of a complex process leading to successful DNA replication. Since the development of an *in vitro* replication system using the simian tumour virus SV40 in the late 1980s, it has been possible to study protein-protein interactions in eukaryotic DNA replication. Replication Factor C (RFC) is a five protein complex involved in initiating and regulating new DNA synthesis by loading another protein, called PCNA, onto DNA. RFC binds the replication site of DNA, and then using ATP-hydrolyses as energy sources RFC opens the ring shaped PCNA molecule and clamps it around the DNA strand. The loaded PCNA tethers other proteins to DNA, and as a result, PCNA forms an important platform for recruiting DNA modifying enzymes, such as DNA polymerases, to sites of replication or repair. Through stable anchoring to DNA *via* PCNA, DNA polymerases efficiently synthesize new DNA (**Chapter 1**).

Due to its obvious importance in DNA replication and repair, much of the studies on RFC have focused on its mechanism of PCNA loading. The work described in this thesis however, deals with DNA binding by the N-terminal portion of the large subunit of RFC, which has no apparent role in PCNA loading. This region of RFC (amino acids (aa) 403-480) shares sequence homology with the family of BRCT (*BRCA1 C-Terminal*) domains, which consist of roughly 90 aa's and are found throughout eukaryotic and prokaryotic DNA replication and repair proteins. Although the aa sequence of BRCT domains is poorly conserved, they share a highly conserved three dimensional protein fold. It is perhaps because of this sequence diversity that a variety of functions are now known to be associated with the members of this family. Such functions include acting as an adaptor module interacting with other proteins, peptide binding module and a DNA binding module (**Chapter 1**).

The first part of this thesis describes both the region of the large subunit of RFC and the structural determinants of the DNA that are required for productive interaction. In order to determine the optimal DNA binding region, several protein fragments encoding of the large subunit of RFC were generated and subsequently assessed for their double stranded DNA (dsDNA) binding. dsDNA binding is achieved only when the RFC constructs retained an intact BRCT domain and extra amino acids N-terminal to the

domain. This “BRCT region”, encompassing residues 375-480 of the RFC large subunit (also called p140(375-480)), forms a tight 1:1 protein-DNA complex ( $K_d \sim 10\text{nM}$ ) that binds specifically to the 5' phosphorylated end of dsDNA (**Chapter 2**). Only a single type of protein-DNA complex was observed. Surprisingly, however, when the protein fragment spanned an additional 60 amino acids at the C-terminus of the BRCT region, the DNA binding specificity was altered to bind indiscriminately to dsDNA, resulting in multiple types of protein-DNA complexes.

The replacement of several amino acids of the BRCT region using site-directed mutagenesis revealed a requirement for not only the few highly conserved amino acids, but also some non-conserved amino acids, for DNA binding. The non-conserved amino acids are found within the N-terminal sequence, while the conserved amino acids are exclusively found in the BRCT domain part of the BRCT region (**Chapter 2**). The conserved residues are also shared among BRCT domain from the bacterial DNA ligases, which have also been implicated in DNA binding. Furthermore, those conserved amino acids are positional equivalents of those used in binding to the phosphate-moiety by the tandem BRCT domains from BRCA1 when bound to a phospho-peptide ligand. Therefore it was speculated that the residues responsible for phosphate binding are conserved between the dsDNA and the phospho-peptide binding BRCTs.

To gain better insight as to how the protein-DNA complex is formed, it was essential to elucidate the three dimensional structure of the DNA-protein complex. For the structural elucidation, multidimensional NMR experiments were employed (**Chapter 4**). From the resulting NMR spectra, we obtained essentially complete  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  resonance assignments of the protein moiety of the 19 kDa p140(375-480)-DNA complex. Prediction of the secondary structures based on the chemical shift indices and on the pattern of backbone NOEs from the NMR data indicated the presence of a consensus BRCT fold with an extra  $\alpha$  helix within the aa's N-terminal to the BRCT domain.

Based on the chemical shift assignment of human RFC p140(375-480) bound to dsDNA (**Chapter 4**), with the aid of the computer program CANDID/CYANA, we calculated an ensemble of 20 structures of the protein (**Chapter 5**). The protein consisted of a well-defined core, corresponding to the consensus BRCT fold, and an N-terminal  $\alpha$  helix, whose spatial orientation with respect to the core of the protein was less well defined.

The structure alignment between the RFC BRCT domain and the distantly related N-terminal BRCA1 BRCT domain (of the tandem domain structure) revealed a remarkable conservation of amino acids that specifically interact with the phosphate of the phosphopeptide bound to the BRCA1 BRCT domain. However, a lack of data from the DNA undermined the NMR-based structure determination of the complete protein-DNA complex. We therefore obtained a model of the protein-DNA complex using the empirical protocol, HADDOCK, which drives docking of molecules using ambiguous restraints derived from mutagenesis (**Chapter 3**), amino acid residue conservation, structural conservation and ambiguously assigned intermolecular NOEs (**Chapter 5**).

In our HADDOCK generated model of the protein-DNA complex, the BRCT region is bound to the DNA with the positively charged surface of the BRCT domain interacting with the 5' phosphate of the DNA while the N-terminal helix forms several interactions with the phosphate backbone, as well as the bases, of the DNA major groove. The model supports all of the biochemical and NMR spectroscopic data. DNA binding of mutants designed based on the model complex structure further supported the validity of this model (**Chapter 5**).

Although the approaches taken in this study did not successfully generate a three-dimensional structure of the protein-DNA complex based purely on NMR data, the evidence from other studies strongly supports the empirically determined model. Thus, if the model is indeed correct, then the remarkable conservation of the mechanism of recognition of the 5'-phosphate 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-dependent DNA binders (**Chapter 6**).



# Samenvatting

Iedere keer dat een cel deelt moet het DNA materiaal eerst verdubbeld worden. Deze aanmaak van een nieuwe streng DNA wordt verzorgd door een serie moleculen genaamd eiwitten. Dit zijn moleculen die bestaan uit een lange keten van achter elkaar geschakelde kleinere moleculen: aminozuren. Zowel de samenstelling van de aminozuren als de wijze waarop de totale keten wordt opgevouwen tot een 3-dimensionale structuur maken samen dat ieder eiwit een bepaalde functie kan uitoefenen.

Vanaf eind jaren 80 is er uitgebreid studie gedaan naar de manier waarop sommige van deze eiwitten met elkaar samen werken om de DNA replicatie mogelijk te maken. Eén van die eiwitten wordt RFC genoemd, kort voor 'Replication Factor C' en staat centraal in dit proefschrift. Het is opgebouwd uit een aantal afzonderlijk onderdelen of 'subunits' en speelt een belangrijke rol aan het begin van het replicatie proces doordat het een ander eiwit, PCNA, naar het DNA brengt. Het PCNA vormt vervolgens als het ware een ring om een deel van het DNA heen waaraan weer andere eiwitten bevestigd kunnen worden. Tot nog toe zijn veel van de wetenschappelijke studies vooral gericht geweest op het onderzoeken van dit mechanisme (**Hoofdstuk 1**).

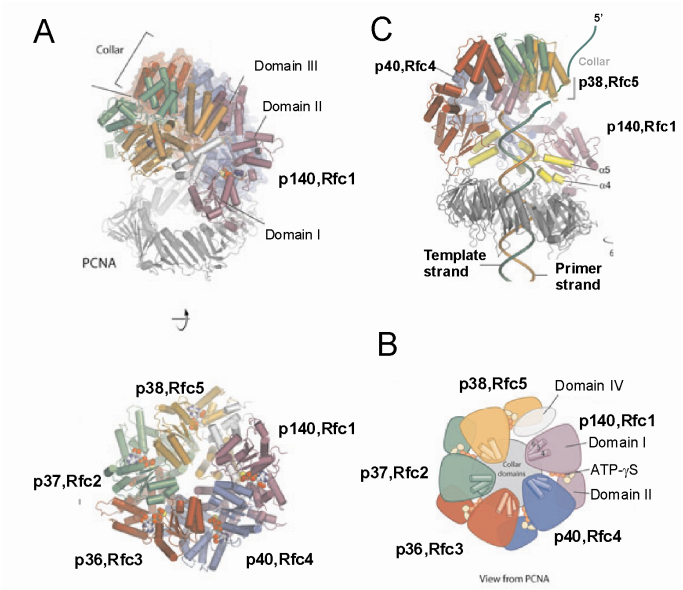
In dit proefschrift is daarentegen juist gekeken naar de interactie tussen RFC en een ander gedeelte van het DNA. Een regio van het RFC eiwit lijkt qua vorm sterk op leden van de familie van de zogenaamde BRCT eiwitten, eiwitten die een rol spelen bij het ontstaan van borstkanker. Hoewel al deze eiwitten een min of meer gelijke structuur hebben zijn er grote onderlinge variaties in de samenstelling van aminozuren. Dit zou kunnen verklaren waarom van deze eiwitten, ondanks hun gelijkenis, wordt gedacht dat ze een groot aantal verschillende functies kunnen uitoefenen. Eén van deze functies, de binding van dubbelstrengs DNA (dsDNA) door RFC, is het voornaamste onderwerp van dit proefschrift.

In het eerste gedeelte van dit proefschrift is onderzocht welke factoren er een rol spelen bij de manier waarop RFC bepaalde structurele elementen van het DNA herkent. Er is gekeken hoe deze interactie geoptimaliseerd kan worden door een aantal verschillende eiwit fragmenten te produceren en vervolgens de sterkte van de binding tussen deze fragmenten en het DNA te vergelijken. Het blijkt dat om een goede binding te vormen, het BRCT-

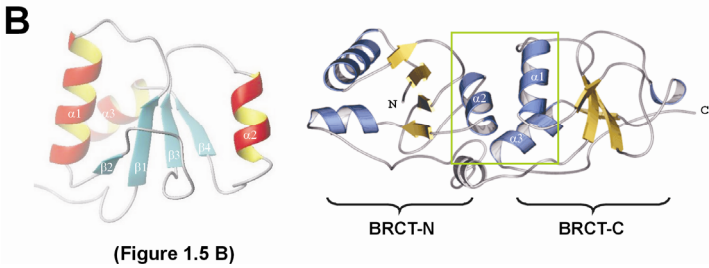
achtige gedeelte van het eiwit alleen niet voldoende is maar dat de fragmenten nog een aantal aminozuren langer moeten zijn. Wanneer dit het geval is kan het fragment een stevig complex vormen met één specifiek deel van het DNA waaraan een zogenaamde ‘fosfaatgroep’ is bevestigd (**Hoofdstuk 2**). Verrassend genoeg echter, zien we dat wanneer het fragment nog langer wordt gemaakt, er verschillende complexen tegelijk gevormd kunnen worden waarbij niet langer maar slechts één gedeelte van het dsDNA betrokken is.

Door lokaal aminozuren te vervangen door andere aminozuren kan er gekeken worden welke residuen uit de BRCT regio van RFC essentieel zijn voor binding van DNA (**Hoofdstuk 3**). Door de zo gevonden aminozuren te vergelijken met die in andere, vergelijkbare, eiwitten waarvan uit de literatuur bekend is wat hun functie is, komen we tot de conclusie dat deze aminozuren specifiek de fosfaatgroep aan het DNA herkennen.

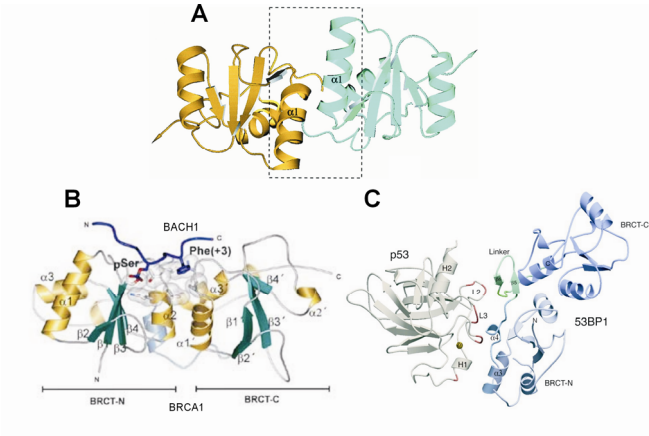
Om goed te begrijpen hoe het complex van DNA en RFC gevormd wordt, is het belangrijk om de structuur er van te kennen. In het tweede gedeelte van dit proefschrift is daarom getracht de structuur van het complex dat gevormd door dsDNA en RFC te bepalen. De techniek die hiervoor gebruikt is heet NMR (nuclear magnetic resonance). Omdat een deel van het complex vrij flexibel blijkt te zijn, was het niet mogelijk om op grond van uitsluitend NMR data de volledige structuur te bepalen. Om deze reden is daarnaast gebruikt gemaakt van het computerprogramma HADDOCK dat diverse vormen van informatie over de afzonderlijke aminozuren combineert om, samen met de NMR data, te komen tot een voorspelling van de structuur van het gehele complex (**Hoofdstuk 4 en 5**). Dit model blijkt uitstekend in staat om de experimentele resultaten, zoals die beschreven zijn in de eerdere hoofdstukken, te verklaren. Uiteindelijk kan dit model van grote waarde blijken bij het identificeren van andere leden van de BRCT familie die in staat zijn DNA te binden (**Hoofdstuk 6**).



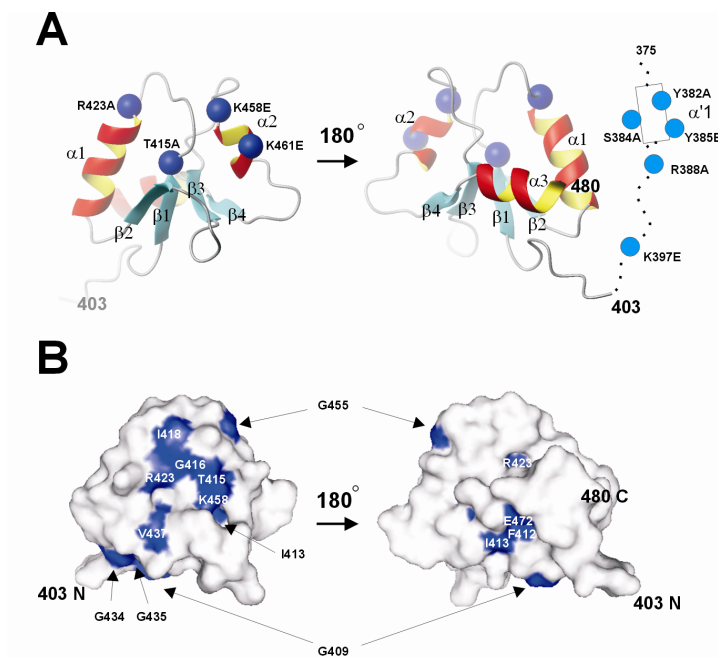
(Figure 1.3)



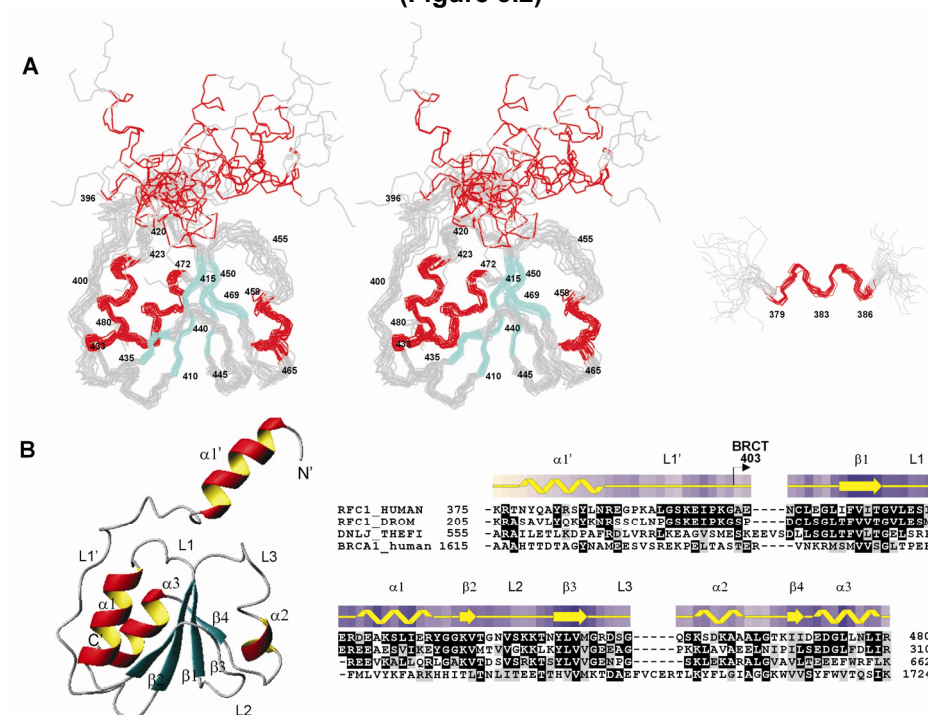
(Figure 1.5 B)



(Figure 1.6)



**(Figure 3.2)**



**(Figure 5.2)**

## Figure Captions

(Figure 1.3) Yeast trRFC-PCNA complex. N-terminally truncated Rfc1 was used to form the trRFC complex. **(A)** Spiral assembly of AAA+ ATPase modules of RFC subunits (bottom). Each AAA+ ATPase module is formed by Domain I and II **(B)**. The ATP binding site (where ATP- $\gamma$ S is bound) is located at the subunit interface which is comprised of the Walker A and B motifs of one subunit and the SRC motif of the adjacent subunit. **(C)** Domain III of one subunit packs against domain II from its neighbours to form a “Collar” (Top). **(C)** The proposed model of primer-template recognition by RFC. The 3' end of the primer strand (orange) is physically blocked by the “Collar” of RFC while the 5' template ssDNA (green) can escape through the wedge shaped gap between p38 and p140. Figures were adopted from the original publication by Bowman et al (31).

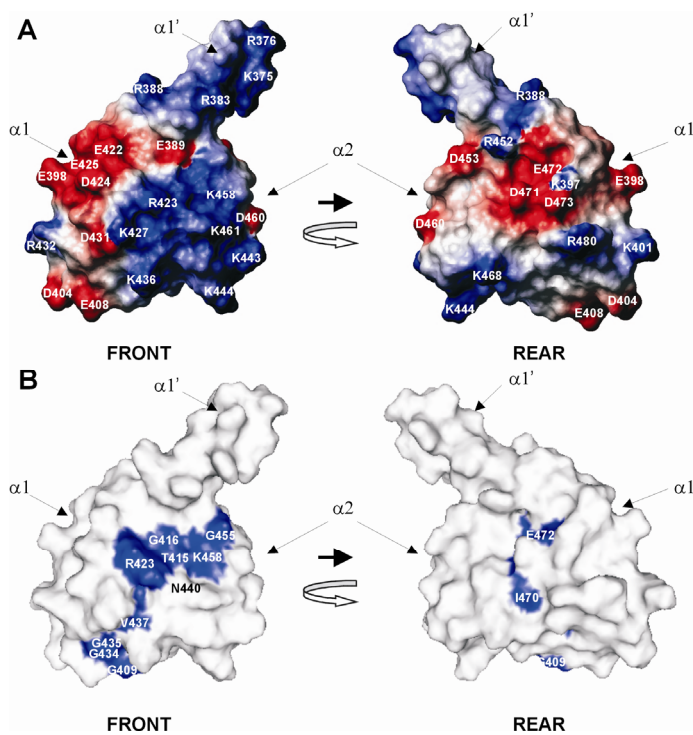
(Figure 1.5) **(B)** Ribbon representation of the XRCC1 C-terminal BRCT (left) and the tandem BRCT domains of BRCA1 (right). Each BRCT unit consists of a  $\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4\alpha 3$  topology. The interface between the tandem BRCT domains (-N and -C denoting N- and C-terminus) is highlighted with a box (right).

(Figure 3.2) Homology model of the BRCT domain from RFC p140 (403–480).

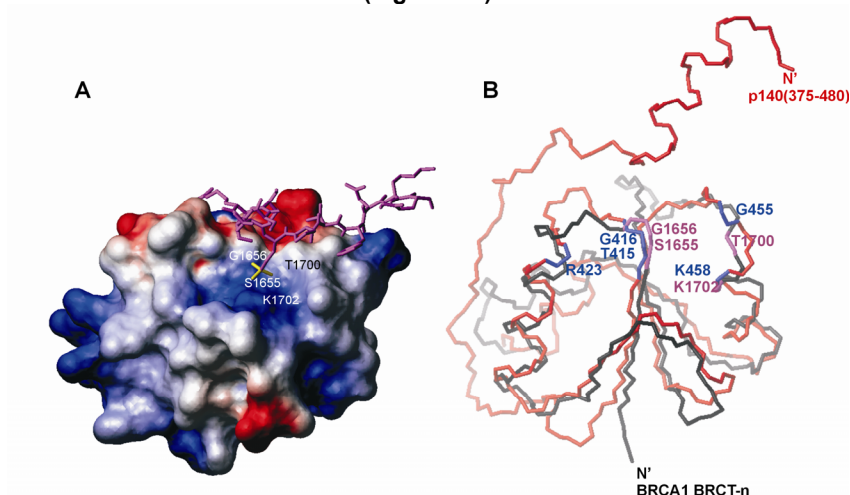
**(A)** Ribbon presentation of the model BRCT domain. The model was based on the NMR structure of the BRCT domain (PDB code: 1IB7) from the NAD<sup>+</sup> dependent DNA ligase which has 33 % amino acids identity with RFC p140 and was generated by 3D-JIGSAW (19). The residues where structural homology was not available are schematically indicated in the figure. The predicted helix between residues 379 and 386 is presented as a box and the loop (387–402) is presented as a dotted line. The blue balls represent the C $\alpha$  locations of the point mutations created in this study. **(B)** Surface presentation of the model BRCT domain. The blue-shaded areas represent the accessible surface occupied by the conserved amino acids defined from the amino acid sequence alignment (Figure 3.1) using AMAS (17). The total accessible surface area (ASA) were calculated using MOLMOL (21).

(Figure 5.2) Structure of RFC p140(375–480) when bound to dsDNA. **(A)** Left; Stereoview of a superposition of the backbone (N, C $\alpha$  and C') atoms for the 20 lowest-energy structures of RFC p140(375–480).  $\alpha$ -helices and  $\beta$ -strands are colored in red and cyan respectively. Right; superposition of the backbone of residues 379–386 demonstrating the well defined helix  $\alpha 1'$  the N-terminus. **(B)** On the left, sequence alignment of the p140(375–480) with the homologous region of RFC p140 from *Drosophila melanogaster* (RFC1\_DROM), NAD<sup>+</sup> dependent ligase of *Thermus filiformis* (DNLJ\_THEFI) and human BRCA1 BRCT-n. The secondary structure of human RFC p140(375–480) was determined by PROCHECK-NMR(26) and the level of solvent accessibility of the residues is colored from light (accessible) to dark blue (not accessible). The loop regions are labeled as L1', L1, L2 and L3. On the left a ribbon diagram of the p140(375–480) structure with the lowest-energy is presented. The orientation is equivalent to (A). Figures (A) and (B) (right) were generated using MOLMOL (28)

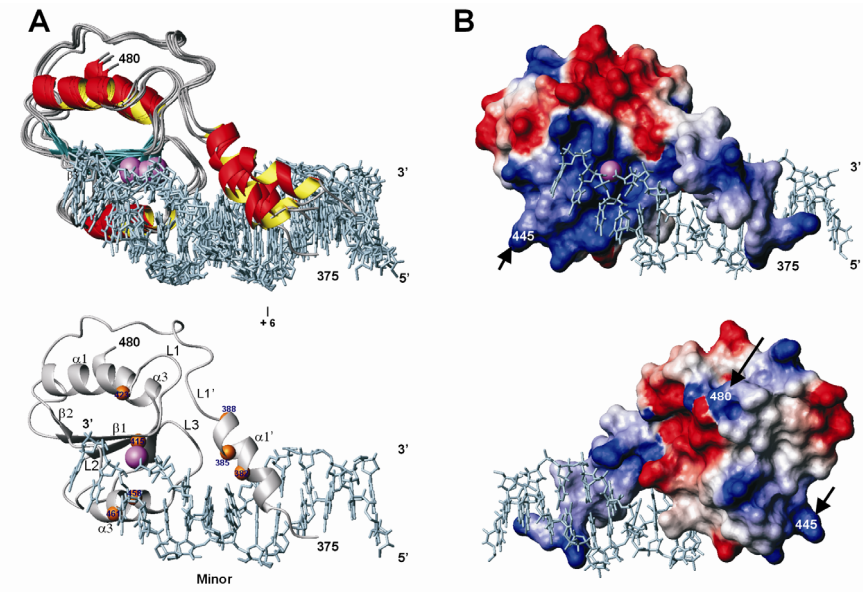




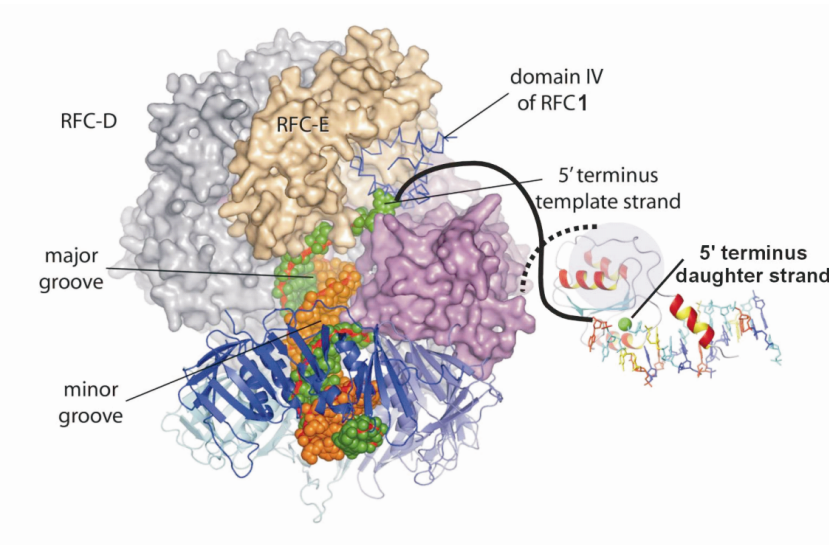
(Figure 5.3)



(Figure 5.4)



(Figure 5.5)



(Figure 6.2)



## Figure Captions

(Figure 5.3) Surface potential of p140(375-480). **(A)** Electrostatic potential of the accessible surface of p140(375-480) is shown. Negative potential is colored in red and positive potential in blue. The residues contributing to the surface charge are indicated on the surface. **(B)** Surface accessible residues that are conserved (defined in Figure 3.1) are colored with blue onto the structure of p140 (375-480). N440 is not a conserved residue. An intermolecular NOE transfer between N440 and unassigned DNA proton was observed (Table 5.3). The “front” and “rear” images were generated by 180° rotation around the Z-axis. The electrostatic potential surface was calculated using MolMol (28) and the residue conservation map (B) was created with Pymol (33).

(Figure 5.4) Structure comparison between the BRCT domains

**(A)** Electrostatic surface presentation of the N-terminal BRCT (BRCT-n) of BRCA1 (PDB:1T29) in complex with a phosphoserine peptide (in magenta). The C-terminal BRCT domain is not directly involved in the phosphate binding and therefore has been deleted from this figure for clarity. Positive potential is shown in blue and negative potential in red. The amino acid residues forming the pocket that binds the phosphate moiety (in yellow) of phosphoserine are indicated on the surface. Phosphate is directly hydrogen bonded by S1655, G1656 and K1702. T1700 is a conserved residue which forms a hydrogen bond to the side chain oxygen of S1655 stabilizing the S1655 side chain conformation. **(B)** Backbone superposition of p140 (375-480) (red) and the BRCT-n from BRCA1 (black). The orientation of the BRCT-n is identical to that of A. The backbone C $\alpha$ , N and C' of the proteins are presented in red for p140(375-480) and in black for the BRCA1 BRCT-n. The conserved residues of p140(375-480) are presented in blue and the residues essential for the phosphate-moiety recognition of BRCA1 BRCT-n are presented in magenta (36-38). The overlay regions of p140 and (BRCA1) are 409-417(1649-1657), 421-452(1660-1691) and 453-479(1697-1723).

(Figure 5.5) Docking model of the p140 (375-480)-dsDNA complex generated by HADDOCK. **(A)** Ribbon representation of ensemble of the five lowest energy structures from the cluster 7. The 5'phosphate is indicated by a magenta sphere. One of the ensemble structures in which the active residues defined in HADDOCK are presented in yellow spheres (bottom). **(B)** Electrostatic surface potential presentation of p140 (375-480) bound to dsDNA. The upper figure has the same orientation as in A while the lower has been rotated 180° around the Z-axis (bottom). The positive and negative charged surfaces are colored in blue and red respectively. Exposed hydrophobic residues are white or slightly colored **(C)** DNA binding activity of R480A and K445A. The mutations were designed as tests of the model of the Protein-dna complex. The wildtype DNA binding activity of the R480A and K445A mutants suggests that the “rear” side and the bottom of p140(375-480) (indicated with an arrow in **(B)**) do indeed not contact the DNA. **(D)** The sequence of the dsDNA used in the HADDOCK docking. The nucleotides with phosphates and bases in contact with the protein are boxed.

(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.

## The list of publications since the year 2000

1. Kobayashi,M., Kidd,D., Hutson,E., Grafton,J., McNulty,S. and Rumsby,M. (2001) Protein kinase C activation by 12-0-tetradecanoylphorbol 13-acetate in CG-4 line oligodendrocytes stimulates turnover of choline and ethanolamine phospholipids by phospholipase D and induces rapid process contraction. *J.Neurochem.*, **76**, 361-371.
2. Kobayashi,M. and Siegal,G. (2005) <sup>1</sup>H, <sup>15</sup>N and <sup>13</sup>C resonance assignments of the BRCT region of the large subunit of human Replication Factor C. *J.Biomol.NMR*, **31**, 183-184.
3. Kobayashi,M., Figaroa,F., Meeuwenoord,N., Jansen,L.E. and Siegal,G. (2006) Characterization of the DNA binding and structural properties of the BRCT region of human replication factor C p140 subunit. *J.Biol.Chem.*, **281**, 4308-4317.
4. Kobayashi, M., AB, E., Bonvin, A, M., and Siegal, G. Solution structure of the DNA bound BRCT region of human replication factor C p140 subunit. *In preparation*

## **Curriculum Vitae**

Masakazu Kobayashi was born on 2<sup>nd</sup> of February 1977 in Kofu-City, Japan. He lived in Japan until his move to the Netherlands in the mid 1992, where he continued further secondary education at Het Rijlands lyceum in Oegstgeest. After obtaining International Baccalaureate degree, he moved to a small northern city of York in England where he followed three years of Bachelor on Biochemistry course and further one year of Masters on Biomolecular science at University of York. The last five months of the master's course were spent in European Molecular Biology Laboratory in Heidelberg, where his mind was set to follow a research in the structural biology. In year 2000, under the supervision of Dr. Gregg Siegal, he began working as a PhD student of Metalloprotein group in Leiden Institute of Chemistry. The structural basis of DNA recognitions achieved by the distant class of BRCT domain from Replication factor C were explored using NMR and other biochemical methods. Since the fall of 2005, he works as a post-doc researcher at the R & D department of Galapagos Genomics N.V.

## Acknowledgement

During my study at Leiden, I encountered various people from different scientific disciplines. These invaluable interactions contributed in completion of this work and allowed me to scratch the upper surface of “mind boggling” scientific subjects. For this, I would like to acknowledge following people:

Alexander Bonvin and Eiso AB (Univ. Utrecht) for their supports on the structure calculation and HADDOCK model (Chapter 5),

Nico van Neeuland (Univ. Leiden) for the large-scale synthesis of the oligonucleotides,

Lars Jansen (Univ. Leiden) for useful discussions over the development of assays.

Francis Figaroa (Univ. Leiden) and Martin van Wijk for their contribution to the work described in Chapter 2 and 3,

Rutger Diederix, Jon Warrall, Peter Crowley, Alex Volkov, Armond Tepper, Miguel Prudêncio and Mike Machczynski (former Metprot members) for sharing interesting scientific discussions

Emanuela Lonardi for being an exceedingly supportive wife throughout this project as well as a good fellow scientist by critical readings of my writings,

Chris Lowe for sharing mentally and physically devastating non-scientific nights spent in non-faculty clubs around NL,

Laura Andolfi for being also a good friend outside the lab,

Thyra de Jongh for the help with “Samenvatting” and for giving me valuable suggestions regarding to the printing of this thesis,

Johan Hollander, Fons Lefeber and Kees Erkelens for answering my endless demands for NMR recordings,

and finally the members of MOLGEN group for letting me use their facilities and their interactions during their weekly meeting.