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

Kobayashi, Masakazu

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

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 (*BR*AC1 *C*-*T*erminus) 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

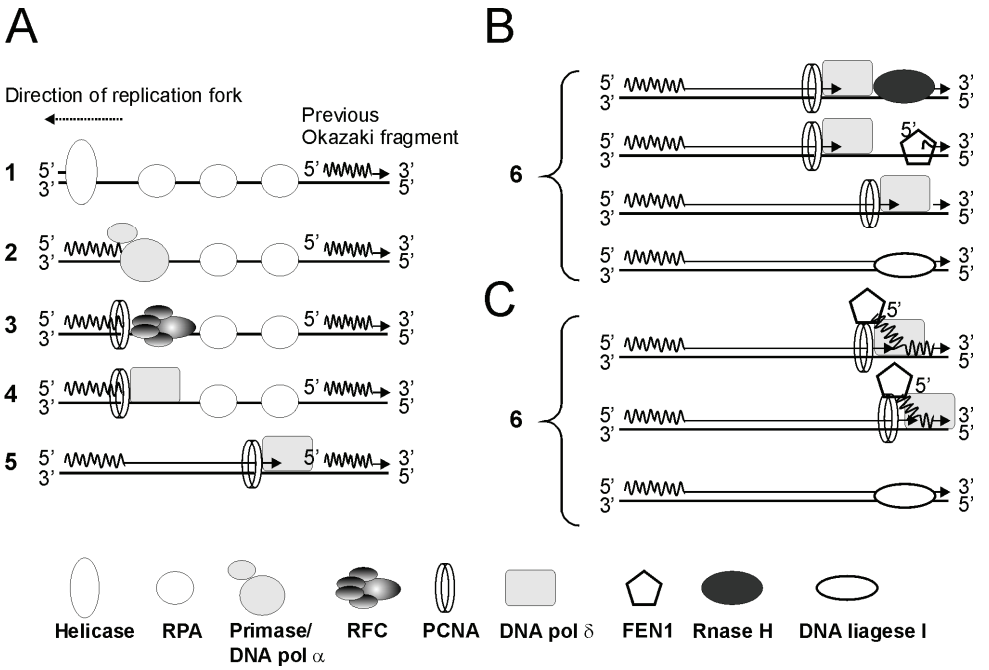
Every eukaryotic cell that divides must pass on two equal complements of the genetic material to the two daughter cells. The discovery of the duplex structure of DNA by Watson and Crick (1;2) led to the notion that each strand of the duplex is used as a template for the synthesis of new DNA, e.g. "DNA 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' → 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 α (pol α). Subsequently a stretch of DNA is synthesized by the polymerase subunit of DNA pol α (Figure 1.1A Step2). DNA synthesis by pol α is

tightly limited to approximately 20 deoxynucleotides, by Replication Factor C (RFC), which binds the 3' end of the primer/template displacing pol α from DNA and initiating the so-called "polymerase switch" (Figure1A 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 steps 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 δ to the 3' primer/template junction. Although DNA pol δ is a poorly processive enzyme by itself, due to its weak interaction with DNA, PCNA bound DNA pol δ 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 ε, plays as important a role as DNA pol δ in DNA

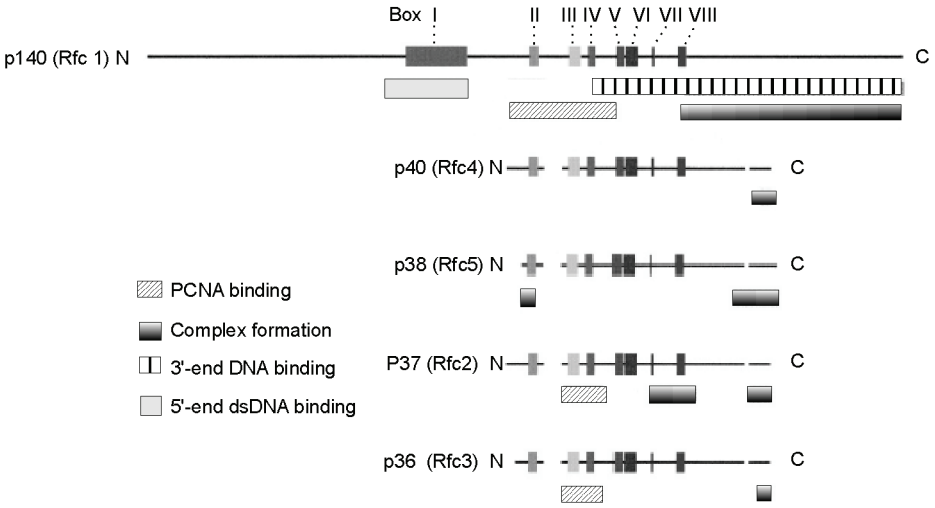
replication. During S-phase of the cell cycle, the association of pol ε at the initiation site of DNA replication has been shown in yeast cells (13;14). The use of mutator DNA polymerases in Yeast suggests that the pol δ and ε replicate opposite template strands at the replication folk (15). Recent studies using immunodepletion of either DNA pol δ or ε in *Xenopus* egg extracts indicate that DNA pol δ is essential for the completion of lagging strand synthesis (16) leaving pol ε 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 α followed by the RFC/PCNA meditated polymerase switch to DNA pol δ, 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 α/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 δ and the nick is sealed by DNA ligase I. The second model involves strand displacement synthesis by pol δ 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 δ and not present in pol ε, further supporting a role for pol δ 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 δ 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 δ (9). This action prevents pol δ 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 δ 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 ε 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 ε in leading strand synthesis, while that of pol σ in lagging strand synthesis. In the both models of lagging strand synthesis, most of the DNA portion made by Pol α is likely not excised, as follows the observation that the mutant Pol α 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. cervisiae* (20).

Replication Factor C (RFC)

As described earlier, RFC plays an essential role in tight regulation of pol α activity through polymerase switching in order to promote high fidelity DNA replication by pol δ. 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 β- and γ-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 p38. 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).

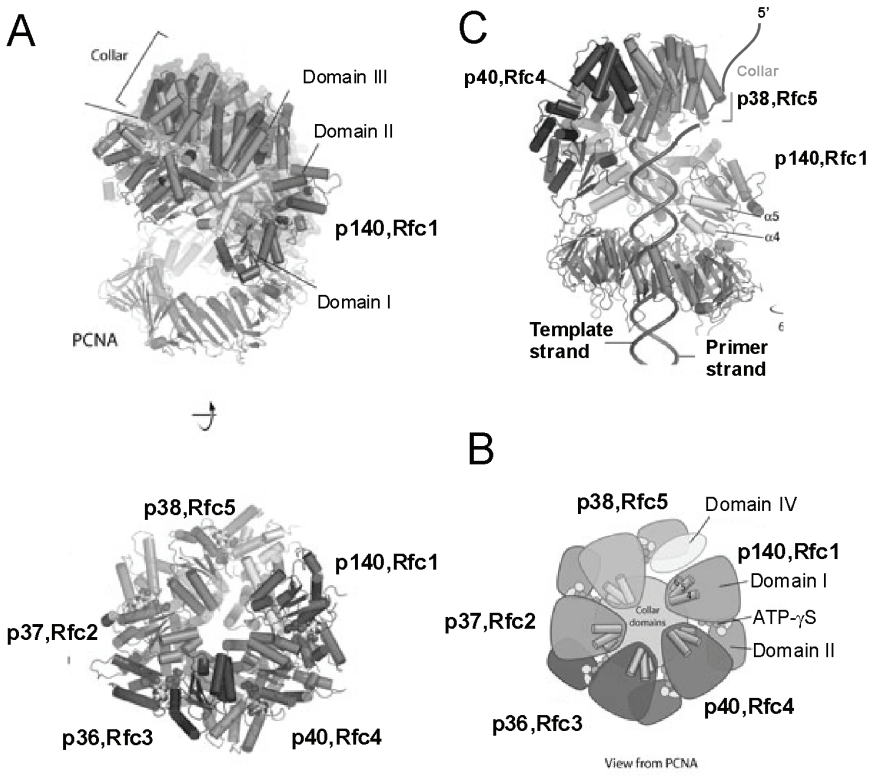
The eukaryotic p140 subunit, contains amino acid sequences in the N-terminal portion that are unique to this subunit including the Box I region of roughly 80 amino acids. Box I contains sequences of high similarity to prokaryotic NAD^+ dependent DNA ligases and some eukaryotic poly (ADP-ribose) polymerases. 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 δ by 2 – 5 times (26;29). A reconstituted yeast trRFC, with Rfc1 (rfc1-ΔN) 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-ΔN 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 δ 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-γS (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-γS 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-γS 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-γS, 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 grove 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).



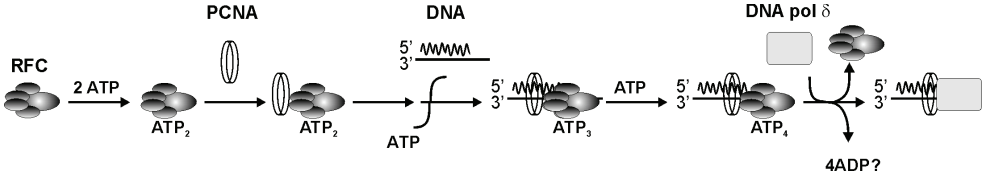
(Figure 1.3) Yeast trRFC-PCNA complex. N-terminally truncated Rfc1 was used to form the trRFC complex. **(A)** Spiral assembly of AAA+ ATPase modules of RFC subunits (bottom). Each AAA+ ATPase module is formed by Domain I and II **(B)**. The ATP binding site (where ATP-γS is bound) is located at the 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 δ and pol ε 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 δ 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 δ (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_{core}) 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. $Rad17/RFC_{core}$ 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. $Rad17/RFC_{core}$ 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 $Rad17/RFC_{core}$ 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_{core}- 9-1-1 complex its 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_{core} (Rad24/RFC_{core}) 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_{core} 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_{core} 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	PCNA	DNA replication
(RFC1 yeast)		
Rad17	Rad9-Rad1-Hus1	Damage checkpoint
(yeast Rad24)	(Rad17-Mec3-Ddc1)	
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_1 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_{core} forms a seven subunit complex with Ctf18, Dcc1 and Ctf8 linked to sister chromatid cohesions (50). Although *Δctf18*, *Δdcc1* and *Δctf8* 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_{core}) and five subunits (Ctf18-RFC_{core}) 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_{core} 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_{core} efficiently unloads PCNA from the primer-template DNA coated with RPA in ATP dependent manner (53). During the end of G_1 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_{core} 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_1/S transition, mitogen activated cdk4-cyclin D and cdk6-cyclin D complexes phosphorylate the retinoblastma 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_1, S and G_2 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 thecell cycle.

Regulatory kinase	phosphorylation site	Effects of phosphorylation	Cell cycle	ref
cdk2-cyclin A	PCNA binding domain of p140	???	G_1/S	(58)
CaMKII	PCNA binding domain of p140	Inhibits RFC-PCNA interaction *in vitro*	S/G_2	(59)
unknown	Thr406 of p140	Inhibits RFC-PCNA interaction *in vitro*	S	(60)
Cdc2-cyclin B	p140	Dissociation of RFC complex	G_2/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_1) 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 MutSα to the mismatch and subsequent recruitment of MutLα (reviewed in (67)) . In *E. coli*, the MutH endonuclease is then activated, creating a nick in the newly synthesized strand either upstream or downstram of the mismatch. A specific exonuclease than 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 EXOI (68). Discrimination between nascent and template DNA strands by EXOI is thought to be mediated by the specific orientation of loaded PCNA which interacts strongly with EXOI (68). Meanwhile the suppression of 5'to 3' hydrolysis activity of EXOI 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 EXOI when the nick is located 3' to the mismatched pair (68). As mentioned before, yeast expressing the genomic mutant rfc1-ΔN 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.

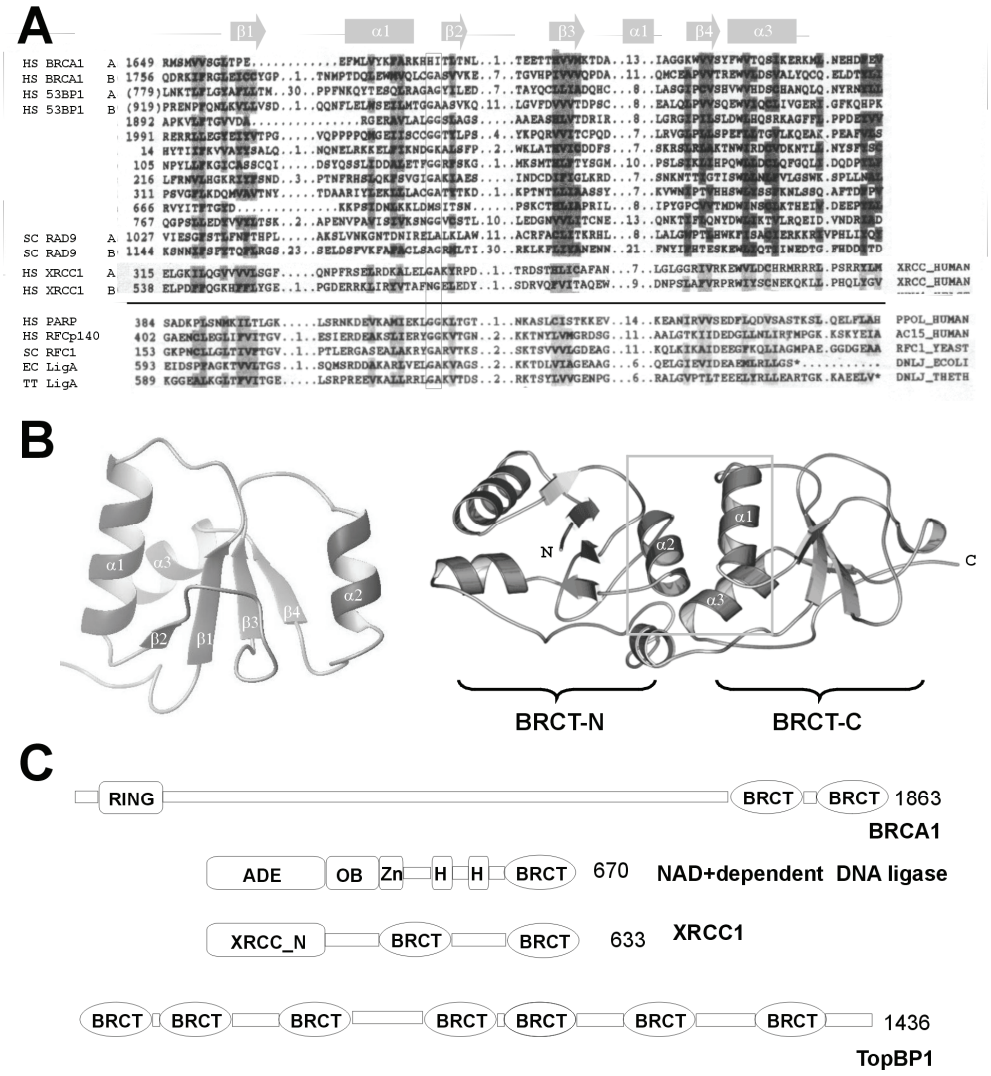
BRCT domain and RFC

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 (**BR**CA1 **C**-**T**erminus) was first identified as a tandem repeat of roughly 90 amino acids at the C- terminus of the Brca1 (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^+ dependent DNA ligases and eukayotic 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^+ dependent DNA ligase (76) confirmed conservation of the same overall protein fold; a compact architecture composed of four parallel β - strands forming the core, flanked on one side by two α - helices and on the other by a single α - helix. The architecture of the β-sheet and helices α1/α3 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 sapience*), Conserved residues are shaded in dark. The conserved GG repeat is highlighted in a box. SC (*Saccharomyces cervisiae*), 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 β1α1β2β3α2β4α3 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+ 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 α1 - helix and the β2 - 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 α2 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+ 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^+ 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 α2 – helix of the N-terminal domain and the α1 – α3 – 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 α1 – and α3 – 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 β-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 phosphoryation dependent	(84-86)
	TopBp1			Likely similar to S. p. Rad4 – Rad9 interaction	
	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	Potential targets in Peptide library		Target serine phosphoryation dependent	(82)
	DNA ligase IV				
	BIRD1				
	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	Transcription regulation		(90-93)
		CtIP			
		LMO4			
	TopBP1	E2F1	DNA repair/checkpoint		(94)
	53BP1 (tandem)	p53	DNA repair	The inter-domain linker interact with p53	(71;74)
	BRCT carrier	BRCT partner			
BRCT-BRCT heterodimer	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	Hyrophobic and salt-bridge interaction between the $\alpha 1$ - $\alpha 1$.	(97;98)
BRCT-BRCT homodimer	S. p. Crb2 S. c. Rad9	N.A	Damage response	unknown	(99;100)

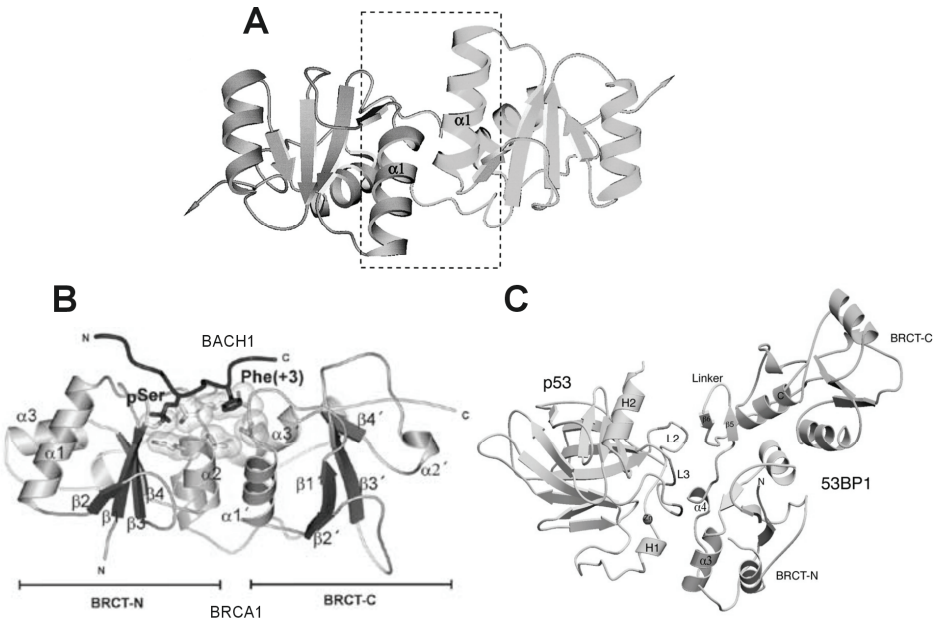
BRCT domains as a protein-protein interaction module and their cellular roles

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 β 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 α1 - α1 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 α1 – 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 α1 helices of the two domains (Figure 1.6A) while the latter pairs together by interactions between α2 of the N-terminal BRCT and the α1/α3 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 dimerisaton has been shown to activate Rad53, a kinase required for cell cycle arrest in response to DNA damage. Rad9 is hyperphosporylated 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 dimmer 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 α1 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 beween 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 repaeat 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^{990}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, carryies 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 β - 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 α3 - 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+ 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+ or ATP. NAD+ 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+. Second, the adenylate moiety is transferred to the 5'-terminal phosphate of the nick. Third, the phsophodiester 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+ 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+ 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+ 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+ 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 identity more members of BRCT family with potential DNA binding functions.

NMR as a tool for Structure determination

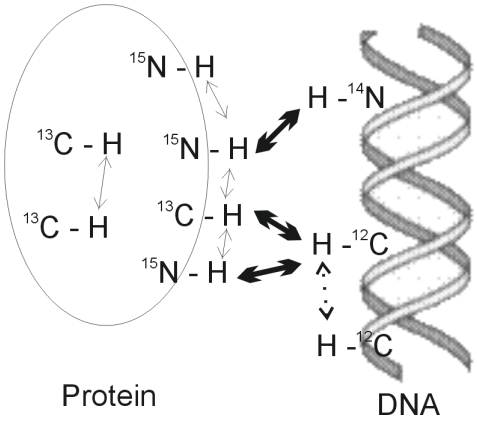
At present X-ray (or neutron) diffraction and Nuclear Magnetic Resonance spectroscopy are the only means to determine the atomic resolution structure of biomacromolecules. For now, NMR can not compete with the accomplishments of X-ray diffraction in the structure determination of supramolecular assemblies 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}N and ^{13}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 of 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 sillico 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 (Kd ≤ 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 maybe 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}N, ^{1}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}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 (Kd $\leq 10^{-5}$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 resonaces 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 α - 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 restrains. 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.Commun.* **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., Coffer, 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 Syvaoja, 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