Structural Characterization
of the DNA Repair Protein Complex
SbcC-SbcD
of *Thermotoga maritima*

Derk Julius Bemeleit

aus

Bremen

München, 2008

Erklärung

Diese Dissertation wurde im Sinne von § 13 Abs. 3 der Promotionsordnung vom 29. Januar 1998 von Herrn Prof. Dr. Karl-Peter Hopfner betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfsmittel erarbeitet.

München, am ....

....

Derk Bemeleit

Dissertation eingereicht am   ....

1. Gutachter: Herr Prof. Dr. Karl-Peter Hopfner

2. Gutachter: Herr Prof. Dr. Patrick Cramer

Mündliche Prüfung am   14.2.2008

Table of Contents

Index of Figures

Index of Tables

1   Summary

DNA damage poses a considerable threat to genomic integrity and cell survival. One of the most harmful forms of DNA damage are double-strand breaks that arise spontaneously during regular DNA processing like replication or meiosis. In addition, they can also be induced by a variety of DNA damaging agents like UV light, cell toxins or anti-cancer drugs. Failure of the rapid repair of these breaks can lead to chromosomal rearrangements and ultimately tumorigenesis in humans. In response to these genomic threats, a highly developed DNA repair network of protein factors has evolved, where the Mre11/Rad50/Nbs1 (MRN) complex is sought to play a key role in sensing, processing and repair of DNA double-strand breaks. Orthologs of Mre11 and Rad50, but not Nbs1, are found in all taxonomic kingdoms of life, suggesting that Mre11 and Rad50 form the core of this complex.

In this work structural studies were performed to decipher the overall architecture and the interaction of SbcC and SbcD, the bacterial orthologs of Rad50 and Mre11. Using X-ray crystallographic and small angle X-ray scattering techniques the crystal as well as the in solution structures of the *Thermotoga maritima* SbcC ATPase domain in complex with full-length SbcD were solved. The crystal and in solution structure match well fortifying the calculated models that reveal an open, elongated complex with dimensions of approximately 210 Å * 75 Å * 65 Å. The heterotetrameric protein assembly consists of two SbcD molecules that homodimerize at domains I to form the central portion of the complex. Located at the outer areas of this homodimer domains II are arranged close to lobe II of SbcC building a small protein-protein interface. The C-terminal domains III of SbcD are connected to domains II *via* a flexible linker and associate through hydrophobic interactions with the coiled-coils of SbcC. These arrangements in combination with earlier findings lead to a model where upon ATP-binding the complex performs a conformational switch resulting in a ring-shaped structure. This conformation would bear a central cavity to harbor DNA strands that can be processed by the inwards oriented nuclease active sites of SbcD.

2   Introduction

2.1   DNA Damage and Repair

In 1953 James Watson and Francis Crick recognized DNA as the genetic information coding molecule (Watson and Crick, 1953). This highly important element of life is continuously threatened by toxic chemicals, radiation or frequent by-products of metabolism. To keep the DNA intact, a multitude of repair mechanisms or ways of mitigating the damages have emerged during evolution. Their impairment or failure is the central cause for a broad range of disease states.

This chapter gives an overview of the most prominent DNA damages and repair mechanisms with a focus on damages leading to DNA double-strand breaks (DSBs) and how cells cope with them.

2.1.1   DNA Damages

There are mainly two major classes of DNA damages: endogenous and environmental based ones. On the one hand, the cellular milieu itself can pose a risk to the integrity of DNA. Endogenous metabolic (by)products like reactive oxygen species harm all the primary components of DNA, i.e. bases, sugars and phosphodiester linkages. But also essential processes such as DNA replication can inflict DNA damage. On the other hand, environmental agents like UV light, genotoxic compounds (e.g. free radicals in cigarette smoke) or ionizing radiation can penetrate cells and organisms, thereby damaging the DNA therein.

2.1.1.1   Endogenous DNA Damages

DNA continually gets in contact with water and reactive oxygen species (ROS) and so far more than 100 hydrolytic and oxidative DNA modifications have been identified (Cadet *et al.*, 1997). Four of the five bases normally incorporated in DNA (adenine, cytosine, guanine and 5-methylcytosine) possess exocyclic amino groups. Loss of these groups, *deamination*, occurs spontaneously in pH- and temperature-dependent reactions of DNA (Lindahl, 1993) and results in the conversion of cytosine to uracil, adenine to hypoxanthine, guanine to xanthine or 5-methylcytosine to thymine (Friedberg *et al.*, 2006). This loss of the amino group due to reaction with ROS changes the pairing properties of the base, which might lead to mutations during semiconservative synthesis of DNA (Lindahl, 1974).

However, not only amino groups can get lost, whole bases may also be cleaved off by hydrolysis of the N-glycosyl bond. The sugar-phosphate backbone remains intact, but the result is an abasic site, which also constitutes an intermediate during base excision repair. Depending on the missing base these events are denoted as *depurination* or *depyrimidation*.

Additional endogenous DNA damages are the *incorporation of incorrect bases* during replication and *incorporation of nucleotide precursors* like 8-oxo-dGTP instead of dGTP, which can lead to replication errors like nucleotide mismatches (Friedberg *et al.*, 2006).

2.1.1.2 Environmental DNA Damages

As mentioned above there is a broad variety of environmental DNA damages like the ultraviolet fraction of sunlight, genotoxic chemicals and ionizing radiation.

UV light induced damages are very prominent since every living organism on the surface of the earth is exposed to sunlight. The most frequent photoproducts are cyclobutane pyrimidine dimers (CPDs) and (6-4) photoproducts characterized by covalent linkages between adjacent pyrimidines (Franklin *et al.*, 1985; Setlow, 1968).

Chemical agents, especially monofunctional alkylating agents like the highly abundant methyl chloride or methyl bromide pose another threat to DNA. Alkylating agents are electrophilic compounds that covalently modify the nucleophilic centers of the nucleic acid, namely the nitrogens of the bases (Singer, 1975; Singer, 1986). Bifunctional toxins like mitomycin or the well known anti-cancer drug cisplatin cause DNA intra- and interstrand links and hence constitute complete blocks to DNA replication and transcription. But also DSB can result from interactions with genotoxic agents. The radiomimetic natural antibiotic Bleomycin induces hydrogen removal of DNA leading to abasic sites and a mixture of strand breaks, with 10 % double-stranded ones (Friedberg *et al.*, 2006). Camptothecin and etoposide as topoisomerase inhibitors can also lead to DSBs. Topoisomerases are DNA nicking and closing enzymes. Their inhibitors arrest the enzymes in an intermediate DNA processing stage where the DNA is broken. Collision of an approaching DNA replication fork with the enzyme-inhibitor-DNA complex then results in DSBs (Froelich-Ammon and Osheroff, 1995).

Also continually abundant like UV radiation is *ionizing radiation*, which main sources are cosmic radiation and radionuclides naturally occurring on the planet like potassium-40. They may lead to the formation of excited and ionized molecules, which can cause direct and indirect damage to the DNA. Direct damage occurs by ionization of bases and sugars, while indirect damage is inflicted *via* reactions of nucleic acids with radicals formed by radiolysis of water like hydroxyl radicals. These radicals can either depurinate or depyrimidate DNA, resulting in abasic sites or directly attack the DNA sugar backbone (Breen and Murphy, 1995) leading to strand breakage (in ~20% of the

cases). Since a single track of radiation causes many ionizations, induction of a DSB can be the result of a localized attack of each DNA strand by two or more hydroxyl radicals (Ward, 1985; Ward, 1990). This is also true if one strand is broken by a radical and another break occurs within about 10 bp (Michael and O'Neill, 2000).

2.1.1.3   Biologically Required Breaks

The DSBs mentioned above are perceived as lesions and dealt with promptly using one or the other repair mechanism described below. Additionally, DSBs appear in a programmed fashion as obligatory intermediates during certain biological processes.

The vertebrate immune system uses *V(D)J recombination* to generate diversity of immunoglobulins and T cell receptors (Krejci *et al.*, 2003). During this event, DSBs are generated at specific recombination signal sequences by the Rag1/Rag2 protein complex (McBlane *et al.*, 1995). The cleavage results in blunt signal ends and hairpin coding ends. Whereas the signal ends can be joined directly, the coding ends need to be opened first. A pair of non-homologous end joining factors, Artemis/DNA-PKcs (DNA-dependent protein kinase catalytic subunits), likely carry out this hairpin opening, which will be discussed in more detail later. The following processing at the opened coding ends further increases the variety of joined products. The completion of V(D)J recombination is mediated by functions of various NHEJ factors (Grawunder *et al.*, 1998). After interaction with antigen, the affinity and specificity of the different immunoglobulins generated by V(D)J recombination are increased through somatic hypermutation, a process called "affinity maturation" (Siskind and Benacerraf, 1969). Furthermore, "class switching" modifies the constant region of the immunoglobulin (Stavnezer, 1996). Activation-induced cytidine deaminase (AID) has been suggested to act in initiating DSB formation in this class switching recombination (Petersen *et al.*, 2001).

In yeast, DSBs appear during meiosis after DNA replication. These *meiotic double-strand breaks* are present at hotspots that are located in each of the 16 chromosomes (Martini and Keeney, 2002). Generation of these breaks is mediated by a protein complex containing the topoisomerase II-like protein Spo11, which is suggested to

introduce the DSBs at meiotic hotspots. Formation of the meiotic DSBs leads to recombination between homologous chromosomes until their segregation in meiosis I. Accordingly, blocking of the machinery that creates the meiotic DSBs not only abolishes recombination, but also results in chromosomal nondisjunction during meiosis I (Krejci *et al.*, 2003).

Another process in which DSBs are needed is *mating type switching*. The switching of mating type in *S. cerevisiae* is initiated by the introduction of a site-specific DSB at the MAT locus by the HO endonuclease. Once formed, the DSB break is used by the recombination machinery for conducting a gene conversion event that results in the switching of one mating type to the other (a or α) *via* one of the gene cassettes at HML and HMR, which store the a and α mating type information, respectively (Krejci *et al.*, 2003).

2.1.2 DNA Repair

Genomic integrity, and therefore cell survival in general, heavily depends on the accurate and efficient repair of DNA lesions. In response to the above-described vast number of genomic threats, a highly sophisticated and intricate DNA repair network of protein factors has evolved. These mechanisms either reverse the DNA damage or excise the damaged elements. Since there are nearly uncountable agents threatening the DNA in numerous ways many DNA repair mechanisms have evolved. The main mechanisms of DNA repair or mitigations of inflicted damages are listed in table 1 and in the following chapter DSB repair systems will be described in more detail.

Table 1: Biological responses to DNA damage (Friedberg *et al.*, 2006)

Reversal of base damage

Excision of damaged, mispaired or incorrect bases

Breaks in the backbone

    Base excision repair (BER)

    Nucleotide excision repair (NER)

    Transcription-coupled nucleotide excision repair (TC-NER)

    Alternative excision repair (AER)

    Mismatch repair (MMR)

Strand break repair

    Single-strand break repair (SSBR)

    Double-strand break repair (DSBR)

Tolerance of base damage

    Translesion DNA synthesis (TLS)

    Postreplicative gap filling

    Replication fork progression

Cell Cycle checkpoint activation

Apoptosis

2.1.2.1   DNA Double-Strand Break Repair

Double-strand breaks are interesting DNA lesions because they can either be harmful or beneficial to organisms as mentioned above. DSBs, caused by a variety of reactions, can promote genome rearrangements that initiate carcinogenesis or apoptosis (Hoeijmakers, 2001). But, in contrast, they can also be beneficial when they occur in a controlled way during specialized events that demand genome sequences to be rearranged, such as during development of the immune system or generation of genetic diversity in meiosis (Maizels, 2005; Schatz and Spanopoulou, 2005; Whitby, 2005). The multitude of repair mechanisms evolved, counteract the threatening effects of unwanted DSBs and promote the favorable effects of programmed DSBs. This chapter gives a basic outline of the double-strand break repair (DSBR) pathways (figure 1) including a more detailed

review of the two major pathways involving the Rad50/Mre11 complex, namely non-homologous end joining (NHEJ) and homologous recombination (HR).
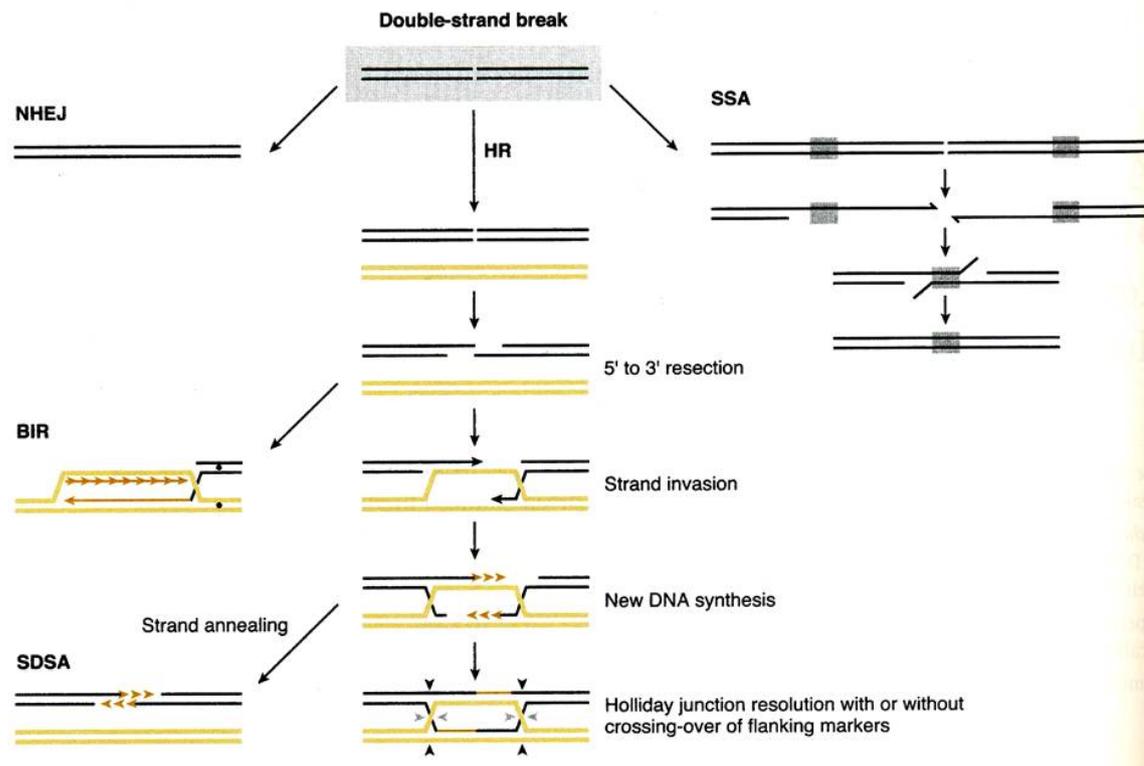
Figure 1: Eukaryotic double-strand break repair pathways.
NHEJ ligates broken DNA ends without the participation of a homologous partner. In the alternative pathways 5'→3' single-strand resection follows a DSB. In case of break-induced replication (BIR), strand invasion can result in formation of a replication fork and an elongated gene conversion tract. Synthesis-dependent strand annealing (SDSA) partially synthesizes DNA using a homologous chromosome or chromatid as template, followed by reannealing of invading strands where no crossover takes place. In HR, after single-strand resection the resulting 3' ends are recombinogenic and can invade a homologous template, to initiate new DNA synthesis. Two Holliday junctions (HJs) are formed and are resolved independently by cutting the crossed or noncrossed strands, resulting in crossover or noncrossover products. Single-strand annealing (SSA) can occur when a DSB appears between or within two direct repeats. Resection of the DSB ends produces two complementary single strands that are annealed. After resection of the nonhomologous 3' ends and new DNA synthesis, ligation restores two continous strands (Friedberg *et al.*, 2006).

The most probable model of *break-induced replication* (BIR) (Paques and Haber, 1999) presumes invasion of the 3' end originating from a processed DSB into a homologous double-stranded molecule as can be seen in figure 1. The displaced strand does not pair back but instead serves as the lagging-strand template of a newly established replication

fork. Semiconservative replication is initiated and may proceed until the end of the chromosome. Several pathways of BIR exist and proteins of the MRN complex, Mre11, Rad50 and Nbs1 (or Xrs2 in yeast), are needed, although the genetic requirements for BIR are not identical in different studies (Friedberg *et al.*, 2006; Malkova *et al.*, 1996; Malkova *et al.*, 2005).

During *synthesis-dependent strand annealing* (SDSA) the extended 3' single strand that invades from one DSB end after new DNA synthesis gets displaced and anneals with the complementary second DSB end. This mechanism might also apply to mitotic recombination events in human (Friedberg *et al.*, 2006).

If a DSB occurs between two flanking homologous regions, repair of the broken chromosome by *single-strand annealing* (SSA) is very efficient and leads to a deletion containing a single copy of the repeated sequence (Lin *et al.*, 1984; Lin *et al.*, 1985; Paques and Haber, 1999). At sites of breakage 5' → 3' bidirectional degradation occurs until homologous single-stranded regions are exposed. After annealing, non-homologous 3' overhangs are resected and the remaining gaps are filled by repair synthesis and subsequent ligation.

In case there are no repeat regions flanking the DSB and the DSB occurs in a single-copy sequence, as is commonly observed in G_1 haploid cells, *non-homologous end joining* takes place. Here DNA ends directly get ligated without requirement of strand exchanges or the availability of homologous DNA. A "clean" two-ended DSB with either blunt ends or very short 5' or 3' complementary overhangs requires no intense processing (Baumann and West, 1998) and is repaired by the NHEJ core components Ku70, Ku80, DNA-dependent protein kinase catalytic subunit (DNA-PKcs), XRCC4, XLF and DNA ligase IV (figure 2). The Ku70/80 heterodimer forms a ring with a hole that fits DNA ends, explaining its DNA end binding preferences (Walker *et al.*, 2001). It also recruits DNA-PKcs, which might have an effect on positioning of the DNA ends (DeFazio *et al.*, 2002; Spagnolo *et al.*, 2006). DNA-PKcs additionally stimulates Ku70/80 to move about one helical turn inward from the end (Yoo and Dynan, 1999), thereby permitting other proteins access to the free DNA end. The last step in rejoining is mediated by DNA ligase IV (Lees-Miller and Meek, 2003), which is associated with a XRCC4 homodimer (Lees-Miller and Meek, 2003; Modesti *et al.*, 2003) that stabilizes

the ligase and stimulates its adenylation and ligase activity. Another factor, XLF, interacts with the DNA ligase IV/XRCC4 complex, but its precise function in NHEJ has not been identified to date (Ahnesorg *et al.*, 2006; Wyman and Kanaar, 2006). However, cells of patients carrying mutated *XLF* genes are strikingly radiosensitive and DSBR defective (Buck *et al.*, 2006).
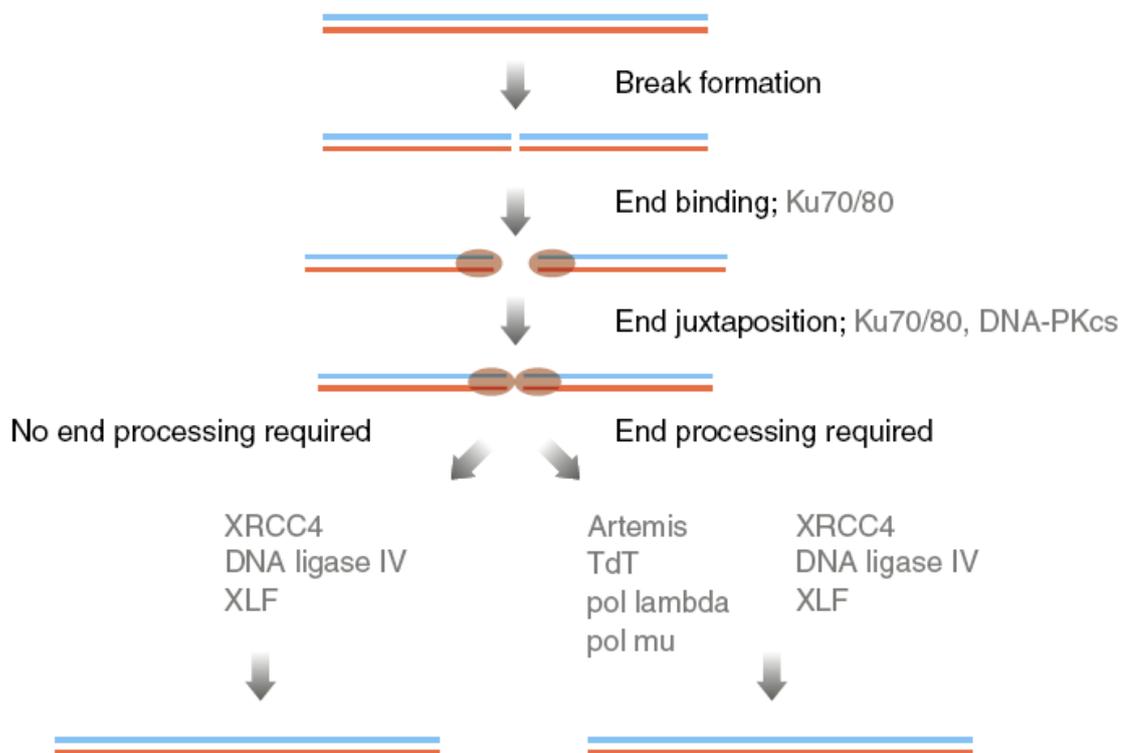
Figure 2: Double-strand break repair through non-homologous end joining.
The Ku70/80 heterodimer binds to DNA ends and associates with DNA-PKcs that performs juxtaposition of the ends. If no end processing is required, XRCC4, DNA ligase IV and XLF are sufficient to ligate the ends. In more complicated cases of NHEJ the core components are necessary as well but additionally Artemis, TdT, pol lambda and pol mu modify the ends (Wyman and Kanaar, 2006).

Frequently, the terminal residues have unusual structures that are not directly ligatable since they are not proper substrates for DNA ligases, which require 3'-hydroxyl and 5'-phosphate groups. Reactive oxygen species resulting from radiation, as mentioned above, cause base and sugar damage. These modified ends require nucleolytic processing and DNA synthesis to remove and replace nonligatable modified nucleotides and incompatible single-strand overhangs (figure 2). Artemis as a multifunctional

nuclease interacts with DNA-PKcs and in cooperation is able to cleave DNA structures like 3'- and 5'-overhangs, hairpins, flaps and gaps (Ma *et al.*, 2002; Ma *et al.*, 2005b). These nuclease activities may result in diverse ends like blunt ends but also overhangs and if ligation is not coordinated properly this may lead to a small single-strand gap. Three polymerases, terminal deoxynucleotidyl transferase (TdT) and the translesion polymerases pol mu and pol lambda, take care of these distinct substrates. Pol mu and pol lambda are able to fill gaps and TdT can add untemplated nucleotides to DNA ends (Ma *et al.*, 2005a; Ma *et al.*, 2004; Nick McElhinny *et al.*, 2005; Nick McElhinny and Ramsden, 2004). After processing the ends are ligated, but sequence information can be lost upon rejoining, making NHEJ an error prone repair pathway (Wyman and Kanaar, 2006).

The second major DSBR process is *homologous recombination*, which is generally accurate since an undamaged sister chromatid is used as repair template (figure 1). As a first and very important task broken ends need to be kept in close proximity to their repair partners. In cell biological experiments, the eukaryotic complex Rad50/Mre11 is one of the first factors detected at DNA double-strand breaks (Lisby *et al.*, 2004). A function for organizing broken ends in the nucleus is suggested by the reduced clustering of DNA ends in Rad50/Mre11 complex-deficient cells (Aten *et al.*, 2004). Rad50/Mre11 form oligomeric complexes on linear DNA and these oligomers act like molecular velcros (figure 3). In humans these oligomers, bound to different DNA molecules, interact through up to 500 Å long coiled-coils thereby tethering broken DNA ends (de Jager *et al.*, 2001; Hopfner *et al.*, 2002; van Noort *et al.*, 2003). DNA binding acts as a conformational switch in the Rad50/Mre11 complex that favors interaction among coiled-coil apexes of different complexes and thus enhances DNA tethering as could be shown by single-molecule imaging experiments (Moreno-Herrero *et al.*, 2005). *In vivo* studies demonstrate the importance of interaction among the Rad50 coiled-coil apexes. When they were replaced with a ligand inducible dimerization domain, rescue of Rad50 function became ligand dependent (Wiltzius *et al.*, 2005). More biochemical and structural characteristics of this complex will be discussed in chapter 2.2 but it should be mentioned at this point that, in contrast to most cases, molecular details of keeping broken ends close are less well understood for bacteria than for eukaryotes.
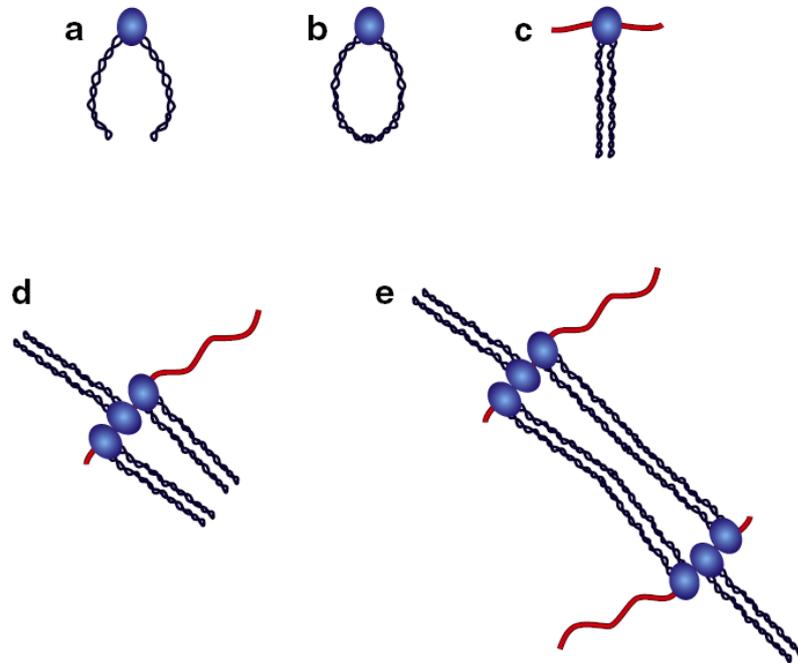
Figure 3: Tethering of broken DNA molecules by the Rad50/Mre11 complex.
Rad50 contains a globular ATPase domain that is connected by long coiled-coils. (a) The Rad50/Mre11 complex consists of two Rad50 molecules, dimerized at their globular domains, and two Mre11 molecules bound to them. (b) The coiled-coils are flexible and their apexes can interact through a zink-hook structure. In this formation where the coiled-coils interact within the same complex, Rad50's biological function, tethering broken DNA ends through intercomplex interactions, is prevented. (c) The complex binds DNA (red line) with its globular domain, inducing thereby conformational change of the coiled-coils to inhibit interaction within the same complex. (d) Complex oligomerization will occur on linear DNA near the ends. (e) Oligomers tether DNA molecules through binding *via* the zink-hooks to keep them in close proximity before repair (Wyman and Kanaar, 2006).

The second step in HR, after having ensured that the broken DNA ends are held together, is processing of the ends into single-stranded tails with 3'-hydroxyl overhangs. Here the mechanism of action is much better examined in bacteria, where processing is performed by the RecBCD helicase/nuclease machine, which additionally associates with the RecA recombinase (Spies *et al.*, 2005; Spies and Kowalczykowski, 2006). RecBD, the two helicases, pull both of the DNA strands that were split by RecC to the nuclease domain of RecB, located further downstream, which processes one strand or the other (Dillingham *et al.*, 2003; Singleton *et al.*, 2004). Finally, the RecBCD enzyme loads RecA protein onto the 3' tail to initiate recombination by means of the RecA pathway (Anderson and Kowalczykowski, 1997b). Details of how nuclease access

switches from one DNA strand to the other upon encountering a Chi sequence, known as recombination hotspot, remain to be clarified. However, for eukaryotes much less is known since the equivalent machinery there has not been identified so far (Wyman and Kanaar, 2006). Yeast based genetic evidence shows that end processing involves at least the Rad50/Mre11 complex but does not prove that this is in fact the nuclease (Lee *et al.*, 1998). It has been been speculated that the yeast Sae2 protein could perform the long sought-after end resection activity. Very recently, it could be shown that the human Sae2 homolog CtIP is indeed a DNA-resection promoting enzyme that physically and functionally interacts with Mre11 (Sartori *et al.*, 2007). Of course, the so far described nuclease activities of Mre11 would create a single-stranded end with the incorrect polarity (Trujillo *et al.*, 1998), but this is also the case for RecBCD activity before modulation by passing a Chi sequence in DNA (Anderson and Kowalczykowski, 1997a). The mechanism of Rad50/Mre11 may well be similarly modulated by interaction with or addition of another component. It could also be the case that Rad50/Mre11 may be less directly involved in end resection but rather act as an important cofactor for another nuclease (Wyman and Kanaar, 2006). The actors in end resection, the mechanisms controlling their action and coordination of this event with the rest of homologous recombination in eukaryotic systems remain to be clarified. There are probably several possibilities to deal with the different circumstances in which DSBs can be created.

For any type of break to be repaired by HR the DSB needs to be processed so that the DNA reveals a single-stranded stretch suitable to load a recombinase. However, as is the case for NHEJ, the exact end structure may guide specific processing to allow eventual repair that will require using the DNA end either as a primer for synthesis or as a ligation partner. Thus, the DNA end either needs a clean 3'-hydroxyl or a 5'-phosphate group. Ends with other chemical structures have to be specifically modified. The various end processing requirements are well reflected in the demand for different specific factors in diverse experimental systems. For example, depending on the type of treatment applied to create DNA breaks, the ends may need no processing at all (site-specific endonucleases) or may need to be modified to remove covalently bound proteins (topoisomerase inhibitors) or may need resection of chemical groups left when the backbone break occurs in the ribose ring (ionizing radiation). Rad50/Mre11's

appearance early in DSBR and its nuclease activities, allowing possible processing of several different end structures, point towards Rad50/Mre11 playing a role here. SbcCD, the *E. coli* protein complex structurally related to Rad50/Mre11 (de Jager *et al.*, 2004), is a potent nuclease that likely processes aberrant DNA structures, including proteins covalently attached to DNA ends (Connelly *et al.*, 2003). It is important to identify the eukaryotic factors required to modify specific end damage because different end-processing factors might direct or require different activities in downstream steps (Wyman and Kanaar, 2006).

The third step in HR is the loading of a recombinase onto processed DNA ends. This enzyme can catalyze the exchange of base-paired partners between two DNA molecules. The bacterial RecBCD complex recruits the RecA recombinase onto 3'-ended single-stranded DNA prepared by the helicase/nuclease. This is similar to the RecFOR proteins that load RecA onto single-stranded DNA regions created when damage that needs to be repaired by HR is encountered during replication (Morimatsu and Kowalczykowski, 2003; Webb *et al.*, 1997). In yeast, Rad52 is the crucial player early in recombination by supporting recombinase filament formation (Krogh and Symington, 2004). Its homolog in mammalian cells is BRCA2. Several cell biological, biochemical and structural studies implicate BRCA2 to be involved in controlling the activity of the eukaryotic recombinase Rad51 by loading it onto single-stranded DNA (Galkin *et al.*, 2005; Shin *et al.*, 2004; Shivji and Venkitaraman, 2004). *In vitro* studies of Rad51 indicate that it supports functions in the homology search and strand-pairing stages that are very similar to those of RecA. Purified yeast Rad51 shows ssDNA-dependent ATPase activity (Sung, 1994) and in presence of ATP, Rad51 forms a helical filament with ssDNA and dsDNA. Like RecA, Rad51 can catalyze complete strand exchange between single-stranded and double-stranded homologous DNA in an ATP- and Mg^{2+}-dependent manner *in vitro* (Friedberg *et al.*, 2006), where three stages of the strand-pairing and exchange reaction can be distinguished (Krejci *et al.*, 2003). In the first stage, *presynaptic filament formation*, Rad51 binds to ssDNA. For this step, ssDNA and not dsDNA binding of Rad51 is the important activity because precoating of dsDNA with Rad51 prevents the following pairing reaction (Sung and Robberson, 1995). A conformational change in Rad51 induced by ATP binding but not necessarily hydrolysis allows efficient nucleoprotein complex formation (Namsaraev and Berg, 1998). Then,

heteroduplex DNA is generated by *synapsis* initiated from the 5'end of the incoming strand and *homologous pairing* concomitant with strand displacement. Once constituted, the regions of heteroduplex formation are extended by branch migration and, *in vitro*, this process stops at regions of heterologous DNA longer than six base pairs (Namsaraev and Berg, 2000; Sung and Robberson, 1995). These DNA structures resulting from stand invasion and joint molecule formation have crossed DNA strands, called Holliday junctions. To finish repair after recombination, the DNA strands have to be disentangled or cut by structure-specific nucleases, the Holliday junction resolvases. The RuvABC complex or RecG are responsible for this task in bacteria (West, 2003) and in eukaryotes the search for an equivalent enzyme has led to identification of a complex of two Rad51 paralogs, XRCC3 and Rad51C, that seems to be associated with resolvase activity (Liu *et al.*, 2004). Alternatively, the crossed DNA strands resulting from joint molecule formation can be divided by the combined action of a helicase and a topoisomerase (Heyer *et al.*, 2003; Wu and Hickson, 2003).

2.2   The Mre11 – Rad50 – Nbs1 complex

As indicated in the previous chapters, the Mre11/Rad50/Nbs1 complex (MRN complex) functions in various areas of repair and cellular regulation associated with DSBs. This complex contains three proteins Mre11, Rad50 and Nbs1 (Xrs2 in yeast). Mre11 and Rad50 build the core of the complex. The Rad50-Mre11 family of proteins includes gp46 (Rad50) and gp 47 (Mre11) of bacteriophage T4, SbcC (Rad50) and SbcD (Mre11) of bacteria and Rad50 and Mre11 of eukaryotes and archaea (Hopfner *et al.*, 2000a; Sharples and Leach, 1995; Zhuo *et al.*, 1994). As mentioned before and described in more detail in the following, the complex is involved in telomere maintenance, cell cycle responses to DNA damage and DSB repair by both homologous recombination and non-homologous end joining (D'Amours and Jackson, 2002).
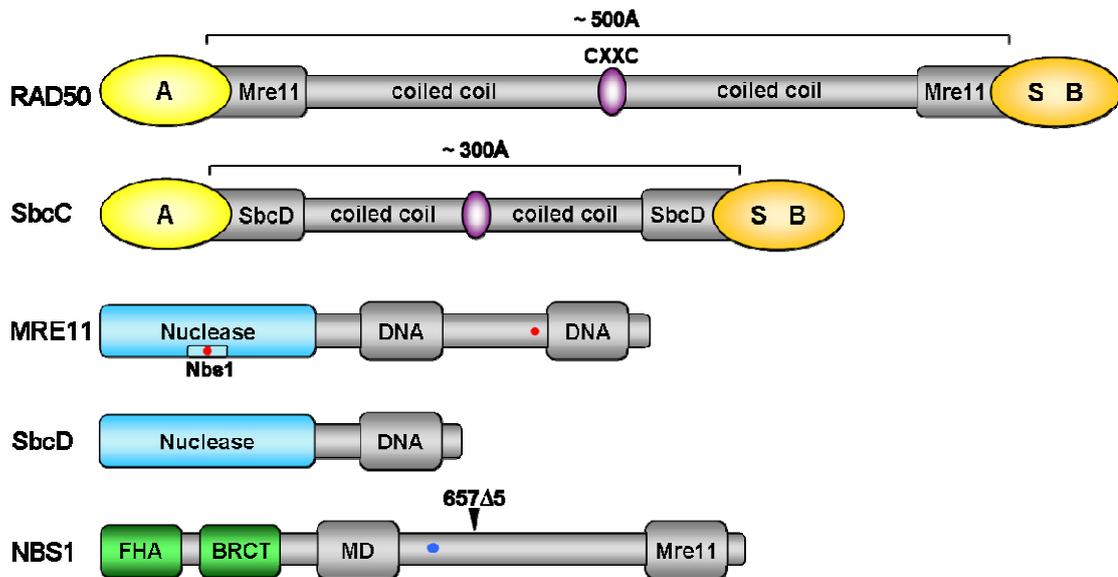
Figure 4: Functional domains of the MRN complex.
Rad50 in eukaryotes and archaea and SbcC in bacteria comprise one Walker A (yellow) and one Walker B (orange) motif carrying the ATPase activity, two heptad repeats folding into an antiparallel coiled-coil domain and a central Cys-X-X-Cys (CXXC, purple) motif forming a zinc hook that facilitates a metal-dependent interaction between Rad50 molecules. The C-terminal signature motif (orange) contains an adenylate kinase activity required for tethering of different DNA molecules. The SbcD/Mre11 binding sites are located adjacent to the Walker domains as parts of the coiled-coils that span ~500 Å in humans and ~300 Å in prokaryotes. Mre11 possesses a conserved phosphoesterase domain at the N-terminus (blue) that bears the nuclease active site and a potential interaction loop with Nbs1 in mammals. The positions of human ataxia telangiectasia-like disorder (ATLD) mutation sites (red dots) are indicated as well as additional DNA binding sites. SbcD contains one DNA binding region and lacks the Nbs1 binding site. Nbs1, apparent in eukaryotes only, possesses a forkhead associated (FHA, green) and breast cancer C-terminus (BRCT, green) domain at its N-terminus, which interacts with phosphoproteins. The central mutation domain (MD, grey), containing the mutations that are found in Nijmegen breakage syndrome is shown as well as the mutation found in most NBS patients (657del5), giving rise to a N-terminally truncated protein lacking the FHA and BRCT domains. The checkpoint phosphorylation site S343 is indicated by a blue dot.

2.2.1   Structural and Biochemical Properties

Rad50 is a split ABC-type ATPase, which contains two heptad repeats in its center that fold into coiled-coils, bringing the two N- and C-terminal ATPase motifs Walker A and B in close proximity (figure 4) (de Jager *et al.*, 2001). These motifs also interact with two Mre11 monomers, forming a globular head domain that interacts with DNA (de Jager *et al.*, 2001). The apex of the coiled-coils of Rad50 contains a CXXC motif, representing one half of a Zn^{2+}-binding site. This motif has been named the zinc hook and its crystal structure has been solved (Hopfner *et al.*, 2002). Purified Rad50 peptides

covering this region of the protein dimerize in the presence of Zn^{2+}. The other half of the binding site is provided by a nearby MRN complex of opposite orientation. The association between two Rad50 molecules can either form a large loop or can occur between two complexes bound on two different DNA molecules (figure 5). Both modes of association have been seen by electron microscopy (Hopfner *et al.*, 2002). Exchange of a single cysteine residue within this motif increases ionizing radiation sensitivity in yeast, indicating the physiological importance of the hook structure (Hopfner *et al.*, 2002). ATP binding promotes the dimerization between two catalytical Rad50 domains, which is crucial for efficient DNA binding (Hopfner *et al.*, 2000b) and thus for all functions of the MRN complex (Alani *et al.*, 1990). Upon DNA binding via the globular head, the intracomplex interaction is prevented and the intermolecular configuration is favored, which allows DNA tethering (Moreno-Herrero *et al.*, 2005). In addition to ATP binding and hydrolysis, the conserved signature motif in the C-terminal ATPase domain of Rad50 bears an adenylate kinase activity, which is required for efficient tethering between different DNA molecules, independently of DNA binding (Bhaskara *et al.*, 2007). These combined observations have led to the proposal that the molecular bridging by the MRN complex *via* the Rad50 coiled-coils constitutes the molecular basis of the MRN complex's influence on recombinational DNA repair.
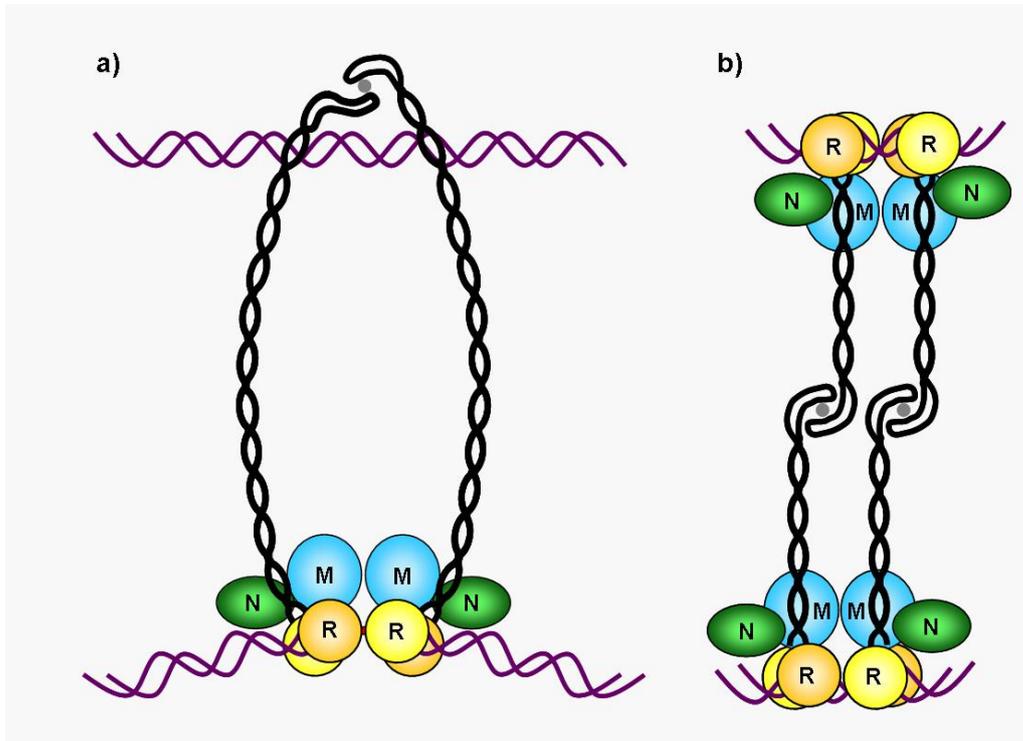
Figure 5: Proposed architecture model of the zinc linked MRN complex with DNA.
a) One MRN complex (Rad50$_2$/Mre11$_2$/Nbs1$_2$) links the sister chromatids to both DNA ends (purple) using combined Rad50, Mre11 and Nbs1 DNA-binding-sites (Chen *et al.*, 2001). b) As deduced from electron microscopy studies, two MRN complexes can interact with each other *via* their hooks and bound zinc molecules (grey) (Hopfner *et al.*, 2002). The position of Nbs1 and the exact localization of DNA are speculative.

Mre11, derived from meiotic recombination 11, is a DNA-binding protein that possesses a conserved phosphoesterase domain at the N-terminus (figure 4). Human and budding yeast Mre11 have ATP-stimulated ssDNA endonuclease activity, 3'-5' dsDNA exonuclease activity as well as weak DNA unwinding (stem-loop opening) activity, all of which require Mg^{2+} ions (Paull and Gellert, 1998; Trujillo *et al.*, 2003; Usui *et al.*, 1998). The observed 3'-5' proofreading-like exonuclease activity was at first surprising because genetic and physical analyses of meiotic recombination and mitotic DSB repair had suggested that the MRN complex is involved in 5'-3' exonucleolytic resection leading to 3'-tails. However, hypomorphic mutations that abolish the nuclease activity of Mre11 do not reduce most forms of DSB repair in mitotic cells, indicating the complex's contribution to that process is largely independent of its nuclease activity, but Mre11 is likely involved in the processing of DNA secondary structures and misfolded

DNA ends (Bressan *et al.*, 1999; Connelly *et al.*, 2003; Farah *et al.*, 2002; Krogh *et al.*, 2005; Lewis *et al.*, 2004; Llorente and Symington, 2004; Lobachev *et al.*, 2002; Moreau *et al.*, 1999; Paull and Gellert, 1999; Trujillo and Sung, 2001; Tsubouchi and Ogawa, 1998). Mre11 also contains two DNA-binding domains, a conserved one in the center of the protein and a less conserved one at the C-terminus (Usui *et al.*, 1998).

Findings in the *E. coli* homologs of Rad50 (SbcC) and Mre11 (SbcD) suggest that they can directly remove covalently attached proteins from DNA ends (Connelly *et al.*, 2003), an activity that could explain the function of the MRN complex in removal of covalently attached Spo11 from meiotic DSBs (Borde *et al.*, 2004; Keeney, 2001; Moreau *et al.*, 1999). The processing of protein-bound DNA ends needs ATP and proceeds by introducing a DSB close to the end. It seems that the MRN complex can process nearly all types of DNA ends and hairpins in an ATP-stimulated fashion, resulting in clean 3'-OH ends for subsequent ligation in HR as mentioned before. The role of Rad50's ATP hydrolysis in the 3'-5' exonuclease activity is not fully deciphered, but it seems that ATP controls mainly the binding of DNA to the MRN complex (Connelly *et al.*, 2003; de Jager *et al.*, 2002; Paull and Gellert, 1999; Trujillo and Sung, 2001). Additionally to DNA end recognition, ATP-induced conformational changes in the Rad50 ABC domains might partially melt DNA to enable end processing by Mre11 (Paull and Gellert, 1999).

Whereas Mre11 and Rad50 are strongly conserved in bacteria (SbcCD), archaea and eukaryotes, the third member of the complex is less well conserved among species. Nevertheless, Nbs1 has a functional homolog, Xrs2, in yeast. The two proteins share some common motifs: (1) a N-terminal FHA domain, involved in protein-protein interactions, which binds to phosphorylated histone H2AX (Kobayashi *et al.*, 2002), (2) phosphorylation sites for the checkpoint protein Atm (Tel1 in yeast) and (3) a conserved C-terminal domain involved in interactions with Mre11 (Kobayashi *et al.*, 2004). Both Nbs1 and Xrs2 proteins enhance Mre11 nuclease activity *in vitro* (Carney *et al.*, 1998; Desai-Mehta *et al.*, 2001; Paull and Gellert, 1999; Tauchi *et al.*, 2001). In patients with the 657del5 allele suffering from Nijmegen breakage syndrome (see chapter 2.2.7) the amount of mutated Nbs1 is reduced and hMre11p is mislocalized to the cytoplasm (Carney *et al.*, 1998; Tauchi *et al.*, 2001). One of the roles of Nbs1 might thus be to

maintain an intranuclear localization of Mre11. Another function of Nbs1 is to activate ATM autophosphorylation specifically in response to DNA damage induced by ionizing radiation (Difilippantonio *et al.*, 2005).

2.2.2   Features of the Bacterial SbcCD (Rad50-Mre11) Complex

As mentioned before Rad50 and Mre11 have related proteins in all kingdoms of life and the orthologs in bacteria are SbcC (Rad50) and SbcD (Mre11). Since in this work the structure of SbcCD of *Thermotoga maritima* is under investigation, this chapter emphasizes on features of Rad50-Mre11 orthologs that were identified in bacteria or show distinct behavior in bacteria.

The *sbcC* and *sbcD* genes of *E. coli* were originally isolated as genetic suppressors, which allowed growth of a recombination-deficient mutant (Lloyd and Buckman, 1985). In wild-type *E. coli* the encoding proteins prevent the replication of DNA molecules containing long palindromic DNA sequences. By comparison, *sbcCD* mutant strains propagate these replicons more readily (Chalker *et al.*, 1988; Gibson *et al.*, 1992). These and further observations led to the hypothesis that SbcCD acts as a DNA endonuclease (Leach, 1994). In deed, SbcCD was shown to function as an ATP-dependent 3'-5' dsDNA exonuclease and hairpin endonuclease (Connelly *et al.*, 1997; Connelly *et al.*, 1998; Connelly and Leach, 1996). SbcD on its own works as an ATP-independent single-strand endonuclease (Connelly and Leach, 1996) but all the activities stated before require Mg^{2+} as a cofactor. SbcCD is inactive on circular dsDNA and has an affinity for DNA molecules that contain double-stranded ends (Connelly *et al.*, 1999; Connelly and Leach, 1996). As mentioned earlier SbcCD processes a diverse array of DNA ends like single-stranded overhangs, blunt ends or molecules covalently sealed at both ends by hairpin loops (Connelly *et al.*, 1999; Connelly *et al.*, 1998). These hairpins get cleaved at the 5' side of the loop to yield products with 5' phosphate and 3' hydroxyl ends. Early density transfer and demethylation studies have indicated that DNA replication is required before SbcCD can recognize DNA palindromes (Lindsey and Leach, 1989; Shurvinton *et al.*, 1987), which then may lead to hairpin formation. These findings have led to the proposal that SbcCD collapses replication

forks by attacking hairpin structures that occur on the lagging strand template and that broken replication forks are repaired by homologous replication (Connelly and Leach, 1996; Leach, 1994). In addition to hairpin structures DNA binding proteins can mask DNA ends so that these are not freely available for enzymatic processing. SbcCD was shown to be able to remove DNA end-bound protein by introducing a double-strand break (Connelly *et al.*, 2003). With each substrate, the initial nuclease activity is seen at DNA ends. However, it is not clear what feature of a DNA duplex or end is recognized by SbcCD.

In contrast to the *in vitro* activities of the corresponding eukaryotic proteins ATP binding is essential for the exonuclease and hairpin endonuclease activities of *E. coli* SbcCD and *P. furiosus* Rad50-Mre11. Nevertheless, ATP binding clearly plays a crucial role in eukaryotes as mutations within the Walker A nucleotide-binding motif cause strains of *S. cerevisiae* to be as sensitive to MMS as a null strain (Alani *et al.*, 1990).

As stated earlier in the previous chapter 2.2.1 and as can be predicted from the amino acid sequences (figures 35 and 36) Rad50-Mre11 orthologs share major structural properties (figure 4). The sequence of SbcC or Rad50, respectively, in between the bipartite ATPase domain, which contains the Walker A at the amino and the B-type motif at the carboxy terminus, is not highly conserved at the amino acid sequence level but its calculated secondary structure is conserved and striking. The 550 to 1000 amino acid residue central portion of Rad50 orthologs is predicted to have a high probability of forming an α-helix (figures 13 and 16) interrupted by the metal binding CXXC motif (Hopfner *et al.*, 2002). However, the degree to which the intervening amino acid sequences are predicted to form coiled-coils varies among the species. *P. furiosus* Rad50 would have very homogenous coiled-coils (figure 16) whereas the coiled-coils of SbcC of *T. maritima* are interrupted by regions with a low probability to form such coils. These discontinuities in the predicted coiled-coil structure are apparent in human Rad50 as well and have been shown to correlate with increased local flexibility in these elongated structures (van Noort *et al.*, 2003).

In addition to different coiled-coil probabilities, the complexes vary in the distribution of different architectural conformations in respect to multimerization state and the way the Rad50/SbcC coiled-coils are connected. As visualized by scanning force

microscopy, *S. cerevisiae* and human Rad50-Mre11 complexes most frequently exhibit heterotetrameric forms (de Jager *et al.*, 2004). These arrangements consist of a large globular domain, including the Rad50 ATPase domains and Mre11, from which two coiled-coils protrude that are either in open or closed formation (conformations III and IV in table 2). Additionally, multimers containing two to ten heterotetrameric complexes that associate *via* their globular heads (conformation V in table 2) occur at low frequency. Bacterial and archaeal complexes, *E. coli* SbcCD and *P. furiosus* Rad50-Mre11 complexes, show different arrangements. A large portion consists of heterodimers comprising one SbcC/Rad50 and one SbcD/Mre11 molecule, a conformation that has not been observed in eukaryotes (conformation I in table 2). Similar to *S. cerevisiae* and human the two stated prokaryotic complexes form heterotetramers though mainly *via* interaction of the tips of their coiled-coil regions (conformation II in table 2).

The large multimers containing several heterotetrameric complexes might be precursors of large oligomers that form on DNA as observed for the human Rad50-Mre11 complex (de Jager *et al.*, 2001; de Jager *et al.*, 2002). However, the relevancies of the differences being observed among the Rad50-Mre11 orthologs are not clear to date. The biological impact of the differences in regard to their continuous coiled-coil formation and the proposed related flexibility as well as the distinct architectural arrangements yet needs to be defined.

Table 2: Distribution of different conformations of Rad50-Mre11 orthologs (de Jager *et al.*, 2004).

	I	II	III	IV	V
Human R/M	0	0	32	48	20
S. cerevisiae R/M	0	23	28	43	6
P. furiosus R/M	45	21	30	3	1
E. coli SbcCD	21	73	1	1	4[a]

The distribution of different architectural conformations, observed by scanning force microscopy, is indicated for each Rad50-Mre11 ortholog as percentage of the total number of complexes observed. The complexes are grouped into five categories: I) heterodimer, II) heterotetramer associating through the apexes of coiled-coils, III) heterotetramer associating through both the apexes of the coiled-coils and the globular domains, IV) heterotetramer associating through the globular domains and V) oligomers consisting of multiple heterotetramers. The different arrangements are illustrated in cartoons above the columns where the Rad50 ortholog ATPase domains are indicated by spheres and the Mre11 ortholog by an oval. [a] Oligomers of *E. coli* SbcCD exhibit a different architecture. Multimerization was not observed to occur between the two globular domains within a heterotetramer but rather between two different heterotetramers (de Jager *et al.*, 2004).

Recent work examined the localization and regulation of *E. coli* SbcCD (Darmon *et al.*, 2007). SbcC was shown to colocalize with SeqA, a protein that binds DNA at replication forks (den Blaauwen *et al.*, 2006; Molina and Skarstad, 2004), whereas SbcD was evenly distributed throughout the cytoplasm (Darmon *et al.*, 2007). It was suggested either that SbcC and SbcD can act independently of each other or that the regulation of function is achieved by assuring that the majority of the two proteins are not associated at a certain time point. SbcC could function in recognizing any of the above mentioned SbcD substrates and modulating the nuclease activity of SbcD. SbcC might constantly scan the replication fork for misfolded DNA and bind to SbcD only when DNA repair is needed. The fact that SbcD is evenly distributed inside the cell and does not form foci at the replication forks was explained by a possible too small amount or too quick mechanism of action of SbcD for visualization by fluorescence microscopy. However, it needs to be mentioned that in these studies the natural level of SbcCD in general was too low to be observed by fluorescence microscopy and that the localization results are based on overexpressed GFP fusion proteins, which might not ideally reflect the SbcCD positions under native conditions. Notably, the overexpressed

SbcC and SbcD proteins in *E.coli* and *B. subtilis* localize similarly, but the spatial distribution and the regulation of the natural protein levels vary (Mascarenhas *et al.*, 2006; Meile *et al.*, 2006). At native expression levels in 2% of *B. subtilis* cells SbcC localized as discrete foci, whereas SbcC in *E. coli* was not observable (Mascarenhas *et al.*, 2006). In addition, the SbcC and SbcD protein levels in *B. subtilis* were induced when mitomycin C, a DNA cross-linking agent, was applied to the cells, while it had no effect on *E. coli* sbcC and sbcD transcription (Mascarenhas *et al.*, 2006). The lacking SbcC foci at a natural expression level in *E. coli* were explained by a lower production of SbcC and a correlating lower, not detectable fluorescence signal (Darmon *et al.*, 2007). Whereas the distinct regulation patterns of the two organisms were suggested to be due to different functions of the complex (Darmon *et al.*, 2007). *B. subtilis* in contrast to *E. coli* possesses Ku- and ligase IV- like proteins that are involved in the non-homologous end-joining DNA repair pathway (Weller *et al.*, 2002) that is described in 2.1.2.1 and the SbcCD orthologs Rad50-Mre11 in eukaryotes are thought to play a role in this pathway (Aravind and Koonin, 2001; Doherty *et al.*, 2001; Pastwa and Blasiak, 2003). However, biochemical data proving these hypotheses are not present to date indicating the need for further studies of this highly versatile protein complex.

2.2.3   Rad50 – Mre11 Interaction

Although Mre11 binds to a variety of DNA structures, it was concluded that the active site sterically excludes dsDNA based on the substrate specificity of the exonuclease activity. Mre11 dimerizes and interacts with the coiled-coil region of Rad50 immediately adjacent to the ATP-binding site (Hopfner *et al.*, 2001), where it enlarges the DNA-binding surface of Rad50 in the heterotetrameric complex of $(Rad50)_2(Mre11)_2$ (figure 5). Electron micrographs and atomic force microscopy characterize the complex as a coiled-coil helical bundle with a protruding single globular DNA-interacting head (de Jager *et al.*, 2001; Hopfner *et al.*, 2001) (figure 5). The third component, Nbs1/Xrs2, apparently binds as a monomer, but the exact localization remains to be determined. Mre11 homodimerization is the initial driving force for MRN complex formation, whereas ATP-binding to Rad50 results in even

tighter association through Rad50 dimerization (Hopfner *et al.*, 2001; Hopfner *et al.*, 2000b). Similar to other ABC ATPases, it has been suggested that subsequent ATP hydrolysis leads to conformational changes that initiate altered substrate access for Mre11, its translocation and finally its dissociation from DNA (Friedberg *et al.*, 2006).

2.2.4   ATM and Checkpoint Activation

Introduction of breaks into DNA leads to the activation of a number of systems that recognize, signal and repair these breaks. The DSB-induced checkpoint response is triggered mainly by the ataxia-telangiectasia mutated (ATM) kinase, whose activation initiates a signaling cascade that leads to cell cycle arrest, DNA repair and/or apoptosis (Shiloh, 2003). ATM phosphorylates several proteins in response to DSBs that are involved in G1, intra-S-phase and G2/M checkpoints like p53 (Banin *et al.*, 1998) and BRCA1 (Xu *et al.*, 2002). ATM autoactivation and the phophorylation of downstream substrates is MRN dependent (Lee and Paull, 2004). Nbs1 itself is a substrate of ATM and thus the MRN complex works both upstream and downstream of ATM. For example, Nbs1 phosphorylation at S343 is required for ATM dependent phosphorylation of histone H2AX (Lee and Paull, 2004). In contrast, phosphorylation of p53 requires the presence of the MRN complex but not Nbs1 phosphorylation (Lee and Paull, 2004). Upon introduction of DSBs ATM autophosphorylates itself at Ser1981 and this phosphorylation leads to ATM dimer dissociation into monomers (Bakkenist and Kastan, 2003). The molecular basis for the MRN – ATM interaction has not been revealed so far, but MRN seems to physically interact with ATM, resulting in direct stimulation of ATM kinase activity (Lee and Paull, 2004). One possibility could be a structural switch of Nbs1 altering the binding surface properties induced by phophorylation of S343.

2.2.5   MRN Complex in Replication

Studies with cell-free *Xenopus* egg extracts demonstrate a clear requirement for MRN in avoiding replication-associated DSBs (Costanzo *et al.*, 2001). In eukaryotes, MRN is needed for the propagation of inverted repeats (Lobachev *et al.*, 2002) and in *E. coli* SbcCD was shown to be necessary for the processing of hairpins that arise during lagging strand synthesis and at inverted repeats as well (Connelly and Leach, 1996). The bacteriophage T4 homologs of Rad50/Mre11, gene products 46/47, are required for recombination-induced replication (Bleuit *et al.*, 2001; George *et al.*, 2001; Kreuzer, 2000). The MRN complex localizes not only to progressing replication forks but also to origins of replication, where it interacts through its Nbs1 subunit with E2F1 (Maser *et al.*, 2001). E2F1 is a transcription factor that is required for the localization of origin recognition proteins at the correct DNA sites. MRN is loaded onto nascent sister chromatids in S-phase (Mirzoeva and Petrini, 2003). A hypothetical model for the replication-linked architectural functions of MRN is depicted in figure 6. Replication over nicked or damaged DNA could specifically lead to a separated nascent sister chromatid. MRN could be loaded onto chromosomes before replication or onto nascent sister chromatids during replication. A possible model might be the entrapment of a sister chromatid into the coiled-coil ring, which could allow translocation by diffusion along DNA. In case of a break, MRN could specifically locate to the DNA end thereby joining the end to the sister chromatid (Assenmacher and Hopfner, 2004). Conformational changes of Nbs1 might lead to activation of ATM as mentioned above. Phosphorylation of Nbs1 by ATM will promote an active S-phase checkpoint complex that will phosphorylate downstream targets. If replication continues a DSB will be created. Another role of MRN could be the cohesion of both DNA ends with each other as in NHEJ, but the precise mechanism is so far unclear, though certain results indicate that such a clustering of MRN and ATM at DNA ends occurs *in vitro* (Costanzo *et al.*, 2004).
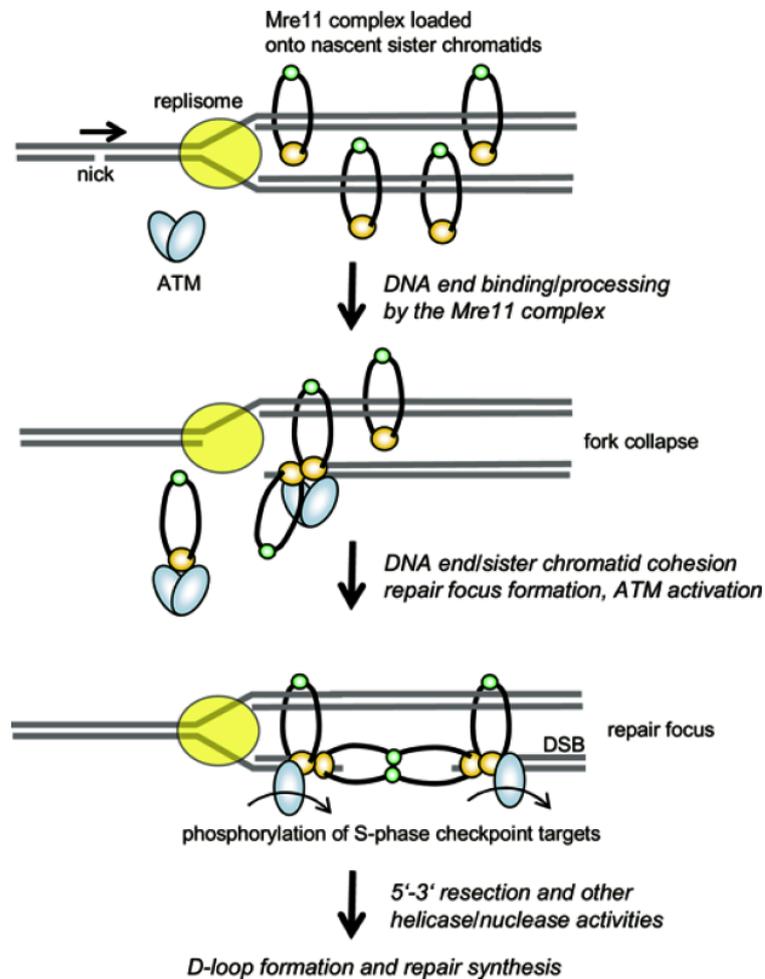
Figure 6: Replication fork-associated model of MRN.
Replication by the replisome (yellow) over DNA nicks could lead to a detached nascent chromatid. MRN complexes (orange heads, black coiled-coils, green hooks) loaded onto nascent chromatids in the wake of replication forks, could capture the free DNA end and degrade any secondary structures. Additionally, DNA-bound or free MRN might activate ATM (light blue). Following ATM activation, it phosphorylates Nbs1 and, after assembly of foci, the ATM/MRN complexes could phosphorylate S-phase checkpoint targets. Repair proceeds by interaction with downstream factors such as Werner (WRN) helicase/nuclease or other nucleases. Digestion of ends then leads to strand exchange and D-loop formation, followed by repair synthesis (Assenmacher and Hopfner, 2004).

2.2.6   MRN Complex in Telomere Maintenance

To the long list of cellular processes in which the MRN complex is involved, the participation of MRN in telomere maintenance needs to be mentioned. Treatment of a normal telomere as a DSB, possibly leading to end-to-end fusions, must absolutely be avoided in a living cell. A small fraction of the human telomere-capping protein TRF2

coimmunoprecipitates with all proteins of the MRN complex (Zhu *et al.*, 2000) (figure 7). Importantly also other factors implicated in NHEJ like the Ku proteins are found at telomeres. Mre11 and Rad50 have been shown to localize to telomeres throughout the cell cycle by indirect immunofluorescence, while Nbs1 seems to associate with telomeres specifically in S phase (Zhu *et al.*, 2000). The lethality conferred by targeted deletions of vertebrate MRN components may be related to a telomere defect (Luo *et al.*, 1999; Xiao and Weaver, 1997). Additionally, budding yeast, *Drosophila* and *Arabidopsis* MRN mutants show defects in telomere maintenance, telomere fusions, and, in some cases, the possibly associated effect of accelerated cellular senescence (Bi *et al.*, 2004; Boulton and Jackson, 1998; Ciapponi *et al.*, 2004; Gallego and White, 2001; Kironmai and Muniyappa, 1997; Nugent *et al.*, 1998).

Possibly, the MRN complex interacts with the specific DNA structure of telomeres (de Lange and Petrini, 2000). Human telomeres form a so-called T-loop structure in which the 3' single-stranded tail folds back into the duplex portion of the telomere and forms a local D-loop. Structures like this must be resolved in a controlled fashion to allow for telomerase activity. Possibly, not unlike its assumed role in homologous recombination, the MRN complex plays a role in establishing, maintaining or resolving such structures (Friedberg *et al.*, 2006).
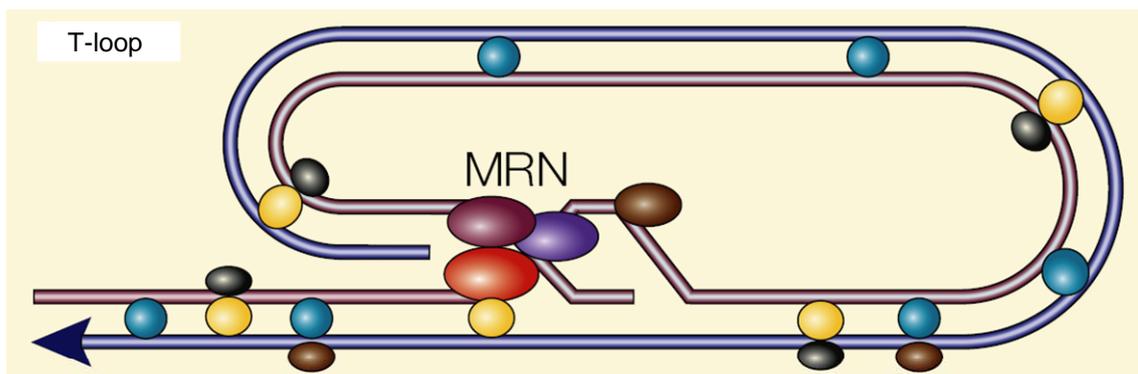
Figure 7: Telomere capping in humans.
The DNA structure of a telomere and its binding proteins are depicted. Mammalian telomeres are protected by TRF2, perhaps through its ability to form T-loops. Loss of TRF2 (yellow) results in cell cycle arrest and end-to-end ligation of telomeres. Commonly, ssDNA-binding protein POT1 (brown) binds to single-stranded G-rich telomeric DNA *in vitro*, its *in vivo* binding site could be the D-loop of human telomeres. The red line indicates the G-rich telomeric repeat strand that extends as a 3′ overhang beyond the double-stranded region of the telomere. The MRN complex is depicted in purple blue (Mre11), magenta (Rad50) and red (Nbs1). Further factors are indicated by additional colors (TRF1 (blue) and RAP1(black) (de Lange, 2001; Lustig, 2003).

Studies in *Drosophila*, however, suggest a different model, since in this organism telomeres are maintained without telomerase but rather through a transposition-like mechanism. Telomere shortening appears not to be required for the frequent telomeric fusions observed in MRN mutants. Instead, the failure to target two capping proteins to telomeres can explain their phenotype (Bi *et al.*, 2004; Ciapponi *et al.*, 2004).

2.2.7   MRN Complex and Diseases

Null mutations in components of the MRN complex are lethal in higher eukaryotes (Gorski *et al.*, 2004; Luo *et al.*, 1999; Yamaguchi-Iwai *et al.*, 1999; Zhu *et al.*, 2000) and hypomorphic mutations in Nbs1 and Mre11 cause the genome instability syndromes Nijmegen breakage syndrome (NBS) and Ataxia telangiectasia-like disease (ATLD), respectively (Carney *et al.*, 1998; Matsuura *et al.*, 1998; Stewart *et al.*, 1999; Varon *et al.*, 1998). A hypomorphic mutation in Rad50, originally identified in yeast (Alani *et al.*, 1990), the so-called separation-of-function point mutation (Rad50S), results in partial embryonic lethality and cancer susceptibility in mice (Bender *et al.*, 2002). 5' - 3' exonucleolytic single-strand degradation that results in 3' single-stranded tails typically observed in wild-type cells is absent in *rad50S* mutants (Sun *et al.*, 1989). This shows that even minor modifications of the MRN complex can have significant effects on the integrity of the genome, emphasizing an important role in genome maintenance.

2.3   Structure Determination

Protein function is strongly related to its structure. The knowledge of the atomic structure of macromolecules is essential for a fundamental understanding of their function and a prerequisite for structure based functional studies. At present, four different methods are available to study the three dimensional structure of proteins: nuclear magnetic resonance spectroscopy (NMR), electron microscopy (EM), X-ray crystallography and small angle X-ray scattering (SAXS). NMR can provide atomic resolution information of proteins in solution, but is still limited by the size of the

proteins (usually 30 kDa up to 70 kDa) that can routinely be studied. EM, in particular cryo-EM, can provide valuable low resolution information (~10 Å down to so far ~5 Å) on proteins in solution through single particle analysis. However, cryo-EM requires very large protein complexes (>250 kDa) or unusual symmetry for effective analyses. X-ray crystallography provides high resolution atomic structures of crystallized proteins without size limitations, while SAXS is more broadly able to provide low resolution structural information for either rigid or flexible proteins in solution to complement crystallographic analyses. The advantage of SAXS over crystallography is that the samples need not be crystalline, however, owing to the random orientation of dissolved or partially ordered molecules, spatial averaging occurs, which leads to a loss of information. In this work application of X-ray crystallography reveals the molecular structural details of the SbcCD complex and the results are fortified using SAXS for the in solution structure. These two techniques are briefly described in the following, while more detailed information can be found in textbooks (e.g. Blow, 2002; Drenth, 1999; Massa, 2002; McPherson, 2001).

2.3.1   X-ray Crystallography

As described by the Abbe limit, two distinct points can only be recognized as individuals if the wavelength of the electromagnetic radiation used for examination is in the range of the distance of both. For the determination of protein structures the wavelength has to be in the range of 1 Å since the typical length of a Cα-Cβ bond is 1.497 Å (Engh, 1991). In general electromagnetic waves with a wavelength between 0.1 and 10 Å are denoted as X-rays. In regard to protein structure solution these rays need to be highly energetic and are generated by rotating copper anodes or synchrotrons. Since the scattering power of a single molecule in solution is insufficient to generate a detectable signal, the protein under investigation needs to be crystallized. Crystals are highly ordered structures, where a unit cell containing the protein is periodically repeated in a three dimensional lattice. The waves scattered from different atoms in a crystal interfere and, depending on the phase difference, amplify or damp each other. If the waves are in-phase, meaning their shift is proportional to 2π, the signal enhances

leading to a diffraction pattern on the detector. The reflections are determined by Bragg's law (equation 1) to a set of parallel lattice planes in the reciprocal space and are characterized by the Miller indices h, k, and l.

$$n \cdot \lambda = 2 \cdot d \cdot \sin \theta \quad (1)$$

A signal can only be detected if the distance d and the angle θ between the planes and the incident beam follow this law, where n is an integer and λ the wavelength of the X-rays.

The intensity of a reflection (h, k, l) is proportional to the square of the structure amplitude $|F_{hkl}|$. The structure factor F is a mathematical description of how the electrons of the crystal scatter the incident beam (equation 2).

$$F_{hkl} = \sum_{j=1}^{N} f_j \cdot \exp\left[2\pi \cdot i(hx_j + ky_j + lz_j)\right] \cdot \exp\left[-B_j \cdot \sin^2 \theta / \lambda^2\right] \quad (2)$$

The atomic scattering factor f_j describes the scattering power of an atom j with the coordinates x, y, z and is dependent on the atom type. The Debye-Waller factor (B-factor), the last term of the equation, represents the effect of thermal disorder.

Since X-rays get diffracted by the electrons of the atoms the structure factor is related to the repetitive electron density pattern of the crystal by a Fourier transformation (FT). By direct Fourier transformation it is possible to compute the structure factors out of the electron density. Reversely, structure factors can be used to calculate electron density by inverse FT (equation 3).

$$\rho_{xyz} = \frac{1}{V} \sum_{hkl} F_{hkl} \cdot \exp\left[-2\pi \cdot i(hx + ky + lz)\right] \quad (3)$$

The intensities of the reflections contain the information of the amplitudes of the structure factors $|F_{hkl}|$. In addition to the amplitude $|F_{hkl}|$, the structure factor F_{hkl} (equation 4) consists of the phase α of a scattered wave, which cannot be recorded. This missing phase information is referred to as the "phase problem" in crystallography.

$$F_{hkl} = |F_{hkl}| \cdot \exp\left[i \cdot \alpha_{hkl}\right] \quad (4)$$

Obtaining the phase information is essential and can be achieved in different ways. If the structure of a similar molecule is available, it can be used for phase determination by molecular replacement. For *de novo* phasing, heavy atom methods like isomorphous replacement (SIR/MIR) or anomalous dispersion (SAD/MAD) can be utilized. Direct methods can be used for very small molecules with less than ~1000 atoms in the asymmetric unit. In this work a single wavelength anomalous diffraction (SAD) experiment using selenium has been performed.

To calculate the phase angles in *de novo* structure determination the Patterson function is an important tool. It is a Fourier transformation of the measured intensities (its squared reflection amplitudes $|F_{hkl}|^2$) that do not depend on phases (phase angle = 0°) (equation 5).

$$P(uvw) = \frac{1}{V} \sum_{hkl} |F_{hkl}|^2 \exp[-2\pi \cdot i(hu + kv + lw)] \quad (5)$$

The Patterson unit cell is given in (u, v, w) and has the same dimensions as the real unit cell (x, y, z). The Patterson function results in the so-called Patterson map of interatomic distance vectors. Endpoints of the vectors represent the multiplied electron density and so their height depends on the electron number of the respective atoms, which makes it utilizable to determine the relative position of heavy atoms in a unit cell.

If the incident beam possesses an energy close to the eigenfrequency of an atom some photons are absorbed and re-emitted at lower energy (fluorescence) or at the same energy with a phase delay (anomalous dispersion). In case of anomalous dispersion the atomic scattering factor, which is normally given for wavelengths far from the absorption edge of an atom, gains an anomalous contribution that is composed of a real and an imaginary part (equation 6 and figure 8).

$$f_{ano} = f_0 + \Delta f + if'' = f' + f'' \quad (6)$$
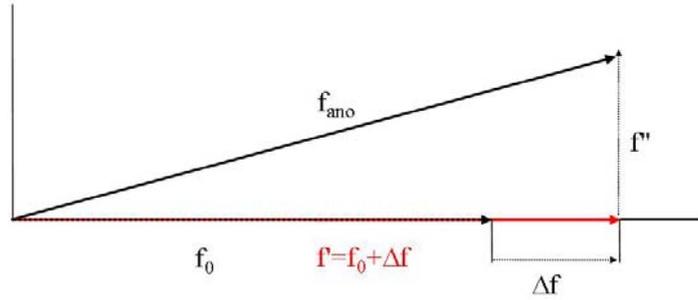
Figure 8: The anomalous atomic scattering factor f$_{ano}$.
The anomalous contribution to the atomic scattering factor is composed of a real part Δf and an imaginary component f''. The phase of the imaginary part is always shifted by 90°.

Elements normally occurring in biological molecules such as carbon, oxygen and nitrogen do not have their absorption edges in the range of X-ray wavelengths used in diffraction experiments. For this reason heavy atoms like selenium, mercury or platinum that show detectable anomalous scattering at X-ray wavelengths are introduced into the protein. Because the actual energy where the transition occurs depends on the environment of the heavy atoms, the required wavelength needs to be determined for every experiment by a fluorescence scan.

The scattering is then a sum $|F_{PH}|$ of the normal scattering from light atoms $|F_P|$ and scattering from heavy atoms $|F_H|$ with a normal $|F_{HN}|$ and an anomalous $|F_{HA}|$ contribution (equation 7).

$$F_{PH} = F_P + F_{HN} + F_{HA} = F_P + F_H \quad (7)$$

In normal scattering the structure factors $F_P(h,k,l)$ and $F_P(-h,-k,-l)$ have the same amplitudes and opposing phases (Friedel mates) according to Friedel's law, but anomalous scattering causes violation of Friedel's law and $F_{PH}(h,k,l)$ and $F_{PH}(-h,-k,-l)$ (also called Bijvoet pair) do not have the same magnitudes.

$$\Delta|F_{PH}|_{ano} = \left(\left|F_{PH(hkl)}\right| - \left|F_{PH(-h-k-l)}\right|\right) \cdot \frac{f'}{2f''} \quad (8)$$

A Patterson map calculated with the squared Bijvoet differences ($\Delta|F_{PH}|_{ano}$) results in a map showing only interatomic distance vectors between the anomalous scatterers and allows localization of the heavy atom substructure in the unit cell. This allows

determination of the structure factor F_{HA} (amplitude and phase angle) of the heavy atoms and the calculation of its contribution to the structure factor F_H in SAD experiments. F_{PH} (equation 9) allows calculation of F_P and determination of the protein phase angles.

$$F_{PH} = |F_P| \cdot \alpha_P + |F_H| \cdot \alpha_H \quad (9)$$

In a SAD experiment, compared to MAD, only the peak anomalous wavelength is recorded, where f'' is maximal. This is advantageous especially for radiation sensitive crystals, since radiation damage can cause a MAD experiment to fail. In the absence of other crystallographic phase information, anomalous diffraction data collected at a single wavelength cannot produce unimodal phase-probability distributions. This inability to provide unique phase choices is termed the SAD phase ambiguity. However, in most cases it is possible to resolve the phase ambiguity by solvent flattening and non-crystallographic symmetry averaging.

2.3.2   Small Angle X-ray Scattering

Generally, small angle X-ray scattering (SAXS) and X-ray crystallography are similar techniques. Both methods apply X-ray beams to irradiate a sample and the intensities of the scattered (SAXS) or diffracted (crystallography) X-rays are measured by an X-ray detector. The fundamental differences between solution X-ray scattering and X-ray crystallography solely are due to the effects of taking target molecules from solution (SAXS) and placing the sample into a crystal lattice (crystallography).

When matter is irradiated by an X-ray beam, the primary interaction is the elastic scattering of X-rays by electrons. X-rays thus scattered by spatially correlated electron pairs constructively and destructively interfere in accordance with Bragg's law (as explained in more detail in chapter 2.3.1). The longer the distance of the electron pair, the smaller the angle at which interference effects are observable. Traditionally, SAXS was defined by scattering at angles where the small angle cosine and sine approximations apply, but practically SAXS analysis can use scattering at angles up to the primary scattering peak of water at a momentum transfer of ~2 $Å^{-1}$ which thereafter

covers all other signals. The momentum transfer s is described in equation 10 where λ is the wavelength of X-rays and θ is the scattering angle relative to the incident beam (figure 9).

$$s = \frac{4\pi}{\lambda} \sin\left(\frac{\theta}{2}\right) \quad (10)$$
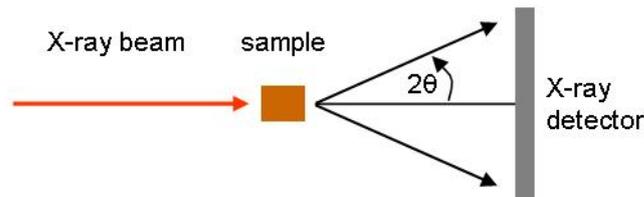
Figure 9: The scattering angle θ.
Both SAXS and X-ray crystallography involve placing a sample (brown) into a highly collimated X-ray beam (red) and measuring the scattered X-rays. The angle of any scattered position with the direct beam is 2 θ (Putnam *et al.*, 2007).

For proteins tumbling in solution, all configurations of the macromolecules are sampled during the experiment and the scattering profile is radially symmetric around the primary beam. As the scattering of buffer and protein solution are very similar, the solvent scattering needs to be subtracted from sample scattering. The scattering profile depends on size and shape of the molecule under investigation and is unique for every protein.

As mentioned earlier, X-ray scattering from protein molecules is a function of their electron density. The scattering profile of a buffer-subtracted macromolecular solution is the Fourier transform of the spatially averaged auto-correlation function of the electron density (Patterson function) (equation 11).

$$I(s) = 4\pi \int_0^{D_{max}} \langle \rho(\vec{r}) * \rho(-\vec{r}) \rangle \frac{\sin(sr)}{sr} r^2 dr \quad (11)$$

I represents the intensity, ρ(r) the electron density, D_{max} the maximum dimension of the protein and r is the vector describing the spatial position. The Patterson function multiplied by r^2 is called the pair distribution function or p(r) (equation 12).

$$p(r) = r^2 \langle \rho(\vec{r}) * \rho(-\vec{r}) \rangle \quad (12)$$

The p(r) plot shows a distribution of electrons that are in distance r from each other. From the p(r) distribution the maximum intraparticle distance (D_{max}), the radius of gyration (R_g) as well as the overall shape (globular or elongated) can be determined (Koch *et al.*, 2003; Tsutakawa *et al.*, 2007). A comparison of the pair distribution function with the Patterson function used in X-ray crystallography is depicted in figure 10.
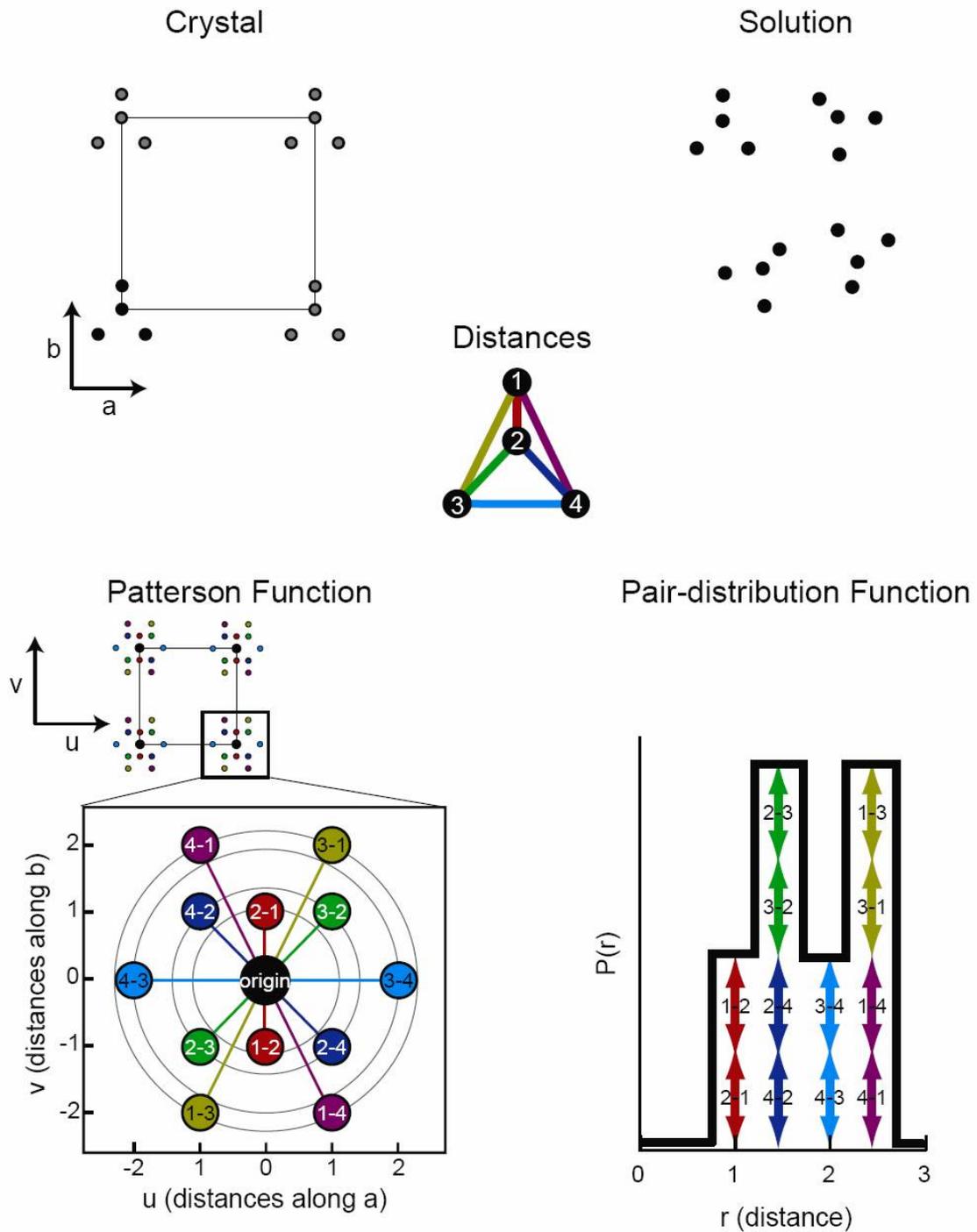
Figure 10: Comparison of the Patterson and the pair distribution function.
A theoretical molecule of four atoms is placed in an arbitrary two-dimensional crystal in solution. The Patterson function comprises cross peaks for all the interatomic distances in the crystal and these cross peaks in the u, v plane are indicated by circles and contain directional information about their positions in the crystal. Cross peaks between symmetry mates are not shown. The pair distribution function resolves only distances but not directions within each scattering unit. Therefore all equivalent distances in the molecule summate (Putnam *et al.*, 2007).

The structure calculation based on SAXS data is always an *ab initio* approach. The program GASBOR starts with a sphere which is large enough to represent the whole protein (Svergun *et al.*, 2001). Then this sphere is filled with dummy atoms, where each amino acid is represented by one bead. The algorithm can assign two states for each dummy atom: belonging to a) the protein or b) the solvent. The theoretical scattering curve is then calculated and compared with the measured data. Progressive trial and error assignment of the beads is performed and the theoretical curve is calculated again. The difference between experimental and theoretical curves can be used as measure of the quality of the fit. With every iteration loop the bead model gets closer to the real structure.

3   Objectives

DNA gets continually threatened by extracellular toxins but also by a diverse range of intracellular metabolic by-products. DNA damage repair plays an essential role in the maintenance of genomic integrity. Since DNA double-strand breaks represent one of the most severe damages their repair is highly important. The Mre11/Rad50/Nbs1 complex is sought to play a key role in this repair. Mre11 and Rad50, but not Nbs1, are present in all kingdoms of life and are therefore suggested to build the core of this complex.

The structures of the Rad50 ATPase domain and Mre11 containing two of its three domains have been solved as individual molecules. However, their interactions as well as their relative spatial orientations have been unclear so far. Additionally, the Mre11 full-length structure is not known so far.

Aim of this PhD thesis was to gain insight into the three-dimensional structure of the SbcCD complex, an ortholog of the eukaryotic Rad50/Mre11 complex that can provide important insights into its interactions and its proposed mode of action. Therefore SbcCD should be cloned, expressed, purified and crystallized to be able to perform X-ray crystallographic studies in order to obtain high resolution structural information. In addition, the resulting crystallographic structure should be validated by small angle X-ray scattering as a method determining the structure in solution. After examination of the protein structure the role of residues suggested to be important for biological function should be confirmed by *in vivo* assays of wild-type and mutant proteins.

4   Materials

4.1   Chemicals

All common chemicals were obtained from Merck (Darmstadt, Germany), Roth (Karlsruhe, Germany) and Sigma (Deisenhofen, Germany), unless otherwise stated. Crystallization screens and crystallization tools were from Hampton Research (Aliso Viejo, USA), Nextal Biotechnologies (Montreal, Canada; now QIAGEN, Hilden, Germany) or Jena Bioscience (Jena, Germany). RP-HPLC purified oligonucleotides were ordered from Thermo Electron Corporation (Ulm, Germany) and Biomers (Ulm, Germany), respectively. Enzymes and nucleotides for molecular biology were from Fermentas (St- Leon-Rot, Germany).

4.2   Molecular Biology Methods

Standard molecular biology methods like polymerase chain reaction (PCR), cleavage of phosphodiester bonds in DNA by restriction endonucleases, dephosphorylation of DNA ends by alkaline phosphatase, ligation of DNA ends, amplification of plasmid DNA and agarose gel electrophoresis were carried out according to standard protocols (Sambrook, 1989).

Genomic DNA was prepared with DNAzol® reagent (Molecular Research Center, Cincinnati, USA) according to the manual and plasmid DNA isolation was performed using the NucleoSpin®-Plasmid Quick Pure Kit (Macherey-Nagel, Dueren, Germany). DNA fragments were extracted from agarose gels with the NucleoSpin®-Extract II Kit (Macherey-Nagel, Dueren, Germany). DNA-sequencing was carried out by Medigenomix (Martinsried, Germany).

4.3   Enzymes, Standards and Chromatographic Material

ACCUZYME™ DNA Polymerase	Bioline, Luckenwalde, Germany
BenchMark Protein Ladder	Invitrogen, Karlsruhe, Germany
Calf Intestine Alkaline Phosphatase	Fermentas, St. Leon-Rot, Germany
GeneRuler 1kb DNA Ladder	Fermentas, St. Leon-Rot, Germany
HiTrap Q HP	GE Healthcare, Uppsala, Sweden
Ni-NTA-Agarose	Qiagen, Hilden, Germany
Protein Molecular Weight Marker	Fermentas, St. Leon-Rot, Germany
Restriction Endonucleases	Fermentas, St. Leon-Rot, Germany
Selenomethionine	Calbiochem, Schwalbach/Ts., Germany
Superdex S200	GE Healthcare, Uppsala, Sweden
Sephacryl S300	GE Healthcare, Uppsala, Sweden
Source 15 S	GE Healthcare, Uppsala, Sweden
T4 DNA Ligase	Fermentas, St. Leon-Rot, Germany
Taq DNA Polymerase	Fermentas, St. Leon-Rot, Germany

Unless stated otherwise, all enzymes and reagents were used according to the manufacturer's instructions.

4.4   Oligonucleotides

Table 3: List of oligonucleotides used for cloning.
RP-HPLC purified DNA was ordered by Thermo Electron Corporation (Ulm, Germany) and Biomers (Ulm, Germany), respectively. Sequences are given in 5' → 3' direction.

tmr50fwd1	AAAAAAAACATATGCGCCCTGAACGCCTCACCGTTAGAAACTTTCTCGG
tmr50rev2	CCGCACCGCCCGCACCGCCTGAGGATATCTCGTTCTCCAGCT
tmr50fwd3	GGCGGTGCGGGCGGTGCGGGCGGTAGTCTGGAGAAGAAACTAAAAGAGA
tmr50rev4	AAAAAAAAGCGGCCGCTCACTCATTCACCACAACTCCTC
tmr50EQrev2	GTGTCGAGACTGGAAAACCCTTGATCGATGAA
tmr50EQfwd3	CTGGACGCGTTCTTCATCGATCAAGGGTTTTC
tmmreL7M1fwdNotI	AAAAAAAAGCGGCCGCTAACTTTAAGAAGGAGATATACAATGAAAATTCTGCACACATCCGACTGGCACCTC
TmM11K337fwdNotI_2	AAAAAAAAGCGGCCGCTAACTTTAAGAAGGAGATATAGTAATGAAAGAAGAACTGGACAAGCTTGATT
tmmrerev_6xHis_Bpu_2	TTTTTTGCTCAGCTTAGTGGTGGTGGTGGTGGTGGCCGGCCTCACTCTTTTTCACCTCAT
tmmrerev.3	TTTTTTGCTCAGCTCAGGCCTCACTCTTTTTCACCTCAT
tmmre11H88Qrev2	CTTCCAATCCTGGTTTCCTGGAAGTACAACCAC
tmmre11H88Qfwd3	CAGGAAACCAGGATTGGAAGGGGTTGAAACTCTT
pfur50fwd1	AAAAAAAACATATGAAACTGGAACGCGTGACTGTGAAAAACTTTAGGTCTCAT
pfr5eqn	GATTTTAGATCAACCAACGCCTTATTTAGATGAGGAG
pfr5eqc	ATAAGGCGTTGGTTGACTAAAATCAGGAGGCTTATCTC
pfur50rev4	AAAAAAAAGCGGCCGCCTAAGAGACCACCTCCACCTTG
pfum11notIfwd	AAAAAAAAGCGGCCGCTAACTTTAAGAAGGAGATATAGTAATGAAATTTGCTCACTTAGCCGATATTCAT
pfum11m1i344bpurev	TTTTTTGCTCAGCTTATATTTTTCCACTTTCTTCGTCTGT
T7 promotor	TAATACGACTCACTATAGGG
T7 terminator	TATGCTAGTTATTGCTCAG

4.5   *E.coli* strains and Plasmids

Table 4: List of bacterial strains and plasmids.

E.coli strain	Genotype	Source
XL1 blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F′ proAB lacIqZΔM15 Tn10 (Tetr)]	Stratagene, Heidelberg, Germany
Rosetta (DE3)	F⁻ *ompT hsdS*B(rB⁻ mB⁻) *gal dcm* (DE3) pRARE2 (CamR)	Novagen, Madison, USA
B834(DE3)	F⁻ *ompT hsdS*B(rB⁻ mB⁻) *met gal dcm* (DE3)	Novagen, Madison, USA

Plasmid	Source
pET21b(+)	Novagen, Madison, USA
pET29b(+)	Novagen, Madison, USA

5   Methods

5.1   Cloning

Genes of interest were amplified by polymerase chain reaction (PCR) from *Thermotoga maritima* MSB8 or *Pyrococcus furiosus* genomic DNA, respectively.

Thermotoga maritima is a non-spore forming, rod-shaped eubacterium with an optimal growth temperature of ~80°C. It is phylogenetically very important, because it appears to be one of the deepest lineages in eubacteria. Since its genome has the highest percentage (24%) of genes similar to archeal genes, it is an ideal system to study proteins apparent in the distinct kingdoms of life. Its genome has a size of 1861 kbp and comprises 1928 genes. *Pyrococcus furiosus* is an extremophilic species of archaea. It grows between 70ºC and 103ºC, with an optimum temperature of 100ºC and belongs to the phylum euryarchaeota. 2228 genes are encoded by a 1908 kbp chromosome.

Oligonucleotides used as PCR-primers were designed to have melting temperature of 68 °C according to the Wallace rule. In general, oligonucleotides used in this thesis have a complementary part with a melting temperature between 66°-70°C according to the Wallace rule, an attached endonuclease restriction site and a 7-9 poly(A) overhang to assure efficiency of restriction endonuclease cleavage. If two genes were cloned into one vector, the forward primer for the downstream gene contained an additional ribosome binding site. The upstream gene (*sbcC*) was cloned with *Nde*I/*Not*I and the downstream gene (*sbcD*) with *Not*I/*Bpu*1102I. A 50 µl PCR reaction contained 600 µM of each of the four dNTPs, 2.5 mM $MgCl_2$, 0.5 pM each of forward and reverse primer, and about 1 ng of template DNA. Each thermocycling program consisted of 25-30 cycles, whereas times and temperatures of denaturation, annealing and elongation were modified according to special requirements of the polymerase used in different amplifications. Analysis and purification of PCR products was performed by mixing with loading dye followed by separation of the DNA fragments in a 1% agarose gel at 10 V/cm with 1x TAE as running buffer. To visualize DNA, 0.01% (v/v) of a 10 mg/ml ethidium bromide stock solution was added to the gel and UV light was applied using a

standard ultraviolet transilluminator (λ = 254 nm, Eagle Eye, Stratagene, LaJolla, USA). Bands of interest were cut out and DNA was extracted using the NucleoSpin Extract II Kit.

<u>6x loading dye solution</u>: 10 mM Tris-HCl (pH 7.6), 0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% (v/v) glycerol, 60 mM EDTA

<u>50x TAE</u>: 2 M Tris (pH 8.0), 1 M Acetate, 50 mM EDTA

Mutations were introduced by PCR-based site-directed mutagenesis (Ho *et al.*, 1989). Here, two specially designed complementary oligonucleotides encoding the desired mutation were used to generate two DNA fragments with overlapping ends in separate PCR reactions. These two PCR products were purified by agarose gel electrophoresis and 0.5 µl of each fragment was used as template in a second PCR reaction. In this second reaction, the two DNA fragments serve as template to amplify the full-length gene with the incorporated nucleotide exchange.

For the *Thermotoga maritima* (Tm) SbcC construct, the N- (M1-S190) and C-terminal (S686-E852) fragments were connected using an eight amino acid flexible linker (GGAGGAGG). In case of *Pyrococcus furiosus* (Pfu) Rad50, the N-terminus (M1-E186) was linked to the C-terminus (K713-S882) *via* four amino acids (GSGS). A mutation (E823Q) was introduced into the Walker B motif to deplete ATPase activity. The length of the coiled-coils was designed based on the known Rad50cd structure (Hopfner *et al.*, 2001; Hopfner *et al.*, 2000b) and secondary structure prediction. In the following, these SbcC/Rad50 constructs are denoted as SbcC and PfR50NC. Pfu Mre11 remained full-length, while Tm SbcD was modified at the N-terminus (removal of the first six amino acids and replacement of L7 by M7) since alignment of the sequence deposited in the NCBI database (http://www.ncbi.nlm.nih.gov) showed weak homology in this region compared to SbcD/Mre11 from other bacteria and archaea. SbcC is deposited in the National Center for Biotechnology Information (NCBI) database under the accession number Q9X1X1 (locus tag TM1636), TmSbcD under AAD36702 (locus tag Tm1635), PfRad50 under AAL81291 (locus tag PF1167) and PfMre11 under Q8U1N9 (locus tag PF1166).
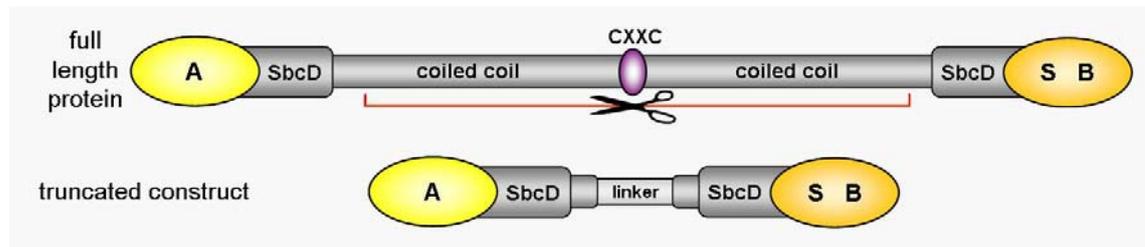
Figure 11: SbcC/Rad50 construct overview.
The coiled-coils were truncated to facilitate crystallization. An artificial amino acid linker was inserted to connect the N- and C-terminal part of the protein.

In the course of generating the crystal setup construct for SbcC in complex with DNA, the truncated SbcC construct as described before was cloned into a bicistronic pET29 vector with the C-terminal His-tagged coding region of SbcD (Lys343-Ala385) located downstream.

The PCR amplicons were digested using the corresponding restriction endonucleases and buffers according to the manufacturer. Digested products were purified with NucleoSpin® Extract II kit. Cleaved vectors were additionally treated with calf intestine alkaline phosphatase to prevent religation. For ligation, a five- to tenfold excess of the digested DNA fragment was incubated with linearized vector and T4 ligase in a 10 μl reaction following instructions of the manufacturer.

Preparation of chemical transformation competent *E. coli* cells was performed according to Hanahan (Hanahan, 1983). 200 ml LB media was inoculated with 2 ml of an overnight culture of the desired bacterial strain. The cells were grown for about 3 hours at 37°C to an OD_{600} of approximately 0.4, cooled down to 4°C on ice and centrifuged for 10 minutes at 3000 rpm and 4°C. The pellet was washed with 30 ml of TBI, incubated for 10 minutes and centrifuged again as above. This pellet was resuspended in 4 ml of TBII, aliquoted and flash frozen in liquid nitrogen. Washing was performed at 4°C and the competent cells were stored at -80°C.

LB media: 1% (w/v) Bacto Trypton, 0.5% (w/v) Yeast Extract, 0.5% (w/v) NaCl, pH 7.0 (adjusted with 2 N NaOH)

TBI: 100 mM KCl, 50 mM $MnCl_2$, 30 mM potassium acetate, 10 mM $CaCl_2$, 15% (v/v) glycerol, pH 5.8 (adjusted with 0.2 M Acetate)

<u>TBII</u>: 10 mM MOPS, 10 mM KCl, 75 mM CaCl$_2$, 15% (v/v) glycerol, pH 7.0 (adjusted with 2 N NaOH)

For transformation, 50-100 µl of the competent *E. coli* cells were mixed with 10 µl of the ligaton reaction or 1 µl of purified plasmid DNA (40 ng/µl) and annealed on ice for 15 minutes according to Sambrook (Sambrook, 1989). The cells then were incubated at 42°C for 55 seconds and chilled on ice for two minutes. 600 µl of fresh LB medium were added, followed by incubation at 37°C for 45 minutes in a thermo shaker. For selection the cells were streaked out on LB agar plates containing the appropriate antibiotic (1:1000) and incubated over night at 37°C. Plasmid DNA was isolated from a 5 ml overnight culture of a single colony using the NucleoSpin®-Plasmid Quick Pure Kit. DNA-sequencing of clones was carried out by Medigenomix (Martinsried, Germany).

<u>LB Agar</u>: LB media with 1.5% (w/v) agar

<u>Ampicillin</u> (Na-salt): 100 mg/ml in water

<u>Kanamycin</u>: 50 mg/ml in water

<u>Chloramphenicol</u>: 50 mg/ml in 100% ethanol

<u>Tetraycline</u>: 12.5 mg/ml in 100% ethanol

5.2   Bioinformatics and Protein Parameters

DNA and protein sequences were found using the NCBI database (http://www.ncbi.nlm.nih.gov) and homology searches were performed using the NCBI Basic Local Alignment Search Tool (BLAST) server (http://www.ncbi.nlm.nih.gov/BLAST). Chemical and physical parameters of the recombinant proteins like molecular weight, the theoretical isoelectric point (pI) and extinction coefficients were calculated using ProtParam (Gasteiger *et al.*, 2003) (www.expasy.org/tools/protparam.html) from the ExPASy Proteomics Server. Protein concentrations were calculated using the absorption at 280 nm and the theoretical

molecular weight of the heterodimeric SbcCD/Rad50-Mre11 complex. Protein secondary structure prediction was performed by the PSIPRED Protein Structure Prediction Server (http://bioinf.cs.ucl.ac.uk/psipred/) (Bryson *et al.*, 2005; Jones, 1999; McGuffin *et al.*, 2000) and COILS was used to calculate the probability of a coiled-coil conformation (Lupas *et al.*, 1991). Sequence alignment was carried out with ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html).

5.3   Protein Expression

For the over expression of recombinant proteins, competent *E. coli* Rosetta (DE3) cells were transformed with plasmids containing the gene(s) of interest. Cells were grown at 37°C in LB medium spiked with the appropriate antibiotics to an OD_{600} of 0.6-0.8. Protein expression was induced with 0.2 mM IPTG and carried out for 4-5 hours at 37°C. Cells were harvested by centrifugation at 4°C and the pellets were flash frozen in liquid nitrogen and stored at -80°C.

Selenomethionine-substituted protein was generated using the methionine auxotrophic *E. coli* B834 (DE3) strain. Cells were transformed with the pRARE (Cm^R) plasmid isolated from the Rosetta (DE3) strain and with a plasmid containing the gene(s) of interest. Cells were grown in 2000 ml LB medium supplemented with the appropriate antibiotics at 37°C to an OD_{600}=0.6-0.8. Cells were centrifuged and the pellet was resuspended in the same volume of LeMaster´s medium (table 4) supplemented with selenomethionine (125 mg/2000 ml) and antibiotics. Cells were grown to an OD_{600} of 0.8 at 37°C to deplete the medium of any residual methionine. Protein expression was induced with 0.2 mM IPTG and carried out for 4-5 hours at 37°C. Cells were harvested by centrifugation at 4°C and pellets flash frozen in liquid nitrogen and stored at -80°C.

Table 5: LeMaster's medium (LeMaster and Richards, 1985).

• **autoclavable portion 1 for LeMaster's medium (g / 2000 ml)**			
• alanine	• 1.0	• serine	• 4.166
• arginine hydrochloride	• 1.16	• threonine	• 0.46
• aspartic acid	• 0.8	• tyrosine	• 0.34
• cystine	• 0.066	• valine	• 0.46
• glutamic acid	• 1.5	• adenine	• 1.0
• glutamine	• 0.666	• guanosine	• 1.34
• glycine	• 1.08	• thymine	• 0.34
• histidine	• 0.12	• uracil	• 1.0
• isoleucine	• 0.46	• sodium acetate	• 3.0
• leucine	• 0.46	• succinic acid	• 3.0
• lysine hydrochloride	• 0.84	• ammonium chloride	• 1.5
• phenylalanine	• 0.266	• sodium hydroxide	• 1.7
• proline	• 0.2	• dibasic potassium phosphate	• 21.0

• **non-autoclavable solution 2**	
• glucose	• 20.0 g
• magnesium sulfate heptahydrate	• 0.5 g
• iron sulfate	• 8.4 mg
• sulfuric acid (concentrated)	• 16.0 µl
• thiamin	• 10.0 mg

All amino acids were reagent-grade L-enantiomers purchased from Sigma-Aldrich. After autoclaving of solution 1, it was cooled down to 37°C and filter-sterilized solution 2 (200 ml of solution 2 / 2000 ml of solution 1) was added. Selenomethionine was dissolved in sterile H_2O and added to the medium (100 mg / 2200 ml).

5.4   Protein Purification

Cells having expressed the truncated *Thermotoga maritima* SbcC construct and SbcD were resuspended in 500 mM NaCl, 5 mM EDTA and 20 mM Tris (pH 7.5) followed by sonication. Insoluble proteins and cell debris were removed by centrifugation for 20 minutes at 40000 g. The supernatant was heated up to 60°C for 10 minutes to denature endogenous *E. coli* proteins and subsequently cleared by centrifugation as performed above. 2.5 M KCl were added to the protein located in the supernatant and this solution was then loaded onto a Phenyl HP column pre-equilibrated with buffer 1 at 10°C. Buffer 2 was used for elution. Since most of the protein did not bind to the column, the hydrophobic interaction purification procedure was repeated at room temperature. The bound protein was eluted with buffer 2 and the peak elution fractions were pooled. For subsequent processing the protein solution's conductivity was reduced to 7.0 mS by dilution with ultrapure water. After dilution the protein was loaded onto a HiTrap Q HP anion exchange column equilibrated with buffer 3 and then eluted with a linear gradient of 10 column volumes from 50 mM to 1 M NaCl (buffer 4) in the same buffer. Peak fractions were pooled, concentrated and afterwards applied onto a Superdex 200 size-exclusion column rinsed with buffer 5. Again, peak fractions were pooled and concentrated to the desired concentration. Selenomethionine substituted protein was purified analogously with addition of 2 mM DTT to all buffers.

Buffer 1: 3 M KCl, 1 mM EDTA, 20 mM Tris pH 7.5

Buffer 2: 1 mM EDTA, 20 mM Tris pH 7.5

Buffer 3: 50 mM NaCl, 1 mM EDTA, 20 mM Tris pH 7.5

Buffer 4: 1 M NaCl, 1 mM EDTA, 20 mM Tris pH 7.5

Buffer 5: 200 mM NaCl, 1 mM EDTA, 5% (v/v) Glycerol, 20 mM Tris pH 7.8

Pyrococcus furiosus Rad50NC and Mre11 (PfuR50NCM11) expressing *E. coli* cells were lysed by heat shock. First the cell pellet was broken up using a liquid nitrogen pre-cooled pestle and mortar followed by 3 minutes heat shock through 95°C hot buffer A. The suspension was cooled down on ice-water to 4°C and then spinned down for 15 minutes at 40000 g. Then a heat denaturation step at 75°C for 15 minutes with subsequent centrifugation as above was performed. To remove nucleic acids, proteins were precipitated by addition of ammonium sulfate to the supernatant to 80% of saturation. After 1 hour incubation at 4°C, precipitated proteins were pelleted by centrifugation for 15 minutes at 20000g and resuspended in buffer B prior loading on a S300 gel filtration column to remove residual DNA molecules. The peak fractions were pooled and KCl was added to a 2.5 M final concentration. A Phenyl HP column was equilibrated with buffer C and then loaded with the protein at room temperature. Following separation, bound proteins were eluted with buffer D. The sample then was dialyzed with buffer E, which was also used to equilibrate a HiTrap S HP column. The protein then was loaded onto this cation exchanger and eluted with a linear gradient from 100% buffer E to 100% buffer F over 10 column volumes. Peak fractions were pooled and, after concentration, applied onto a Superdex 200 26/60 size exclusion column rinsed with buffer G. After this last separation step, the peak fractions were concentrated again.

Buffer A: 500 mM NaCl, 10 mM EDTA, Lysozyme, PMSF, 50 mM NAH_2PO_4 pH 7.4

Buffer B: 1.5 M NaCl, 1.5 M EDTA, 20 mM Tris pH 7.0

Buffer C: 3 M KCl, 1 mM EDTA, 20 mM Tris pH 7.5

Buffer D: 1 mM EDTA, 10% Glycerol, 20 mM Tris pH 7.5

Buffer E: 50 mM NaCl, 1 mM EDTA, 20 mM Tris pH 7.0

Buffer F: 1 M NaCl, 1 mM EDTA, 20 mM Tris pH 7.5

Buffer G: 200 mM NaCl, 1 mM EDTA, 5% (v/v) Glycerol, 20 mM Tris pH 7.8

Protein concentration was carried out by centrifugation (900 g, 4°C) using Amicon Ultra-15 Centrifugal Filter Units (Millipore, Billerica, USA). Concentrations were determined by UV absorption at 280 nm and the corresponding extinction coefficient calculated by ProtParam as stated above.

5.5   SDS-PAGE

Protein samples were analyzed by polyacrylamide gel electrophoresis (SDS-PAGE) with 15% polyacrylamide gels (acrylamide-bisacrylamide (37.5:1) (Laemmli, 1970) using the vertical Mini-PROTEAN 3 System (Bio-Rad, Munich, Germany) and 1x TGS as running buffer. Before loading, samples were mixed with 4x loading dye and boiled at 95°C for 3 minutes. Gels were routinely stained with Coomassie staining solution and destained with water.

4x loading dye: 0.11 M Tris (pH 6.8), 16% (v/v) glycerol, 4% (w/v) SDS, 5% (v/v) β-mercaptoethanol, 0.05% (w/v) bromophenol blue

10x TGS buffer: 0.5 M Tris (pH 8.3), 1.9 M glycin, 1% (w/v) SDS

Coomassie stain: 50% (v/v) ethanol, 7% (v/v) acetic acid, 0.2% (w/v) Coomassie Brilliant Blue R250

5.6   Analytical Size Exclusion Chromatography

The molecular weight or the hydrodynamic radius, respectively, of protein complexes were estimated by analytical gelfiltration chromatography using a Superdex 200 26/60 gelfiltration column. Later, when available, a high resolution Superose 6 PC 3.2/30 column connected to an Ettan LC System was used (GE Healthcare, Freiburg, Germany). Prior to analysis the respective column was equilibrated with the corresponding protein buffer. For calculating the molecular sizes, the column was calibrated with Gel Filtration Standard (Bio-Rad, Munich, Germany). The stoichiometry

of the complexes and the molecular weight of single proteins were additionally estimated by SDS-PAGE.

5.7   Electrophoretic Mobility Shift Assay

DNA binding activities of PfR50NCM11 were tested using an electrophoretic mobility shift assay (EMSA), where complexes of protein and DNA migrate through a native polyacrylamide gel more slowly than unbound oligonucleotides.

Both DNA strands were purchased with a Cy3-label at the 5'-end. After denaturing at 90°C, annealing with equimolar amounts of each strand was performed by slowly cooling down the oligonucleotide mixture to 4°C. For ensuring the equimolarity and the highest possible proportion of double-stranded oligonucleotides, the annealed strands where loaded on a native 12% polyacrylamide gel. The gel was run at 130 Volts in pre-cooled 1x TAE buffer at 4°C to prevent melting of the double-strands. The double-stranded oligonucleotide bands were cut out of the gel and the gel pieces where put into small plastic bags and shock-frozen in liquid nitrogen. Following freezing, the gel pieces where manually broken up with mortar and pestle, thawed and shock-frozen again. Several cycles of this procedure ensured a high yield of purified oligonucleotides. The crushed gel pieces where resolved in 600 µl TE buffer by shaking over night at 4°C. The next day the suspension was centrifuged for 2 minutes at 11000 g and the DNA concentration of the supernatant was calculated by UV absorption at 260 nm and the corresponding extinction coefficient provided by the manufacturer.

TE buffer: 1 mM EDTA, 10 mM Tris pH 7.5

DNA with a final concentration of 50 nM was added to the reaction buffer, which is similar to the gel filtration buffer of the protein. Then protein was pipetted to the buffer/DNA mix and incubated for 1 hour at room temperature in the dark. Out of these 20 µl, 15 µl were loaded on a native 6% polyacrylamide gel. Gels were run in 0.5x TBE-Buffer for 2 hours at 120 V at 4°C in the dark. Reactions products were analyzed by fluorescence imaging using the Storm® gel and blot imaging system (GE Healthcare, Freiburg, Germany). The lowest visible band corresponds to single-stranded

DNA, the band above to double-stranded DNA and the upper band to the shifted protein-bound oligonucleotide.

10x TBE buffer: 0.02 M EDTA, 0.89 M boric acid, 0.89 M Tris pH 7.5

Oligonucleotide 1 ("16fCy3"):   5' Cy3-AAA AGC AAA TTG CCT T

Oligonucleotide 2 ("12rCy3"):   5' Cy3-GGC AAT TTG CTT

5.8   Anisotropy

The measurement of polarized emission of fluorescence allows the observation of rotational motions in fluorescence during the lifetime of the excited state. Because the rotation of macromolecules depends among other factors on their size and shape, several kinds of information may be extracted. Here polarized-emission measurements were used to detect changes in molecular size based on binding of protein to a fluorescently labeled DNA oligonucleotide.

The first step in these measurements is the excitation of fluorophores. Vertically polarized light is used to excite a population of molecules whose absorption dipole is oriented in the vertical direction. It is generated using a polarizer in the excitation path. The second step is molecular rotation. The molecule, once excited, may rotate during the lifetime of the excited state and this rotation will depolarize the fluorescence emission. Measurement of the polarized emission components allows calculation of the type and extent of rotational motions of the molecule. The third step is measurement of emission. The polarized components of fluorescence emission are measured using a polarizer in the emission path. Measurements of anisotropy are derived from the intensities of the vertically and horizontally polarized components of the fluorescence emission. The last step is calculation, where the extent and type of rotational behavior can be determined from the magnitude of the vertically and horizontally polarized emission components. Anisotropy is a ratio defined as the linearly polarized component's intensity divided by the total light intensity.

Anisotropy experiments were carried out at 22°C using the FluoroMax-P fluorescence spectrometer (HORIBA Jobin Yvon, Edison, NJ, USA). Double-stranded oligonucleotides diluted in 700 µl gel filtration buffer used in protein purification to a final concentration of 50 nM were pipetted into a quarz cuvette. Subsequently, 15 µM protein solution was titrated in the cuvette and after each addition of protein the solution was stirred three times for three seconds using a magnetic stirrer. After 2 minutes and 30 seconds anisotropy was measured three times directly after another and the mean value was used to determine the binding affinity. Exposure was 1 second and for each data point the sample was measured at least 10 times or until the standard deviation did not exceed 2%. Before each experiment an emission and excitation scan were performed to obtain the peak maxima, at which the sample was excited or where emission was measured, respectively. Oligonucleotides were the same as for the EMSA.

5.9   Crystallization

Protein crystallization requires large amounts of pure and homogeneous protein. The protein solution slowly gets concentrated to a supersaturated state and during this process crystals may grow. Several crystallization techniques, including microdialysis, batch crystallization, hanging drop or sitting drop vapor diffusion are available, though vapor diffusion has become the most popular one. For solving the SbcCD structure, crystals were grown by sitting drop vapor diffusion (figure 12). Here, crystallization reagents are given into the reservoir of a crystallization plate. Reservoir solutions typically consist of buffer solution, salt and precipitant. A small drop of concentrated protein sample mixed with reservoir solution is set on a platform in vapor equilibration with the reservoir. Since the reagent concentration in the drop is lower than in the reservoir, water vapor leaves the drop. Thereby, the protein solution is slowly concentrated to a supersaturated state. In many cases, the protein will form aggregates and precipitate out of solution. Under certain conditions, stable nuclei may form and crystals growth takes place.
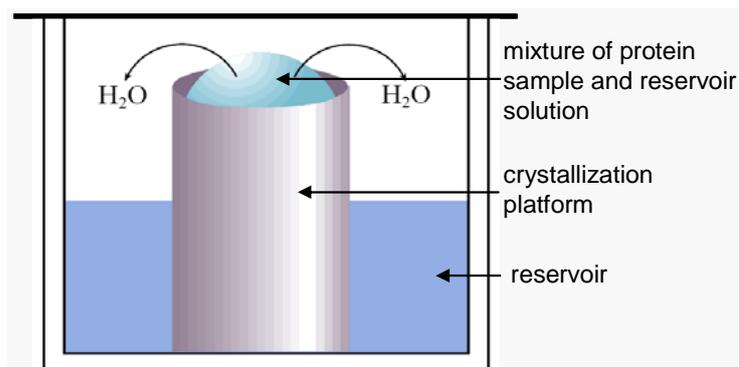
Figure 12: Principle of protein crystallization by sitting drop vapor diffusion.
Principle of protein crystallization by sitting-drop vapor diffusion (adopted from http://www.hamptonresearch.com).

First screening for crystallization conditions was performed in CrystalEX 96 well sitting drop plates (Corning, New York, USA) using commercially available sparse matrix screens from Hampton, Jena Bioscience and Nextal. Setups were pipetted either by a Hydra II semi-automatic protein crystallization robot (0.2 µl + 0.2 µl) (Matrix Technologies Corporation, Hudson, USA) or by hand mixing of 1 µl protein solution with 1 µl crystallization buffer. Additionally it was tried to co-crystallize the protein complex together with DNA that was tested positively in advance by EMSA regarding its binding abilitiy to the protein complex. To improve crystal quality the conditions of initial hits were refined by modifying concentrations of the protein and other components of the crystallization buffer, temperature, pH, drop size, mixing ratio of the drop and ratio between drop and reservoir volume. Refinements were set up in 24 well plates either in sitting drop Cryschem plates (Hampton Research, Aliso Viejo, USA) or hanging drop EasyXtal Tool plates (Qiagen, Hilden, Germany).

5.10   X-ray Data Collection and Structure Determination

Prior to data collection, crystals were cryo-protected, mounted in nylon loops and flash frozen in liquid nitrogen. Redundant single-wavelength anomalous dispersion (SAD) experiments were carried out at theoretical selenomethionine peak wavelength at beamline PX at the Swiss Light Source (Villigen, Switzerland) with selenomethionine-

containing crystals. Data from 360 images (1° rotation) were integrated and scaled with XDS and XSCALE (Kabsch, 1993). Using the program package autoSHARP (Bricogne *et al.*, 2003) all 20 selenomethionine sites have been located to calculate the phases. Refinement of the native data were performed with CNS (Brunger *et al.*, 1998) and included solvent flattening, grouped B-factor refinement, bulk solvent corrections, simulated annealing and positional refinement. The quality of the refinement was controlled by calculating the R_{work} and R_{free} of the model. The R_{free} was calculated for 5% of the reflections omitted from the refinement.

5.11   SAXS Sample Preparation, Data Collection and Structure Determination

As well as for crystallization sample preparation for SAXS measurements is crucial. The protein stocks stored at -80°C were gel filtrated after thawing with a Superdex 200 column prior further concentration to remove possible protein aggregrates. Proteins then were concentrated by centrifugation (850 g, 4°C) using Amicon Ultra-15 Centrifugal Filter Units (Millipore, Billerica, USA). Concentrations of the protein solutions were determined by UV absorption at 280 nm and the corresponding extinction coefficient derived from primary sequence analysis using ProtParam (Gasteiger *et al.*, 2003). The flow-trough of the concentrator filters was used as the buffer value for the latter SAXS experiments. At the beginning and at the end of each experiment, the buffer was used to calculate the blank value, since the buffer also significantly scatters X-rays and a buffer blank is used to subtract scatter from the buffer solution as well as eliminate any systematic errors in the data. Following to the first buffer measurement, a protein sample was measured twice to calculate possible radiation damage, even though 2 mM DTT was added to the protein buffer during the preparative gel filtration run to reduce radiation damage. When it was assured that the protein was not affected visibly by the X-rays, different concentrations of the protein were measured to make sure that proteins remain monodisperse at high protein concentrations during data collection. Later, the highest quality data measurement was used for structure determination. Each measurement consumed about 100 µl of highly pure protein with concentrations ranging

from 3.5 to 21.8 mg/ml (corresponding to 41 – 257 µM). In case of measurements containing poorly-hydrolyzable ATP-analogons, the analogons were added before the concentration steps to a final concentration of 1 mM in addition to 5 mM of magnesium chloride.

The equipment used for measuring SAXS data is very similar to standard protein crystallography. The predominant difference is that the beam stop and detector are often a meter or more away from the sample in order to spread the SAXS signal over a larger area for detection and for separation of the signal from the primary beam. SAXS experiments performed in this thesis were carried out either at the EMBL X33 beamline at the DORIS storage ring (EMBL, Hamburg, Germany) or at the SIBYLS beamline at the Advanced Light Source at the Lawrence Berkeley National Laboratory (Berkeley, California, USA). Energy sources for the X33 and SIBYLS synchrotron radiation are bending magnets. At the X33 beamline the sample is separated from the detector with a 2.4 m flight tube under vacuum to minimize air scatter. The two dimensional detector is a Mar345 image plate with online readout and the X-rays used had a wavelength of 1.5 Å, covering the range of momentum transfer $0.012 < s < 0.45$ Å^{-1}. The wavelength was adjusted with a horizontal focusing triangular Si (111) crystal monochromator. Exposure time of each measurement was 10 seconds. In case of the SIBYLS beamline a dual double crystal monochromator adjusted the wavelength. It consists of double Si (111) crystals as the first set and double multilayer Mo/B$_4$C crystals as the second set of crystals. The detector was a MAR 165 image plate.

After measuring, all data manipulations were performed with PRIMUS (Konarev *et al.*, 2003) included in the ATSAS package (Konarev *et al.*, 2006). The forward scattering, $I(0)$, and the radius of gyration, R_g, were computed by using the Guinier approximation (Guinier, 1939), assuming that at very small angles ($s * R_g < 1.3$) the intensity is represented as $I(s) = I(0) \exp(-(sR_g)^2/3)$. These parameters were also calculated from the entire scattering patterns by using the indirect Fourier transform method (Moore, 1980) as implemented in GNOM (Svergun, 1992), which also provides the distance distribution function, $p(r)$, of the particle.

For comparing the X-ray with the SAXS model and to evaluate SAXS data quality, the scattering from the atomic coordinates of the crystal structure was calculated with the

program CRYSOL (Svergun *et al.*, 1995) and compared with the measured SAXS data. To create the SAXS low resolution envelope the *ab initio* algorithm GASBOR was utilized, comparing calculated scattering profiles from thousands of configurations of a chain of dummy residues to the measured scattering curve (Svergun *et al.*, 2001). GASBOR was run 10 times and the results were averaged for a final shape using DAMAVER (Volkov and Svergun, 2003), which superimposed all models and determined the common envelope containing all models. In general the accuracy of the final model varies from protein to protein depending on its shape, size and if symmetry is enforced. In this thesis, based on the determined X-ray structure, a two-fold symmetry was given to calculate a more reliable model, though one-fold symmetry based models were computed as well (data not shown). Finally, DAMAVER was employed to align X-ray and SAXS based models. DAMAVER is based on the program SUPCOMB, which aligns two arbitrary models by minimizing a normalized spatial discrepancy (NSD) (Kozin and Svergun, 2001). All pairs of models were compared and the model with the lowest mean value of NSD was chosen as the reference. Models with the mean NSD value exceeding twice the dispersion of NSD were considered as outliers and discarded (Volkov and Svergun, 2003).

5.12   Crystal Staining

Protein:DNA complex crystals were SYBR Gold (Invitrogen, Karlsruhe, Germany) stained to ensure DNA incorporation. SYBR Gold stain is a proprietary, unsymmetrical cyanine dye that binds to RNA or to single- or double-stranded DNA and exhibits two fluorescence excitation maxima when bound to DNA, one centered at ~300 nm and one at ~495 nm. SYBR Gold staining was chosen due to its high fluorescence quantum yield of the dye-nucleic acid complexes (~0.7) and the dye's large fluorescence enhancement upon binding to nucleic acids (~1000-fold) (Tuma *et al.*, 1999). The staining solution was prepared by adding SYBR Gold to an aliquot of the reservoir solution to a final concentration of 5x. 1 µl of this staining solution was added to the well containing the crystals. After incubation for three hours the crystals were washed three times in reservoir solution lacking the dye. A standard transilluminator microscope with the

appropriate filter was used to examine the staining that was documented with a separate manual camera. Crystals of similar size consisting of protein but lacking DNA were used as negative controls.

Dr. Katja Lammens contributed to the crystal structure determination and the *in vivo* studies in yeast were performed by Emanuel Clausing (data not shown). The crystal structure refinement and the yeast studies have not completely been finished at the point of submission of this thesis.

6   Results

6.1   Cloning, Expression and Purification

6.1.1   *Thermotoga maritima* SbcCD

The gene encoding full-length SbcD as well as the construct encoding for truncated SbcC were amplified from *Thermotoga maritima* MSB8 genomic DNA and cloned bicistronically into the pET-29b vector as described in 5.1. The length of the coiled-coils were designed based on the known *Pyrococcus furiosus* Rad50cd structure (Hopfner *et al.*, 2001; Hopfner *et al.*, 2000b) and coiled-coil prediction using the COILS program (Lupas *et al.*, 1991) (figure 13). The coiled-coils could not be truncated totally since biochemical studies indicated Mre11 binding to this flexible region (Hopfner *et al.*, 2001).
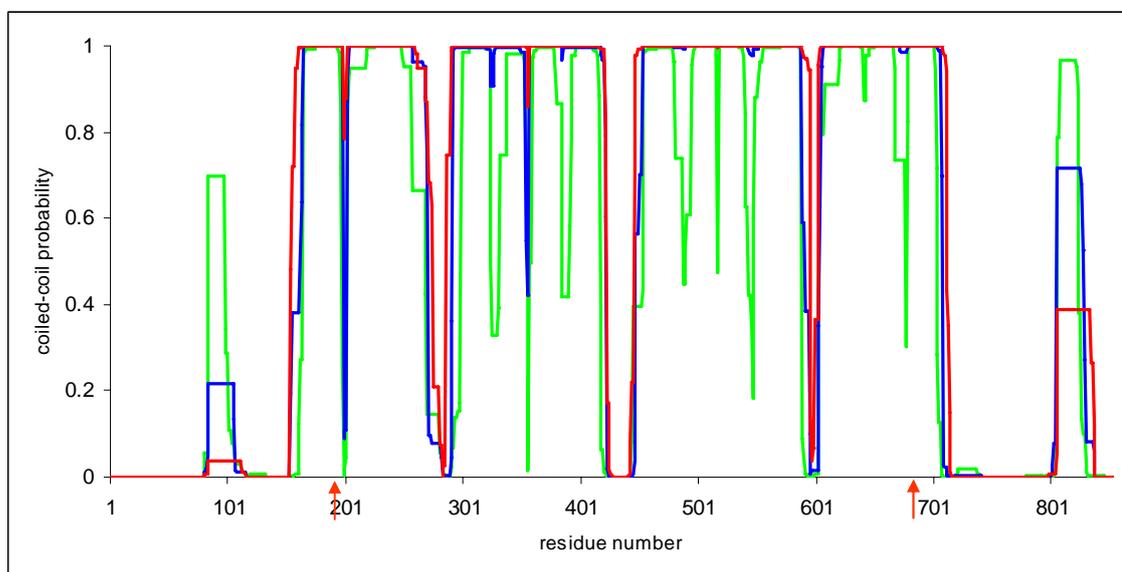
Figure 13: Coiled-coil prediction for full-length *Thermotoga maritima* SbcC.
Alpha-helical coiled-coil probability for full-length SbcC calculated with COILS assuming a window of 14 (green), 21 (blue) and 28 (red) residues. Red arrows indicate the ends of the 2 peptide chains of SbcC.

E. coli Rosetta (DE3) cells were transformed with the resulting expression plasmid and the SbcCD construct was co-expressed. The heterotetrameric complex has an overall size of 170 kDa, consisting of two truncated SbcC (41 kDa) and two SbcD (44 kDa) molecules.

SbcCD was purified as specified in 5.4. After heat denaturation, the proteins were loaded onto a Phenyl HP column. The protein was further purified using a Resource Q anion exchange column and by size exclusion chromatography with a HiLoad 26/60 Superdex 200 column (figure 14). The peak representing the purified protein complex was quite broad, but the mean elution volume was ~173 ml. This elution volume corresponds to a molecular weight of about 155 kDa, being in accordance with a heterotetrameric SbcC$_2$SbcD$_2$ complex.
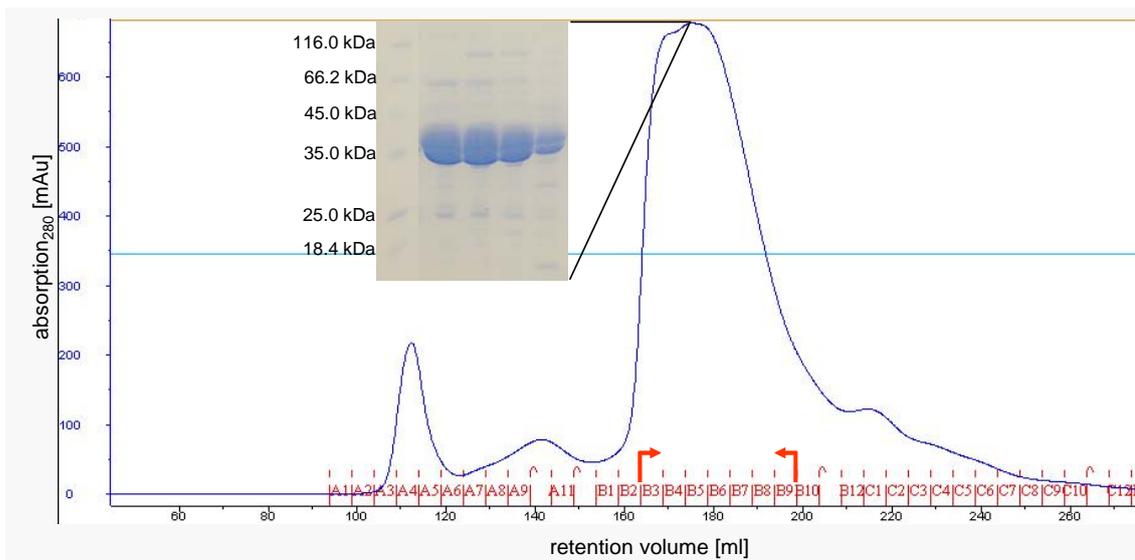
Figure 14: Elution profile of SbcCD from the Superdex 200 size exclusion column.
The major peak corresponds to purified SbcCD. SDS-PAGE analysis of the peak fractions is shown as inlet.

For more detailed information an additional analytical size exclusion run was performed on a high resolution Superose 6 PC 3.2/30 column connected to an Ettan LC system (figure 15). Based on these results, the Stoke's radius of SbcCD corresponded to a globular protein with a molecular weight of roughly 138 kDa. Since SbcCD is an extended and potentially very flexible molecule, as becomes evident in sections 6.4 and 6.5, and not a spherical one, the calculated Stokes radius might be smaller than the

effective radius. Thus, the shape and the expected flexibility of SbcCD might influence gel filtration analysis, though a smaller retention volume was expected. Two different species of conformers that might have caused different retention volumes were not detected, indicated by a single sharp peak. The R squared value of the protein standard trend line of 0.97 indicated a possible deviation at calculating the molecular size. However, bovine gamma-globulin with a molecular weight of 158 kDa, which is included in the gel filtration standard solution, eluted at 1.63 ml compared to 1.59 ml for SbcCD. This supported the S200 gel filtration result, suggesting a heterotetrameric complex.
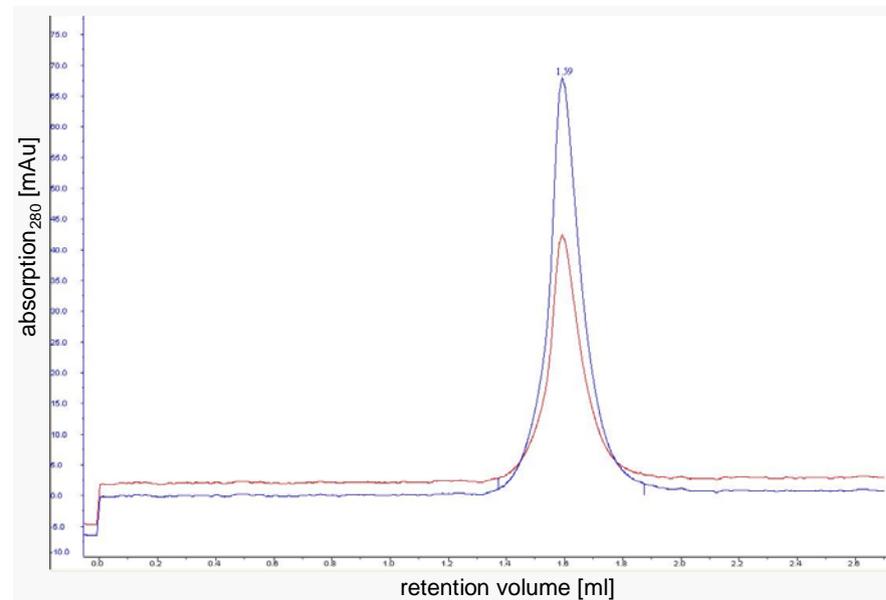
Figure 15: Analytical size exclusion chromatography of SbcCD.
The elution profile of SbcCD on a high resolution Superose 6 PC 3.2/30 column is shown. The retention volume is plottet against the relative absorbance units at 280 nm. Protein (blue) and DNA absorbance (red) was measured at 280 nm and 260 nm, respectively. The protein eluted at 1.59 ml.

6.1.2   *Pyrococcus furiosus* Rad50NCMre11

Pyrococcus furiosus R50NCM11 was cloned and expressed like SbcCD as described in 5.1 and 6.1.1. PfRad50 has a theoretical molecular weight of 41 kDa and Mre11 of 49 kDa, resulting in a 91 kDa heterodimeric or 182 kDa heterotetrameric complex. As for

the *Thermotoga maritima* construct, the PfR50NC design was based on structural information and biochemical indications of PfRad50cd and Rad50cd-Mre11 interactions (Hopfner *et al.*, 2001; Hopfner *et al.*, 2000b) as well as on coiled-coil predictions (figure 16).
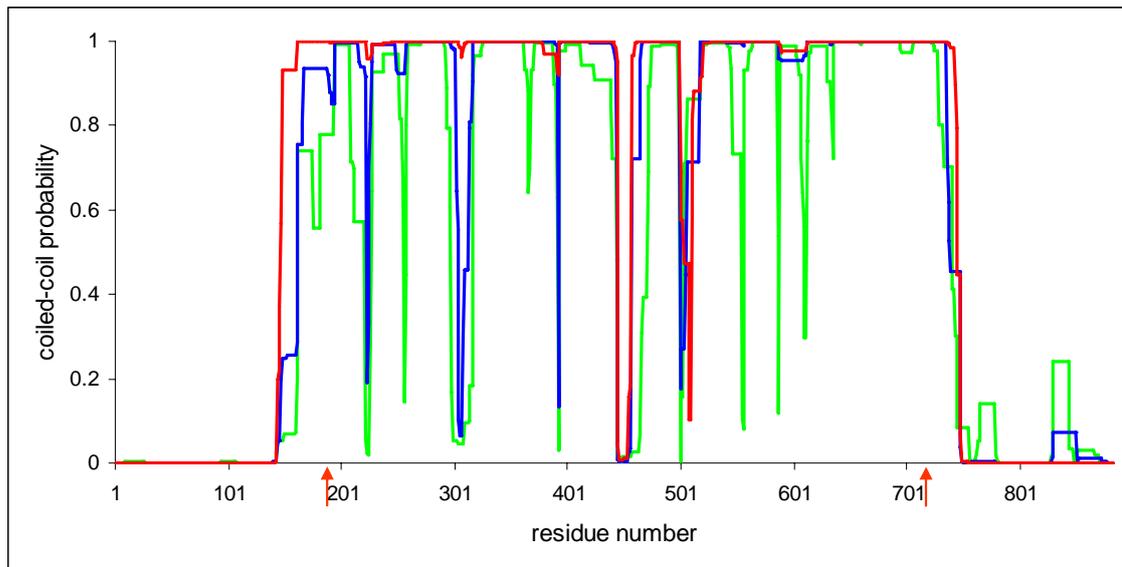
Figure 16: Coiled-coil prediction for full-length *Pyrococcus furiosus* Rad50.
The plot shows the probability of a given amino acid residue being present in an alpha-helical coiled-coil domain determined by COILS assuming a window of 14 (green), 21 (blue) and 28 (red) residues. Red arrows indicate the termini of the 2 peptide chains of R50NC.

After heat denaturation and removal of nucleic acids using ammonium sulfate precipitation, high salt gel filtration and hydrophobic interaction chromatography, endogenously co-expressed *E. coli* proteins where separated using a HiTrap S HP cation exchanger. As a final step, a gel filtration was performed using a Superdex 200 26/60 column (figure 17). PfR50NCM11 eluted after ~161 ml corresponding to the Stoke's radius of a globular protein with a molecular weight of about 230 kDa compared to the theoretical size of 181 kDa for the heterotetrameric complex.
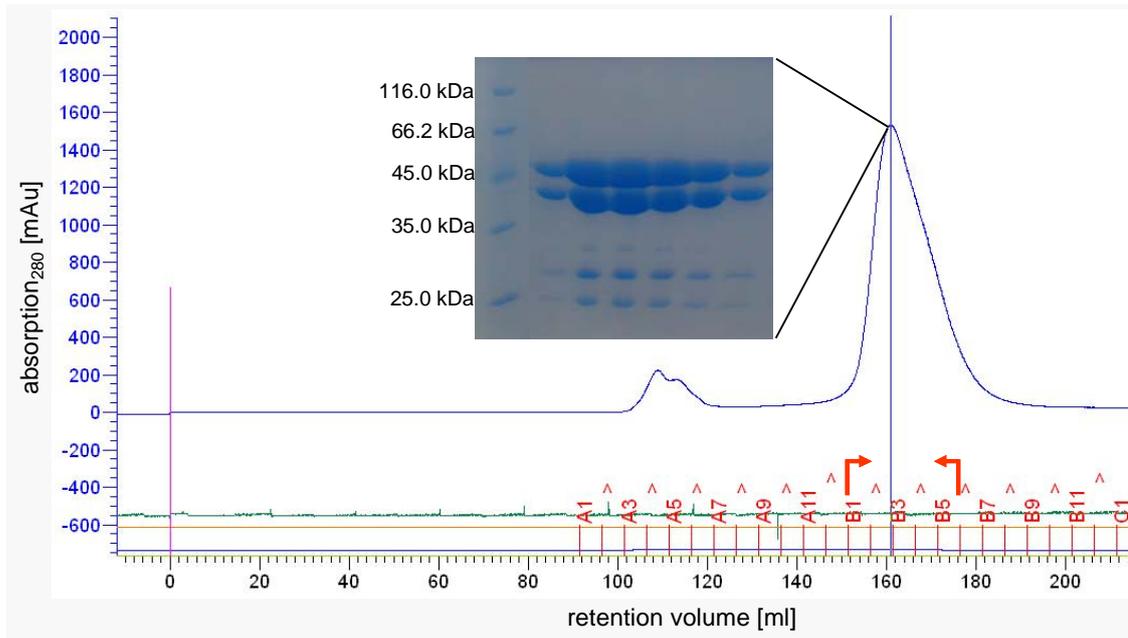
Figure 17: Elution profile of PfR50NCM11 from the Superdex 200 size exclusion column.
The major peak corresponds to purified PfR50NCM11. SDS-PAGE analysis of the peak fractions
is shown as inlet.

6.2   Biochemical Studies

6.2.1   Electrophoretic Mobility Shift Assay

To characterize DNA binding abilities of Rad50/Mre11 and to identify suitable DNA
oligomers for protein:DNA co-crystallization, an electrophoretic mobility shift assay
(EMSA) was performed using double-stranded Cy3-labeled oligonucleotides and
PfR50NCM11. A 50 nM DNA solution was incubated with different amounts of protein
for 1 hour at room temperature, followed by separation of protein:DNA complexes,
dsDNA and ssDNA through running a native polyacrylamide gel. The EMSA clearly
shows PfR50NCM11 binding to DNA. In regard to the quality of DNA binding of
PfR50NCM11 its preference to dsDNA could be observed since the dsDNA band
intensity diminishes with increasing protein concentration which is not the case for the
ssDNA band intensity (figure 18).
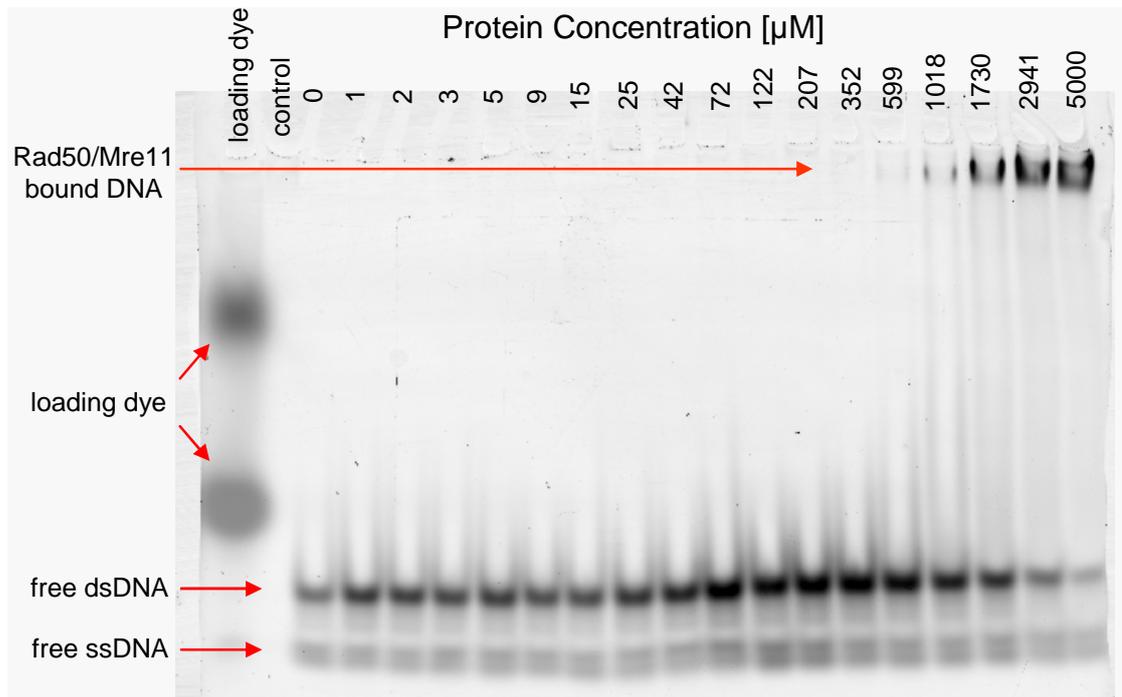
Figure 18: Electrophoretic mobility shift assay.
Binding of PfR50NCM11 to dsDNA can be observed from the upper most band. Two distinct bands of free DNA appear as indicated by the arrows.

6.2.2   Anisotropy

For more detailed quantification of the DNA binding abilities of PfR50NCM11 anisotropy experiments have been carried out. Changes of the rotational behaviour of the fluorescently labeled double-stranded oligonucleotides upon titration of protein have been measured (for theoretical and practical details see 5.8). Anisotropy measurements confirm the EMSA results and quantify the binding constant K_D to 859 nM (±94 nM) indicating a relative weak protein:DNA interaction (figure 19). DNA oligos showing clear binding abilities to R50NCM11 were chosen for protein:DNA co-crystallization.
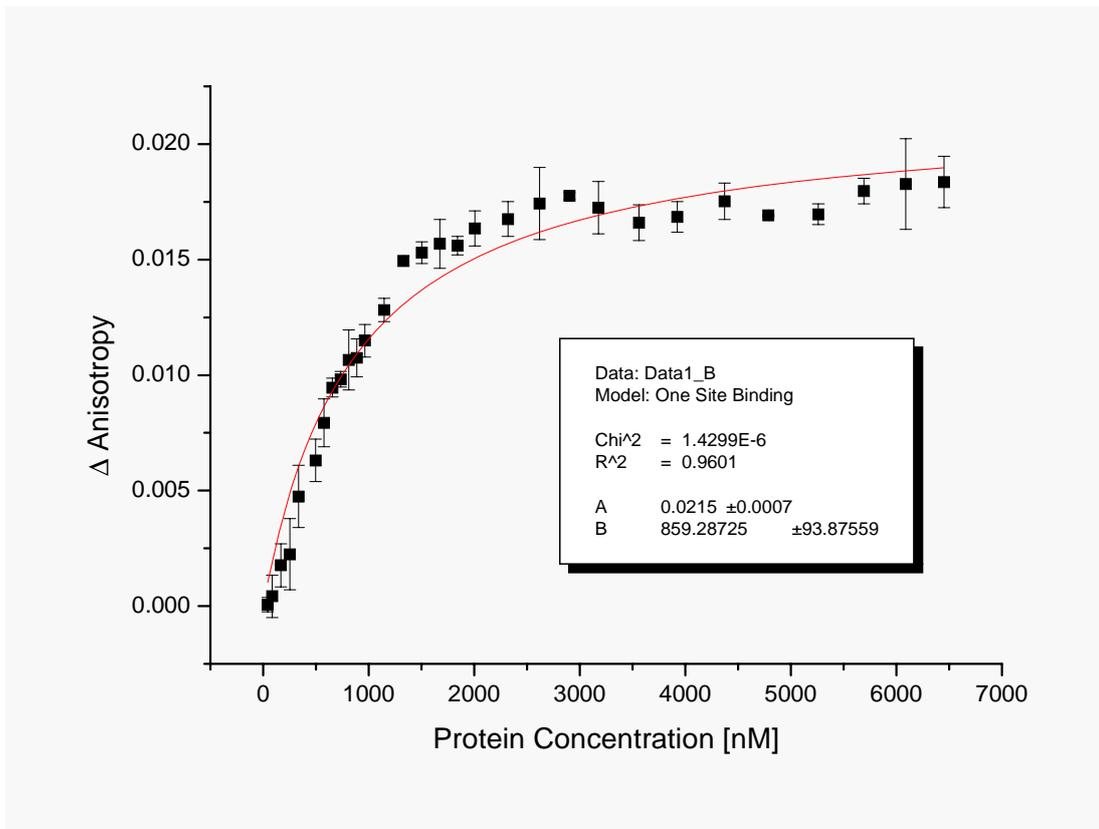
Figure 19: Anisotropy measurements determine PfR50NCM11 binding to DNA.
Anisotropy measurements quantify PfR50NCM11 binding abilities to double-stranded oligonucleotides. The binding constant K_D is 859 nM (± 94 nM), indicating the concentration where 50% of the DNA is protein bound. Each data point results from three measurements. Error bars indicate the standard deviation. The curve was fitted based on a one site binding model applying the equation $y = (B_{max} * x)/K_D + x)$, where y is the total binding, B_{max} the maximum specific binding to be fit, x the protein concentration and K_D the equilibrium association constant.

6.3   X-ray Crystallography

6.3.1   Crystallization

For initial crystal setups with commercial screens, the Hydra II semi-automatic protein crystallization robot (Matrix Technologies Apogent Discoveries, Hudson, USA) was used to spot 0.2 + 0.2 µl drops. Small, plate shaped crystals grew in the crystallization condition Nextal Classic H9 (10% w/v PEG 6000, 5% v/v MPD, 0.1M HEPES pH 7.5, protein:DNA = 1:1.2, 9.6 mg protein/ml) (figure 20A). In order to reproduce and enlarge the obtained crystals, initial crystallization conditions were refined manually. 2

µl + 2 µl drops were pipetted into the same wells. The reservoir solution composition and volume as well as the protein concentration were varied. Crystallization setups with and without DNA were performed but resulted in no difference in crystal shape indicating that DNA was not incorporated (DNA binding data and crystals not shown here).

Thermotoga SbcCD was crystallized by sitting drop vapour diffusion using 24-well Cryschem plates (Hampton Research, Aliso Viejo, USA) (figure 20B).

After trying overall more than 65000 different crystallization conditions, including various source organisms, protein constructs and additives, diffracting crystals were obtained with 400 µl of reservoir solution containing 9% (w/v) PEG-6000, 5% (v/v) MPD, 1 mM TCEP and 0.1 M HEPES pH 7.9. After mixing 2 µl of protein solution (9.6 mg/ml in 20 mM TRIS/HCl pH 7.8, 200 mM NaCl, 0.1 mM EDTA, 5% Glycerol, 2 mM DTT) with 2 µl of reservoir solution, crystals were grown for three days at 18°C leading to a maximum size of about 90 µm * 90 µm * 25 µm. These crystals already consisted of selenomethionine containing proteins.

As a cryoprotectant 2,3-butanediol was added to a final concentration of 15% (v/v) to the reservoir solution. After a few seconds of incubation in this solution the crystals were flash-frozen in liquid nitrogen.
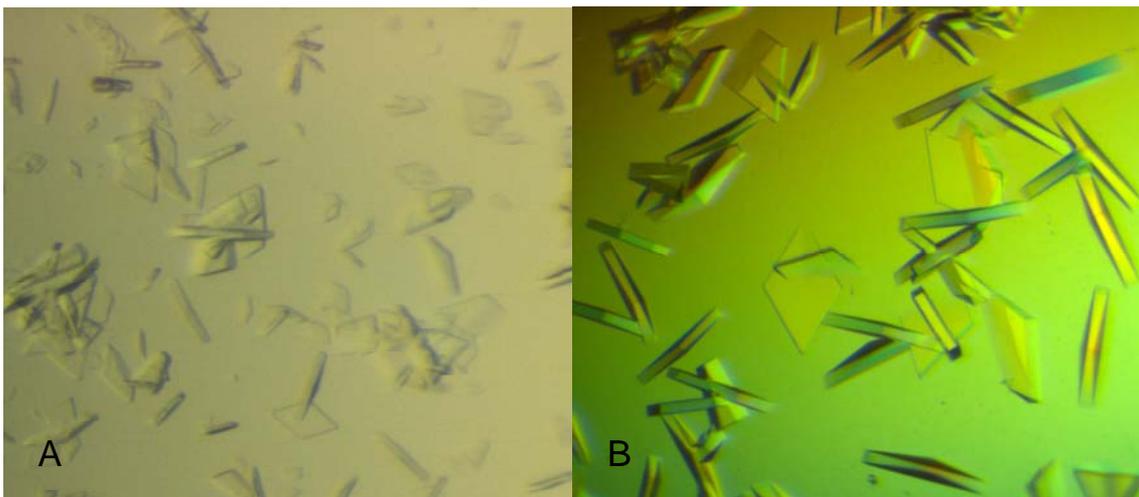
Figure 20: Photographs of the *T. maritima* SbcCD construct crystals.
A) Initial crystals of SbcCD were obtained from Nextal Classic H9 (10% w/v PEG 6000, 5% v/v MPD, 0.1M HEPES pH 7.5, protein:DNA = 1:1.2, 9.6 mg protein/ml). B) Diffracting crystals with a maximum size of about 90 µm * 90 µm * 25 µm were generated in the following condition: 9% (w/v) PEG-6000, 5% (v/v) MPD, 1 mM TCEP and 0.1 M HEPES pH 7.9.

6.3.2   Data Collection

Diffraction data of the selenomethionine containing *T. maritima* SbcCD crystals were collected at PX beamline (SLS, Villigen, Switzerland) with a mar225 mosaic CCD detector at 100 K. 360 frames with 1° oscillation were recorded for the data set. Phase determination was performed by a single wavelength anomalous dispersion (SAD) experiment at the selenium K edge. For high resolution SAD datasets the peak wavelength needed to be chosen, determined experimentally by a fluorescence scan for each selenomethionine containing crystal prior to data collection (figure 21). Data for the peak wavelength at 0.9774 Å (12.6852 keV, f´-7.21, f´´ 5.34) were collected to 3.4 Å (figure 22 and appendix VI).
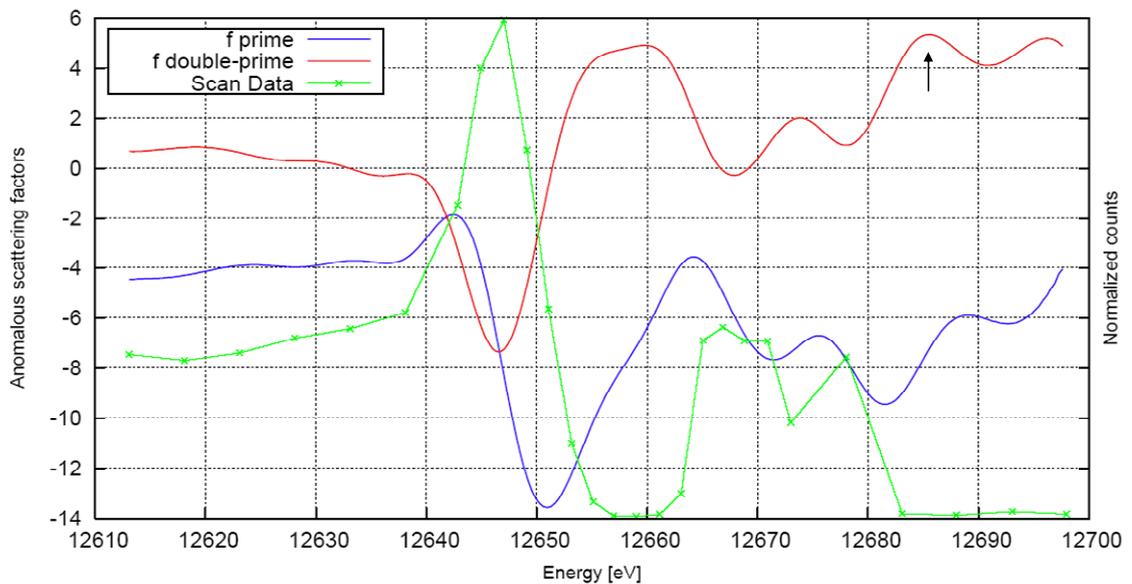
Figure 21: Fluorescence scan for selenomethionine K(1s) edge.
A fluorescence scan on selenomethionine containing *T. maritima* SbcCD crystal around the selenium K-edge (12.6578 keV) was performed. Values of f', f'' and the common scan data are plotted against the X-ray energy (eV). The peak wavelength used in the SAD experiment is indicated by an arrow.
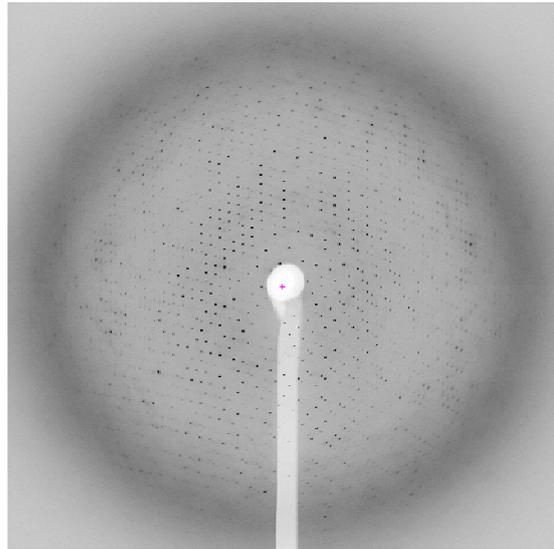
Figure 22: Diffraction pattern of the *T. maritima* SbcCD construct.
The crystal diffracted to a resolution of 3.4 Å at the PX beamline (SLS, Villigen, Switzerland).

6.3.3   Structure Determination and Refinement

All data were processed with XDS and XSCALE (Kabsch, 1993). Data were indexed and scaled in I222 and one heterotetramer ($SbcC_2SbcD_2$) was found in the asymmetric unit. The number of molecules per asymmetric unit was determined by the Matthews coefficient using the unit cell parameters (105 Å * 187 Å * 300 Å), the twofold symmetry and the molecular weight of the *T. maritima* SbcCD construct (85 kDa per heterodimer). The solvent content was about 70%. Theoretically the asymmetric unit should contain 20 selenomethionines and all of these anomalous scatterers could be located using SHARP (Bricogne *et al.*, 2003). The SHARP calculated phases were improved by solvent flattening running SOLOMON (CCP4, 1994) and lead to an interpretable electron density represented in figure 23.
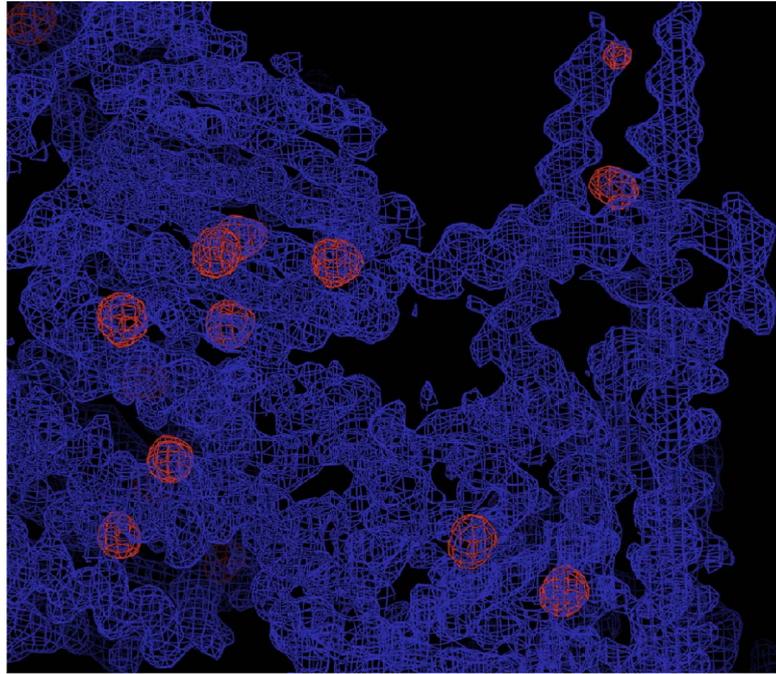
Figure 23: Single anomalous dispersion map of SbcCD of *Thermotoga maritima*.
Single anomalous dispersion (SAD) map of SbcCD at 3.4 Å resolution (blue mesh) with superimposed anomalous difference fourier map (red mesh). Using autoSHARP, all 20 selenomethionine sites have been located. In the upper right section the two produding coiled-coils of SbcC can be recognized as well as the interacting C-terminal part of SbcD.

Phases to 3.4 Å required manual model building using MAIN (Turk, 1992) and Coot (Emsley and Cowtan, 2004). The presence of the selenomethionine molecules could be used as sequence markers and helped in the manual model building process. The model was additionally fitted by rigid body refinement and improved by grouped B-factor and energy minimization refinement using CNS (Brunger *et al.,* 1998). The final model of SbcCD contains residues Met1-Glu187 and Lys689-Val850 of SbcC and Lys8-Gly92, His94-Pro140, Lys148-Glu184, Arg196-Arg336 and Glu344-Ala385 of SbcD and was refined to a final R factor of 26.3 (R_{free} 30.1). More details of the refined model are listed in table 6 at appendix VI. The final X-ray crystallography structure is depicted and analyzed in chapter 6.5.1.

6.3.4   Crystallization and Data Collection of SbcC in Complex with DNA

In addition to solving the crystal structure of SbcCD it was tried to crystallize the same complex in association with DNA. Initial screening was performed using commercial screens as described above but suitable crystal formation of the DNA-bound complex could not be observed, neither for SbcCD nor SbcD of *T. maritima*. In case of SbcC associated with the C-terminal domain III of SbcD crystals could be generated by vapour diffusion using 2 μl + 2 μl drops in 24-well plates (figure 24). This protein construct was complexed with DNA oligonucleotides that were assayed before and showed association with Rad50Mre11/SbcCD (see chapters 5.7 and 5.8). The crystals grew at 18°C in a solution containing 0.1 M magnesium acetate, 5% PEG 8000, 1 mM TCEP and 0.1 M sodium acetate pH 5.4 with a ratio of protein:DNA = 1:1.5 and a final protein concentration of 10.0 μg/μl. Since high solvent contents for the overnight growing previous crystals of the same construct were observed shrinking of the large crystals (150 μm * 150 μm * 650 μm) was attempted by adding PEG400 to a final concentration of 20% (v/v) to the reservoir solution 12 hours prior to harvesting. These crystals were generated using protein expressed by *E. coli* Rosetta cells grown in selenomethionine containing media. As cryoprotectants 2,3-butanediol or MPD were added to final concentrations of 10-15% (v/v) to the reservoir solution. After a few seconds of incubation in this solution the crystals were flash-frozen in liquid nitrogen. The DNA-bound state of the proteins in the crystal was assayed by SYBR Gold (Invitrogen, Karlsruhe, Germany) nucleic acid staining with a separate portion of crystals (figure 24).

Diffraction data of 120 degrees with 1 degree oscillation were collected at the theoretical high remote wavelength at 0.9801 Å the PX beamline (SLS, Villigen, Switzerland). The crystals diffracted to a limiting resolution of 3.9 Å (figure 24 and table 7 at appendix VII) but generated only very weak anomalous signals, which was not sufficient for phase and structure determination in this case.
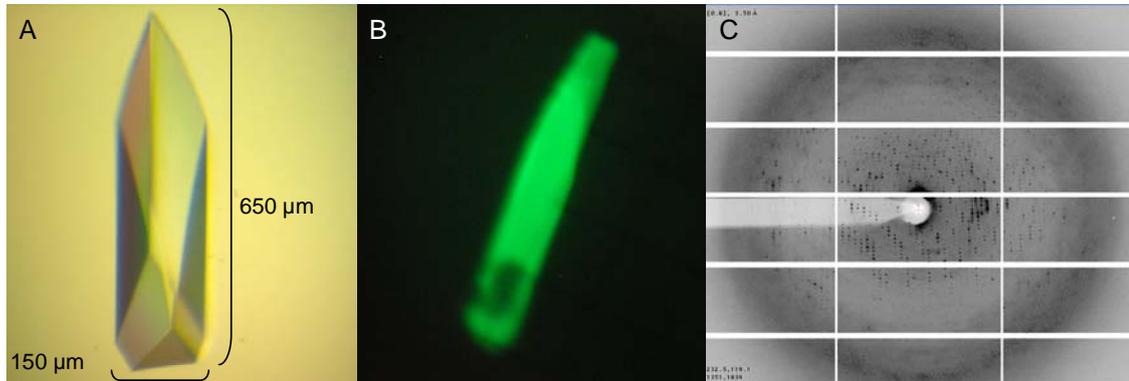
Figure 24: Crystallization of *T. maritima* SbcC in complex with DNA.
DNA oligonucleotides assayed positively in Rad50 binding by anisotropy (for *P. furiosus*) and EMSA (for *P. furiosus* and *T. maritima* (data not shown)) were used for crystallization. A) The crystals grew in 0.1 M magnesium acetate, 5% PEG 8000, 1 mM TCEP and 0.1 M sodium acetate pH 5.4 with a ratio of protein:DNA = 1:1.5 and a final protein concentration of 10.0 µg/µl. B) SYBR Gold staining was performed to verify the DNA-bound state of SbcC and the stained crystal was photographed using a standard transilluminator with the appropriate photographic filter. C) The protein:DNA crystals diffracted to a limiting resolution of 3.9 Å, in this case insufficient for structure determination.

6.4   SAXS

6.4.1   Data Collection and Structure Determination of the Apo SbcCD Complex

Small angle X-ray scattering (SAXS) was performed to investigate the structure of the *Thermotoga maritima* SbcCD complex construct in solution at concentrations ranging from 3.5 mg/ml to 21.8 mg/ml. The best data set was recorded at a concentration of 21.8 mg/ml yielding strong signal and no aggregation. The scattering pattern of SbcCD calculated using the ATSAS program package and the CRYSOL computed theoretical scattering pattern based on the atomic X-ray crystallography structure are presented in figure 25. Comparison of these two curves indicate a very high similarity between the crystal and the in solution structure.
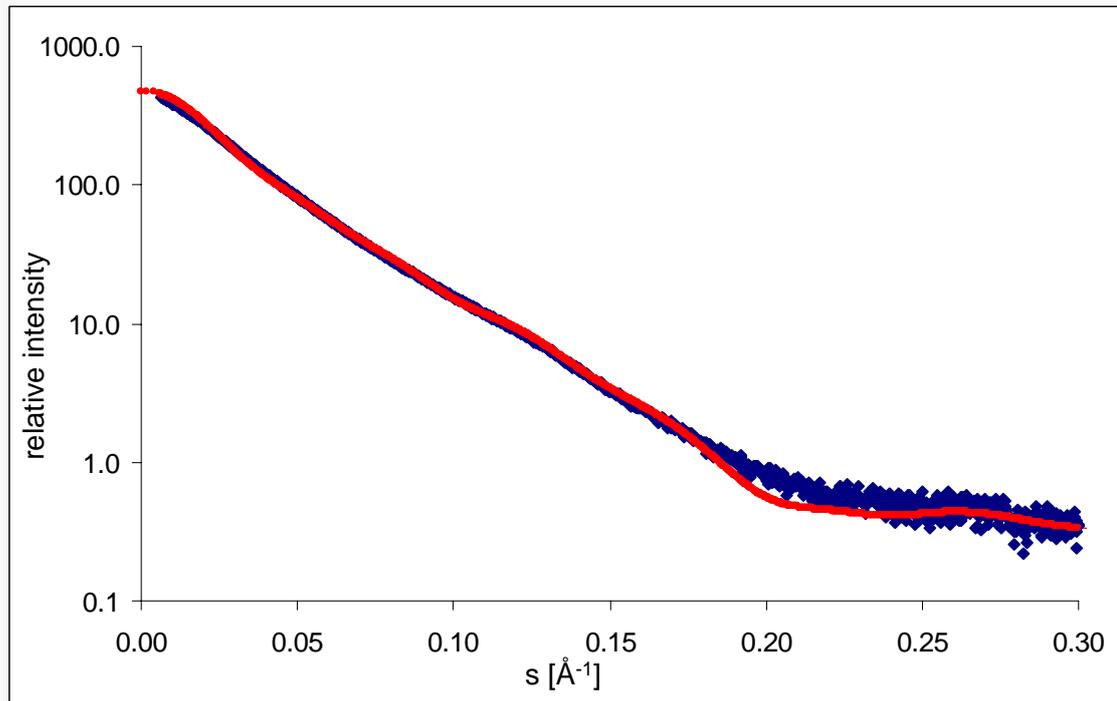
Figure 25: Buffer-subtracted scattering patterns of the SbcCD construct of *T. maritima*.
Experimental X-ray scattering pattern (blue) is compared with the computed theoretical scattering pattern based on the atomic X-ray crystallography structure (red) demonstrating a very good fit. The plot displays the scattering intensity as a function of momentum transfer $s = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle between the incident and scattered radiation and λ is the X-ray wavelength.

The scattering data indicates a radius of gyration (R_G) of about 59 Å calculated by GNOM with a given maximum dimension of the protein (D_{max}) of ~210 Å based on curve analysis of the p(r) plot. The radius of gyration is the square root of the average squared distance of each scatterer from the center of the particle that is scattering the X-rays. In general R_G can be described as the radius of a hypothetical hard sphere that diffuses as the same rate as the molecule of interest similar to the hydrodynamic radius relevant in size-exclusion chromatography (Putnam *et al.*, 2007). The distance distribution function p(r) was calculated using GNOM as well and provides direct information about the distances between electrons in the scattering particles in the sample (for details see sections 2.3.2 and 5.11). The shape of the curve already suggests an elongated molecule structure (figure 26).
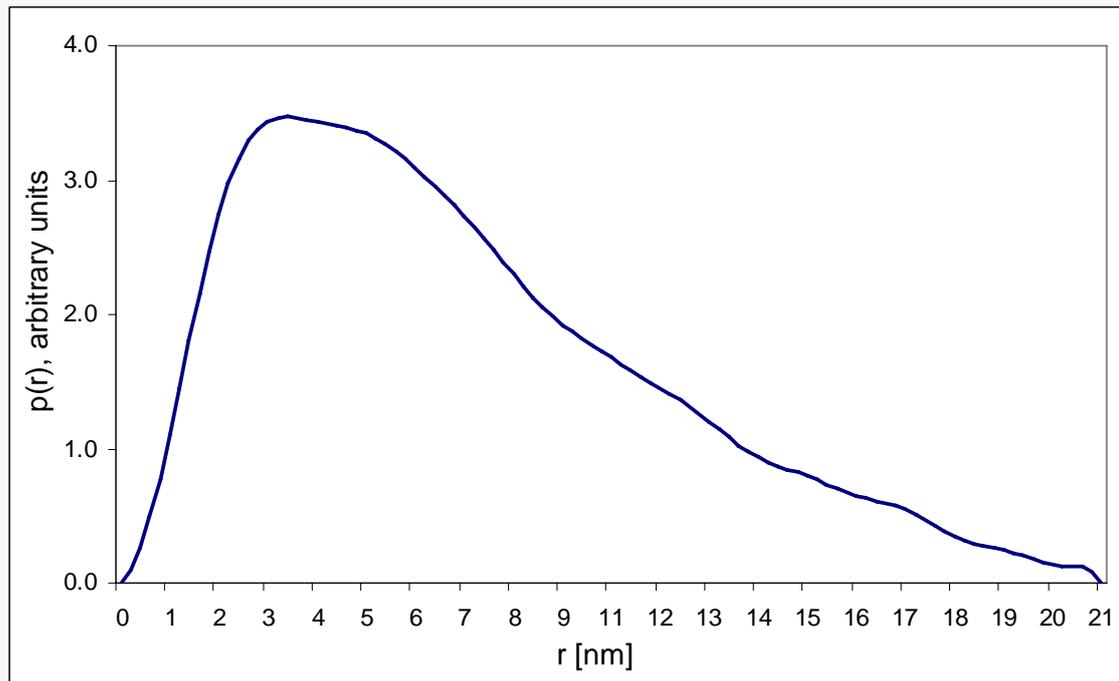
Figure 26: Pair distribution function of the SbcCD construct in solution.
The pair distribution plot shows the distribution of electrons that are in distance r from each other.
The extended shape of the curve with an early peak and a long tail clearly indicates an elongated
protein complex (for details refer to sections 2.3.2 and 5.11).

After processing the scattering data with GNOM, the modeling program GASBOR,
which represents the particle as a collection of dummy residues, was used to calculate
the overall conformations of the SbcCD construct *ab initio*. The simulation starts with
randomly positioned residues and utilizes simulated annealing to find a chain-
compatible spatial distribution of dummy residues inside the search volume. The model
was first constructed without symmetrical constraint being applied (data not shown), but
to obtain a more detailed model, a two-fold symmetry, as determined by the X-ray
structure, was given. The models constructed without symmetry restriction have an
elongated shape as well as the P2-symmetry constrained models, though they do not
match perfectly. To evaluate the creditability and accuracy of the reconstructed model,
DAMAVER, which aligns all reconstructed models and removes outliers below a given
cutoff volume, was employed. The different calculated models are shown in figure 27.
Out of 10 structures 9 meet the qualification criteria and were used for further model
building. The models meeting the quality criteria were used to compute the final SAXS

model by merging using DAMAVER. The final SAXS structure of the SbcCD construct is depicted in figure 37 in chapter 6.5.2.1.
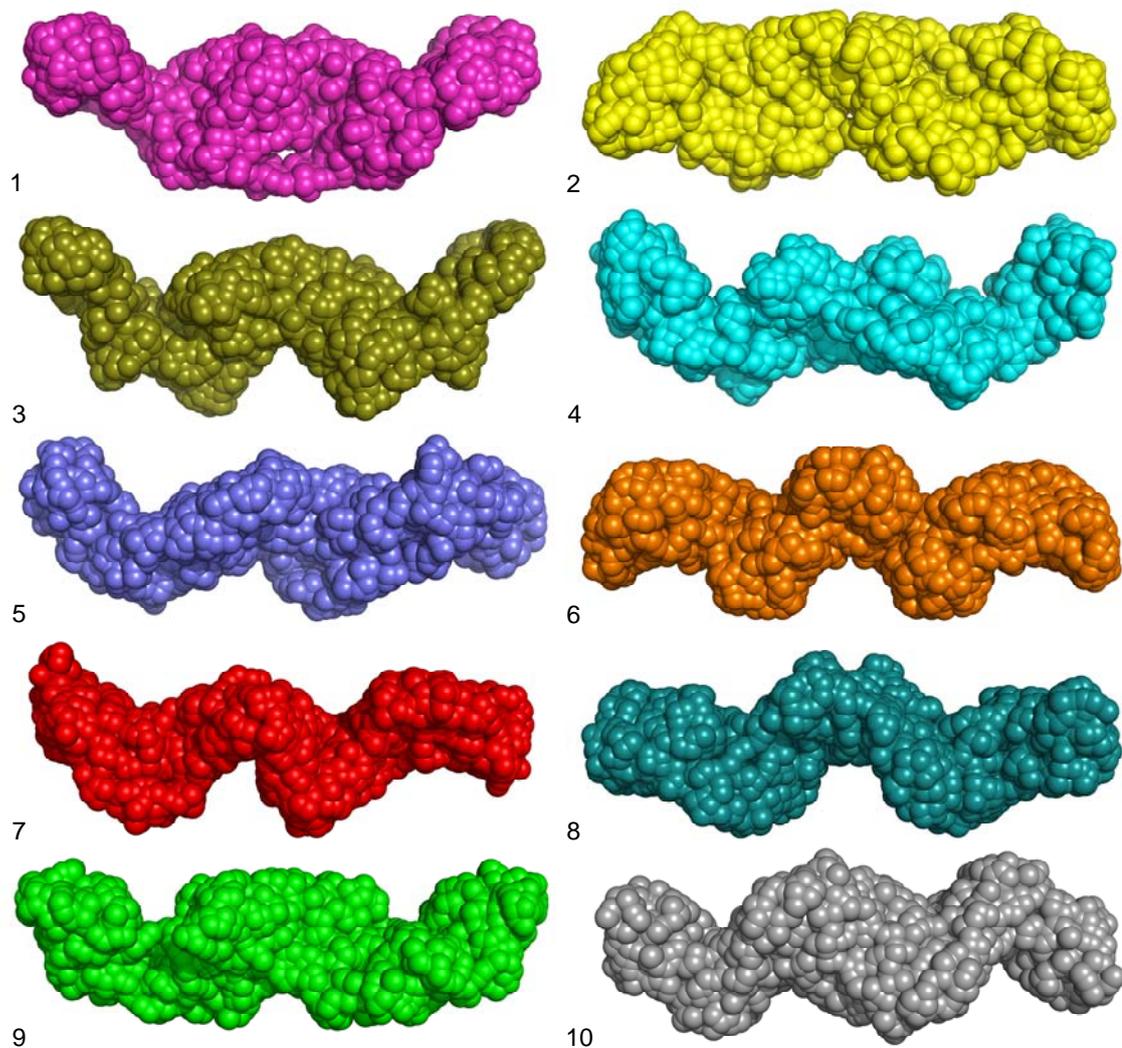
Figure 27: Calculated SbcCD *ab initio* structures using DAMAVER.
10 different in solution structures have been computed using the DAMAVER program suite. Structure 9 exhibits the lowest mean value of normalized spatial discrepancy and is chosen as the reference, whereas model 7 was flagged as an outlier and therefore neglected in further structure determination (see methods section 5.11 for more details). All models were aligned to the reference using DAMAVER.

6.4.2   Data Collection and Structure Determination of the Nucleotide-Bound SbcCD Complex

In addition to in solution studies of the elongated SbcCD complex, the structure of the ATP-analog-bound form was examined. Therefore the same construct was expressed and purified like the sample in absence of ATP-like molecules. Prior to the protein concentration steps, ATPγS as a poorly-hydrolyzable ATP-analog was added to a final concentration of 1 mM as well as magnesium chloride with a final concentration of 5 mM. The protein then was concentrated and small angle X-ray experiments were performed at concentrations of 5 mg/ml and 10 mg/ml. The best dataset was recorded at 5 mg/ml and the scattering curve is presented in figure 28. This data was collected at the SYBILS beamline in Berkeley compared to the ATPγS-unbound SbcCD complex described in the previous chapter, which was measured at the EMBL X33 beamline in Hamburg. The processing procedure was equal in both cases though hardware experimental setup and data collection strategy differs between the two beamlines. Compared to the 59 Å radius of gyration (R_G) of the SbcCD apo in solution structure the ATPγS-bound form exhibits an R_G of 47 Å with a given maximum dimension of the protein (D_{max}) of ~200 Å based on curve analysis of the pair distribution function. This observed reduction of particle size in addition to the overall shape change of the pair distribution graphs (figure 28) between the two states indicate a conformational change dependent upon ATPγS binding.
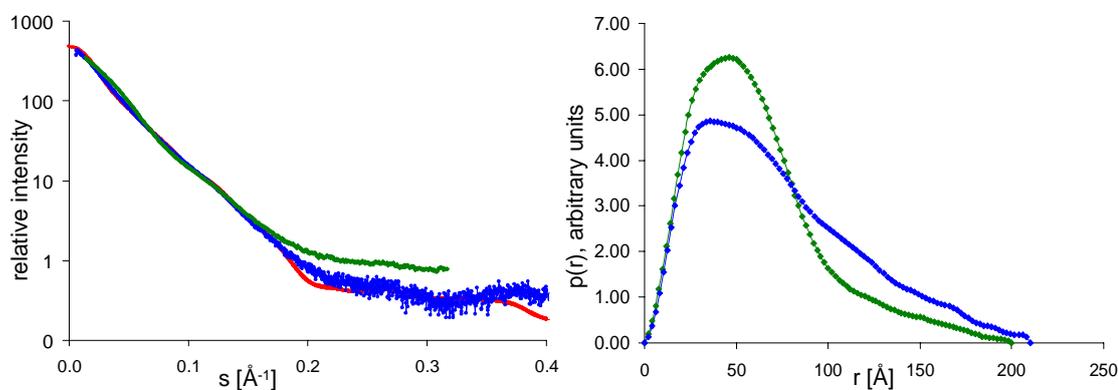
Figure 28: Comparison of the scattering profile and pair distribution function of SbcCD in the ATPγS-bound and -unbound forms.
On the left side the scattering profiles of SbcCD in ATPγS unbound (blue), ATPγS-bound form (green) and the computed theoretical scattering pattern based on the atomic X-ray crystallography structure (red) are compared. The relative scattering intensity is plotted against the function of momentum transfer $s = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle between the incident and scattered radiation and λ is the X-ray wavelength. The corresponding pair distribution functions are depicted applying the same color code on the right. The more bell shaped curve of the ATPγS-bound form (green) indicates a more globular protein structure compared to the nucleotide free form.

As for the apo structure three dimensional models were calculated using the modeling program GASBOR as shown in figure 29. SUPCOMB cross-correlation calculations implemented in DAMAVER were applied to identify models not fitting the quality criteria. No outliers were observed and all models were utilized for computing the final averaged model of the ATPγS-bound SbcCD complex, which is depicted in figure 39 in chapter 6.5.2.2.
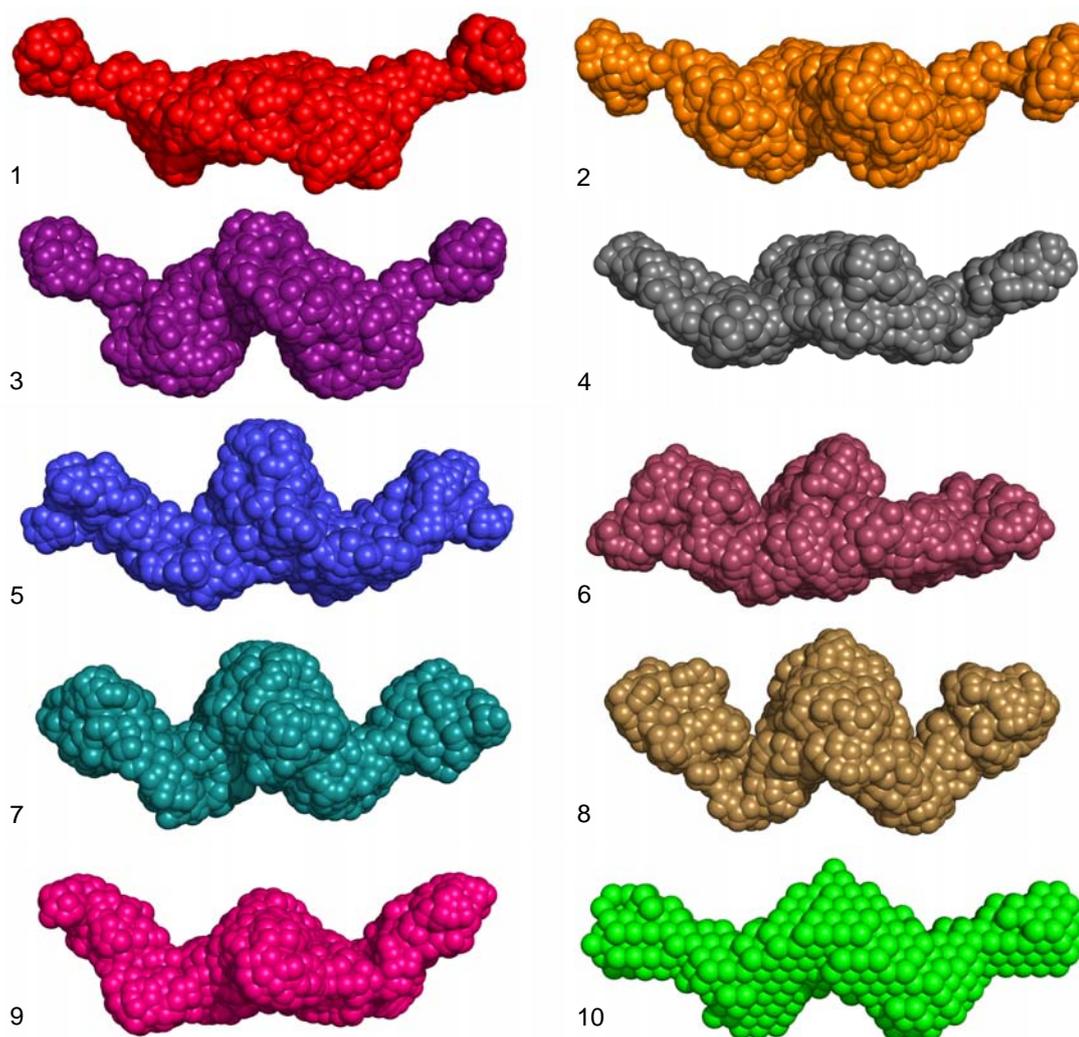
Figure 29: GASBOR algorithm based models of the ATPγS-bound SbcCD of *T. maritima*.
10 different in solution structures have been computed using GASBOR implemented in the
DAMAVER program suite. Structure 10 exhibits the lowest mean value of normalized spatial
discrepancy and is chosen as the reference. Due to low spatial discrepancies within the range of
tolerance no structure was neglected in further structure determination (see methods section 5.11
for more details). All models were aligned to the reference using DAMAVER.

6.5   Structural Analysis

6.5.1   X-ray Crystallography

6.5.1.1   Overall Structure of SbcCD
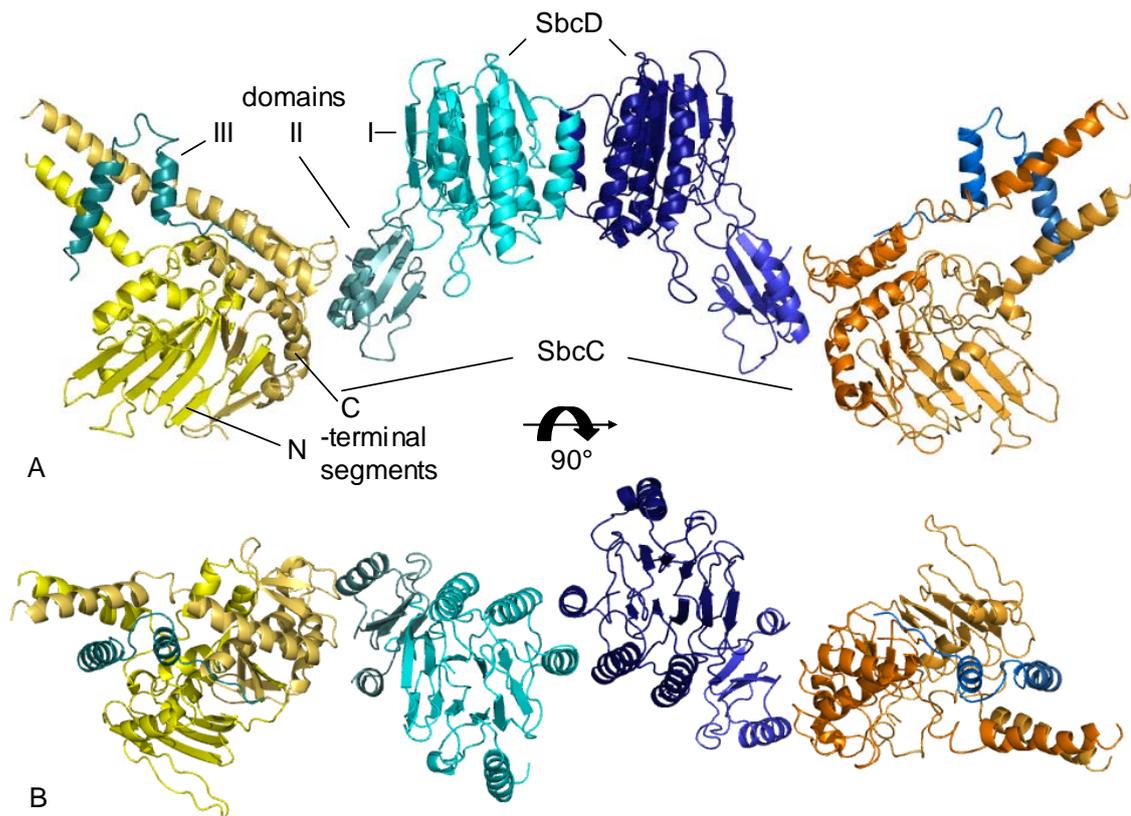
Figure 30: Ribbon presentation of the SbcCD structure of *T. maritima*.
SbcCD crystallized with one molecule in the asymmetric unit. The heterotetramer consists of two Rad50 (molecule 1 (yellow) and molecule 2 (orange)) and two SbcD molecules (molecule 1 (cyan) and 2 (blue)). The shadings aid in distinguishing between different intramolecular segments. In terms of color, SbcC is subdivided in the N- and C-terminal parts and SbcD in domains I, II and III. A) The front and B) the top view are presented using PyMOL (DeLano, 2002).

The *Thermatoga maritima* SbcCD construct is an elongated, w-shaped molecule complex with dimensions of approximately 210 Å * 75 Å * 65 Å. It is a heterotetramer consisting of two SbcC and two SbcD molecules where the latter build the central portion that is flanked by SbcC (figure 30).

The SbcC construct is formed by two lobes that fold into a single ellipsoidal domain with concave and convex sides (figure 31). Lobe I is mainly formed by the N-terminal segment of the protein and comprises the central α helix A wrapped in an antiparallel β-sheet consisting of the strands 7A, 6, 5, 4, 1 and 2. Helix A is connected via helix B to strand 4 in this sheet. Sheet 7A could not be observed due to a resolution of 3.4 Å, but seems to be apparent in an unstructured loop based on comparison with the *Pyrococcus furiosus* Rad50 structure (Hopfner *et al.*, 2000b). From the end of the sheet the peptide chain continues *via* helices B to E to the coiled-coil starting at helix F. The eight amino acid linker is not visible in this structure due to its flexibility. The incoming coiled-coil leading to helix G1, is connected to a two-stranded antiparallel β sheet by helix G2. The β sheet contains strands 8 and 9. This sheet, together with the β sheet consisting of strands 10, 11, 3 and 12, sandwich the helices G2, H, I and J, primarily forming lobe II. The nucleotide binding domain (NBD), constituted by the Walker A motif (P-loop), is positioned in lobe I close to the interface of both lobes and can link binding and hydrolysis to conformational changes in lobe II.

SbcD contains three domains, where domain I and II interact at the active site (figure 31). Domain I consists of two parallel mixed β sheets (3, 2, 1, 10, 11 and 4, 5, 6, 7, 9, 8 respectively), which are surrounded by five α helices (a, b, c, d and e). The nuclease active site is located in domain I in the bottom groove. Domain II is connected to domain I *via* a peptide chain containing helix f. Domain II comprises a 3-stranded β sheet (11, 12 and 13) and three α helices (g, h and i). Helix I protrudes from domain II and indicates the start of the linker to domain III. Domain III, as the C-terminal part of SbcD, contains the helices j and k and forms the interaction domain with SbcC. Due to missing electron density, the linker could not be modeled in its entire length and helices f and i of domain I and II, respectively, were not recognized by PyMOL as such. These helices were identified by structure alignment with a 2.2 Å resolution structure of domains I and II of *Thermotoga maritima* SbcD deposited in the protein data bank (PDB-ID 2Q8U, JCSG and Wilson, 2007).
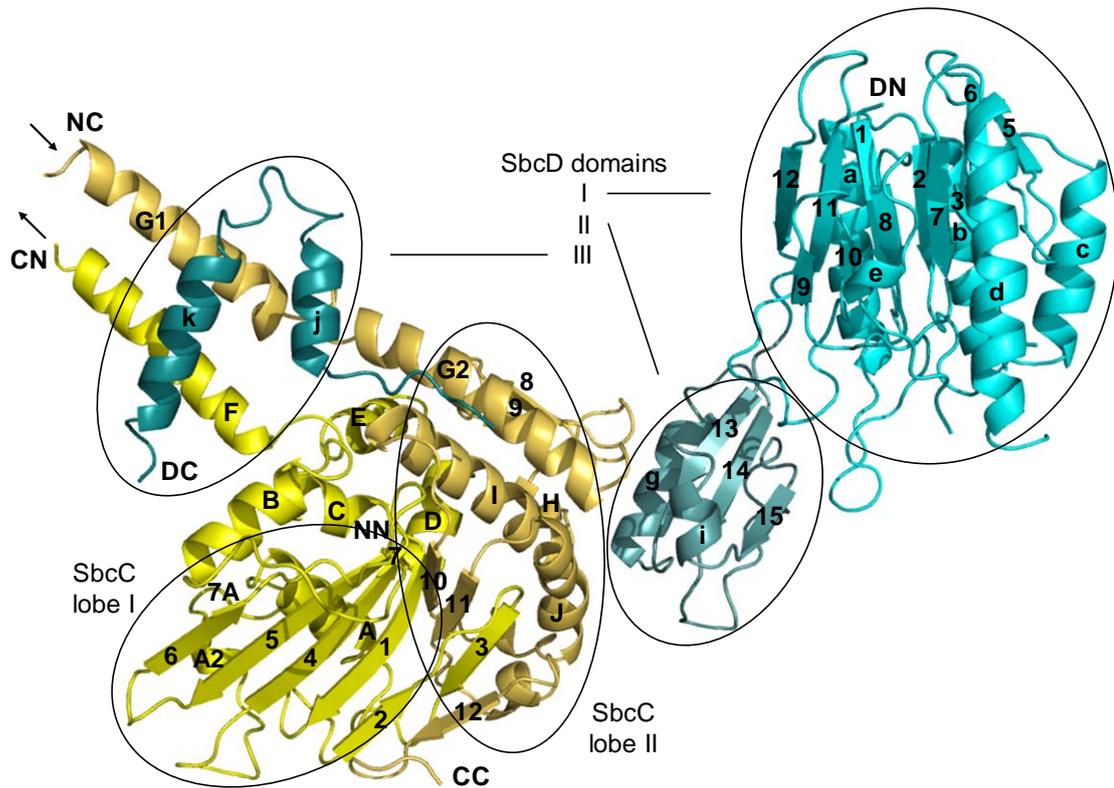
Figure 31: Ribbon model of a SbcCD construct heterodimer.
The model with annotated secondary structure elements shows SbcC and SbcD in more detail. SbcC exhibits a two lobal fold typical for a member of the ABC-ATPase superfamily, constituted by intermolecular association of the N-terminal (yellow) and C-terminal (yellow-orange) segments. The N-terminal part heads from NN to CN and the C-terminal part from NC to CC. Arrows indicate the protruding and incoming coiled-coils. SbcD domains I (cyan), II (light teal) and III (dark teal) are depicted and labeled as well. The N-terminus of SbcD is labelled DN and the C-terminus DC, respectively. Capitals denominate helices of SbcC and lower case letters helices of SbcD.

6.5.1.2   SbcD Interface

SbcD dimerizes and thereby forms the central part of the heterotetrameric SbcCD complex. Dimerization is based on interaction of hydrophobic surface clusters of domains I of the SbcD molecules. Involved in this interaction are mainly the helices b and c (figure 32). Helix b contains 9 hydrophobic residues (Val68, Val69, Ala70, Leu71, Leu74, Leu75, Leu78, Met81 and Met82) from which residues Leu71, Leu75, Leu78 and Met82 point to the opposing helix. Out of the 7 hydrophobic amino acids (Leu99, Leu101, Phe102, Gly103, Phe105, Val106 and Ile109) comprised in Helix c,

Leu101, Phe102 and Phe105 build up the dimerization area for the facing SbcD molecule.
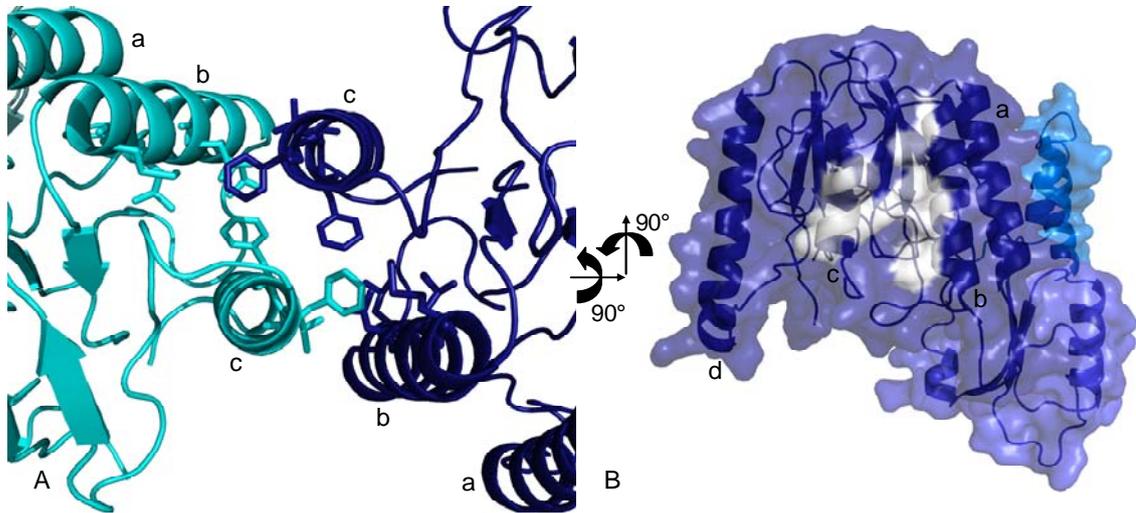
Figure 32: SbcD dimerization site.
The interaction of the two SbcD molecules is represented. A) Residues of the interface helices b and c likely involved in dimerization are represented from the top as stick models. Figure B) shows the interface with a transparent surface and ribbon representation of the amino acid chain. The residues highlighted in A) create a hydrophobic surface patch (white) at the interface of the SbcD molecules. SbcC is omitted for better recognition. The color code is as before.

6.5.1.3   SbcC-SbcD Interaction

SbcD interacts with SbcC *via* its domain III (figure 33). A 17 amino acid linker (Val334-Asp350) at the C-terminus of SbcD, that connects domains II and III, and helices j and k, folded back against each other, exhibit numerous hydrophobic residues. These amino acids of SbcD and especially the ones directed towards helices of SbcC, namely Leu346 and Leu349 (linker), Phe352, Leu354, Phe355 and Leu359 (helix j) and Leu369, Leu370, Ile372, Leu373, Leu376, Leu377 and Val380 (helix k) generate an interaction surface appropriate for binding to SbcC. The SbcC-SbcD interplay region of SbcC consists mainly of the coiled-coils close to the globular head domain. The coil protruding from the N-terminal region, represented as helix F, harbors the hydrophobic amino acids Phe164, Leu170, Leu173, Leu176, Leu177, Met181 and Leu184 pointing towards helices j and k of SbcD. Residues of the C-terminal coil probably involved in binding of SbcD are leucins 701, 703, 704, 708 and Phe709 of helix G1 as well as Phe714, Phe718 and Ala725 of helix G2. Based on their adjacency to the hydrophobic

residues of the C-terminal linker of SbcD (Val334-Asp350) amino acids Val786 and Ala787 of SbcC also seem to be important for the interaction of SbcC with SbcD.
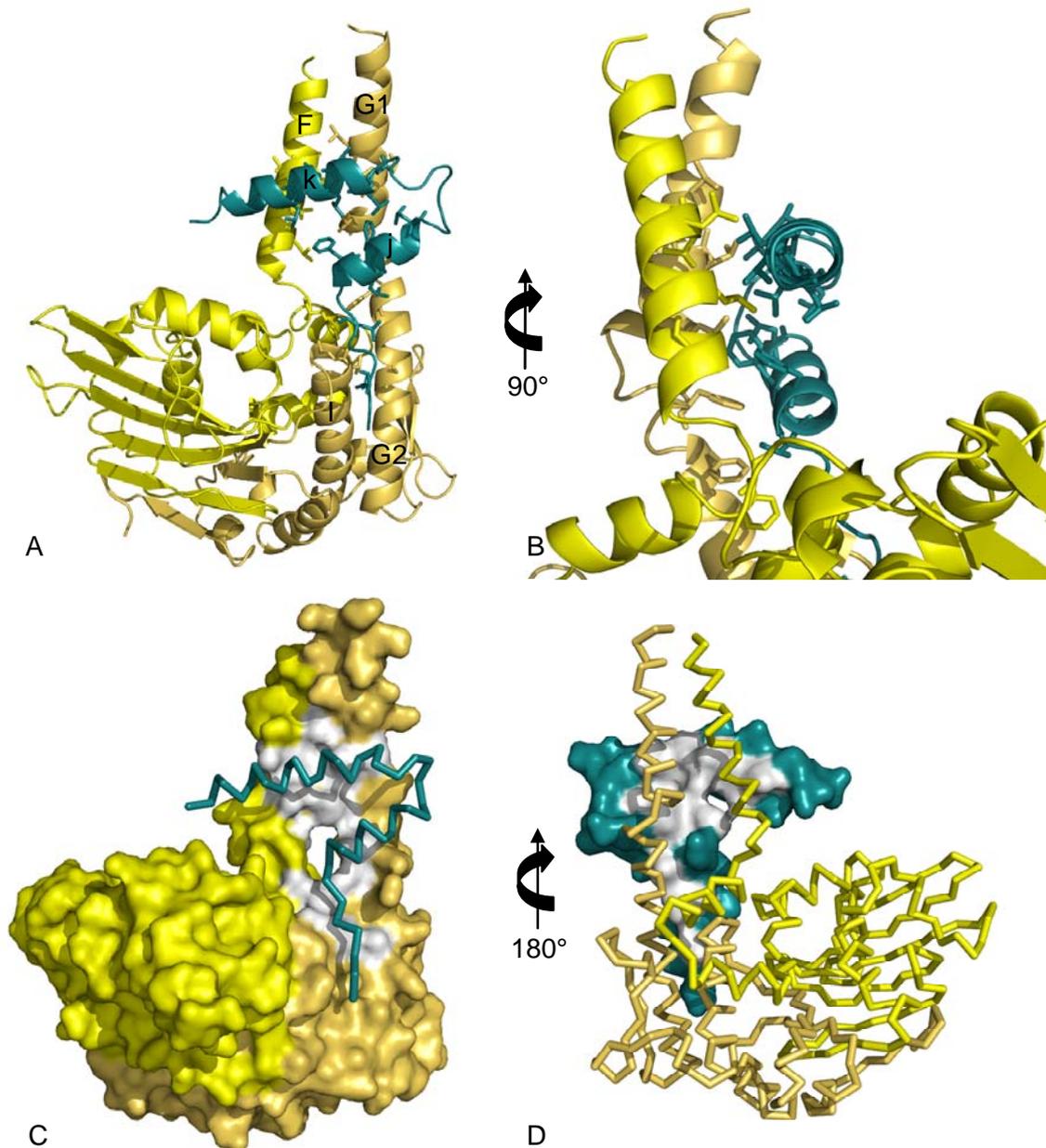
Figure 33: Interaction of *Thermotoga maritima* SbcC with SbcD.
This figure displays the C-terminal interaction domain III of SbcD, consisting of a linker region, helix j and k as well as the opposing binding part of SbcC, namely helices F, G1, G2 and I. The hydrophobic residues implicated in protein-protein interaction are depicted as stick models in A) and B) (color code as above). Below, the surfaces C) of SbcC and D) domain III of SbcD are modelled and hydrophobic residues close to their binding partner are highlighted. Both molecules are oriented in a fashion to place emphasis on their interaction areas. The respective binding partner is ribbon represented applying the color code as used previously. The amino acids of SbcC highlighted in the upper figures compose a hydrophobic dimerization region ideal for binding to the C-terminus of SbcD.

6.5.1.4   Comparison of *Thermotoga maritima* SbcCD with *Pyrococcus furiosus* Rad50 - Mre11

The sequence and structure of eubacterial SbcC and SbcD from *Thermotoga maritima* was compared with the archaeal orthologs of *Pyrococcus furiosus* (Rad50: PDB-ID 1II8, Hopfner *et al.*, 2001; Mre11: PDB-ID 1II7, Hopfner *et al.*, 2001). As expected, the bacterial domains are structurally quite similar to their corresponding archaeal parts (figure 34).

Concerning SbcC/Rad50, the bacterial and archaeal proteins show the same overall fold and harbor the same motifs. *P. furiosus* comprises the three additional very small β-sheets 7 (Glu107-Lys109), 10 (Lys785-Arg787) and 14 (Ser876-Val880) as becomes evident when displaying the structures with PyMol. A structure based sequence alignment of *T. maritima* SbcC and *P. furiosus* Rad50 constructs was generated with MSDfold (http://www.ebi.ac.uk/msd-srv/ssm/) using the PDB files as input and edited manually with GeneDoc (Nicholas and Nicholas, 1997) (figure 35). The high structural similarity with a root-mean-square distance (RMS distance) of 1.41 Å was calculated by comparing residues Met1-Glu187 and Lys689-Val850 with amino acids Met1-Ile184 and Glu715-Ser882 from *T. maritima* and *P. furiosus*, respectively, using LSQMAN (Kleywegt, 1996).

In case of SbcD/Mre11, there are slightly more differences apparent between the bacterial and archaeal structures. Again, the overall fold is rather similar as depicted in figure 34, but *P. furiosus* Mre11 contains three additional β-sheets. Sheets 5 (Leu119-Leu124), 12 (Val236-Trp240) and 13 (Phe245-Tyr249) are apparent in the archaeal protein but not in bacterial SbcD. Nevertheless, with a RMS distance of 1.44 Å the overall structural similarity is quite high as was calculated for SbcD/Mre11 using residues Lys8-Ser327 (lacking domain III) of *T. maritima* and amino acids Lys2-Ile333 of *P. furiosus*. A structure based sequence alignment is depicted in figure 36.
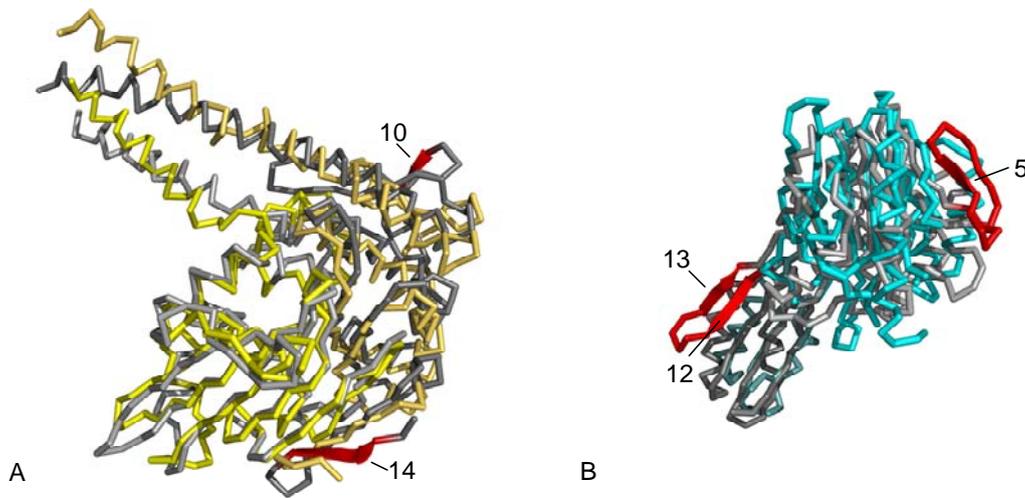
Figure 34: Superposition *T. maritima* SbcCD and *P. furiosus* Rad50 and M11.
Superposition of backbone ribbons of A) *T. maritima* (yellow and orange) and *P. furiosus* (light and dark gray) SbcC/Rad50 and B) *T. maritima* (cyan and deep teal) and *P. furiosus* (light and dark grey) SbcD/Mre11. Varying secondary structure elements between the two species are highlighted in red. Rad50 exhibits only minor differences whereas *P. furiosus* Mre11 contains two regions harboring three additional β-sheets. Since for *P. furiosus* Mre11 the structure of the C-terminal domain III is not known the corresponding domain of *T. maritima* was neglected in this alignment.

Figure 35: Structure based sequence alignment of SbcC/Rad50 (next page).
A structure based sequence alignment was performed using MSDfold with default settings and *T. maritima* SbcC and *P. furiosus* Rad50 PDB files as input coordinates. Rad50 of *S. cerevisiae* was aligned based on the ClustalW2 algorithm and edited manually afterwards. Amino acids forming β-sheets are printed in blue, whereas residues being organized in α-helices are depicted in red and amino acids not visible in the structure are displayed in grey. Arrows (β-sheets) and bars (α-helices) emphasize the secondary structure elements of *T. maritima* and are labeled according to the nomenclature in the text. Secondary structures recognized by PyMOL in *P. furiosus* but not in *T. maritima* are highlighted in orange. Differences due to insertions are accentuated in magenta. Numbering of these two general differences was performed according to the literature where the *P. furiosus* structures were presented at first (Hopfner *et al.*, 2001; Hopfner *et al.*, 2000b). Important motifs are underlayed in grey and residues mutated in *S. cerevisiae* plate survival assays are depicted in green. The remaining coloring is consistent with previous figures.
The Walker A motif located in lobe I represents the nucleotide binding site. Two ATP molecules lie sandwiched between Walker A and signature motifs of opposing globular domains of Rad50. The signature motif binds to an ATP γ-phosphate oxygen *via* the serin sidechain oxygen and the glycine mainchain nitrogen. Thus the signature motif is a sensor for an ATP γ-phosphate in the opposing molecule and can link ATP hydrolysis to structural changes. The active site Mg^{2+} ion binds to ATP β- and γ-phosphate oxygens and to serin (Walker A) and the conserved glutamine (Q-loop) sidechain oxygens, plus two water molecules. One of these binds to asparagine and glutamine in the Walker B motif, where the glutamine is suggested to polarize the water molecule. The dimerization loop (D-loop) binds to one active site Mg^{2+} ion (*via* interaction of the Walker B motif with the hydrated Mg^{2+} ion) and present the attacking water in the other active site. The D-loops likely link ATP-hydrolysis in one Rad50 globular domain to ATP-hydrolysis of the second Rad50 molecule in the dimer (Hopfner *et al.*, 2000b).

```
              1                 2        3              A
                                                20              40
TmSbcC  ----MRPERLTVRNFLGLKNVDIEFQSGITVVEgPNGAGKSSLFEAISFALFG-------   49
PfRad50  ----MKLERVTVKNFRSHSDTVVEFKEGINLIiGQNGSGKSSLLDAILVGLYWpl-----   51
ScRad50  msaiyklsiqgirsfdsndretiefgkpltlivgmngsgkttiieclkyattgdlppnsk   60
                                          Walker A

              A2              4           5                    6
                  60              80
TmSbcC  NGI-RYpnsydYVnrnaVDGTARLVFQFERGGKRYEIIREINALQr----------KHNA   98
PfRad50  RIKdIKk---dEFtk-vGARDTYIDLIFEKDGTKYRITRRFLKGysS--------GEIH   98
ScRad50  ggvfihd---pkitg-ekdiraqvklaftsanglnmivtrniqllmkkttttfktlegql  116

              7           B           C       7    D           E
              100             120             140
TmSbcC  KLSEILeNGKKAAIA-AKPTSVKQEVEKILgiEHRTFIRTVFLPqGEidKLlISPPSEIT  157
PfRad50  AMKRLV-GNEWKHVTePSSKAISAFMEKLI--PYNIFLNAIYIrQGQidAIL-ESDEARE  154
ScRad50  vainns-gdrstlstrsleldaqvplylgv--pkaileyvifchqedslwpl-sepsnlk  172
                                                    Q-loop

                      F                              G1
              160             180     188         689         700
TmSbcC  EIISDVFQskETLEKLEKLLKEMKKLENei(coiled-coil)kklkemsdeyNnldLL  704
PfRad50  KVVREVLN-lDKFETAYKKLSELKKTINNri(coiled-coil)-eikdlekakdfteEL  729
ScRad50  kkfdeifq-amkftkaldnlksikkdmsvdi(coiled-coil)-elqtrsfvtddidvy 1130

                      G2                  8
                      720             740
TmSbcC  RKYlfdksnfsRYFTGRVLEAVLKRTKAYLDILTNG-RF-DIDFDDEK------------  750
PfRad50  IEKvkk---ykALAREAALSKIGELASEIFAEFTEG-KYsEVVVRAEe---------nk  775
ScRad50  skalds---aimkyhglkmqdinriidelwkrtysgtdidtikirsdevsstvkgksyny 1187

              9   10   H           I               10
                  760             780             800
TmSbcC  gGFIIKDWGIErPARGLSGGERALISISLAMSLAEVASGRLDAFFIDEGFssldtenKEK  810
PfRad50  vRLFVVWEGKeRPLTFLSGGERIALGLAFRLAMSLYLAGEISLLILDepTPYlD----EE  831
ScRad50  rvvmykqdveldmrgrcsagqkvlasiiirlalsetfgancgvialdepttnld----ee 1243
                      Signature motif            Walker B  D-loop

              J               11      K               12
                  820                             840
TmSbcC  IASVLKELERL---------NkVIVFITHdrefSEAF----------DRklRITG----  846
PfRad50  RRRKLITIMER-----YlkkipQVILVSH----DEELKDaa------dhviRISLengss  876
ScRad50  nieslakslhniinmrrhqknfqlivithdekflghmnaaaftdhffkvkrddrqksqie 1303

              14

TmSbcC  GVVVNE---   852
PfRad50  KVEVvs---   882
ScRad50  wvdinrvty  1312
```

Figure 35: Structure based sequence alignment of SbcC/Rad50 (above).
Figure description and color code are as explained at the previous page.

Figure 36: Structure based sequence alignment of SbcD/Mre11 (next page).
Figure description and color code are as explained at the previous page.

```
                    1                              a                    2
           1               20                    40
TmSbcD   --MINLKELKILHTSDWHLGVTSWTSSRPVD-RREELKKALDKVVEEAEKREVDLILLTG   57
PfMre11  --------MKFAHLADIHLGYEQ----FHKPQREEEFAEAFKNALEIAVQENVDFILIAG   48
ScMre11  mdypdpdtirilittdnhvgyne-----ndpitgddswktfhevmmlaknnnvdmvvqsg   55


                         b                           3
              60              80
TmSbcD   DLLHSRNNPSVVALHDLLDYLKRM-------MRT-APVVVLPGNHDW------------   96
PfMre11  DLFHSS-RPSPGTLKKAIALLQIP-------KEHSIPVFAIEGNHDRTQR--------G   91
ScMre11  dlfhvn-kpskkslyqvlktlrlc(77-111)fnisipvfgisgnhddasgdsllcpmdi  139
              Mre11 dimerization                    nuclease site


                     c            4        5            5        6
                100                            120
TmSbcD   ----KGLKLFGNFVTSISSDITFVMSF--------------------EPVDVEAKRGQKVR  133
PfMre11  PSVLNLLEDFG--------LVYVIGMRKEKVENEYLTSERLGNGEYLVKGVY----KDLE  139
ScMre11  lhatglinhfg----------kviesdkikvvpllfqkg------stklaly----glaa  179
              Mre11 dimerization


                               d                  7        e1
              140             160              180
TmSbcD   ILPFPYPDESEALRKNEGDFRFFLESRL-NKLYEEALKKEDFAIFMGHFTVEGLAGYAGI  192
PfMre11  IHGMKYMSS-----AWFEANK---EILK-RLFRPTD-----NAILMLHQGVREVSEARGE  185
ScMre11  vrder-lfr-----tfkdggv---tfevptmregew-----fnlmcvhqnht---ghtnt  222


                  e         8          9         10           f
                200          220                   240
TmSbcD   EQGREIIINRALIPSVVDYAALGHIHSFREIQK-----QPLTIYPGS--LIRIDFGEE--  243
PfMre11  DYF---EIGLGDLPEGYLYYALGHIHKRYETSYSG----SPVVYPGS--LERWDFGDYEV  236
ScMre11  aFlp--eqflpdfld---mviwghehecipnlvhnpiknfdvlqpgssvatslceaeaqp  277


           12                 13         11        12         13
TmSbcD   -----------------------ADEKGAVFVELKRGEPPRYERIDASPLPLKTLYYK  278
PfMre11  RYEW------------DGIKFKERYGVNKGFYIVED-----FKPRFVEIKVRPFIDVKIK  279
ScMre11  kyvfildikygeapkmtpipletirtfkmksislqd-----vphlrphdkda-tskylie  331


                     g                    14
              280                   300
TmSbcD   KI---DTSALKSIRDFCRN----------FPGYVRVVYEE----------------DSG  308
PfMre11  GSEEEIRKAIKRLIPLIP-----------KNAYVRLNIGWRKP----------------F  312
ScMre11  qveemirdaneetkqkladdgegdmvaelpkplirlrvdysapsntqspidyqvenprrf  391


                h        15       i                              j
              320                       340
TmSbcD   ILPDLMGEI-----DNLVKIERKSRREIEEVLRES--------PEEFKEELDKLDYFELF  355
PfMre11  DLTEIKELLN----VEYLKIDTWRI-KERTDEES-------GKIGLPSDFFTEFELKII  359
ScMre11  snrfvgrvangnnvvqfykkrspvt--rskksgingtsisdrdveklfsesggelevqtl  449
              TEV insertion site              SbcC interaction


                       k
              360
TmSbcD   KEYLKKREENHEKLLKILDELLDEVKKSEA-----------------------------  379
PfMre11  DILGEKDFDDFDYIIKLITEGKVEEEGPLEEAVKKVSEEKGKTVRQKIESIPKKKRGTLD  419
ScMre11  vndllnkmqlsllpevglneavkkfvdkdektalkefisheisnevgilstneeflrtdd  509
         SbcC interaction              SbcC interaction


TmSbcD   ----------------
PfMre11  SWLGGAR--------- 426
ScMre11  aeemkal(517-692)
```

6.5.2   Small angle X-ray Scattering

### 6.5.2.1   Small angle X-ray Scattering of the Apo SbcCD Complex

Figure 37: Averaged SbcCD *ab initio* structure determined by SAXS.
The final averaged *ab initio* model of the SbcCD construct is depicted from A) the front and B) the top as spheres models. The protruding coiled-coils of SbcC and SbcD as the core portion are clearly recognizable. The maximum particle diameter is 210 Å.

Comparison of the atomic structure determined by X-ray crystallography and the low resolution model calculated based on small angle X-ray scattering demonstrates that the crystal structure is not affected by artifacts resulting from the crystal packing. The alignment was performed with DAMAVER with the crystal structure as the reference. The two models fit very well as becomes apparent in figure 38. The only larger regions not included completely in the SAXS envelope are lobes I of the SbcC molecules. This might be explained by the natural flexibility of these domains connected *via* the coiled-coil with the C-terminal part of SbcC. Another part, namely helix d of the second SbcD

molecule of the crystal structure, is not included in the SAXS sphere model as well. This 18 amino acid long helix might be flexible due to its lateral position, which might explain why it does not fit ideally into the SAXS model. The more detailed 2.2 Å structure of SbcD (JCSG and Wilson, 2007) shows are slightly more inwards centered helix d.
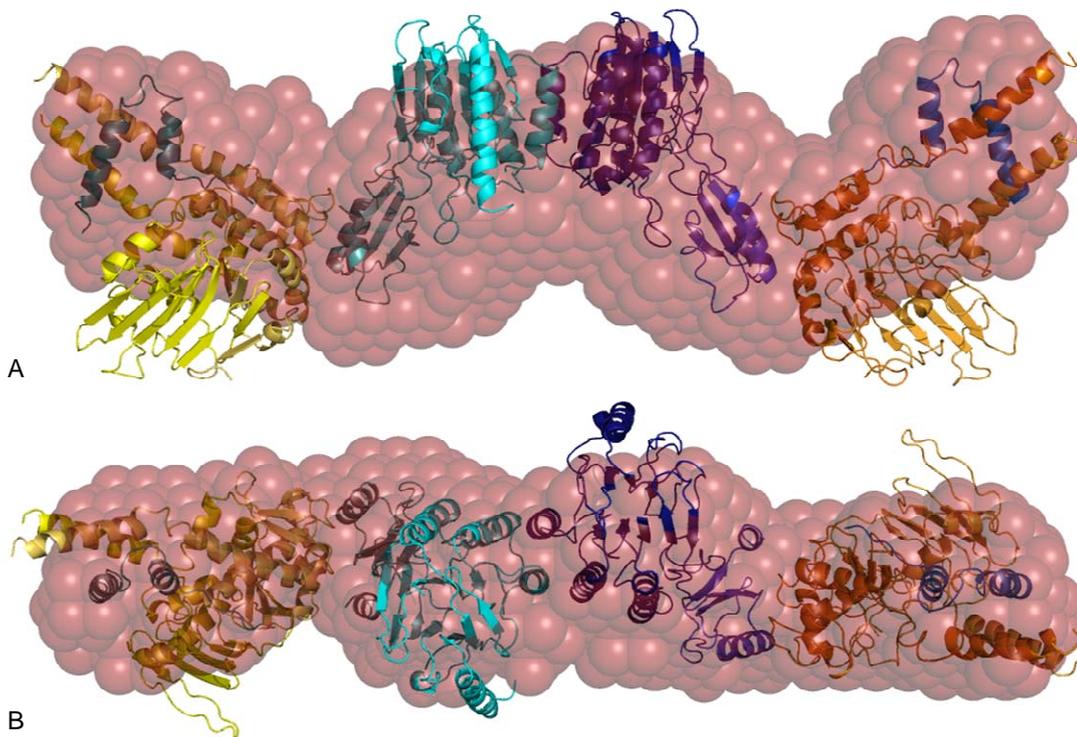

Figure 38: Superposition of the SAXS and crystal structures of SbcCD.
The in solution model of SbcCD (transparent red) and the atomic crystal structure (colored as before) are superposed and depicted from A) the front and B) the top. Superposition was carried out using DAMAVER.

6.5.2.2   Small angle X-ray Scattering of the Nucleotide-Bound SbcCD Complex

As can already be deduced from the pair distribution plot (figure 28) the molecule shapes of the ATPγS free and ATPγS-bound SbcCD complexes are different though still similar. In figure 39 the in solution structure of SbcCD spiked with ATPγS has been compared with the crystal structure based on the DAMAVER alignment algorithm. The two structures exhibit a similar overall w-shape. As it is true for the nucleotide free

form, the N-terminal part of the bipartite ATPase domain does not overlay with the ATPγS-bound in solution structure. In contrast to the nucleotide unbound form, the N-terminal parts of the coiled-coils (helices F) are not superimposed as well. Remarkably, when regarded from the top, the central part consisting of SbcD is more globular possessing a diameter of ~65 Å compared to ~50 Å of the ATPγS-unbound conformation. In addition, the crystal and the nucleotide free form are slightly more extended than the ATPγS-bound model (~210 Å *versus* ~200 Å).
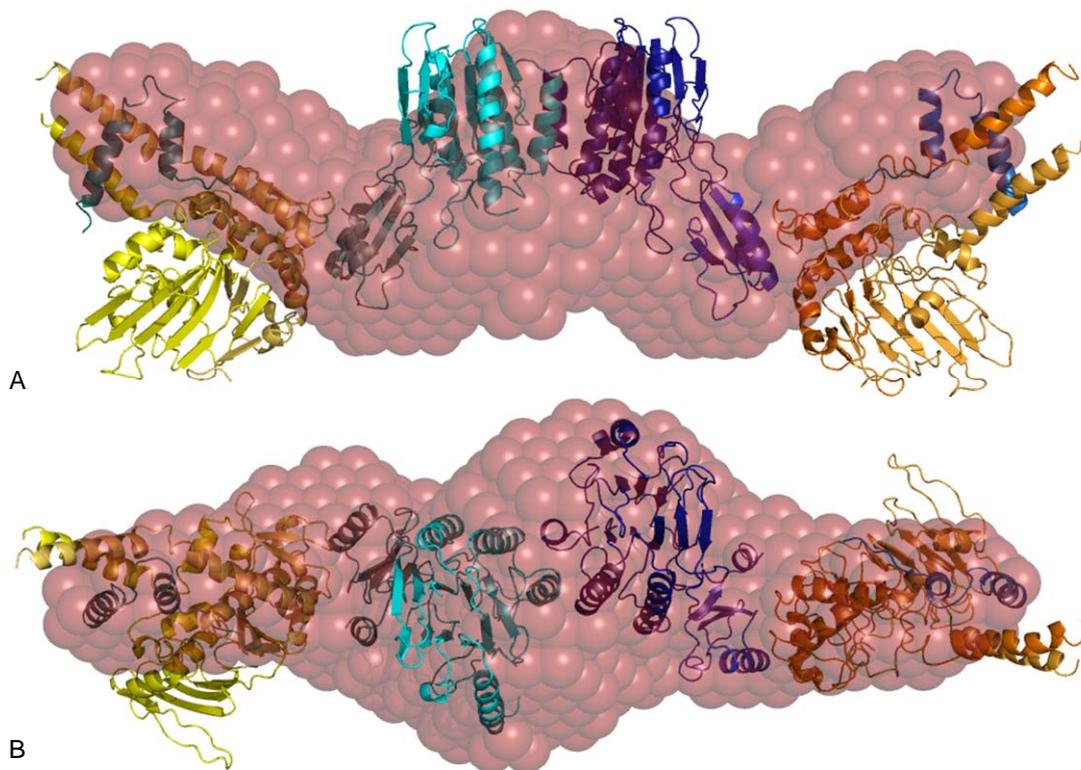
**Figure 39: Overlay of the in solution ATPγS-bound and crystal structures of TmSbcCD.**
The SAXS derived model of ATPγS-bound SbcCD (transparent red) and the atomic crystal structure (colored as before) are overlaid and depicted from A) the front and B) the top. Superposition was carried out using DAMAVER.

7   Discussion

DNA double-strand breaks are among the most harmful forms of DNA damage. Failure of repair of DSBs can lead to chromosomal rearrangements, loss of chromosome arms and tumorigenesis in humans (Khanna and Jackson, 2001; van Gent *et al.*, 2001). As such, sophisticated DSB repair mechanisms have evolved in all organisms. Response to DSBs requires efficient recognition of the damaged DNA and precisely orchestrated activation of cell-cycle checkpoint controls. In these processes the Mre11/Rad50/Nbs1 complex plays a central role, which is underscored by the fact that null mutations in any of the three proteins lead to embryonic death in mice (Luo *et al.*, 1999; Xiao and Weaver, 1997; Zhu *et al.*, 2001). Additionally, hypomorphic mutations in Mre11 and Nbs1 cause the genome instability syndromes Nijmegen breakage syndrome (NBS) and Ataxia telangiectasia-like disease (ATLD), respectively. The so-called separation-of-function point mutation in Rad50 results in partial embryonic lethality and cancer susceptibility in mice (Bender *et al.*, 2002; Carney *et al.*, 1998; Stewart *et al.*, 1999; Varon *et al.*, 1998).

The MRN complex plays three critical roles in double-strand break sensing, stabilization, signaling and effector scaffolding: 1) quick establishment of protein:nucleic acid tethering scaffolds for the recognition and stabilization of DSBs, 2) initiation of DSB sensing, cell-cycle checkpoint signaling cascades and establishment of epigenetic marks *via* the ATM kinase and 3) functional regulation of chromatin remodeling in the vicinity of a DSB (Williams *et al.*, 2007). Mre11 and Rad50 are found in all kingdoms of life, whereas Nbs1 is a eukaryotic-specific protein, which has evolved in part to regulate cell-cycle checkpoints (Carney *et al.*, 1998). This suggests that Mre11 and Rad50 form the core of this highly important DSBR complex acting as a multipurpose tether that directly bridges unconnected DNA ends and chromatin structures (Chen *et al.*, 2001; de Jager *et al.*, 2001; Hopfner *et al.*, 2002; Moreno-Herrero *et al.*, 2005).

To gain insights into the function of the Rad50/Mre11 complex the crystal structure of the bacterial ortholog complex consisting of SbcC and SbcD of *Thermotoga maritima*

was solved. *T. maritima* as source organism was chosen since it is a thermophilic bacterium. It grows at about 80°C indicating that its proteins are heat stable and therefore better to handle and to separate from endogenously expressed proteins from the *E. coli* expression system. Additionally, protein expression analysis of SbcCD/Rad50-Mre11 constructs from a variety of organisms showed that the *T. maritima* protein complex yields one of the highest protein amounts. A higher expression level was only observed for *Pyrococcus furiosus* constructs that unfortunately were not crystallizable. In regard to crystallization there also appears another major challenge in case of SbcC/Rad50. The N- and the C-terminal globular parts are connected by an anti-parallel coiled-coil that measures up to 600 Å (in human) or 300 Å (in *P. furiosus*) as seen in electron microscopy (Hopfner *et al.*, 2001). This extended structure makes full-length SbcC/Rad50 unsuitable for crystallization since the coiled-coils induce a high flexibility. This conformational heterogeneity prevents crystallization where high homogeneity is a prerequisite for proper three dimensional arrangements of the proteins. To overcome this obstacle a protein construct was designed where the major part of the coiled-coils was truncated and an artificial eight amino acid linker between the remaining N- and C-terminal coiled-coil was introduced. The length of the construct was based on coiled-coil prediction using the program COILS (Lupas *et al.*, 1991) and the known archaeal Rad50 structure (PDB-ID 1II8, Hopfner *et al.*, 2001). Since it was the aim of this work to solve the crystal structure of SbcC in association with SbcD, the coiled-coils could not be truncated in their entire length. It was indicated by co-purification and guanidinium-HCl dissociation that the Mre11 interacting region of Rad50 is located at the coiled-coils close to globular domains (Hopfner *et al.*, 2001). The remaining coiled-coil parts were not ideal for crystallization but necessary for the interaction of SbcC with SbcD. In case of SbcD, the first six amino acids were omitted and leucine at position 7 was replaced by methionine as the first amino acid since sequence alignment with other organisms showed weak homology in this region. The resulting constructs (SbcC: M1-S190-linker-S686-E852; SbcD: M7-A385) were cloned in a modified pET-29 vector enabled for bicistronic expression. The complex expressed in *E. coli* remained associated during the entire purification procedure.

The SbcCD complex was crystallized and the structure was solved by *de novo* phasing to a limiting resolution of 3.4 Å. The structure gives an impression of the overall architecture of these two proteins. It was suggested by electron microscopy and ultracentrifugation data and the calculated solution radius of gyration determined by small angle X-ray scattering of *P. furiosus* Rad50-ATPase/Mre11 complex that Rad50 and Mre11 form a single heterotetrameric DNA processing head (Hopfner *et al.*, 2001). This conception was now shown to be true by X-ray crystallography. The SbcCD complex consists of two SbcC and two SbcD molecules. SbcD homodimerizes at a hydrophobic surface patch. This homodimer builds the core subunit of the heterotetrameric complex. At the outer parts of SbcD, i. e. domains III, the SbcC molecules are arranged resulting in an elongated w-shape of the whole complex with dimensions of approximately 210 Å * 75 Å * 65 Å.

The dimerization of SbcD is driven by hydrophobic interactions of helices b and c. These helices located at the N-terminal phosphoesterase domains I form a four helix bundle across the dimerization interface. Residues of helix b of molecule one interact with residues of helix c of molecule two and *vice versa*, including the amino acids Leu71, Leu75, Leu78, Met82 (helix b) and Leu101, Phe102 and Phe105 (helix c). A similar arrangement was observed in the N-terminal structure of *P. furiosus* Mre11 (Hopfner *et al.*, 2001). It was speculated that the Mre11 dimerization observed in *P. furiosus* could be a result of crystal packing rather than an interaction of physiological importance. The Mre11 hydrophobic surface cluster was proposed to be a potential Rad50 binding site. One reason which argues against this theory is the observation of the same SbcD/Mre11 interface in different crystal forms. To exclude a coincidence that dimerization occurred during crystal packing, small angle X-ray studies were performed. These studies support the crystal structure of the SbcCD complex of *T. maritima* strengthening the findings in regard to SbcD/Mre11 dimerization. To provide further indications of the physiological importance of this homodimerization, *in vivo* studies in *Saccharomyces cerevisiae* were performed by collaborators (data not shown). Relevant hydrophobic residues apparent in the four helix interacting bundle as identified by structure based sequence alignment were exchanged by polar basic amino acids (Leu145Arg and Phe149Arg). In plate survival assays wild-type and mutants were treated with methyl methanesulfonate (MMS), which methylates the DNA bases leading

to strand breaks to study the sensitivity to DNA damage. Additionally, western blot analysis was performed to confirm equal abundance of mutated constructs. Mutation of Leu145Arg lead to a phenotype corresponding to the deletion of Mre11 and Phe149Arg resulted in a higher sensitivity to MMS compared to wild-type Mre11. This indicates the biological significance of both residues in dimerization and that leucine 145 contributes to the interaction in a higher extent as phenylalanine 149.

Sticking to new insights regarding SbcD/Mre11, the C-terminal portion needs to be mentioned. So far, the crystal structure illustrating domains I and II of *P. furiosus* Mre11 was solved (Hopfner *et al.*, 2001). In this work the structure of SbcD containing all three domains is presented. Domains I and II of *T. maritima* are quite similar compared with those of *P. furiosus* and the minor differences will be addressed later in this chapter. Domains I and II are connected to domain III by a 17 amino acid linker (Val334-Asp350). This linker could not entirely be modeled due to lacking electron density, which indicates its flexible nature. The electron density of the 8 amino acids Arg336-Lys343 was not visible. The low evolutionary conservation and the position apart from the proposed DNA binding sites (Hopfner *et al.*, 2001; Hopfner *et al.*, 2000b) indicate a structural role rather than a catalytical one. To support this hypothesis, the N-terminal part of the coil region in *S. cerevisiae* proposed to correspond to the one in *T. maritima* based on sequence alignment was mutated by E. Clausing as collaborator. Amino acids Lys419-Asn424 were replaced by six different residues (Glu-Asn-Leu-Tyr-Phe-Gln; representing the nuclear inclusion a protein ("TEV protease") restriction site). This exchange originated no difference regarding the phenotype upon MMS treatment underlining a structural linker function of the coil connecting domains II and III of Mre11.

Domain III of the exonuclease consists mainly of the previously mentioned linker, and helices j and k. As can be deduced from the crystal structure in this work, the C-terminal part of the linker adjacent to helix j is arranged along helices G2 and I of SbcC. It continues to the stem of the coiled-coils of SbcC (helices F, G1 and G2) where it fades to helix j. Helix j consists of nine amino acids (Tyr351-Leu359) and is followed by a ten residue long coil that turns around 180 degrees. This enables the subsequent helix k (Leu370-Glu384) to adopt a parallel orientation compared to helix j. Both

helices are attached to the predicted interaction site of Rad50 (Hopfner *et al.*, 2001) located at the coiled-coils close to the globular domain resulting in a four helix bundle. Structural analysis indicates that the linker and the C-terminal helices (j and k) of domain III of SbcD as well as the interacting helices of SbcC (F, G1 and G2 (coiled-coils) and E and I) are involved in SbcC-SbcD association. These secondary structure elements exhibit vast hydrophobic surfaces. In case of SbcD amino acids Leu346 and Leu349 (linker), Phe352, Leu354, Phe355 and Leu359 (helix j) and Leu369, Leu370, Ile372, Leu373, Leu376, Leu377 and Val380 (helix k) point in direction to hydrophobic interaction partners of SbcC (Phe164, Leu170, Leu173, Leu176, Leu177, Met181 and Leu184 (helix F), Leu701, Leu703, Leu704, Leu708 and Phe709 (helix G1), Phe714, Phe718 and Ala725 (helix H) as well as Val786 and Ala787). In respect to the crystal structure these amino acids form the interaction region between SbcC and the C-terminal part of SbcD. Additionally, an interaction site between lobe II of the globular part of SbcC and domain II of SbcD is apparent. Both molecules exhibit uncharged surfaces at these areas with hydrophobic residues Trp758, Gly759 and Ile760 (SbcC) and Leu275 (SbcD) protruding to the opposing protein. However these few residues at this interacting site seem not to play a dominant role in connecting SbcC with SbcD. It was shown that upon ATP binding Rad50 performs conformational changes leading to dimerization of the globular head domains (Hopfner *et al.*, 2000b). In the course of dimerization a rigid interaction between lobe II of SbcC and domain II of SbcD would be hindering suggesting a loose binding at this contact area.

As can be concluded from the bacterial structure, the association function of the linker is probably true for the yeast complex as well. Structure guided deletion of the C-terminal helices of Mre11 in *S. cerevisiae* (Met1-Ser429) resulted in a phenotype corresponding to the deletion of Mre11, indicating domain III to bear an essential function. Structure based sequence alignment in addition to secondary structure prediction suggest the *S. cerevisiae* Rad50-interacting helices of Mre11 between residues Glu443-Lys456 and Glu464-Thr500. In fact, mutation of the aromatic phenylalanine 474, which is highly conserved in eukaryotes, to a polar basic arginine turned out to be lethal (E. Clausing, personal communication). Additionally, another *S. cerevisiae* Mre11 mutant (M1-G559) that lacks 134 amino acids at the C-terminus but

comprises the predicted binding helices is still able to bind Rad50 (Chamankhah *et al.*, 2000) supporting this hypothesis.

Regarding the interaction of SbcC and SbcD or Rad50 and Mre11, respectively, another observation needs to be assessed. As could be noticed in the *P. furiosus* Rad50 structure (PDB-ID 1II8, Hopfner *et al.*, 2001) the helix that composes the C-terminal part of the coiled-coil ranges from the incoming coiled-coil region to the base of the globular region of Rad50. It is minimally bended but the structure remains helical. This is also true for the N-terminal coiled-coil forming helix, however it ends at the top of the ABC ATPase domain. This arrangement is not observed in the SbcC structure of *T. maritima* (figure 40). The N-terminal coil (helix F) exhibits a similar spatial orientation compared to its archaeal ortholog but the structure of the C-terminal coil (helices G1 and G2) adjacent to the globular part differs between *P. furiosus* and *T. maritima*. The helix is kinked by approximately 30 degrees as can be concluded from the structural model represented by PyMOL. The kink spans the amino acids Lys711-Arg716 that also interact with helix j of SbcD. It is now questionable why this kink is apparent. Since there has no other bacterial SbcC structure been published so far it is not possible to compare the structures amongst the same kingdom of life to find out if the distorted helix is common for bacteria in general or for *T. maritima* in particular. Because there is no biological advantage apparently recognizable, the biological reason for the kink remains elusive. A possible explanation could be an induced distortion resulting through interaction with SbcD. The distortion might ease and/or enforce the interaction. To deny or admit this hypothesis it would be useful to solve the structure of *T. maritima* SbcC as a homodimer without SbcD as binding partner.
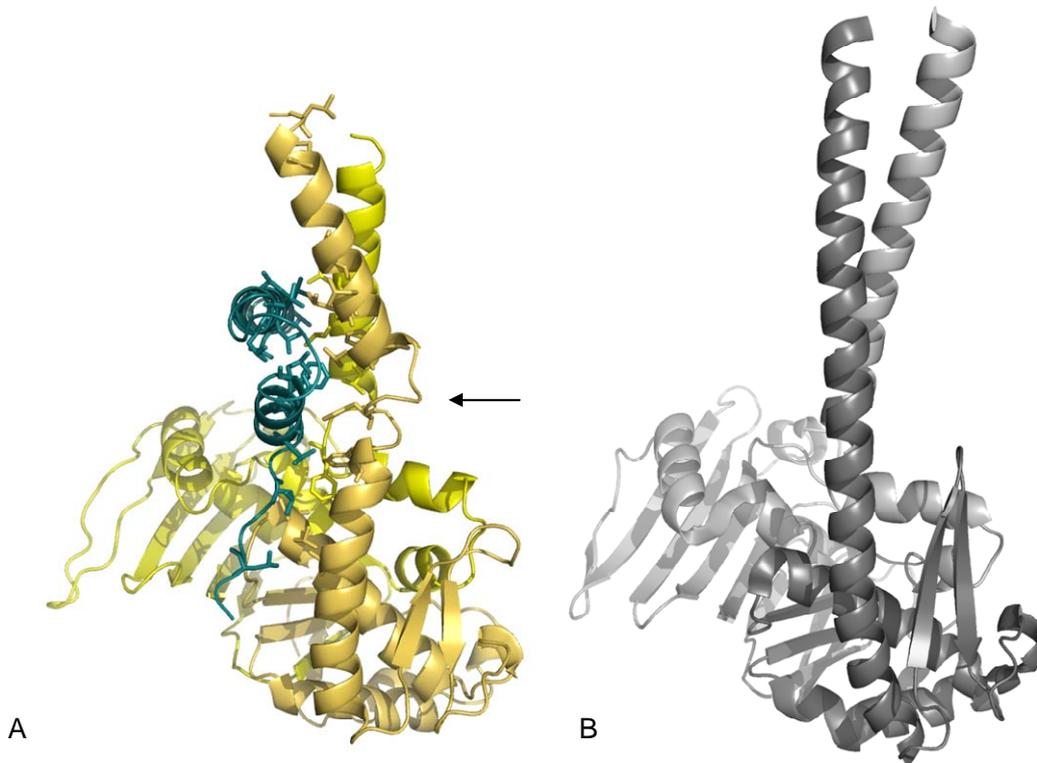
Figure 40: Comparison of the SbcC-SbcD/Rad50-Mre11 interaction region of *T. maritima* and *P. furiosus*.
Ribbon models of B) Rad50 of *P. furiosus* (PDB-ID 1II8, Hopfner *et al.*, 2001) and of A) SbcC associated with the C-terminal domain III of SbcD of *T. maritima* (left) are depicted. The structures are orientated based on the PyMOL alignment algorithm and represented using the color code as before. The kink (arrow) observed in the ribbon model of *T. maritima* SbcC is located between helices G1 and G2 of the C-terminal coiled-coil close to the interacting helix j of SbcD.

Further but minor differences in regard to the comparison of SbcC/Rad50 of *T. maritima* and *P. furiosus* (PDB-ID 1F2T, Hopfner *et al.* 2000) were observed at amino acid residues Glu107-Lys109, Ile760-Arg762 and Glu847-Val850 of *T. maritima*. The corresponding residues of *P. furiosus* exhibit β-sheet formation, which is not the case in the bacterial structure. However, the residues mentioned are arranged very similarly. Additionally, the amino acids of *P. furious* (Lys785-Arg787) corresponding to *T. maritima* Ile760-Arg762 get represented differently between the available *P. furiosus* Rad50 structures (PDB-ID 1F2T (ATPase domain), Hopfner *et al.* 2000 and PDB-ID 1II8 (ATPase domain with truncated coiled-coils), Hopfner *et al.* 2001). The *T. maritima* SbcCD and the *P. furiosus* structure of the ATPase domain with truncated coiled-coils (PDB-ID 1II8) were solved to limiting resolutions of 3.4 Å and 3.0 Å (R_{obs} 0.255, R_{free} 0.294) respectively, compared to the 1.6 Å (R_{obs} 0.187, R_{free} 0.231)

resolution of the *P. furiosus* ATPase domain structure deposited under the PDB-ID 1F2T. The lower resolution structures might not be modeled perfectly in a recognizable fashion for PyMOL but in fact form sheets in these regions.

As for SbcC/Rad50, the difference between the bacterial and the archaeal structure described by the root-mean-square deviation is quite small for SbcD/Mre11 as well (RMSD$_{SbcC/Rad50}$: 1.41 Å; RMSD$_{SbcD/Mre11}$: 1.44 Å). However, in case of SbcC/Rad50 the globular domains under investigation possess nearly the same molecular weight (41 kDa), whereas archaeal Mre11 is larger than its bacterial ortholog. Archaeal full-length Mre11 has a molecular weight of 49 kDa and is thereupon about 4 kDa larger than the corresponding nuclease of *T. maritima*. This size difference (426 *versus* 385 amino acids) is noticeable by three additional β-sheets in *P. furiosus* Mre11 (figure 41). Amino acids Leu119-Arg123 forming sheet 5 (as denoted in Hopfner *et al*. 2001) are located at an exposed position of domain I. Interestingly, this part not apparent in *T. maritima* harbors the Pro162Ser mutation in yeast (mre11(ts)) and is absolutely conserved among eukaryotes (Chamankhah *et al.*, 2000; Chamankhah and Xiao, 1998). It was shown to be a temperature sensitive separation-of-function mutation with respect to DNA repair and telomere maintenance that abolished the ability of Mre11 to associate tightly with Rad50 and Xrs2 as shown by β-galactosidase filter assays (Chamankhah *et al.*, 2000). However, based on structural information derived from this work it is unlikely that this mutation approximately 75 Å apart from the detected SbcCD association site is directly involved in Rad50-Mre11 interaction. A role in Xrs2 association could be possible, which would be supported by the fact, that the sheet containing this residue is not apparent in bacteria and not conserved in archaea, which as prokaryotes do not possess Xrs2/Nbs1. Due to the observation that the Pro162Ser mutation leads to temperature sensitivity of DNA double-strand break repair in yeast a function in DNA interaction could be possible as well. But then the question rises how bacteria perform processing of DNA while lacking this region. Structures of co-complexes of Mre11 with Xrs2 and/or DNA and further biochemical characterization of Mre11(ts) would be helpful to address this issue. In addition to the difference just described the protein region containing β-sheets 12 and 13 in Mre11 of *P. furiosus* (Glu235-Tyr249) is not apparent in *T. maritima* as well. Sequence related these residues belong to domain I but they fold into domain II resulting in a 5-stranded β-sheet and two α-helices composed domain.

This leads to an attachment of domain I to domain II based on three peptide linker chains. In *T. maritima* domain II is formed by three β-sheets (13, 14 and 15) and three helices (g, h and i), where the additional helix i results from a longer construct since *P. furiosus* Mre11 was truncated upstream of the respective helix. The two missing β-sheets of domain II result in a connection to domain I by only one coil. Since domain II partially caps the active site phophodiesterase motif of domain I a role in DNA substrate specificity was suggested (Hopfner *et al.*, 2001). The rotational flexibility between domains I and II in *P. furiosus* observed by comparing the two molecules in the asymmetric unit lead to the hypothesis that this flexibility may enable Mre11's binding to different DNA substrates. With only one connecting peptide linker the flexibility of domain II of *T. maritima* would be higher. As eukaryotic Mre11 is already able to deal with a broad range of DNA structures like hairpins, ssDNA and dsDNA ends and since it has not been reported that bacteria are capable to process different DNA substrates as eukaryotes or archaea are able to an additional biological reason due to a higher flexibility is not obvious. Biochemical characterization of deletion mutants in yeast lacking these β-sheets could give a hint. As pointed out above, the SbcD C-terminus is directly bound to the SbcC coiled-coils. Domain II is located between domain I containing the nuclease active site and domain III as the interacting region with SbcC. Despite a role in defining DNA substrate specificity it might also act as a mechanistic signal transmitter that changes its spatial orientation upon the DNA binding state. However, this signal transduction needs to be carried out in bacteria as well as in higher organisms and does not explain the differences of domains II observed in *P. furiosus* and *T. maritima*.
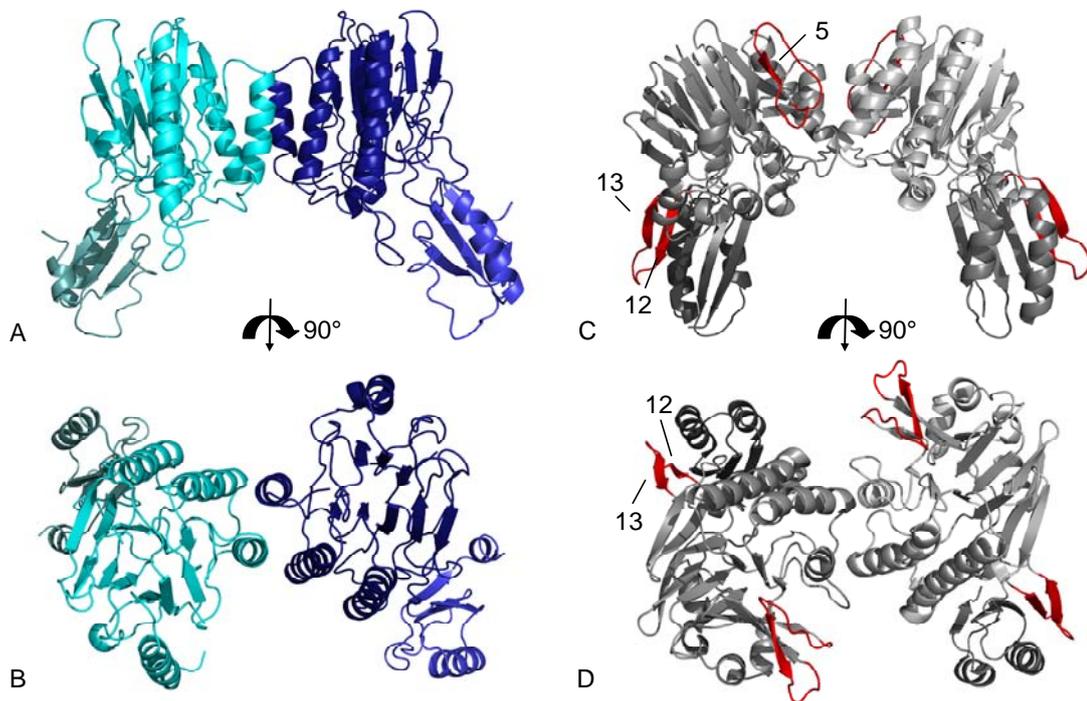
Figure 41: Comparison of *T. maritima* SbcD with *P. furiosus* Mre11.
Crystal structures of domains I and II of the SbcD/Mre11 homodimers of *T. maritima* (A and B) and *P. furiosus* (C and D) are presented. To emphasize the main differences between the archaeal and bacterial protein, regions apparent in *P. furiosus* but lacking in *T. maritima* were highlighted (red). The corresponding yeast Mre11(ts) mutation site (Pro162Ser) is estimated to be located adjacent to the archaeal sheet 5 in downstream direction. For better visualization the front (A and C) and the top view (B and D) are depicted.

Comparison of the experimentally obtained small angle X-ray scattering curve with the theoretical scattering generated from the atomic coordinates of the crystal structure reveals a very high similarity. Analyzing the Guinier plot (data not shown) and the constitutive SAXS electron pair distribution function thereupon indicates that the complex exhibits an elongated shape with maximum particle dimensions of about 210 Å that corresponds very well to the crystallographic model with a maximum diameter of approximately 210 Å as well. Based on the small angle X-ray scattering curve several three dimensional models were calculated and averaged. This averaged model then was superimposed with the crystal structure. The resulting overlay clearly demonstrates the similarity of the in solution and the crystal structure, fortifying the models obtained by different methods. Additionally, the SAXS findings underline that the discussed homodimerization of SbcD is not an effect of crystal packing but rather an arrangement that occurs in solution as well.

Summarizing, this structure allows an atomic resolution view on the SbcCD complex bringing together the so far known individual structures of Rad50 and Mre11. Their spatial arrangements in respect to each other and the identified interaction sites allow a more sophisticated model development of the mechanism of action of this protein complex (figure 42). As discussed in 6.5.1.1 and earlier in this chapter the SbcCD complex exhibits an elongated w-shaped form with a maximum particle diameter of about 210 Å. In previous work it was reported that the N and C termini of one Rad50 molecule stably associate to form a bipartite ATP-binding cassette ATPase domain (Hopfner *et al.*, 2000b). This bipartite ATPase dimerizes in an ATP-dependent manner with a second Rad50 molecule as was shown by analytical ultracentrifugation and dynamic light scattering. Additionally, the crystal structures were solved displaying Rad50 on one hand as a single molecule or on the other hand in the ATP-bound state as a homodimer that sandwiches the nucleotide. Here, we examined the dimerization by small angle X-ray scattering. Based on the observed reduction of the radius of gyration and the maximum particle diameter the complex was found to become more globular upon ATP-binding. However, the calculated models do not support a clear dimerization of the ATPase domains. A reason for this ambiguity can be the technique *per se* since SAXS measures the scattering pattern of the whole protein solution. Since ATPγS was added to the samples prior to their shipment abroad it might got hydrolyzed during this period. As a result this solution might have contained both complex species, the elongated and the dimerized one. Averaging the data then could have led to a model representing an artificial intermediate state. As it is shown by X-ray crystallography in this work, in the DNA- and ATP-unbound state the two homodimerization surfaces of SbcC are about 100 Å apart from each other. These results compared with the individual Rad50 structures are not contradictory but rather more might form the basis for explaining the mechanism of action of the SbcCD/Rad50-Mre11 complex: Upon ATP-binding the globular SbcC/Rad50 parts that flanked SbcD/Mre11 in the ATP-unbound and elongated state perform a large conformational change and homodimerize. This will result in an oval structure where the SbcD/Mre11 dimer is located at one tip and the dimerized SbcC/Rad50 ATPase domains at the opposing other end. The sides are represented by the coiled-coils connected to the SbcD/Mre11 molecules *via* domains II of the nucleases. This model would exhibit a central hole, which would be consistent

with electron micrographs (Hopfner *et al.*, 2001). The nuclease active sites of SbcD/Mre11 would be directed toward the center allowing DNA binding and processing.

It was shown previously that individual Rad50 binds DNA *in vitro* as well. An electropositive surface area on top of the dimerized ATPase domains close to the stem of the coiled-coils was suggested to be an additional DNA binding site (Hopfner *et al.*, 2000b). In the model, which is postulated here, this binding site would be oriented to the central hole as well. The C-terminal linker of domain III of SbcD that is bound to the coiled-coils could function as a bidirectional signal transmitter. Either SbcC/Rad50 ATPase domain dimerization or possible domain II movements of SbcD/Mre11 upon DNA binding could be communicated to the catalytic part of the protein complex partner. Additionally, conformational changes of the SbcCD/Rad50-Mre11 head domain could be transmitted to the coiled-coils resulting in large-scale structural alterations as was observed upon DNA binding by atomic force microscopy (Moreno-Herrero *et al.*, 2005). Initial findings that poorly-hydrolyzable ATP analogs do not alter these dynamic processes (Moreno-Herrero *et al.*, 2005) imply that conformational changes in the Rad50 ATPase do not modulate large-scale MRN conformational rearrangements and DNA tethering. However, studies in *P. furiosus*, *S. cerevisiae* and *H. sapiens* suggest that MRN complexes possess both Rad50 ATPase and adenylate kinase catalytic activities (Bhaskara *et al.*, 2007). Specific inhibitors that block adenylate kinase but not ATPase activity abolish DNA tethering in *Xenopus* cell-free egg extracts, indicating that Rad50 nucleotide-bound states may in deed mediate DNA bridging reactions (Bhaskara *et al.*, 2007; Williams and Tainer, 2007).
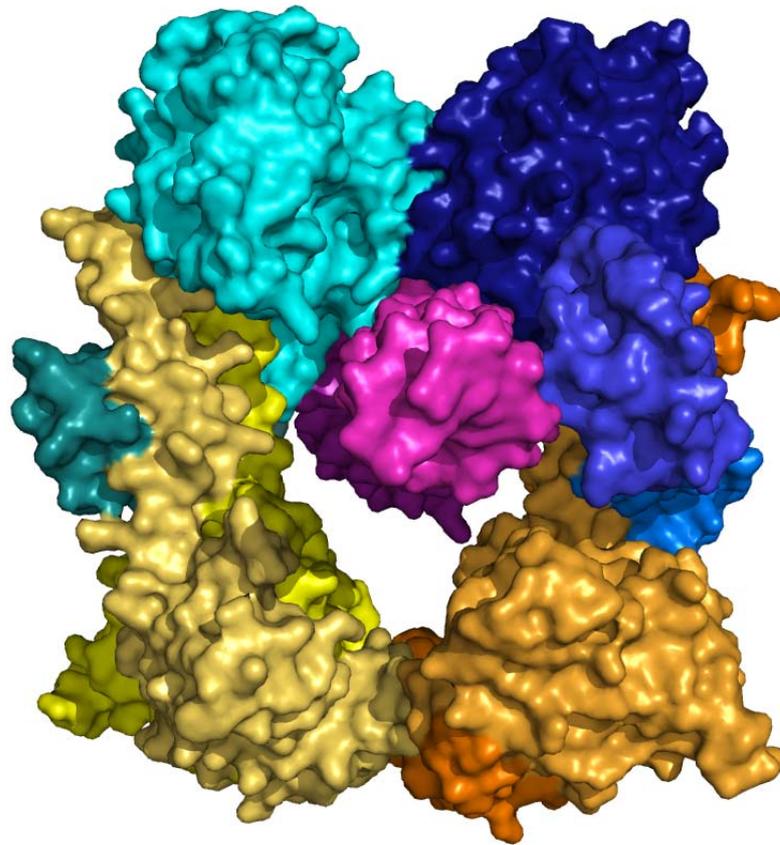
Figure 42: Hypothetical model of SbcCD/Rad50-Mre11 in complex with DNA.
Based on the crystal and SAXS structures solved in this work and on the known Rad50 structure in dimerized conformation (Hopfner *et al.*, 2000b) supplemented by electron microscopy data (Hopfner *et al.*, 2001) an integrative model is postulated. Upon SbcC/Rad50 dimerization a ring-shaped protein complex forms with a cavity suitable to harbor DNA strands requiring nucleotic processing by SbcD/Mre11. The DNA strands are oriented perpendicular to the ring and are represented in light and dark magenta, whereas the other molecules and domains are colored as before.

Overall, the MRN complex plays a central role in DNA double-strand repair based on its diverse functions and features like DNA damage recognition, protein-protein interaction domains, nuclease motifs, ATP-driven conformational switches and an ATPase motor for opening DNA at breaks. The structural connections within the MRN complex act to reduce the release of toxic and mutagenic DNA intermediates and to coordinate molecular handovers in repair to avoid disrupting processes like replication. Participation in multiple pathways underlines the crucial role of the complex. In

complex repair events, the multivalency provided by linking weak interactions together leads to high specificity, even when individual dynamic interactions are of lower specificity (Williams *et al.*, 2007). These weak interactions might be blocked in the MRN complex by small molecule inhibitors or directly targeted by gene therapy strategies as it was shown in human head and neck squamous cell carcinomas (Rhee *et al.*, 2007). Such cancer therapeutics, aiming at impairing DNA double-strand break repair pathways, might be used as supportive agents to sensitize highly proliferating tumor cells to chemotherapy and radiation therapies.

8   Outlook

Still, one of the unanswered questions is how SbcCD/Rad50-Mre11 associate with DNA. It would be interesting to learn more about the overall structure and also about the catalytic activities of this complex. Generating well diffracting and robust crystals will be a challenging task since even the complex lacking DNA molecules required evaluation of about 65,000 different crystallization conditions taking different constructs from different organisms amongst other factors in account. Having solved the structure the obstacles in crystallizing of the complex became evident. Not only the coiled-coils, as were expected, seem to pose a challenge for crystallization but also the flexible linker between domains II and III of SbcD/Mre11 might disturb crystal formation. Additionally, the overall flexibility as suggested to be necessary for biological function might have played a major role in hindering crystallization. The crystals generated consisting of SbcC in complex with DNA did not facilitate structure determination so far and need to be refined as well.

Ongoing experiments try to address the proposed ring-formation using small angle X-ray scattering and X-ray crystallography. Addition of ADP-beryllium fluoride should mimic the ATP-bound state, where beryllium fluoride inhibits the ATPase activity by tightly binding in conjunction with ADP to the active ATPase site of SbcC/Rad50. The active site would act as a template to facilitate coordinated covalent binding between ADP and beryllium fluoride, which would result in trapping the complex in the closed formation. In addition, further yeast mutants are under construction. Exchange of amino acids in yeast likely acting in Mre11 homodimerization and in Rad50/Mre11 interaction but not having been assayed so far should further elucidate the extent of involvement of certain residues and secondary structure elements in protein-protein associations. The physiological importance of the amino acids linking domains II and III of Mre11 is going to be under more detailed examination as well. The physical link between Mre11 and Rad50 should be abolished by dissecting the linker. The amino acids Glu-Asn-Leu-Tyr-Phe-Gln that exchanged residues Lys419-Asn424 represent the restriction site of the nuclear inclusion a protein (NIa, "TEV protease") of the tobacco etch virus. As discussed earlier in this chapter, the simple substitution of residues of the linker resulted

in no phenotypical change in yeast. Induced heterologous expression of the TEV protease should dissect the connection between the nucleolytic active part of Mre11 and Rad50. A change of the phenotype would give hints about the importance of an orchestrated mechanism of action of the Rad50/Mre11 complex compared to individual operation of the two proteins.

Prospective structural research will also deal with the third member of the MRN complex. The structure for Nbs1/Xrs2 is still unknown like the interaction architecture with its binding partners. In addition, in eukaryotes the structural biochemistry of DNA damage responses must be connected to the cell biology of chromatin dynamics. However, discussion of these issues would go beyond the scope of this work but research in this field will remain a hot topic in science.

9   References

Ahnesorg, P., Smith, P., and Jackson, S. P. (2006). XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. Cell *124*, 301-313.

Alani, E., Padmore, R., and Kleckner, N. (1990). Analysis of wild-type and rad50 mutants of yeast suggests an intimate relationship between meiotic chromosome synapsis and recombination. Cell *61*, 419-436.

Anderson, D. G., and Kowalczykowski, S. C. (1997a). The recombination hot spot chi is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. Genes Dev *11*, 571-581.

Anderson, D. G., and Kowalczykowski, S. C. (1997b). The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. Cell *90*, 77-86.

Aravind, L., and Koonin, E. V. (2001). Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. Genome Res *11*, 1365-1374.

Assenmacher, N., and Hopfner, K. P. (2004). MRE11/RAD50/NBS1: complex activities. Chromosoma *113*, 157-166.

Aten, J. A., Stap, J., Krawczyk, P. M., van Oven, C. H., Hoebe, R. A., Essers, J., and Kanaar, R. (2004). Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains. Science *303*, 92-95.

Bakkenist, C. J., and Kastan, M. B. (2003). DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. Nature *421*, 499-506.

Banin, S., Moyal, L., Shieh, S., Taya, Y., Anderson, C. W., Chessa, L., Smorodinsky, N. I., Prives, C., Reiss, Y., Shiloh, Y., and Ziv, Y. (1998). Enhanced phosphorylation of p53 by ATM in response to DNA damage. Science *281*, 1674-1677.

Baumann, P., and West, S. C. (1998). DNA end-joining catalyzed by human cell-free extracts. Proc Natl Acad Sci U S A *95*, 14066-14070.

Bender, C. F., Sikes, M. L., Sullivan, R., Huye, L. E., Le Beau, M. M., Roth, D. B., Mirzoeva, O. K., Oltz, E. M., and Petrini, J. H. (2002). Cancer predisposition and hematopoietic failure in Rad50(S/S) mice. Genes Dev *16*, 2237-2251.

Bhaskara, V., Dupre, A., Lengsfeld, B., Hopkins, B. B., Chan, A., Lee, J. H., Zhang, X., Gautier, J., Zakian, V., and Paull, T. T. (2007). Rad50 adenylate kinase activity regulates DNA tethering by Mre11/Rad50 complexes. Mol Cell *25*, 647-661.

Bi, X., Wei, S. C., and Rong, Y. S. (2004). Telomere protection without a telomerase; the role of ATM and Mre11 in Drosophila telomere maintenance. Curr Biol *14*, 1348-1353.

Bleuit, J. S., Xu, H., Ma, Y., Wang, T., Liu, J., and Morrical, S. W. (2001). Mediator proteins orchestrate enzyme-ssDNA assembly during T4 recombination-dependent DNA replication and repair. Proc Natl Acad Sci U S A *98*, 8298-8305.

Borde, V., Lin, W., Novikov, E., Petrini, J. H., Lichten, M., and Nicolas, A. (2004). Association of Mre11p with double-strand break sites during yeast meiosis. Mol Cell *13*, 389-401.

Boulton, S. J., and Jackson, S. P. (1998). Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. Embo J *17*, 1819-1828.

Breen, A. P., and Murphy, J. A. (1995). Reactions of oxyl radicals with DNA. Free Radic Biol Med *18*, 1033-1077.

Bressan, D. A., Baxter, B. K., and Petrini, J. H. (1999). The Mre11-Rad50-Xrs2 protein complex facilitates homologous recombination-based double-strand break repair in *Saccharomyces cerevisiae*. Mol Cell Biol *19*, 7681-7687.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr D Biol Crystallogr *59*, 2023-2030.

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.*, et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54 (Pt 5)*, 905-921.

Bryson, K., McGuffin, L. J., Marsden, R. L., Ward, J. J., Sodhi, J. S., and Jones, D. T. (2005). Protein structure prediction servers at University College London. Nucleic Acids Res *33*, W36-W38.

Buck, D., Malivert, L., de Chasseval, R., Barraud, A., Fondaneche, M. C., Sanal, O., Plebani, A., Stephan, J. L., Hufnagel, M., le Deist, F.*, et al.* (2006). Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. Cell *124*, 287-299.

Cadet, J., Berger, M., Douki, T., and Ravanat, J. L. (1997). Oxidative damage to DNA: formation, measurement, and biological significance. Rev Physiol Biochem Pharmacol *131*, 1-87.

Carney, J. P., Maser, R. S., Olivares, H., Davis, E. M., Le Beau, M., Yates, J. R., 3rd, Hays, L., Morgan, W. F., and Petrini, J. H. (1998). The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. Cell *93*, 477-486.

CCP4, N. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr *50*, 760-763.

Chalker, A. F., Leach, D. R., and Lloyd, R. G. (1988). *Escherichia coli* sbcC mutants permit stable propagation of DNA replicons containing a long palindrome. Gene *71*, 201-205.

Chamankhah, M., Fontanie, T., and Xiao, W. (2000). The *Saccharomyces cerevisiae* mre11(ts) allele confers a separation of DNA repair and telomere maintenance functions. Genetics *155*, 569-576.

Chamankhah, M., and Xiao, W. (1998). Molecular cloning and genetic characterization of the *Saccharomyces cerevisiae* NGS1/MRE11 gene. Curr Genet *34*, 368-374.

Chen, L., Trujillo, K., Ramos, W., Sung, P., and Tomkinson, A. E. (2001). Promotion of Dnl4-catalyzed DNA end-joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 complexes. Mol Cell *8*, 1105-1115.

Ciapponi, L., Cenci, G., Ducau, J., Flores, C., Johnson-Schlitz, D., Gorski, M. M., Engels, W. R., and Gatti, M. (2004). The Drosophila Mre11/Rad50 complex is required to prevent both telomeric fusion and chromosome breakage. Curr Biol *14*, 1360-1366.

Connelly, J. C., de Leau, E. S., and Leach, D. R. (1999). DNA cleavage and degradation by the SbcCD protein complex from *Escherichia coli*. Nucleic Acids Res *27*, 1039-1046.

Connelly, J. C., de Leau, E. S., and Leach, D. R. (2003). Nucleolytic processing of a protein-bound DNA end by the *E. coli* SbcCD (MR) complex. DNA Repair (Amst) *2*, 795-807.

Connelly, J. C., de Leau, E. S., Okely, E. A., and Leach, D. R. (1997). Overexpression, purification, and characterization of the SbcCD protein from *Escherichia coli*. J Biol Chem *272*, 19819-19826.

Connelly, J. C., Kirkham, L. A., and Leach, D. R. (1998). The SbcCD nuclease of *Escherichia coli* is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. Proc Natl Acad Sci U S A *95*, 7969-7974.

Connelly, J. C., and Leach, D. R. (1996). The sbcC and sbcD genes of *Escherichia coli* encode a nuclease involved in palindrome inviability and genetic recombination. Genes Cells *1*, 285-291.

Costanzo, V., Paull, T., Gottesman, M., and Gautier, J. (2004). Mre11 assembles linear DNA fragments into DNA damage signaling complexes. PLoS Biol *2*, E110.

Costanzo, V., Robertson, K., Bibikova, M., Kim, E., Grieco, D., Gottesman, M., Carroll, D., and Gautier, J. (2001). Mre11 protein complex prevents double-strand break accumulation during chromosomal DNA replication. Mol Cell *8*, 137-147.

D'Amours, D., and Jackson, S. P. (2002). The Mre11 complex: at the crossroads of DNA repair and checkpoint signalling. Nat Rev Mol Cell Biol *3*, 317-327.

Darmon, E., Lopez-Vernaza, M. A., Helness, A. C., Borking, A., Wilson, E., Thacker, Z., Wardrope, L., and Leach, D. R. (2007). SbcCD regulation and localization in *Escherichia coli*. J Bacteriol *189*, 6686-6694.

de Jager, M., Trujillo, K. M., Sung, P., Hopfner, K. P., Carney, J. P., Tainer, J. A., Connelly, J. C., Leach, D. R., Kanaar, R., and Wyman, C. (2004). Differential arrangements of conserved building blocks among homologs of the Rad50/Mre11 DNA repair protein complex. J Mol Biol *339*, 937-949.

de Jager, M., van Noort, J., van Gent, D. C., Dekker, C., Kanaar, R., and Wyman, C. (2001). Human Rad50/Mre11 is a flexible complex that can tether DNA ends. Mol Cell *8*, 1129-1135.

de Jager, M., Wyman, C., van Gent, D. C., and Kanaar, R. (2002). DNA end-binding specificity of human Rad50/Mre11 is influenced by ATP. Nucleic Acids Res *30*, 4425-4431.

de Lange, T. (2001). Cell biology. Telomere capping--one strand fits all. Science *292*, 1075-1076.

de Lange, T., and Petrini, J. H. (2000). A new connection at human telomeres: association of the Mre11 complex with TRF2. Cold Spring Harb Symp Quant Biol *65*, 265-273.

DeFazio, L. G., Stansel, R. M., Griffith, J. D., and Chu, G. (2002). Synapsis of DNA ends by DNA-dependent protein kinase. Embo J *21*, 3192-3200.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System (Palo Alto: DeLano Scientific).

den Blaauwen, T., Aarsman, M. E., Wheeler, L. J., and Nanninga, N. (2006). Pre-replication assembly of E. coli replisome components. Mol Microbiol *62*, 695-708.

Desai-Mehta, A., Cerosaletti, K. M., and Concannon, P. (2001). Distinct functional domains of nibrin mediate Mre11 binding, focus formation, and nuclear localization. Mol Cell Biol *21*, 2184-2191.

Diederichs, K., and Karplus, P. A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol *4*, 269-275.

Difilippantonio, S., Celeste, A., Fernandez-Capetillo, O., Chen, H. T., Reina San Martin, B., Van Laethem, F., Yang, Y. P., Petukhova, G. V., Eckhaus, M., Feigenbaum, L., *et al.* (2005). Role of Nbs1 in the activation of the Atm kinase revealed in humanized mouse models. Nat Cell Biol *7*, 675-685.

Dillingham, M. S., Spies, M., and Kowalczykowski, S. C. (2003). RecBCD enzyme is a bipolar DNA helicase. Nature *423*, 893-897.

Doherty, A. J., Jackson, S. P., and Weller, G. R. (2001). Identification of bacterial homologues of the Ku DNA repair proteins. FEBS Lett *500*, 186-188.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Engh, R. H., R. (1991). Accurate bond and angle parameters for x-ray protein structure refinement. Acta Crystallogr D Biol Crystallogr *A47*, 392-400.

Farah, J. A., Hartsuiker, E., Mizuno, K., Ohta, K., and Smith, G. R. (2002). A 160-bp palindrome is a Rad50.Rad32-dependent mitotic recombination hotspot in *Schizosaccharomyces pombe*. Genetics *161*, 461-468.

Franklin, W. A., Doetsch, P. W., and Haseltine, W. A. (1985). Structural determination of the ultraviolet light-induced thymine-cytosine pyrimidine-pyrimidone (6-4) photoproduct. Nucleic Acids Res *13*, 5317-5325.

Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A., and Ellenberger, T. (2006). DNA repair and mutagenesis, 2nd ed. edn (Washington D.C.: ASM Press).

Froelich-Ammon, S. J., and Osheroff, N. (1995). Topoisomerase poisons: harnessing the dark side of enzyme mechanism. J Biol Chem *270*, 21429-21432.

Galkin, V. E., Esashi, F., Yu, X., Yang, S., West, S. C., and Egelman, E. H. (2005). BRCA2 BRC motifs bind RAD51-DNA filaments. Proc Natl Acad Sci U S A *102*, 8537-8542.

Gallego, M. E., and White, C. I. (2001). RAD50 function is essential for telomere maintenance in Arabidopsis. Proc Natl Acad Sci U S A *98*, 1711-1716.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res *31*, 3784-3788.

George, J. W., Stohr, B. A., Tomso, D. J., and Kreuzer, K. N. (2001). The tight linkage between DNA replication and double-strand break repair in bacteriophage T4. Proc Natl Acad Sci U S A *98*, 8290-8297.

Gibson, F. P., Leach, D. R., and Lloyd, R. G. (1992). Identification of sbcD mutations as cosuppressors of recBC that allow propagation of DNA palindromes in *Escherichia coli* K-12. J Bacteriol *174*, 1222-1228.

Gorski, M. M., Romeijn, R. J., Eeken, J. C., de Jong, A. W., van Veen, B. L., Szuhai, K., Mullenders, L. H., Ferro, W., and Pastink, A. (2004). Disruption of *Drosophila* Rad50 causes pupal lethality, the accumulation of DNA double-strand breaks and the induction of apoptosis in third instar larvae. DNA Repair (Amst) *3*, 603-615.

Grawunder, U., West, R. B., and Lieber, M. R. (1998). Antigen receptor gene rearrangement. Curr Opin Immunol *10*, 172-180.

Guinier, A. (1939). La diffraction des rayons X aux petits angles; application a l'etude de phenomenes ultramicroscopiques. Ann Phys (Paris) *12* 161–231.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol *166*, 557-580.

Heyer, W. D., Ehmsen, K. T., and Solinger, J. A. (2003). Holliday junctions in the eukaryotic nucleus: resolution in sight? Trends Biochem Sci *28*, 548-557.

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene *77*, 51-59.

Hoeijmakers, J. H. (2001). Genome maintenance mechanisms for preventing cancer. Nature *411*, 366-374.

Hopfner, K. P., Craig, L., Moncalian, G., Zinkel, R. A., Usui, T., Owen, B. A., Karcher, A., Henderson, B., Bodmer, J. L., McMurray, C. T., *et al.* (2002). The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. Nature *418*, 562-566.

Hopfner, K. P., Karcher, A., Craig, L., Woo, T. T., Carney, J. P., and Tainer, J. A. (2001). Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. Cell *105*, 473-485.

Hopfner, K. P., Karcher, A., Shin, D., Fairley, C., Tainer, J. A., and Carney, J. P. (2000a). Mre11 and Rad50 from *Pyrococcus furiosus*: cloning and biochemical characterization reveal an evolutionarily conserved multiprotein machine. J Bacteriol *182*, 6036-6041.

Hopfner, K. P., Karcher, A., Shin, D. S., Craig, L., Arthur, L. M., Carney, J. P., and Tainer, J. A. (2000b). Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. Cell *101*, 789-800.

JCSG, J. C. f. S. G., and Wilson, I. A. (2007). Crystal structure of putative exonuclease (TM1635) from *Thermotoga maritima* at 2.20 A resolution (RCSB Protein Data Bank).

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol *292*, 195-202.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Cryst *26*, 795-800.

Keeney, S. (2001). Mechanism and control of meiotic recombination initiation. Curr Top Dev Biol *52*, 1-53.

Khanna, K. K., and Jackson, S. P. (2001). DNA double-strand breaks: signaling, repair and the cancer connection. Nat Genet *27*, 247-254.

Kironmai, K. M., and Muniyappa, K. (1997). Alteration of telomeric sequences and senescence caused by mutations in RAD50 of *Saccharomyces cerevisiae*. Genes Cells *2*, 443-455.

Kleywegt, G. J. (1996). Use of non-crystallographic symmetry in protein structure refinement. Acta Crystallogr D Biol Crystallogr *52*, 842-857.

Kobayashi, J., Antoccia, A., Tauchi, H., Matsuura, S., and Komatsu, K. (2004). NBS1 and its functional role in the DNA damage response. DNA Repair (Amst) *3*, 855-861.

Kobayashi, J., Tauchi, H., Sakamoto, S., Nakamura, A., Morishima, K., Matsuura, S., Kobayashi, T., Tamai, K., Tanimoto, K., and Komatsu, K. (2002). NBS1 localizes to gamma-H2AX foci through interaction with the FHA/BRCT domain. Curr Biol *12*, 1846-1851.

Koch, M. H., Vachette, P., and Svergun, D. I. (2003). Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Q Rev Biophys *36*, 147-227.

Konarev, P. V., Petoukhov, M. V., Volkov, V. V., and Svergun, D. I. (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. J Appl Cryst *39*, 277–286.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003). PRIMUS: a Windows-PC based system for small-angle scattering data analysis. J Appl Crystallog *36* 1277–1282.

Kozin, M. B., and Svergun, D. I. (2001). Automated matching of high- and low-resolution structural models. J Appl Cryst *34*, 33-41.

Krejci, L., Chen, L., Van Komen, S., Sung, P., and Tomkinson, A. (2003). Mending the break: two DNA double-strand break repair machines in eukaryotes. Prog Nucleic Acid Res Mol Biol *74*, 159-201.

Kreuzer, K. N. (2000). Recombination-dependent DNA replication in phage T4. Trends Biochem Sci *25*, 165-173.

Krogh, B. O., Llorente, B., Lam, A., and Symington, L. S. (2005). Mutations in Mre11 phosphoesterase motif I that impair *Saccharomyces cerevisiae* Mre11-Rad50-Xrs2 complex stability in addition to nuclease activity. Genetics *171*, 1561-1570.

Krogh, B. O., and Symington, L. S. (2004). Recombination proteins in yeast. Annu Rev Genet *38*, 233-271.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst *26*, 283-291.

Leach, D. R. (1994). Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. Bioessays *16*, 893-900.

Lee, J. H., and Paull, T. T. (2004). Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex. Science *304*, 93-96.

Lee, S. E., Moore, J. K., Holmes, A., Umezu, K., Kolodner, R. D., and Haber, J. E. (1998). *Saccharomyces* Ku70, mre11/rad50 and RPA proteins regulate adaptation to G2/M arrest after DNA damage. Cell *94*, 399-409.

Lees-Miller, S. P., and Meek, K. (2003). Repair of DNA double strand breaks by non-homologous end joining. Biochimie *85*, 1161-1173.

LeMaster, D. M., and Richards, F. M. (1985). 1H-15N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. Biochemistry *24*, 7263-7268.

Lewis, L. K., Storici, F., Van Komen, S., Calero, S., Sung, P., and Resnick, M. A. (2004). Role of the nuclease activity of *Saccharomyces cerevisiae* Mre11 in repair of DNA double-strand breaks in mitotic cells. Genetics *166*, 1701-1713.

Lin, F. L., Sperle, K., and Sternberg, N. (1984). Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process. Mol Cell Biol *4*, 1020-1034.

Lin, F. L., Sperle, K., and Sternberg, N. (1985). Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences. Proc Natl Acad Sci U S A *82*, 1391-1395.

Lindahl, T. (1974). An N-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues. Proc Natl Acad Sci U S A *71*, 3649-3653.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709-715.

Lindsey, J. C., and Leach, D. R. (1989). Slow replication of palindrome-containing DNA. J Mol Biol *206*, 779-782.

Lisby, M., Barlow, J. H., Burgess, R. C., and Rothstein, R. (2004). Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. Cell *118*, 699-713.

Liu, Y., Masson, J. Y., Shah, R., O'Regan, P., and West, S. C. (2004). RAD51C is required for Holliday junction processing in mammalian cells. Science *303*, 243-246.

Llorente, B., and Symington, L. S. (2004). The Mre11 nuclease is not required for 5' to 3' resection at multiple HO-induced double-strand breaks. Mol Cell Biol *24*, 9682-9694.

Lloyd, R. G., and Buckman, C. (1985). Identification and genetic analysis of sbcC mutations in commonly used recBC sbcB strains of *Escherichia coli* K-12. J Bacteriol *164*, 836-844.

Lobachev, K. S., Gordenin, D. A., and Resnick, M. A. (2002). The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. Cell *108*, 183-193.

Luo, G., Yao, M. S., Bender, C. F., Mills, M., Bladl, A. R., Bradley, A., and Petrini, J. H. (1999). Disruption of mRad50 causes embryonic stem cell lethality, abnormal embryonic development, and sensitivity to ionizing radiation. Proc Natl Acad Sci U S A *96*, 7376-7381.

Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. Science *252*, 1162-1164.

Lustig, A. J. (2003). Clues to catastrophic telomere loss in mammals from yeast telomere rapid deletion. Nat Rev Genet *4*, 916-923.

Ma, Y., Lu, H., Schwarz, K., and Lieber, M. R. (2005a). Repair of double-strand DNA breaks by the human nonhomologous DNA end joining pathway: the iterative processing model. Cell Cycle *4*, 1193-1200.

Ma, Y., Lu, H., Tippin, B., Goodman, M. F., Shimazaki, N., Koiwai, O., Hsieh, C. L., Schwarz, K., and Lieber, M. R. (2004). A biochemically defined system for mammalian nonhomologous DNA end joining. Mol Cell *16*, 701-713.

Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M. R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell *108*, 781-794.

Ma, Y., Schwarz, K., and Lieber, M. R. (2005b). The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps. DNA Repair (Amst) *4*, 845-851.

Maizels, N. (2005). Immunoglobulin gene diversification. Annu Rev Genet *39*, 23-46.

Malkova, A., Ivanov, E. L., and Haber, J. E. (1996). Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. Proc Natl Acad Sci U S A *93*, 7131-7136.

Malkova, A., Naylor, M. L., Yamaguchi, M., Ira, G., and Haber, J. E. (2005). RAD51-dependent break-induced replication differs in kinetics and checkpoint responses from RAD51-mediated gene conversion. Mol Cell Biol *25*, 933-944.

Martini, E., and Keeney, S. (2002). Sex and the single (double-strand) break. Mol Cell *9*, 700-702.

Mascarenhas, J., Sanchez, H., Tadesse, S., Kidane, D., Krisnamurthy, M., Alonso, J. C., and Graumann, P. L. (2006). *Bacillus subtilis* SbcC protein plays an important role in DNA inter-strand cross-link repair. BMC Mol Biol *7*, 20.

Maser, R. S., Mirzoeva, O. K., Wells, J., Olivares, H., Williams, B. R., Zinkel, R. A., Farnham, P. J., and Petrini, J. H. (2001). Mre11 complex and DNA replication: linkage to E2F and sites of DNA synthesis. Mol Cell Biol *21*, 6006-6016.

Matsuura, S., Tauchi, H., Nakamura, A., Kondo, N., Sakamoto, S., Endo, S., Smeets, D., Solder, B., Belohradsky, B. H., Der Kaloustian, V. M.*, et al.* (1998). Positional cloning of the gene for Nijmegen breakage syndrome. Nat Genet *19*, 179-181.

McBlane, J. F., van Gent, D. C., Ramsden, D. A., Romeo, C., Cuomo, C. A., Gellert, M., and Oettinger, M. A. (1995). Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. Cell *83*, 387-395.

McGuffin, L. J., Bryson, K., and Jones, D. T. (2000). The PSIPRED protein structure prediction server. Bioinformatics *16*, 404-405.

Meile, J. C., Wu, L. J., Ehrlich, S. D., Errington, J., and Noirot, P. (2006). Systematic localisation of proteins fused to the green fluorescent protein in *Bacillus subtilis*: identification of new proteins at the DNA replication factory. Proteomics *6*, 2135-2146.

Michael, B. D., and O'Neill, P. (2000). Molecular biology. A sting in the tail of electron tracks. Science *287*, 1603-1604.

Mirzoeva, O. K., and Petrini, J. H. (2003). DNA replication-dependent nuclear dynamics of the Mre11 complex. Mol Cancer Res *1*, 207-218.

Modesti, M., Junop, M. S., Ghirlando, R., van de Rakt, M., Gellert, M., Yang, W., and Kanaar, R. (2003). Tetramerization and DNA ligase IV interaction of the DNA double-strand break repair protein XRCC4 are mutually exclusive. J Mol Biol *334*, 215-228.

Molina, F., and Skarstad, K. (2004). Replication fork and SeqA focus distributions in *Escherichia coli* suggest a replication hyperstructure dependent on nucleotide metabolism. Mol Microbiol *52*, 1597-1612.

Moore, P. B. (1980). Small-Angle Scattering. Information Content and Error Analysis. J Appl Cryst *13*, 168-175.

Moreau, S., Ferguson, J. R., and Symington, L. S. (1999). The nuclease activity of Mre11 is required for meiosis but not for mating type switching, end joining, or telomere maintenance. Mol Cell Biol *19*, 556-566.

Moreno-Herrero, F., de Jager, M., Dekker, N. H., Kanaar, R., Wyman, C., and Dekker, C. (2005). Mesoscale conformational changes in the DNA-repair complex Rad50/Mre11/Nbs1 upon binding DNA. Nature *437*, 440-443.

Morimatsu, K., and Kowalczykowski, S. C. (2003). RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. Mol Cell *11*, 1337-1347.

Namsaraev, E. A., and Berg, P. (1998). Interaction of Rad51 with ATP and Mg2+ induces a conformational change in Rad51. Biochemistry *37*, 11932-11939.

Namsaraev, E. A., and Berg, P. (2000). Rad51 uses one mechanism to drive DNA strand exchange in both directions. J Biol Chem *275*, 3970-3976.

Nicholas, K. B., and Nicholas, H. B. (1997). GeneDoc: a tool for editing and annotating multiple sequence alignment, In Distibuted by the author.

Nick McElhinny, S. A., Havener, J. M., Garcia-Diaz, M., Juarez, R., Bebenek, K., Kee, B. L., Blanco, L., Kunkel, T. A., and Ramsden, D. A. (2005). A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. Mol Cell *19*, 357-366.

Nick McElhinny, S. A., and Ramsden, D. A. (2004). Sibling rivalry: competition between Pol X family members in V(D)J recombination and general double strand break repair. Immunol Rev *200*, 156-164.

Nugent, C. I., Bosco, G., Ross, L. O., Evans, S. K., Salinger, A. P., Moore, J. K., Haber, J. E., and Lundblad, V. (1998). Telomere maintenance is dependent on activities required for end repair of double-strand breaks. Curr Biol *8*, 657-660.

Paques, F., and Haber, J. E. (1999). Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. Microbiol Mol Biol Rev *63*, 349-404.

Pastwa, E., and Blasiak, J. (2003). Non-homologous DNA end joining. Acta Biochim Pol *50*, 891-908.

Paull, T. T., and Gellert, M. (1998). The 3' to 5' exonuclease activity of Mre 11 facilitates repair of DNA double-strand breaks. Mol Cell *1*, 969-979.

Paull, T. T., and Gellert, M. (1999). Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. Genes Dev *13*, 1276-1288.

Petersen, S., Casellas, R., Reina-San-Martin, B., Chen, H. T., Difilippantonio, M. J., Wilson, P. C., Hanitsch, L., Celeste, A., Muramatsu, M., Pilch, D. R., *et al.* (2001). AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. Nature *414*, 660-665.

Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, confirmations and assemblies in solution. Quarterly Reviews in Biophysics.

Rhee, J. G., Li, D., Suntharalingam, M., Guo, C., O'Malley, B. W., Jr., and Carney, J. P. (2007). Radiosensitization of head/neck squamous cell carcinoma by adenovirus-mediated expression of the Nbs1 protein. Int J Radiat Oncol Biol Phys *67*, 273-278.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Vol 1, 2, 3 (NY).

Sartori, A. A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S. P. (2007). Human CtIP promotes DNA end resection. Nature.

Schatz, D. G., and Spanopoulou, E. (2005). Biochemistry of V(D)J recombination. Curr Top Microbiol Immunol *290*, 49-85.

Setlow, R. B. (1968). The photochemistry, photobiology, and repair of polynucleotides. Prog Nucleic Acid Res Mol Biol *8*, 257-295.

Sharples, G. J., and Leach, D. R. (1995). Structural and functional similarities between the SbcCD proteins of *Escherichia coli* and the RAD50 and MRE11 (RAD32) recombination and repair proteins of yeast. Mol Microbiol *17*, 1215-1217.

Shiloh, Y. (2003). ATM and related protein kinases: safeguarding genome integrity. Nat Rev Cancer *3*, 155-168.

Shin, D. S., Chahwan, C., Huffman, J. L., and Tainer, J. A. (2004). Structure and function of the double-strand break repair machinery. DNA Repair (Amst) *3*, 863-873.

Shivji, M. K., and Venkitaraman, A. R. (2004). DNA recombination, chromosomal stability and carcinogenesis: insights into the role of BRCA2. DNA Repair (Amst) *3*, 835-843.

Shurvinton, C. E., Stahl, M. M., and Stahl, F. W. (1987). Large palindromes in the lambda phage genome are preserved in a rec+ host by inhibiting lambda DNA replication. Proc Natl Acad Sci U S A *84*, 1624-1628.

Singer, B. (1975). The chemical effects of nucleic acid alkylation and their relation to mutagenesis and carcinogenesis. Prog Nucleic Acid Res Mol Biol *15*, 219-284.

Singer, B. (1986). O-alkyl pyrimidines in mutagenesis and carcinogenesis: occurrence and significance. Cancer Res *46*, 4879-4885.

Singleton, M. R., Dillingham, M. S., Gaudier, M., Kowalczykowski, S. C., and Wigley, D. B. (2004). Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. Nature *432*, 187-193.

Siskind, G. W., and Benacerraf, B. (1969). Cell selection by antigen in the immune response. Adv Immunol *10*, 1-50.

Spagnolo, L., Rivera-Calzada, A., Pearl, L. H., and Llorca, O. (2006). Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. Mol Cell *22*, 511-519.

Spies, M., Dillingham, M. S., and Kowalczykowski, S. C. (2005). Translocation by the RecB motor is an absolute requirement for {chi}-recognition and RecA protein loading by RecBCD enzyme. J Biol Chem *280*, 37078-37087.

Spies, M., and Kowalczykowski, S. C. (2006). The RecA binding locus of RecBCD is a general domain for recruitment of DNA strand exchange proteins. Mol Cell *21*, 573-580.

Stavnezer, J. (1996). Immunoglobulin class switching. Curr Opin Immunol *8*, 199-205.

Stewart, G. S., Maser, R. S., Stankovic, T., Bressan, D. A., Kaplan, M. I., Jaspers, N. G., Raams, A., Byrd, P. J., Petrini, J. H., and Taylor, A. M. (1999). The DNA double-strand break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia-like disorder. Cell *99*, 577-587.

Sun, H., Treco, D., Schultes, N. P., and Szostak, J. W. (1989). Double-strand breaks at an initiation site for meiotic gene conversion. Nature *338*, 87-90.

Sung, P. (1994). Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. Science *265*, 1241-1243.

Sung, P., and Robberson, D. L. (1995). DNA strand exchange mediated by a RAD51-ssDNA nucleoprotein filament with polarity opposite to that of RecA. Cell *82*, 453-461.

Svergun, D. I. (1992). Determination of the regularization parameter in indirect transform methods using perceptual criteria. J Appl Crystallog *25*, 495–503.

Svergun, D. I., Barberato, C., and Koch, M. H. J. (1995). CRYSOL: a program to evaluate X-ray solution scattering of biological macro-molecules from atomic coordinates. J Appl Crystallog *28*, 768–773.

Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001). Determination of domain structure of proteins from X-ray solution scattering. Biophys J *80*, 2946-2953.

Tauchi, H., Kobayashi, J., Morishima, K., Matsuura, S., Nakamura, A., Shiraishi, T., Ito, E., Masnada, D., Delia, D., and Komatsu, K. (2001). The forkhead-associated domain of NBS1 is essential for nuclear foci formation after irradiation but not essential for hRAD50-hMRE11-NBS1 complex DNA repair activity. J Biol Chem *276*, 12-15.

Trujillo, K. M., Roh, D. H., Chen, L., Van Komen, S., Tomkinson, A., and Sung, P. (2003). Yeast xrs2 binds DNA and helps target rad50 and mre11 to DNA ends. J Biol Chem *278*, 48957-48964.

Trujillo, K. M., and Sung, P. (2001). DNA structure-specific nuclease activities in the *Saccharomyces cerevisiae* Rad50*Mre11 complex. J Biol Chem *276*, 35458-35464.

Trujillo, K. M., Yuan, S. S., Lee, E. Y., and Sung, P. (1998). Nuclease activities in a complex of human recombination and DNA repair factors Rad50, Mre11, and p95. J Biol Chem *273*, 21447-21450.

Tsubouchi, H., and Ogawa, H. (1998). A novel mre11 mutation impairs processing of double-strand breaks of DNA during both mitosis and meiosis. Mol Cell Biol *18*, 260-268.

Tsutakawa, S. E., Hura, G. L., Frankel, K. A., Cooper, P. K., and Tainer, J. A. (2007). Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography. J Struct Biol *158*, 214-223.

Tuma, R. S., Beaudet, M. P., Jin, X., Jones, L. J., Cheung, C. Y., Yue, S., and Singer, V. L. (1999). Characterization of SYBR Gold nucleic acid gel stain: a dye optimized for use with 300-nm ultraviolet transilluminators. Anal Biochem *268*, 278-288.

Turk, D. (1992) Weiterentwicklung eines Programms fuer Molekuelgraphik und Elektrondichte-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklaerungen, PhD Thesis, Technical University, Munich.

Usui, T., Ohta, T., Oshiumi, H., Tomizawa, J., Ogawa, H., and Ogawa, T. (1998). Complex formation and functional versatility of Mre11 of budding yeast in recombination. Cell *95*, 705-716.

van Gent, D. C., Hoeijmakers, J. H., and Kanaar, R. (2001). Chromosomal stability and the DNA double-stranded break connection. Nat Rev Genet *2*, 196-206.

van Noort, J., van Der Heijden, T., de Jager, M., Wyman, C., Kanaar, R., and Dekker, C. (2003). The coiled-coil of the human Rad50 DNA repair protein contains specific segments of increased flexibility. Proc Natl Acad Sci U S A *100*, 7581-7586.

Varon, R., Vissinga, C., Platzer, M., Cerosaletti, K. M., Chrzanowska, K. H., Saar, K., Beckmann, G., Seemanova, E., Cooper, P. R., Nowak, N. J.*, et al.* (1998).

Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. Cell *93*, 467-476.

Volkov, V. V., and Svergun, D. I. (2003). Uniqueness of ab initio shape determination in small-angle scattering
J Appl Cryst *36*, 860-864.

Walker, J. R., Corpina, R. A., and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. Nature *412*, 607-614.

Ward, J. F. (1985). Biochemistry of DNA lesions. Radiat Res Suppl *8*, S103-111.

Ward, J. F. (1990). The yield of DNA double-strand breaks produced intracellularly by ionizing radiation: a review. Int J Radiat Biol *57*, 1141-1150.

Watson, J. D., and Crick, F. H. C. (1953). A Structure for Deoxyribose Nucleic Acid. Nature *171*, 737-738.

Webb, B. L., Cox, M. M., and Inman, R. B. (1997). Recombinational DNA repair: the RecF and RecR proteins limit the extension of RecA filaments beyond single-strand DNA gaps. Cell *91*, 347-356.

Weller, G. R., Kysela, B., Roy, R., Tonkin, L. M., Scanlan, E., Della, M., Devine, S. K., Day, J. P., Wilkinson, A., d'Adda di Fagagna, F., *et al.* (2002). Identification of a DNA nonhomologous end-joining complex in bacteria. Science *297*, 1686-1689.

West, S. C. (2003). Molecular views of recombination proteins and their control. Nat Rev Mol Cell Biol *4*, 435-445.

Whitby, M. C. (2005). Making crossovers during meiosis. Biochem Soc Trans *33*, 1451-1455.

Williams, R. S., and Tainer, J. A. (2007). Learning our ABCs: Rad50 directs MRN repair functions via adenylate kinase activity from the conserved ATP binding cassette. Mol Cell *25*, 789-791.

Williams, R. S., Williams, J. S., and Tainer, J. A. (2007). Mre11-Rad50-Nbs1 is a keystone complex connecting DNA repair machinery, double-strand break signaling, and the chromatin template. Biochem Cell Biol *85*, 509-520.

Wiltzius, J. J., Hohl, M., Fleming, J. C., and Petrini, J. H. (2005). The Rad50 hook domain is a critical determinant of Mre11 complex functions. Nat Struct Mol Biol *12*, 403-407.

Wu, L., and Hickson, I. D. (2003). The Bloom's syndrome helicase suppresses crossing over during homologous recombination. Nature *426*, 870-874.

Wyman, C., and Kanaar, R. (2006). DNA double-strand break repair: all's well that ends well. Annu Rev Genet *40*, 363-383.

Xiao, Y., and Weaver, D. T. (1997). Conditional gene targeted deletion by Cre recombinase demonstrates the requirement for the double-strand break repair Mre11 protein in murine embryonic stem cells. Nucleic Acids Res *25*, 2985-2991.

Xu, B., O'Donnell, A. H., Kim, S. T., and Kastan, M. B. (2002). Phosphorylation of serine 1387 in Brca1 is specifically required for the Atm-mediated S-phase checkpoint after ionizing irradiation. Cancer Res *62*, 4588-4591.

Yamaguchi-Iwai, Y., Sonoda, E., Sasaki, M. S., Morrison, C., Haraguchi, T., Hiraoka, Y., Yamashita, Y. M., Yagi, T., Takata, M., Price, C., *et al.* (1999). Mre11 is essential for the maintenance of chromosomal DNA in vertebrate cells. Embo J *18*, 6619-6629.

Yoo, S., and Dynan, W. S. (1999). Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. Nucleic Acids Res *27*, 4679-4686.

Zhu, J., Petersen, S., Tessarollo, L., and Nussenzweig, A. (2001). Targeted disruption of the Nijmegen breakage syndrome gene NBS1 leads to early embryonic lethality in mice. Curr Biol *11*, 105-109.

Zhu, X. D., Kuster, B., Mann, M., Petrini, J. H., and de Lange, T. (2000). Cell-cycle-regulated association of RAD50/MRE11/NBS1 with TRF2 and human telomeres. Nat Genet *25*, 347-352.

Zhuo, S., Clemens, J. C., Stone, R. L., and Dixon, J. E. (1994). Mutational analysis of a Ser/Thr phosphatase. Identification of residues important in phosphoesterase substrate binding and catalysis. J Biol Chem *269*, 26234-26238.

10   Abbreviations

ABC – ATP binding cassette
AER – alternative excision repair
Å – angstrom
ATM – ataxia-telangiectasia mutated
ATP – adenosine triphosphate
BIR – break-induced repair
BER – base excision repair
bp – base pairs
BRCT – breast cancer C-terminus
CPD – cyclobutane pyrimidine dimers
DNA – deoxyribonucleic acid
DNA-PKcs – DNA-dependent protein kinase catalytic subunit
DSB – double-strand break
DSBR – double-strand break repair
dsDNA – double-stranded DNA
e. g. – *exempli gratia* (for example)
E. coli – Escherichia coli
EDTA – Ethylenediaminetetraacetic acid
EM – electron microscopy
EMSA – electrophoretic mobility shift assay
FHA – forkhead associated
FT – Fourier transformation
g – gram
g – multiples of the earth's gravitational field (centrifugation)
HJ – Holliday junction
HR – homologous recombination
i. e. – *id est* (that is)
K – Kelvin
kDa – kilo Dalton
NER – nucleotide excision repair
NHEJ – non-homologous end joining
Mg – magnesium
MMR – mismatch repair
MMS – methyl methanesulfonate
MRN – Mre11/Rad50/Nbs1 complex
NBD – nucleotide binding domain
nm – nanometer
NMR – nuclear magnetic resonance spectroscopy
NSD – normalized spatial discrepancy
PCR – polymerase chain reaction
Pfu – *Pyrococcus furiosus*
pI – isoelectric point
RMS – root-mean-square
ROS – reactive oxygen species
S – Sievert

S. cerevisiae – Saccharomyces cerevisiae
SAD – single-wavelength anomalous dispersion
SAXS – small angle X-ray scattering
SDSA – synthesis-dependent strand annealing
SSA – single-strand annealing
ssDNA – single-stranded DNA
SSBR – single-strand break repair
T. maritima – Thermotoga maritima
TC-NER – transcription-coupled nucleotide excision repair
TdT – terminal deoxynucleotidyl transferase
TEV – tobacco etch virus
TLS – translesion DNA síntesis
Tm – *Thermotoga maritima*
UV – ultraviolet
Zn – Zinc

Appendix

Thermotoga maritima constructs:

SbcC

MRPERLTVRNFLGLKNVDIEFQSGITVVEGPNGAGKSSLFEAISFALFGNGIRYP
NSYDYVNRNAVDGTARLVFQFERGGKRYEIIREINALQRKHNAKLSEILENGKK
AAIAAKPTSVKQEVEKILGIEHRTFIRTVFLPQGEIDKLLISPPSEITEIISDVFQSKE
TLEKLEKLLKEKMKKLENEISSGGAGGAGGSLEKKLKEMSDEYNNLDLLRKYL
FDKSNFSRYFTGRVLEAVLKRTKAYLDILTNGRFDIDFDDEKGGFIIKDWGIERP
ARGLSGGERALISISLAMSLAEVASGRLDAFFIDEGFSSLDTENKEKIASVLKELE
RLNKVIVFITHDREFSEAFDRKLRITGGVVVNE

Number of amino acids: 365

Molecular weight: 41013.9 Da

Theoretical pI: 7.75

SbcD

MKILHTSDWHLGVTSWTSSRPVDRREELKKALDKVVEEAEKREVDLILLTGDL
LHSRNNPSVVALHDLLDYLKRMMRTAPVVVLPGNHDWKGLKLFGNFVTSISS
DITFVMSFEPVDVEAKRGQKVRILPFPYPDESEALRKNEGDFRFFLESRLNKLYE
EALKKEDFAIFMGHFTVEGLAGYAGIEQGREIIINRALIPSVVDYAALGHIHSFRE
IQKQPLTIYPGSLIRIDFGEEADEKGAVFVELKRGEPPRYERIDASPLPLKTLYYK
KIDTSALKSIRDFCRNFPGYVRVVYEEDSGILPDLMGEIDNLVKIERKSRREIEEV
LRESPEEFKEELDKLDYFELFKEYLKKREENHEKLLKILDELLDEVKKSEA

Number of amino acids: 379

Molecular weight: 43975.4 Da

Theoretical pI: 5.45

<u>C-terminal Construct of SbcD for Co-Crystallization with SbcC in Complex with DNA</u>

MKEELDKLDYFELFKEYLKKREENHEKLLKILDELLDEVKKSEAGHHHHHH

Number of amino acids: 51

Molecular weight: 6351.2 Da

Theoretical pI: 5.93

Pyrococcus furiosus constructs:

<u>Rad50NC</u>

MKLERVTVKNFRSHSDTVVEFKEGINLIIGQNGSGKSSLLDAILVGLYWPLRIKD
IKKDEFTKVGARDTYIDLIFEKDGTKYRITRRFLKGYSSGEIHAMKRLVGNEWK
HVTEPSSKAISAFMEKLIPYNIFLNAIYIRQGQIDAILESDEAREKVVREVLNLDK
FETAYKKLSELKKTINNRIKEGSGSKKEIKDLEKAKDFTEELIEKVKKYKALARE
AALSKIGELASEIFAEFTEGKYSEVVVRAEENKVRLFVVWEGKERPLTFLSGGE
RIALGLAFRLAMSLYLAGEISLLILDEPTPYLDEERRRKLITIMERYLKKIPQVILV
SHDEELKDAADHVIRISLENGSSKVEVVS

Number of amino acids: 360

Molecular weight: 41192.7 Da

Theoretical pI: 8.92

<u>Mre11</u>

MKFAHLADIHLGYEQFHKPQREEEFAEAFKNALEIAVQENVDFILIAGDLFHSSR
PSPGTLKKAIALLQIPKEHSIPVFAIEGNHDRTQRGPSVLNLLEDFGLVYVIGMR
KEKVENEYLTSERLGNGEYLVKGVYKDLEIHGMKYMSSAWFEANKEILKRLFR
PTDNAILMLHQGVREVSEARGEDYFEIGLGDLPEGYLYYALGHIHKRYETSYSG
SPVVYPGSLERWDFGDYEVRYEWDGIKFKERYGVNKGFYIVEDFKPRFVEIKV
RPFIDVKIKGSEEEIRKAIKRLIPLIPKNAYVRLNIGWRKPFDLTEIKELLNVEYLK
IDTWRIKERTDEESGKIGLPSDFFTEFELKIIDILGEKDFDDFDYIIKLITEGKVEEE
GPLEEAVKKVSEEKGKTVRQKIESIPKKKRGTLDSWLGGAR

Number of amino acids: 426

Molecular weight: 49455.6

Theoretical pI: 5.76

Table 6: Crystallographic data collection and model refinement of TmSbcCD.

Crystal	SeMet 11_1
Space group	I222
Unit cell	a=105 Å b=187 Å c=300 Å
	α=90° β=90° γ=90°
Wavelength (Å)	0.9774
Data range (Å)	50.0-3.4
Observations (unique)	161144 (72379)
Completeness (%) (last shell)	91.9 (83.3)[1]
R_{sym}[2] (last shell)	0.082 (0.480)
R_{meas}[3] (last shell)	0.104 (0.605)
I/σI (last shell)	10.6 (2.1)

Refinement	
Data range (Å)	20-3.4
Reflections F>0 (cross validation)	40428
R_{work}[4] / R_{free}[5] (%)	26.34 / 30.11
rmsd bond length (Å) / bond angles (°)	0.0056 / 1.34

[1] anomalous completeness

[2] R_{sym} is the unweighted R value on I between symmetry mates

[3] R_{meas} is the weighted R value on I between symmetry mates (Diederichs and Karplus, 1997).

[4] $R_{work} = \sum_{hkl} \| F_{obs}(hkl) | - | F_{calc}(hkl) \| / \sum_{hkl} | F_{obs}(hkl) |$ for reflections in the working data set

[5] $R_{free} = \sum_{hkl\ testset} \| F_{obs}(hkl\ testset) | - | F_{calc}(hkl\ testset) \| / \sum_{hkl} | F_{obs}(hkl\ testset) |$ for 5% of reflections against which the model was not refined

Table 7: Crystallographic data collection of TmSbcC in complex with DNA.

Crystal	SeMet ak1_3
Space group	C222
Unit cell	a=180 Å b=201 Å c=156 Å α=90° β=90° γ=90°
Wavelength (Å)	0.9801
Data range (Å)	30.0-3.9
Observations (unique)	107784 (45974)
Completeness (%) (last shell)	91.0 (77.5)[1]
R_{sym}[2] (last shell)	0.079 (0.428)
R_{meas}[3] (last shell)	0.102 (0.560)
I/σI (last shell)	9.27 (2.4)

[1] anomalous completeness

Acknowledgements

Curriculum vitae

Personal Details

Name	Derk Julius Bemeleit
Date of Birth	May 27th 1978
Place of Birth	Bremen
Nationality	German

Education

1984-1988	Grundschule Worpswede, Worpswede
1988-1990	Orientierungsstufe Worpswede, Worpswede
1990-1997	Gymnasium Lilienthal, Lilienthal
1997	Allgemeine Hochschulreife

Civilian Service

1997-1998	Rehabilitationsklinik Pyramide, Bremen

Academic Training

1997-2003	Studies in biology, University of Bremen, Bremen, Germany
2001	Research assistant, Center for Bioelectronics, Biosensors and Biochips, Richmond, Virginia, USA
2003	Diploma thesis "Preclinical Investigation of the Mechanism of Action of Novel Platinum Compounds in Malignant Glioma using Microarray Gene Expression Monitoring", School of Engineering at the Virginia Commonwealth University, Richmond, Virginia, USA
2003	Diploma in biology
since 2004	PhD studies in structural biochemistry, Gene Center and Institute of Biochemistry, Department of Chemistry and Pharmacy, Ludwig-Maximilians-University, Munich, Germany