UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Biologie Structurale et Nanobiologie**

Arrêté ministériel : 7 août 2006

Présentée par

« Simone PELLEGRINO »

Thèse dirigée par **« Sean McSWEENEY »**

préparée au sein du **Laboratoire « Structural Biology Group, European Synchrotron Radiation Facility »**
dans **l'École Doctorale de Chimie et Science du Vivant**

Comprendre le rôle de RecN dans la voie de réparation CDB chez *Deinococcus radiodurans*

Thèse soutenue publiquement le **« 28 Fevrier 2012 »**,
devant le jury composé de:

Dr. Bernard Gilquin
Rapporteur
Dr. Ingar Leiros
Rapporteur
Prof. Karl-Peter Hopfner
Membre
Dr. Jean-Luc Ravanat
President
Prof. Suzanne Sommer
Membre
Dr. Sean McSweeney
Directeur de thèse
Dr. Joanna Timmins
Co-encadrant

Abstract

Deinococcus radiodurans is a Gram-positive bacterium known for its extreme resistance to a broad variety of DNA damaging agents. Among these, Ionizing Radiations and desiccation are the most harmful for the cell, since they introduce breaks in the genome. Double Strand Breaks (DSB) are particularly hazardous for the cell and they need to be repaired very efficiently, in order to avoid mutations leading to altered, if not lethal, phenotypes. Homologous Recombination (HR) is the most efficient mechanism by which DSBs are repaired. *D. radiodurans* is able to completely restore its genome in only 3 hours, and it accomplishes the entire process through the RecFOR pathway.

In order to be repaired, DSBs first need to be recognized. The protein believed to be responsible for this important step that takes place soon after the damage occurs in the cell, is RecN. RecN is recruited at the early stages of DNA repair and *in vivo* studies have demonstrated its propensity to localize to discrete foci. *In vitro* studies also suggest that RecN possesses a DNA end-joining activity previously observed for SMC proteins (such as cohesin), which are structurally related to RecN. Several structural studies have been carried out on the SMC-like protein, Rad50, but so far no structural information is available for RecN.

The work presented here focused on the structural characterization of RecN and its constitutive domains. We obtained crystal structures of three partially overlapping constructs of RecN and Small Angle X-ray Scattering was performed on the individual domains and the full-length protein. The study of RecN in solution complemented our crystallographic study and enabled us to build a reliable, atomic model of the full-length protein. Mutations were designed and the mutant RecN proteins were produced in order to characterize the ATP hydrolysis activity of RecN, which is a conserved feature of this family of proteins. Extensive biochemical studies were carried out on wild-type and mutants of both the full-length protein and the single domains, in order to determine the role and function of each of the domains. Our results led us to propose a model for how RecN might recognize DSBs, tether two broken DNA ends and prepare the DNA for subsequent repair by the RecFOR machinery.

Résumé de thèse en français

Deinococcus radiodurans est une bactérie à gram-positive connue pour son extrême résistance à une grande variété d'agents endommageant l'ADN. Parmi ces derniers, les rayonnements ionisants et la dessiccation sont les plus nocifs pour la cellule, car ils introduisent des cassures dans le génome. Les cassures double brin (CDB) sont particulièrement dangereuses et doivent être réparées de façon très efficace, afin d'éviter l'apparition de mutations pouvant mener à la mort de la cellule ou de l'organisme. La recombinaison homologue (RH) est le mécanisme le plus efficace pour la réparation des CDBs. *D. radiodurans* est capable de restaurer entièrement son génome en à peine 3 heures, et elle accomplit la totalité du processus par la voie RecFOR.

Afin d'être réparées, les CDBs doivent d'abord être reconnu. Cette étape importante, qui a lieu peu de temps après l'apparition du dommage dans la cellule, implique la protéine RecN. RecN est recrutée dès les premières étapes de la réparation de l'ADN et des études *in vivo* ont démontré qu'elle avait tendance à se localiser dans des foyers discrets. Des études *in vitro* suggèrent également que RecN favorise l'assemblage de fragments d'ADN, une fonction décrite précédemment pour les protéines SMC (telle que cohesin), qui sont structurellement similaires à RecN. De nombreuses études structurales ont été effectuées sur la protéine de type SMC, Rad50, alors qu'à présent aucune information structurale n'est disponible pour RecN.

Le travail présenté ici a porté sur la caractérisation structurale de RecN et de ses domaines. Nous avons obtenu les structures cristallines de trois constructions (se chevauchant partiellement) de RecN et une étude de diffusions des rayons X aux petits angles a été effectuée sur les domaines séparés de RecN et sur la protéine entière. Les données obtenues en solution ont complété notre étude cristallographique et nous ont permis de construire un modèle atomique de la protéine entière. Des mutations ont été conçues et les protéines mutées ont été produites et utilisées pour la caractérisation de l'activité d'hydrolyse de l'ATP caractéristique de cette famille de protéines. Des études biochimiques approfondies ont été effectuées sur les différentes constructions et mutants de RecN afin de déterminer le rôle de chacun des ses domaines. Nos résultat nous ont permis de proposer un modèle qui explique comment RecN reconnaît les CDB, maintient les deux extrémités de l'ADN, et prépare l'ADN pour la réparation par les protéines RecFOR.

Chapter 1:

Introduction

Summary of chapter 1

This first chapter provides a thorough introduction to the main topic of this work. DNA lesions need to be repaired in order to avoid mutations and cell death. Causes of DNA damage and processes involved in the repair of such lesions are discussed in this chapter. Moreover, a presentation of the process of Homologous Recombination in eukaryotes and prokaryotes, which involves the RecN protein, is provided.

A paragraph describes the extreme radiation-resistant bacterium *Deinococcus radiodurans* and the principal mechanism underlying its extraordinary ability to efficiently repair its DNA. The role of Mre11-Rad50 complex in the repair of double-strand breaks is also discussed, together with an overview of the Structural Maintenance of Chromosomes proteins involved in DNA cohesion. Finally, this chapter presents our current knowledge regarding bacterial RecN proteins.

Résumé du chapitre 1

Ce premier chapitre propose une présentation complète du sujet principal de ce travail. Les lésions de l'ADN doivent être réparées afin d'éviter les mutations et la mort cellulaire. Les causes de ces dommages et les processus impliqués dans la réparation de ceux-ci sont discutés dans ce chapitre. Par ailleurs, une présentation du processus de recombinaison homologue chez les eucaryotes et procaryotes, ce qui implique la protéine RecN, est proposée.

Un paragraphe décrit la bactérie, *Deinococcus radiodurans,* connue pour sa haute résistance aux rayonnements et le mécanisme principal lui permettant de réparer efficacement son ADN. Une discussion du rôle du complexe Mre11-Rad50 dans la réparation de cassures double-brin et une description générale des protéines de la famille SMC (Structural Maintenance of Chromosomes) impliquées dans la cohésion de l’ADN sont également proposées. Pour finir, ce chapitre présente les connaissances actuelles concernant les protéines bactériennes, RecN.

1.1. DNA Damage and Health

1.1.1. Structure of DNA

DNA is the molecule encoding all the genetic information of an organism. Its double helical arrangement was discovered by Watson and Crick in the 1950's (Watson & Crick, 1953). DNA is a polymer of nucleotides, which in turn are made up of a deoxyribose, a phosphate group and one of the four nitrogenous bases: Adenine (A), Thymine (T), Cytosine (C) or Guanine (G). Since all the necessary information for synthesising RNA and proteins is codified in a single DNA strand, it is not surprising that protection of the genome is a real priority for both eukaryotic and prokaryotic cells. Every event that causes modification to the DNA should therefore be avoided. The consequences of DNA damage are rather diverse: short-term effects arise from disturbed DNA metabolism, triggering cell-cycle arrest or cell death. Long-term effects result from irreversible mutations contributing to oncogenesis (Hoeijmakers, 2001).

The intrinsic structure of DNA as a double helix has provided, throughout living organisms' evolution, a suitable substrate for the storage of the genetic information. As a consequence of this structural organization, DNA always carries a backup copy of its genetic information on the opposite strand, so when one strand is damaged, the other one can, when necessary, serve as template for repair. Only a few viruses use single stranded DNA or RNA for storing genetic information. The size of these genomes, most probably, is not big enough for being targeted by the DNA damaging agents (Alberts et al, 2008).

1.1.2. Lesions in DNA

DNA damage occurs in all living cells and in huge amounts, i.e. on average 1 million lesions per human cell per day (Alberts et al, 2008). Such damage might occur at the level of the DNA structure, through a chemical change in the encoded base (Figure 1) or by inducing single or Double Strand Breaks (DSB), or at the level of the genetic code, through alteration of the genetic message due to mutations. Eukaryotic and prokaryotic cells have therefore developed DNA repair mechanisms that efficiently remove the lesions.

Endogenous and exogenous DNA damage (Table 1) play a crucial role in mutagenesis, carcinogenesis and aging (Best, 2009; De Bont & van Larebeke, 2004). While the former can arise as product of conventional cellular metabolism, exogenous damage is produced by external agents. Both types of damage need to be efficiently and rapidly repaired in order to avoid changes in the information encoded in the genome.

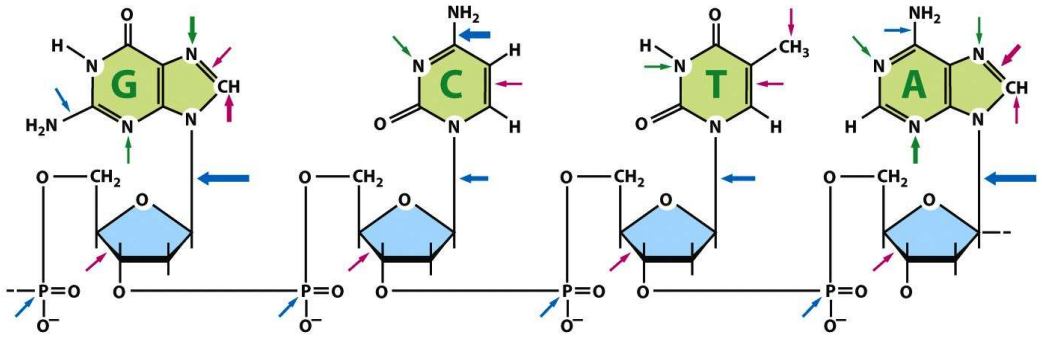


Figure 1: Summary of the modifications that can occur to DNA nucleotides. Arrows indicate the possible sites of attack of oxidative damage (red), hydrolysis (blue) and methylation (green) "Copyright 2008 from *Molecular Biology of the Cell* by B. Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis, LLC."

Despite its very high stability, the structure of DNA can be affected by hydrolysis, oxidation, alkylation (Figure 1) and base pair mismatches. Endogenous, spontaneous hydrolysis leads to the cleavage of the glycosidic bond linking the deoxyribose to the nitrogen base and to the formation of an abasic site (commonly named AP, apurinic/apyrimidinic) (Table 1). This occurs very frequently (50,000-200,000 per genome per day in humans) (Nakamura & Swenberg, 1999). AP sites trigger substitutions of base pairs (Lawrence et al, 1990) and can also result in frameshift mutations along the genome (Jackson et al, 1998). Reactive oxygen species (ROS), such as $O_2^{\cdot-}$, H_2O_2, OH^- and $O_2^{\bullet}$, resulting from metabolic reactions can, in addition to modifying the chemical properties of the DNA bases, also cleave the phosphodiester backbone of the DNA and introduce single- and DSB (De Bont & van Larebeke, 2004). Such damage can also be initiated by radiolysis of DNA. Direct and indirect formation of single strand breaks can occur on the DNA (Figure 2). In the first case the hydrogen abstraction by $OH^{\bullet}$ causes the solvolysis of the 5'- and 3'- phosphates. In the indirect mechanism, damage to the nitrogen base results in a spontaneous or enzymatic hydrolysis of the glycosidic bond, which produce an AP site (Figure 2). A specific lyase will then cleave the phosphate from the nucleotide (Ordoukhanian & Taylor, 2000).

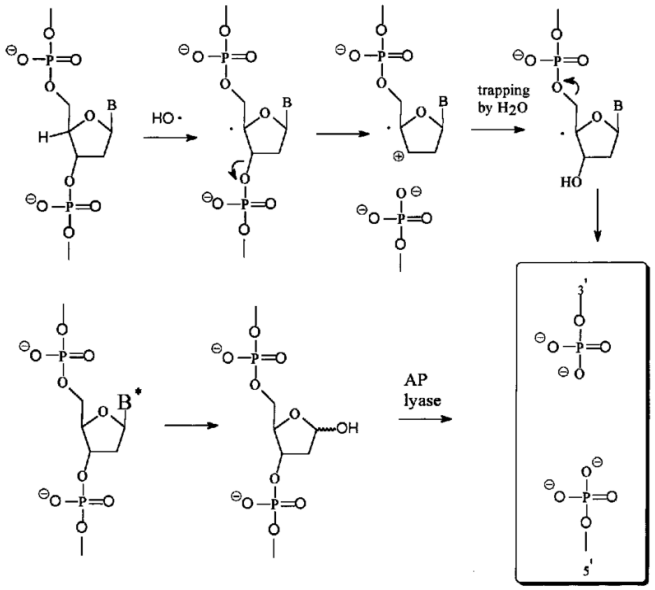


Figure 2: Single strand breaks are the result of hydrolysis. This can be direct (top) or indirect (bottom). Figure taken from (Ordoukhanian & Taylor, 2000).

X-rays, Ultra-Violet (UV) radiation and a broad variety of chemicals represent a source of serious damage to the cellular structure (Hoeijmakers, 2009). One of the most hazardous sources of DNA damage is UV-radiation, which leads to the formation of highly toxic molecules, such as cyclobutane-pyrimidine dimers (CPDs) and 6-4 photoproduct (6-4PP) (Sinha & Hader, 2002). The most dangerous type of DNA lesion is represented by the DSBs, which occur to both DNA strands, therefore leaving no intact genetic template for accurate repair. DSBs, if not properly repaired, can lead to chromosome fragmentation and, finally, to loss of gene structure during cell division (Alberts et al, 2008). Table 1 also reports some of the most common sources of DNA damage and the number of lesions per day/cell is very high for sunlight exposure, which is one of the best examples of how environmental conditions can affect human disease (melanoma in this particular case).

Endogeneous DNA damage	**Dose exposure (Gy)**	**DNA lesions generated**	**Number Lesions/Cell/Day**
Depurination	n/a	AP site	10,000
Cytosine deamination	n/a	Base transition	100-500
Oxidation	n/a	8oxoG	400-1,500
SAM-induced methylation	n/a	7meG	4,000
Exogenous DNA damage	**Dose exposure (Gy)**	**DNA lesions generated**	**Number Lesions/Cell/Day**
Peak hr sunlight	n/a	Pyrimidine dimers, (6-4) photoproduct	100,000
Cigarette smoke		Aromatic DNA adducts	50-1,000
Dental X-rays	0.005	DSBs	0.0002
^{131}I treatment	70-150	DSBs	2.8-6
Airline travel	0.005/hr	DSBs	0.005
Chernobyl accident	300	DSBs	12
Hiroshima atomic bombs	5-4000	DSBs	0.2-160

Table 1: Endogenous and exogenous DNA damage occurring in eukaryotes. The type and number of lesions per cell are reported (Ciccia & Elledge, 2010).

1.1.3. DNA damage and Disease

It has been estimated that fewer than one in 1,000 base changes result in a permanent mutation (Alberts et al, 2008). DNA damage is of particular interest in the field of oncology, since genomic alterations and mutations can lead to cancer and other diseases (a short list is presented in Table 2). Moreover, cellular DNA repair pathways (§ 2) are starting to be viewed as new targets for drug discovery, since the inhibition of such mechanisms can improve the efficacy of current chemotherapeutic drugs that target the replication of cancerous cells (Helleday et al, 2008).

Name	*Phenotype*	*Enzyme or process affected*
Xeroderma pigmentosum (XP)	Skin cancer, UV sensitivity, neurological abnormalities	Nucleotide Excision Repair
XP variant	UV sensitivity, skin cancer	Translesion synthesis by DNA polymerase η
Ataxia telangiectasia (AT)	Leukaemia, lymphoma, γ-ray sensitivity, genome instability	ATM protein, a protein kinase activated by DSBs
BRCA2	Breast, ovarian and prostate cancer	Repair by Homologous Recombination (HR)
Fanconi anemia groups A-G	Congenital abnormalities, leukaemia, genome instability	DNA interstrand cross-link repair
46 BR patient	Hypersensitivity to DNA-damaging agents, genome instability	DNA ligase I

Table 2: List of some of the diseases associated with mutations in DNA repair proteins (Alberts et al, 2008).

1.2. DNA repair pathways

1.2.1. DNA Damage Response

In order to maintain the integrity of genetic information, cells have developed multi-protein complexes able to recognize and repair lesions to DNA while, at the same time, stalling the machinery involved in cell cycle progression. The mechanisms deployed are components of the DNA Damage Response (DDR) (Figure 3). DDR involves a complex kinase signalling cascade, which is very conserved and elaborate in eukaryotes (Harper & Elledge, 2007) and recognizes and repairs all types of DNA lesions, especially DSBs. Detection of DNA damage is a crucial step for repair initiation and in higher eukaryotes, the ataxia telangiectasia mutated (ATM) kinase is one of the leading proteins involved in this process together with the Mre11-Rad50-Nbs1 mediator complex (MRN), which will be described later.

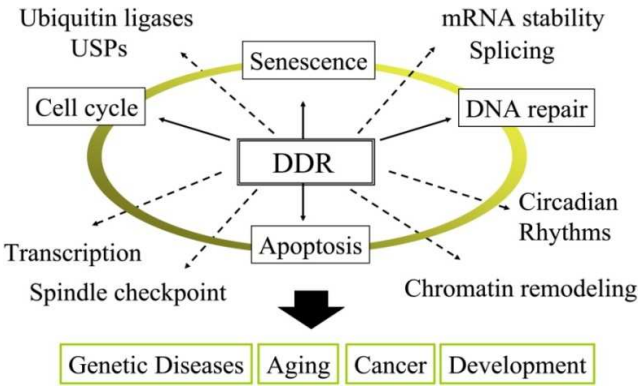


Figure 3: Cellular pathways induced by the DNA Damage Response pathway. Figure taken from (Harper & Elledge, 2007).

1.2.2. Multiple DNA repair pathways

Specific mechanisms have evolved in order to repair different types of DNA damage. Single base mutations are managed by the Base Excision Repair (BER) pathway; mismatches deriving from errors in DNA replication are taken care of by the Mismatch Repair system (MMR); pyrimidine dimers, created by photochemical reactions occurring on the DNA (Goodsell, 2001), are recognized and repaired by the Nucleotide Excision Repair (NER) pathway (Ciccia & Elledge, 2010).

BER and NER (Figure 4), involve the excision of the damage, followed by replacement of the missing nucleotides through DNA polymerase activity that uses the undamaged DNA strand as a template (Alberts et al, 2008). BER, as shown in Figure 4A, recognizes the modified base through the action of specific DNA glycosylases, which create an AP site at the site of base damage. Specific endonucleases then cleave the phosphodiester backbone, leaving a gap to be filled by DNA polymerase. NER (Figure 4B), in contrast, repairs more extensively damaged regions, such as that resulting from excessive exposure to sunlight (causing pyrimidine dimers) or of covalent interaction with carcinogenic hydrocarbons. These kinds of damage lead to distortions in the double strand DNA (dsDNA) structure. Once NER has sensed the topological change, the mechanism of repair begins with a nuclease that cleaves the bulky lesion. The unwanted fragment is then removed, as a consequence of the action of a specific DNA helicase, and the gap refilled by a DNA polymerase (Fuss & Cooper, 2006).

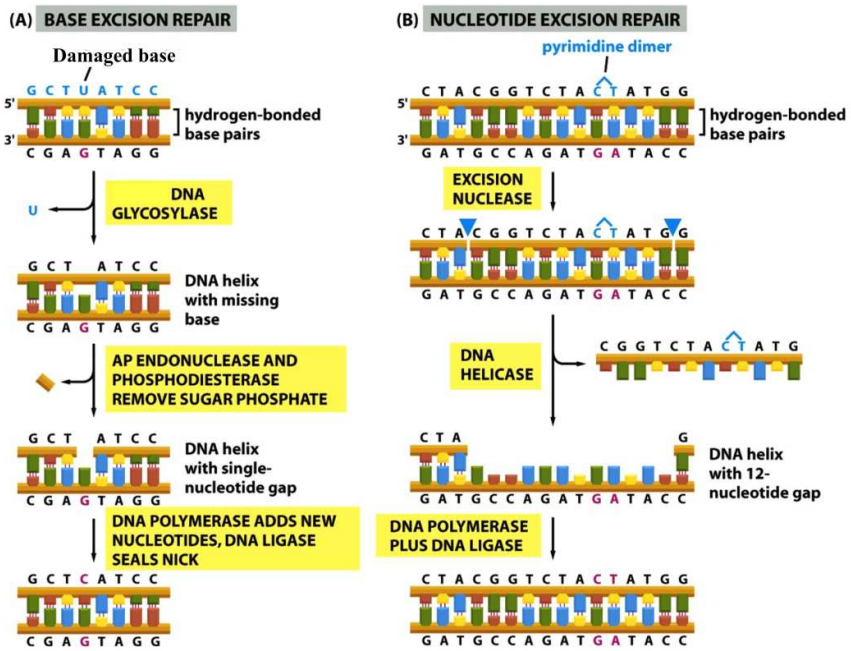


Figure 4: Example of BER (left) and NER (right) pathways. A) In BER a single damaged base is cleaved by the concerted action of specific glycosylases. Once the phosphodiester moiety has been removed, DNA polymerase can fill the gap of the single nucleotide. B) In NER, excision of the damaged site is achieved by cutting the DNA on either sides of the lesion. Helicases unwind the DNA duplex locally and DNA polymerases fill the extended gap (~30 nucleotides). "Copyright 2008 from *Molecular Biology of the Cell* by B. Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis, LLC."

1.2.3. DSB repair pathways

Breaks occurring in the phosphodiester backbones of both DNA strands are particularly hazardous for cells, because they can leave no intact template available in order to repair the DNA by NER or BER (Alberts et al, 2008) (Figure 4). DSBs need to be repaired very efficiently in order to avoid cell-cycle checkpoint arrest, which can in some cases be lethal (Kinoshita et al, 2009) and to avoid breakage of the chromosomes into small fragments leading to loss of gene organization (Alberts et al, 2008). DSBs are processed by two different pathways: Non-Homologous End Joining (***NHEJ)*** and Homologous Recombination (***HR)*** (Figure 5). The main difference between these two mechanisms resides in the fact that NHEJ alters the original DNA sequence (causing either deletions or short insertions), while HR restores the full integrity of the damaged DNA, with no loss of genetic information (Alberts et al, 2008).

1.2.3.1. Non homologous end-joining repair pathway

NHEJ (Figure 5A) is particularly important in repairing DNA that then undergoes DNA replication, since homologous substrate for HR (Figure 5B) is not yet available at this step (Wang et al, 2003). NHEJ has been suggested to be the main DSB repair pathway during the G_1 phase of the mammalian cell-cycle (Burma et al, 2006). It acts by bringing together the broken DNA ends and then inducing strand ligation by a DNA ligase. The full process takes place very rapidly following DNA damage, but, unfortunately, is error-prone (Burma et al, 2006). NHEJ can repair DNA breaks with little or no homology in cases where there are no repeat regions flanking the DSBs, and as a result may lead to possible insertions or deletions in the genome (Wyman et al, 2009). During NHEJ DNA ends are ligated without any requirement for strand exchange or the availability of homologous duplex DNA. In eukaryotic cells, NHEJ is carried out by the **Ku** complex (formed by Ku70 and Ku80), possessing DNA end-bridging activity, a protein kinase (**DNA-PK**) which signals the damage for recruitment of the repair machinery, a potential DNA-end processing enzyme (**Artemis**) and, finally, a ligase complex (**XRCC4-Ligase IV in mammals**) (Burma et al, 2006).

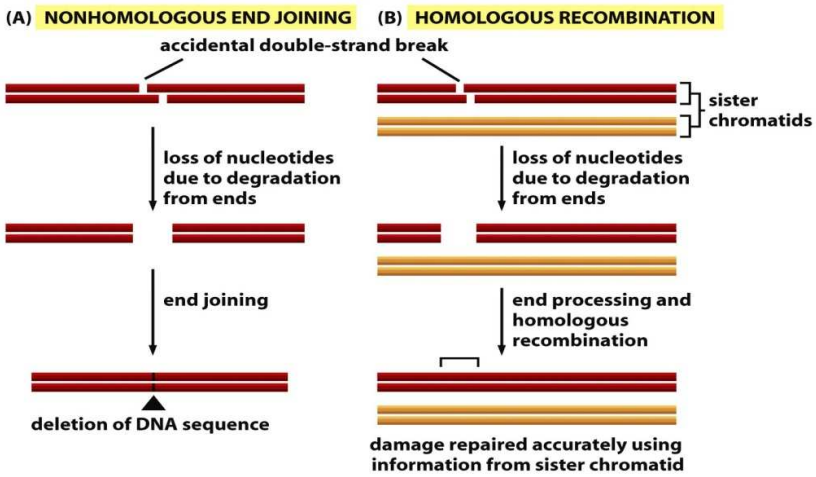


Figure 5: Comparison of NHEJ and HR. A) NHEJ quickly joins broken DNA ends, occasionally removing 1-2 nucleotides, thus leading to mutations. However, this is tolerated by the cell since NHEJ preferably occurs in somatic cells. B) HR, in contrast, repairs the damage with no loss of information from the DNA. "Copyright 2008 from *Molecular Biology of the Cell* by B. Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis, LLC."

In *Saccharomyces cerevisiae*, based on mutational studies that led to impaired end-joining activity, the Mre11-Rad50-Xrs2 complex **(MRX)** is also thought to be involved in NHEJ, (Moore & Haber, 1996). In bacteria, recent findings regarding homologues of the Ku

protein complex and of a mechanism involving a DNA-ligase (**DNA ligase D**) (Della et al, 2004) have started to shed light on the role of NHEJ in repair of bacterial chromosomes (Shuman & Glickman, 2007).

1.2.3.2. Homologous Recombination pathways

DSBs can result in genome rearrangements that initiate carcinogenesis or cell apoptosis (Hoeijmakers, 2001). However, controlled induction of DSBs can also be beneficial in the case of events that demand genome rearrangement, such as the development of the immune system (V(D)J recombination) or during meiosis (Schatz & Spanopoulou, 2005; Yabuki et al, 2005). As seen above (Figure 4 and 5), living organisms have therefore developed efficient mechanisms in order to counteract the hazardous effects of adverse DSBs and, in parallel, promote the favorable effects of controlled DNA rearrangement (Schatz & Spanopoulou, 2005; Yabuki et al, 2005).

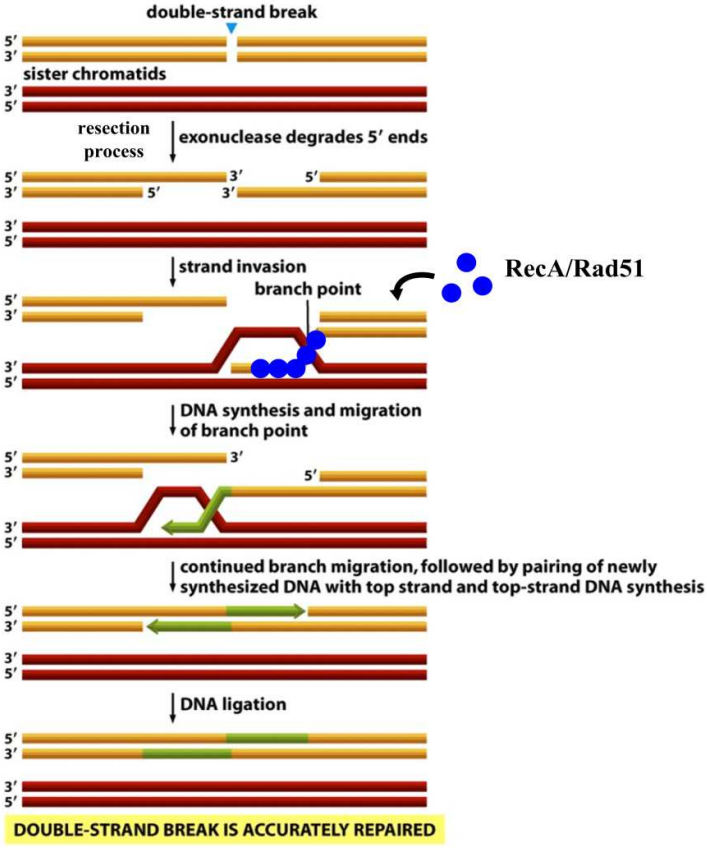


Figure 6: DSB repair by HR. This process takes place when the sister chromatids are still close to each other in order to allow strand invasion to occur. Once the undamaged strand is available, the DNA polymerase can start synthesising new, intact DNA. "Copyright 2008 from *Molecular Biology of the Cell* by B. Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis, LLC."

Homologous Recombination (Figure 5B and 6) is of fundamental importance in three processes: the accurate repair of DSBs, the bypass replication fork arrest caused by the presence of nicks or gaps on ssDNA and, finally, the exchange of genetic information during meiosis (Alberts et al, 2008). Despite the diversity among all kingdoms of life, it has been demonstrated that HR processes are common to all cells (Wyman et al, 2004).

1.3. Homologous Recombination

1.3.1. HR in eukaryotes

A schematic of the HR process is shown in Figure 6. The first step of HR involves the **tethering** of broken DNA ends, necessary for initiating the subsequent processing of such ends into single-stranded tails with 3'-hydroxyl overhangs (***resection*** process). These overhangs are required for **strand invasion** (Figure 6). The Mre11-Rad50-Nsb1 (MRN) complex, in humans, or the MRX complex, in yeast, are responsible for creating these 3'-overhangs suitable for HR (Longhese et al, 2010). Mre11 has both single strand endonuclease and 3'→5' double strand exonuclease activity and it can therefore process the 5'-DNA strand for generation of the required overhang (Williams et al, 2008). The Mre11-Rad50 (MR) complex is one of the first factors detected at DNA DSBs (Lisby et al, 2004). In humans the MR complex forms oligomeric architectures (notably a M_2:R_2 heterotetrameric arrangement) that bind to different DNA molecules, taking advantage of a very long coiled-coil structure (150-600 Å long) (de Jager et al, 2001; Hopfner et al, 2002; van Noort et al, 2003). DNA binding then acts as a conformational switch in the MR complex that favors interaction between coiled-coil apexes (Figure 7) of different complexes and thus enhances DNA tethering as could be shown by single-molecule imaging experiments (Moreno-Herrero et al, 2005). A more detailed description of the MR complex will be given in § 5.5.

Base pairing between the damaged DNA strand and the homologous strand is the major requirement for HR to act efficiently. To fulfil this requisite, it is important that the "afflicted" strand does not base-pair with the complementary "healthy" one, but only with a second DNA molecule (Alberts et al, 2008). In eukaryotes the recombinase **Rad51** protein is loaded onto the ssDNA in order to avoid the self base-pairing of the damaged strand with the complementary one. Functional Rad51 nucleoprotein filament is essential to allow the formation of a hybrid DNA through the pairing of the damaged ssDNA with the complementary strand in a different homologous dsDNA molecule. This ***strand invasion*** process is carried out by Rad54 and other recombinases (Heyer et al, 2010). In eukaryotic cells the process, which involves formation of the Rad51 protein nucleo-filament, has inverted

polarity compared to that in bacteria (Murayama et al, 2008) and occurs in 3' → 5' direction. ***Branch migration*** is the next step of the HR process. An unpaired region of the ssDNA (coloured in green in Figure 6) invades the undamaged duplex DNA and base pairs with the homologous sequence, constituting a branch point. Afterwards, the branch point can move along the DNA, displacing the resident strand, either spontaneously in both directions or in one specific direction, through a protein-directed mechanism (Alberts et al, 2008). The last step, before the DNA polymerase fills the remaining gap and a DNA ligase catalyses the phosphodiester bond formation, involves the ***resolution*** process. During this step, a four DNA stranded cross-shaped architecture, known as ***Holliday junction*** (Holliday, 1964), is cleaved by specific enzymes, which are called *resolvases*. In eukaryotic cells, several proteins are known to act in resolution of Holliday junctions, including Mus81, Yen1 and Slx1-Slx4 (Svendsen & Harper, 2010) (Table 3).

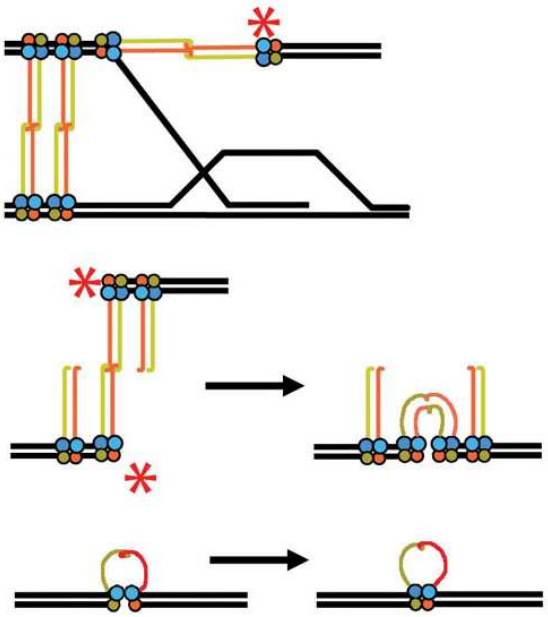


Figure 7: Different models proposed for the interaction of the MR complex with the dsDNA. The architecture of the tail-to-tail MR complexes can suggest a mechanism for bridging sister chromatids. Alternatively, the head domains of two MR complexes can bind separate broken DNA ends, inducing a structural re-arrangement which allows them to be tethered. Finally, circularization of a single MR complex (as supported by electron microscopy studies) could contribute to productive repair of broken DNA ends. DNA tethering is mediated by the coiled-coil domain, which could form a dimeric structure with length up to 1200 Å. Figure taken from (Hopfner et al, 2002).

1.3.2. HR in prokaryotes

1.3.2.1. A stepwise process

In bacteria, once the DSB is recognized, RecQ helicase and RecJ 5'→ 3' exonuclease resections the DNA ends in order to create the 3' overhangs (Amundsen & Smith, 2003). After these processes take place, the loading of a recombinase protein onto processed DNA ends, which catalyzes the exchange of base-paired partners between two DNA molecules, represents the second step of HR (Kim & Cox, 2002). The bacterial homologue of the eukaryotic Rad51 protein is **RecA**, which binds tightly to ssDNA in a cooperative manner, giving rise to a filament of proteins along the strand (Figure 8). Cycles of association/dissociation from the DNA are regulated by ATP hydrolysis, which also confers to RecA the capacity to topologically modify the DNA around it (Cox, 2007). RecA has more than one DNA-binding site and, can therefore bind both the single-stranded and duplex DNA (Figure 8) required for initiating the ***strand-invasion*** process (Lusetti & Cox, 2002). A *heteroduplex* DNA molecule is formed that can be further extended in size by DNA helicases through a process called ***branch migration,*** which gives rise to Holliday junctions. As in eukaryotes, specific proteins cleave these junctions and the DNA gap can then be filled.

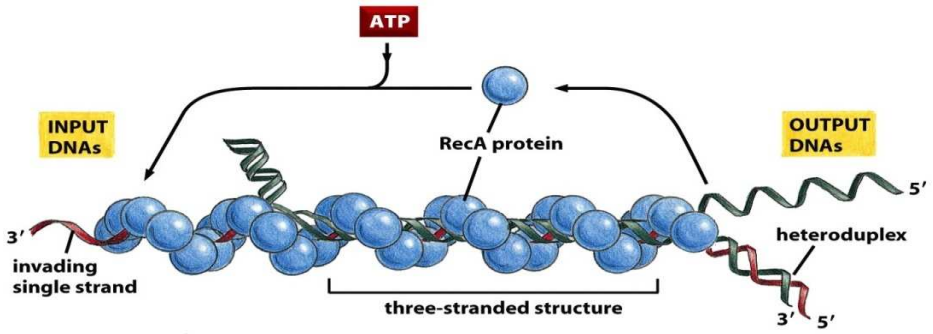


Figure 8: Representation of bacterial RecA, homologous to eukaryotic Rad51 protein, forming nucleoprotein filaments wrapping around the damaged ssDNA and dsDNA of the sister chromatids. The entire process is regulated by ATP hydrolysis. "Copyright 2008 from *Molecular Biology of the Cell* by B. Alberts et al. Reproduced by permission of Garland Science/Taylor & Francis, LLC."

The HR pathway in *E. coli* comprises three important steps: (1) the ***pre-synapsis***, where DNA ends are processed for proficient recombination to occur; (2) the ***synapsis***, leading to the formation of a joint DNA molecule between the newly created DNA end (step 1) and the homologous dsDNA template; (3) the ***post-synapsis*** and resolution, where DNA is repaired and the resolution of the Holliday junctions occurs (Wyman et al, 2004). The proteins participating in these processes in *E. coli* are listed in Table 3.

Stage of HR	Process mediated	Proteins
		E.coli
Pre-synapsis	End-processing	RecBCD, RecQ, RecJ
	Negotiating ssDNA	SSB
	Recombinase loading	RecBCD, RecFOR
Synapsis	Joint molecule formation by strand invasion	RecA
Post-synapsis	Branch migration	RuvAB
	Resolution of crossed DNA strands	RuvC

Table 3: Proteins participating in the various stages of HR, as identified by biochemical studies in bacterial cells (*E. coli*) (Wyman et al, 2004).

1.3.2.2. RecBCD vs. RecF pathway

DSB response in bacteria is rather diverse. In *E. coli*, for instance, two pathways are activated in response to DSBs: RecBCD and, in the case of RecBCD failure, RecF (Spies & Kowalczykowski, 2004). RecBCD is a helicase-nuclease complex in charge of initiation of DSB repair by HR and is regulated by a *cis*-acting DNA sequence named Chi (crossover hotspot instigator) (Figure 9). The mechanism by which RecBCD initiates DSB repair has been proposed to occur in a stepwise fashion (Dillingham & Kowalczykowski, 2008; Singleton et al, 2004): RecB starts to translocate along the DNA duplex to unwind the double strand. The nascent ssDNA subsequently passes through RecC before exiting for being digested by RecB nuclease domain. In RecC a helicase-like domain constitutes the Chi-scanning site, which can recognize this characteristic sequence produced after a DSB occurs. Once the RecC subunit encounters the Chi sequence, further translocation into the nuclease domain is prevented while an attenuation of the nucleolytic activity takes place. RecBCD,

however, continues to translocate along the DNA upon Chi detection (Figure 9), leading to the formation of a ssDNA loop downstream of the recognition sequence.

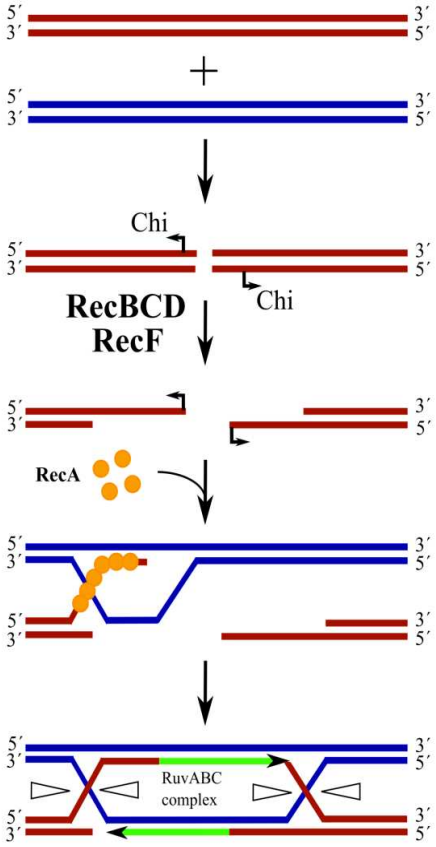


Figure 9: RecBCD catalyzed Homologous Recombination pathway in *E. coli* (Dillingham & Kowalczykowski, 2008). RecBCD activity is stimulated by recognition of the χ sequence.

The holoenzyme complex can then begin the loading of RecA protein. RecB needs to be displaced from the surface of RecC, event that leaves the ssDNA loop available for RecA interaction. The newly generated ssDNA will then constitute the platform for RecA nucleo-protein filament formation (Figure 9). Thus, the function of each subunit can be summarized as follows: RecB is a 3'-5' helicase and a multifunctional nuclease, RecC is involved in Chi sequence recognition and RecD is a 5'- 3' helicase (Dillingham & Kowalczykowski, 2008; Singleton et al, 2004; Spies & Kowalczykowski, 2004; Taylor & Smith, 1999).

The RecF pathway, on which more attention will be focused later, plays a role as a backup mechanism to the RecBCD pathway in some bacteria, such as *E. coli*, while it represents the major DSB repair pathway in others, such as *Deinococcus radiodurans* (Rocha et al, 2005). Both the RecBCD and RecF mechanisms contribute to the loading of RecA onto

the 3' ssDNA overhang in order to create the nucleoprotein filament, which is then used for strand exchange in the HR process (Figure 9) (Morimatsu & Kowalczykowski, 2003; Savir & Tlusty, 2010). In *E. coli*, the resolution process is carried out by RuvABC complex (Table 3), which is also present in *Deinococcus radiodurans* (Sharples et al, 1999).

1.4. *Deinococcus radiodurans*

1.4.1. A radiation-resistant bacterium

Deinococcus radiodurans (DR) is an extremophilic bacterium able to tolerate hundreds of DSBs in its genome. It has been classified as the "world's toughest bacterium" by the *Guinness Book of World Records* (http://www.guinnessworldrecords.com) and nicknamed as "Conan the bacterium" (Huyghe, 1998). DR was first discovered in 1956 during attempts to sterilize canned meat using very high doses of γ-radiation (Anderson et al, 1956). The meat spoiled despite this treatment and the researchers decided to isolate and characterize the bacterium responsible for the spoiling. This was found to be able to survive extremely high doses of radiation (cells in exponential phase survived exposure to doses up to 15 kGy) and was named as *Deinococcus radiodurans* (Battista, 1997). DR has also been shown to withstand desiccation and the attack of chemical agents that damage the DNA. Because of its extraordinary resistance to harsh environmental conditions DR was considered to be of particular interest as a detoxifying agent for digesting toxic inorganic materials, such as mercury and solvent (toluene for instance) in highly radioactive sites (Brim et al, 2000).

DR is a Gram-positive bacterium, even though it shows a similar cellular envelope to those of Gram-negative bacteria, with the presence of an outer membrane after the peptidoglycan layer (Battista, 1997). Its genome was sequenced in 1999 (White et al, 1999) and is composed of 2 chromosomes (2,648,638 and 412,348 base pairs), a megaplasmid (177,466 base pairs) and a small plasmid (45,704 base pairs) and contains approximately 3200 genes encoding ~2000 proteins.

DR is widespread in the natural environment, but is especially found in soils rich in organic nutrients or in very dry areas. The resistance of DR to prolonged periods of dehydration is very important since, like exposure to high doses of radiation, desiccation can lead to a broad variety of DNA damage (Bauermeister et al, 2011). The ability of DR to withstand very high doses of irradiation (5-6 kGy) with no loss of viability (Figure 10) has much intrigued the scientific community, since there is no obvious reason, from the evolutionary point of view, why a living organism should have evolved such resistance. DR's extreme resistance to Ionizing Radiation (IR) has been proposed to have evolved as a side

effect of its ability to tolerate prolonged periods of lack of water in its environment (Mattimore & Battista, 1996).

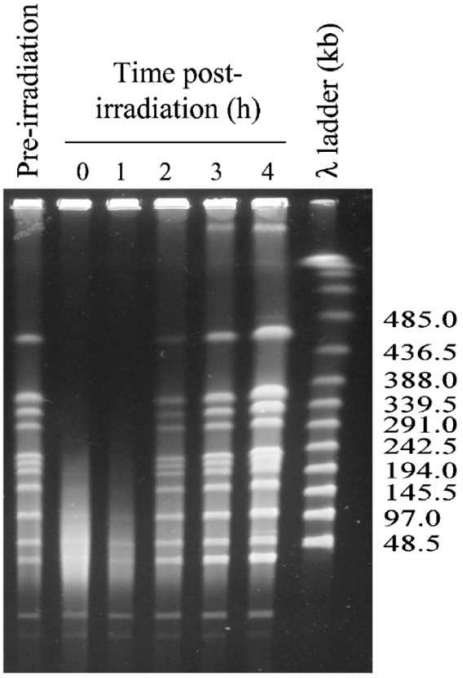


Figure 10: Kinetics of DSB repair in DR exposed to 6,800 Gy γ-irradiation. The full genome is restored 3 hours after exposure to Ionizing Radiation. Figure taken from (Blasius et al, 2008).

DR has developed very efficient mechanisms for maintaining the integrity of its genome, including the very accurate RecF recombinational repair pathway (Bentchikou et al, 2010), which will be discussed in more detail later (§ 1.5.1-2). A number of hypotheses have been proposed to explain the extremophilic nature of DR: highly redundant genetic information may have a pivotal role for a successful genome restoration. In the log phase, DR exhibits 4-10 copies of its genome (Battista, 1997): possessing multiple copies of genetic information ensures the presence of intact DNA at all times for use as template for repair (Daly & Minton, 1995; Daly & Minton, 1997). DR's genome also exhibits a tight structural organization and is condensed as a nucleoid, forming a ring-like structure. Such an architecture is believed to aid the repair of DNA damage caused by irradiation, by, in particular, arresting diffusion of DNA fragments generated by DSBs away from the genome (Zimmerman & Battista, 2005). DR is also able to efficiently export damaged DNA fragments out of the cell through the cell membrane. Exporting damaged fragments may reduce the risk of mutations arising from the reincorporation of damaged bases during DNA synthesis and fragments of DNA may also function as signal molecules for recruitment of the DNA repair

machinery (Battista, 1997). Finally, a high intracellular concentration of Mn^{2+} ions may provide a suitable scavenger for highly reactive oxygen species and free radicals (Slade & Radman, 2011), but its mechanism of action is still only poorly understood.

1.4.2. DSB repair in *Deinococcus radiodurans*

It was previously mentioned (Table 1) that IR introduces hundreds of DSBs into the chromosomes of living organisms, no matter whether these are bacteria, yeast or humans. In DR, DSBs are repaired and the integrity of its genome is restored in a just few hours (Cox et al, 2010) (Figure 10). UV irradiation causes the transcriptional induction of ~60 genes in DR (Tanaka et al, 2004). However, expression of a majority of proteins involved in DSB repair is not induced by irradiation of DR cells, suggesting that constitutive concentrations are sufficient for repairing the DNA damage (Liu et al, 2003; Tanaka et al, 2004).

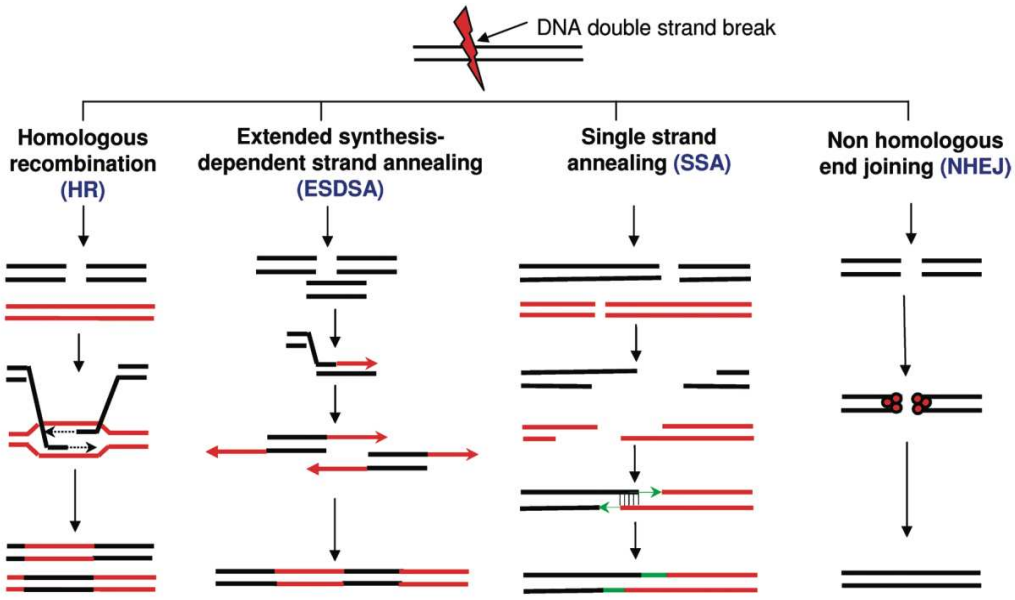


Figure 11: Representation of the DNA repair pathways used by *D. radiodurans* to reassemble its genome following exposure to IR. HR and ESDSA are the most important mechanisms for DSB repair. Single-strand annealing and NHEJ can replace the main HR pathway when this fails. Figure taken from (Blasius et al, 2008).

Two principal mechanisms (Figure 11) are implicated in recognition and repair of DSBs in DR: **Extended Synthesis-Dependent Strand Annealing (ESDSA)** (Radman et al, 2009) and **HR** (Cox et al, 2010; Zahradka et al, 2006). ESDSA has unique features and requires two broken chromosomal copies and a single-round multiplex PCR-like step for production of long overhang fragments suitable for precise annealing (Zahradka et al, 2006). The entire

ESDSA pathway includes resection of DNA broken ends, RecA-mediated strand invasion and abundant DNA synthesis prior to the annealing step. The goal of ESDSA is to start the reassembly of "shattered chromosomes" into larger DNA fragments in order to provide suitable substrates for completion of DNA repair by HR (Zahradka et al, 2006).

1.4.2.1. The RecF pathway

As described above, in *E. coli*, HR is accomplished by the RecBCD complex (Figure 9). In DR, however, there are no homologues of RecB or RecC and the RecF pathway is the principal actor in HR (Bentchikou et al, 2010; Saikrishnan et al, 2009; Wang & Julin, 2004). Indeed, DR cells carrying knockouts of the *recFOR* (individual knockouts Δ*recF*, Δ*recO* and Δ*recR*) and *recA* (Δ*recA*) genes display an increased sensitivity to γ-irradiation (four orders of magnitude), implying that both the RecF complex and RecA are crucial for the viability of DR cells (Bentchikou et al, 2010). However, in the three knockout mutants targeting the RecFOR complex, it was found that the kinetics of DSB repair was comparable to that of the *ΔrecA* mutant, supporting a slow and progressive chromosome reassembly compared to wild-type DR. As a result, DR cells devoid of RecF, RecO or RecR proteins likely repair damage using a RecA-independent mechanism, such as single-strand annealing or NHEJ (§ 3.2). Irradiation studies conducted on *recFOR* and *recA* mutant cells also highlighted a decrease of the fragmentation of DNA, suggesting an active role of these proteins in controlling nucleases for DSB repair. The similarity of the phenotypes of *ΔrecFOR* and *ΔrecA* mutants suggests the RecF pathway is largely dependent on RecA activity (Bentchikou et al, 2010).

1.4.2.2. RecA loading onto stalled replication forks

While the mechanism of action of the RecF pathway is still only poorly understood, current evidence suggests that it is centered on the formation of two principal complexes: RecOR and RecFR, both involved in guiding RecA protein targeting to ssDNA (Figure 12). RecOR is thought to help RecA loading onto ssDNA binding protein (SSB)-coated ssDNA, with RecO displacing SSB from the DNA. RecFR, in contrast, is suggested to be involved in RecA loading onto ssDNA at ssDNA-dsDNA junctions (Bentchikou et al, 2010). A possible model (Figure 12) can be proposed from published results in which RecF binds first to the DNA, possibly in complex with RecR. RecO would then associate with the DNA and facilitate the loading of RecA onto the DNA, through interaction with RecR (Inoue et al, 2008; Sakai & Cox, 2009).

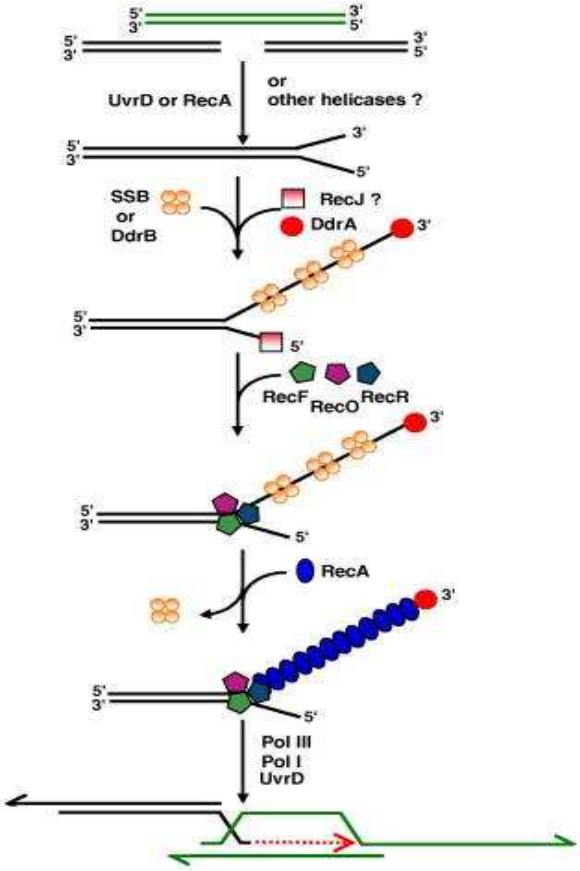


Figure 12: Model of the initiation of DNA DSB repair involving the RecFOR complex in DR. The first step of DSB repair is accomplished by the ESDSA pathway. RecJ begins ssDNA resection in 5' → 3' direction, resulting in a 3'-overhanging ssDNA, as it occurs in *E. coli* (Chow & Courcelle, 2007). DdrA and DdrB proteins, highly induced upon irradiation, are involved in protection and binding of the 3'-ssDNA end, respectively. RecFOR presents the coated-ssDNA to RecA in order to form the nucleoprotein filament, which then triggers strand invasion and subsequent HR steps. Figure taken from (Bentchikou et al, 2010).

1.5. SMC and SMC-like proteins

Structural Maintenance of Chromosomes (SMC) proteins are widespread in all kingdoms of life (Figure 13) and are recognized as a fundamental class of proteins, acting as cohesins or condensins (§ 5.3), that regulate the structural and functional organization of chromosomes. In eukaryotic cells there are at least six members of the SMC family, SMC1-SMC6 (Figure 13), which assemble as heterodimers, while in prokaryotes and archaea there is only one SMC, which assembles as a homodimer (Hirano & Hirano, 1998). In bacteria, additional SMC-like proteins, such as MukB, SbcC and YhaN, are also found. There have been proposed to be involved in chromosome segregation and DNA repair (Graumann & Knust, 2009; Niki et al,

1992). However, it is still an open question whether bacterial SMCs can be considered as the functional counterparts of the eukaryotic condensin and cohesin proteins (Hirano, 2005).

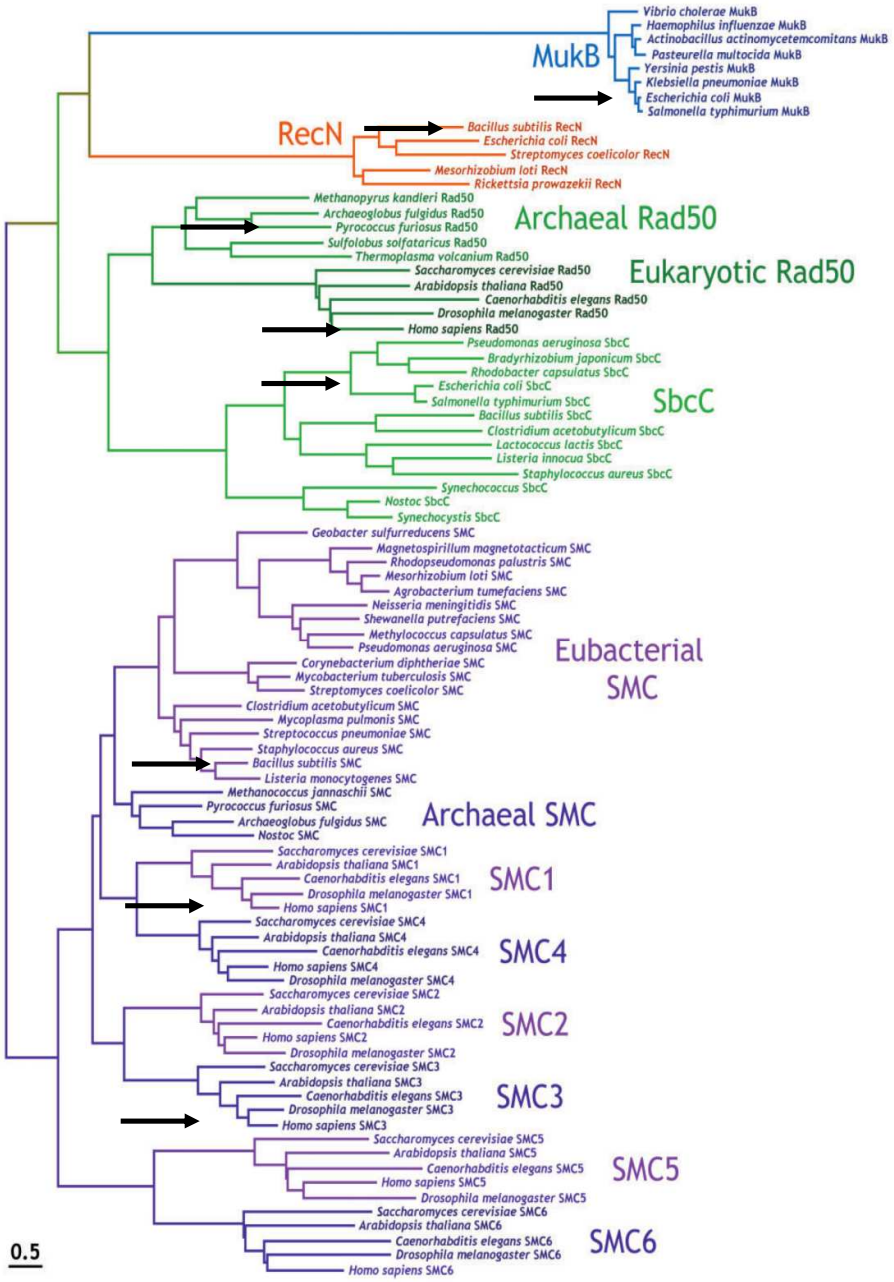


Figure 13: Phylogenetic tree of SMC protein family. Proteins cited in this work are indicated with black arrows. Figure taken from (Cobbe & Heck, 2004).

1.5.1. Architecture of SMC proteins

SMC proteins are related to the ATP-Binding Cassette (ABC) protein family (Hirano, 2002) (Figure 14). However, while most ABC proteins are involved in importing and

exporting ions or small molecules (Jones & George, 2004), SMCs are implicated in DNA repair and cell-cycle checkpoint regulation. Both families of proteins share a globular Nucleotide Binding Domain (NBD), known as the ***head*** domain, which has been shown to bind and hydrolyze ATP (Hollenstein et al, 2007). The NBD of ABC and SMC proteins is principally formed by two motifs, called the Walker-A and Walker-B motifs (Walker et al, 1982), which will be described in detail later (§ 5.4), that form the suitable pocket for ATP binding. In SMC proteins, the two conserved Walker motifs are connected by two long coiled-coils interconnected through a non-helical sequence (*hinge* domain) (Figure 15). The same structural organization is also exhibited by SMC-like proteins, of which ***Rad50***, component of the MR complex (§ 5.5), is one of the most representative members. Structural information on SMC and SMC-like proteins is available (Griese & Hopfner, 2011; Lammens et al, 2004; Li et al, 2010; Lowe et al, 2001) (Figure 14), but at present a complete structure of a member of this protein family has not been obtained.

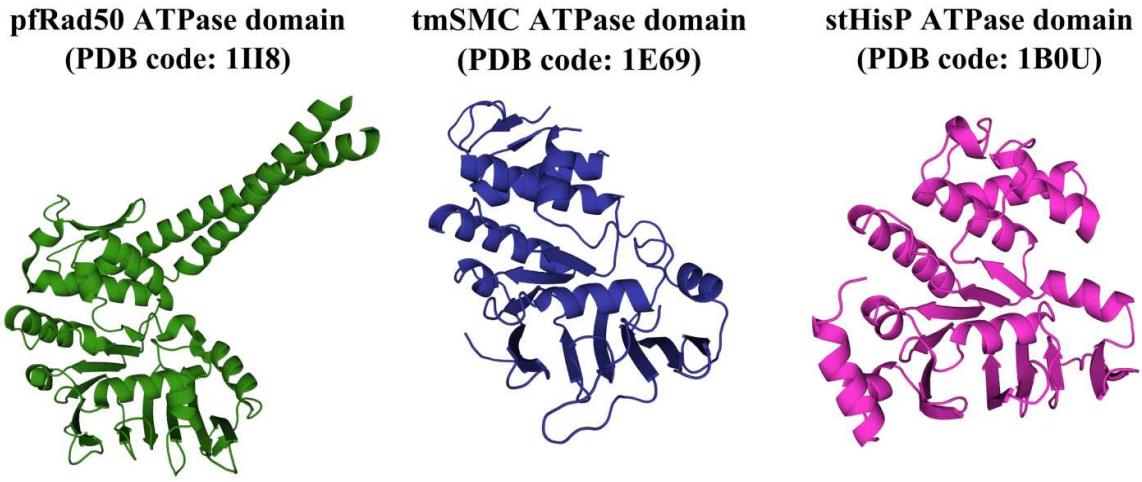


Figure 14: Crystal structures of *Pyrococcus furiosus* Rad50 (pfRad50), *Thermotoga maritima* SMC (tmSMC) and *Salmonella typhimurium* HisP ATPase domains. The three structures are shown in the same orientation with their associated PDB codes.

Based on the structural information available, it has been proposed that a SMC monomer folds back on itself, forming an anti-parallel arrangement of the coiled-coil and, more importantly, leading to the creation of the ATP-binding ***head*** domain (Hirano & Hirano, 2002; Hirano, 2006) (Figure 15). In eukaryotes SMCs are mainly assembled as heterodimers: cohesins, for instance, are formed upon interaction of SMC1 and SMC3 and the entire process is regulated by ATP-binding and hydrolysis (Figure 15). Specific interaction partners, such as

Scc1 and Scc3, can afterwards lead to the formation of a functional DNA-binding complex (Gruber et al, 2003).

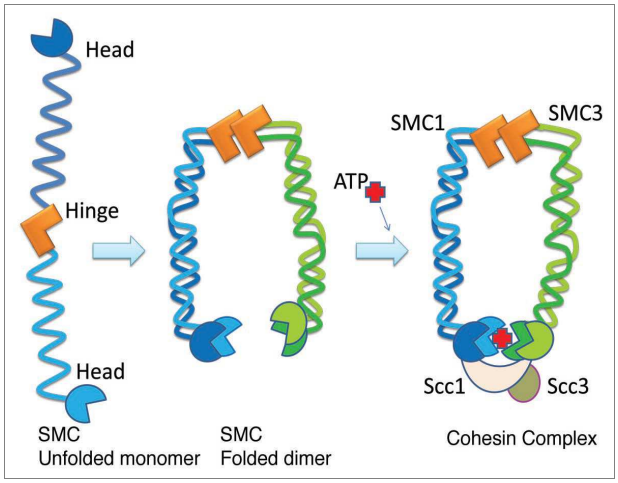


Figure 15: Assembly of an active SMC protein complex in eukaryotic cells (in this particular case, cohesin). Once the heterodimer of SMC is formed, specific protein partners interact with the SMC1-SMC3 structure to constitute an active cohesin complex. Figure taken from (Wong, 2010).

An extensive multi-approach study of SMCs highlighted the highly dynamic and plastic nature of this class of proteins (Hirano, 2005). Cycles of ATP binding and hydrolysis induce engagement or disengagement of the head domains (Figure 15), in order to recycle protein for the next round of DNA-binding (Hirano, 2005).

1.5.2. SMCs and DNA binding

Interactions between SMC head domains can take place either through cycles of ATP-binding and hydrolysis or mediated in an ATP-independent fashion. It has been speculated that eukaryotic SMC proteins are functionally differentiated in order to be involved in different aspects of the cell-cycle. This leads to the formation of various structural architectures like double-sized rings, filaments and rosette-like structures (Hirano, 2006) (Figure 16). However, only double-sized rings have been proposed for cohesins (Figure 17A-B), while the other organizations are likely to be implicated in chromatin condensation (mediated by SMCs known as condensins). A detailed model of the putative SMC-DNA interaction is shown in Figure 16. It has been proposed that SMC dimers are initially present in a closed state in which two ATP molecules are bound to the head domains (step 1). Hydrolysis of ATP induces the disengagement of the head domains and can result in either hinge opening (step 2) or head flipping (step 3). A new cycle of ATP-binding could promote

intramolecular (steps 4 and 6) or intermolecular head domains association (steps 5 and 7). This can result in the formation of multiple possible modes of SMC–DNA interactions via '*topological trapping*' (steps 4 and 6) (Hirano, 2005). Higher-order nucleoprotein complexes can be also generated by ATP-independent SMC-SMC interactions (steps 8–10). Cohesins, for instance, may hold two sister chromatids together (steps 4, 5 and 8), while condensins may organize DNA by one of the mechanisms shown in steps 6, 7, 9 and 10 (Hirano, 2005) (Figure 16).

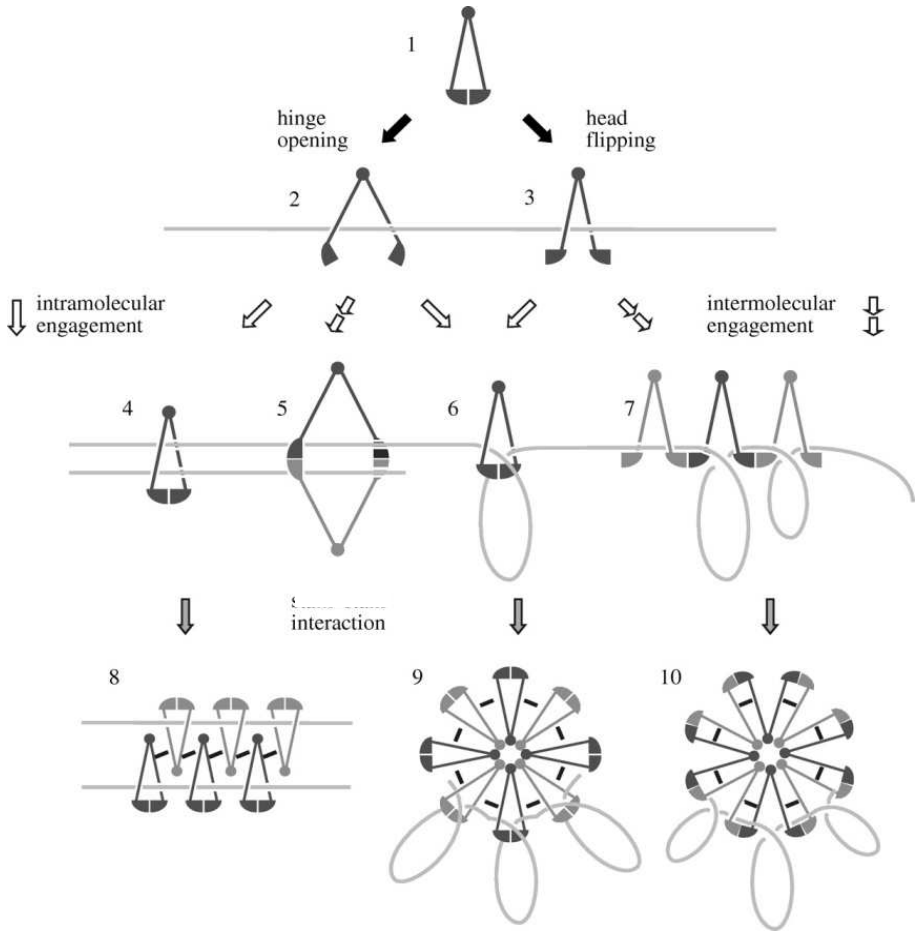


Figure 16: SMCs interact with DNA in a highly dynamic manner. ATP hydrolysis regulates the head-head engagement and subsequent DNA interaction cycles. Inter- and intra-molecular engagement of the head domain of SMC lead to the formation of DNA rosette-like structures, which are essential for cohesion and condensing processes. Figure taken from (Hirano, 2005).

1.5.3. SMC proteins and DNA cohesion

Cohesion is a crucial process involved in the regulation of chromatid separation during meiosis and mitosis. Proteins responsible for holding the sister chromatids together are called

cohesins and they are positioned on many locations along the length of the newly replicated DNA. Eukaryotic cohesin proteins consist of a heterodimer of two SMC proteins, SMC1 and SMC3 (Figure 15), which interact in an ATP-dependent fashion in order to wrap around the DNA, and form ring-like structures (Figure 17). Cohesins are also involved in HR where they are believed to interact simultaneously with the two sister chromatids required in the repair process. Experiments performed in different organisms have shed light on the role of cohesins in providing proximity between the sister chromatids and in stabilizing chromosomes when DSBs cause lack of continuity in the DNA backbone (Watrin & Peters, 2006). However, there are still many open questions regarding the mechanisms of action of cohesins. In particular, the role of ATP hydrolysis is not fully understood (Shintomi & Hirano, 2007).

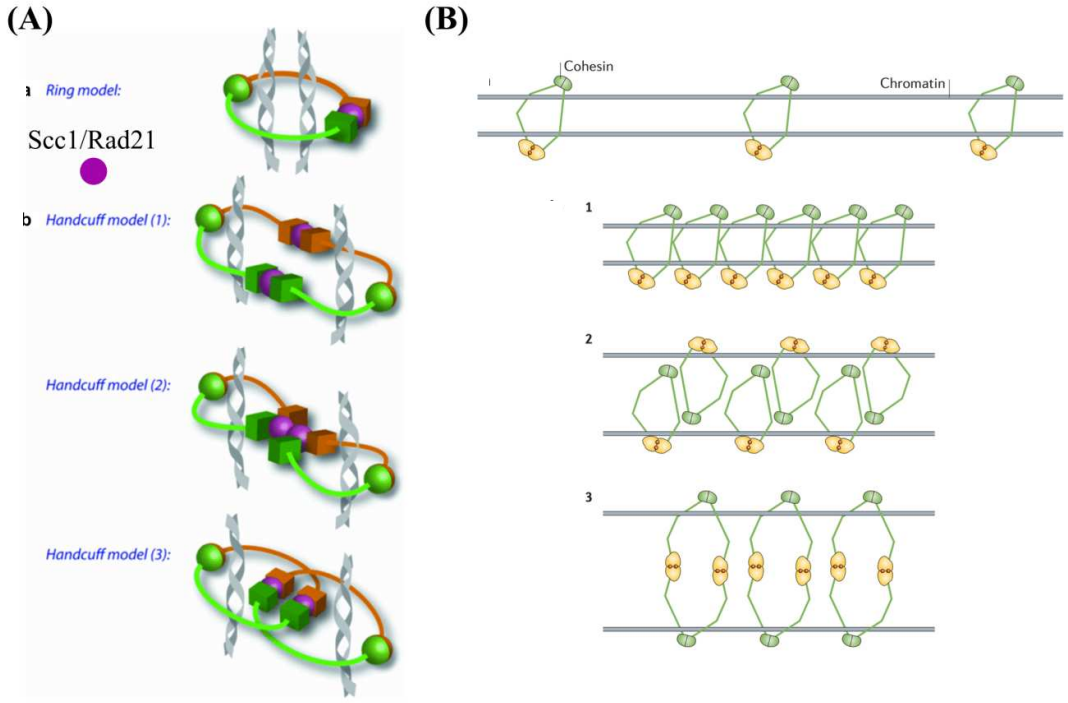


Figure 17: Proposed models of DNA cohesion mechanism. A) Suggested models for the interaction of SMC proteins with DNA (Feeney et al, 2010). The "handcuff" and "ring" models are shown and, so far, there is no clear evidence that favours one model rather than the other. B) Different models of the possible interaction of cohesins with sister chromatids. Figures taken from (Feeney et al, 2010) and (Hirano, 2006), respectively.

There are currently two proposed models describing the possible mode of DNA-binding by SMC proteins, known as the ring (intramolecular) and the handcuff (intermolecular) models (Hirano, 2006) (Figure 17), both of which are supported by *in vitro* studies (Feeney et al, 2010). In the ring model two DNA molecules are encircled by a single SMC dimer.

Evidence for this model comes from Fluorescence Resonance Energy Transfer (FRET) measurements and co-immunoprecipitation experiments during which multiple cohesin complexes were not detected after DNA binding, suggesting that only a single ring is wrapped around the two duplex DNAs (Feeney et al, 2010). The two-ring "handcuff" model, in contrast, proposes that each sister chromatid is encircled by one cohesin-ring and is supported by experiments which revealed self-interaction of Scc1 protein. This would lead to a higher order organization such as a tetrameric assembly in which SMCs can then interact through ATP-binding and, thus, associate two DNA duplexes (Feeney et al, 2010).

Cohesin proteins are loaded onto DNA in an ATP-dependent fashion (Figure 16-17). Dissociation occurs, once the mitosis process takes place, through phosphorylation of one of the interaction partners, notably Scc1, which induces opening of the cohesin ring (Hauf et al, 2005). However, the hinge domain has also been suggested as a potential site of regulation of the opening/closing of the long cohesin arms (Shintomi & Hirano, 2007), leaving even more uncertainty regarding the details of the full process (Figure 17B).

1.5.4. Rad50, a well studied SMC-like protein

ABC transporter and SMC NBDs are characterized by two well conserved motifs, the Walker-A and –B (Figure 18A), and by a specific sequence involved in ATP binding, named the signature sequence (Altenberg, 2003; Davidson et al, 2008). The Walker-A motif contains the typical ***GXXXXGK(T/S)*** sequence, where X can be any amino acid. The Walker-B motif was originally predicted to be ***(R/K)XXXXGXXXXLhhhhD*** (Walker et al, 1982), but has recently been revised to the more common ***hhhhDE*** primary structure (Hanson & Whiteheart, 2005), where ***h*** denotes any hydrophobic residue. The overall arrangement of the NBD architecture is similar to a Rossmann fold (Figure 14), with six parallel β-strands linked to two pairs of α-helices (Hirano et al, 2001; Hirano & Hirano, 2004; Hirano, 2005; Rao & Rossmann, 1973). The Q-loop (coloured in turquoise in Figure 18B) is also involved in coordination of the Mg^{2+} atom.

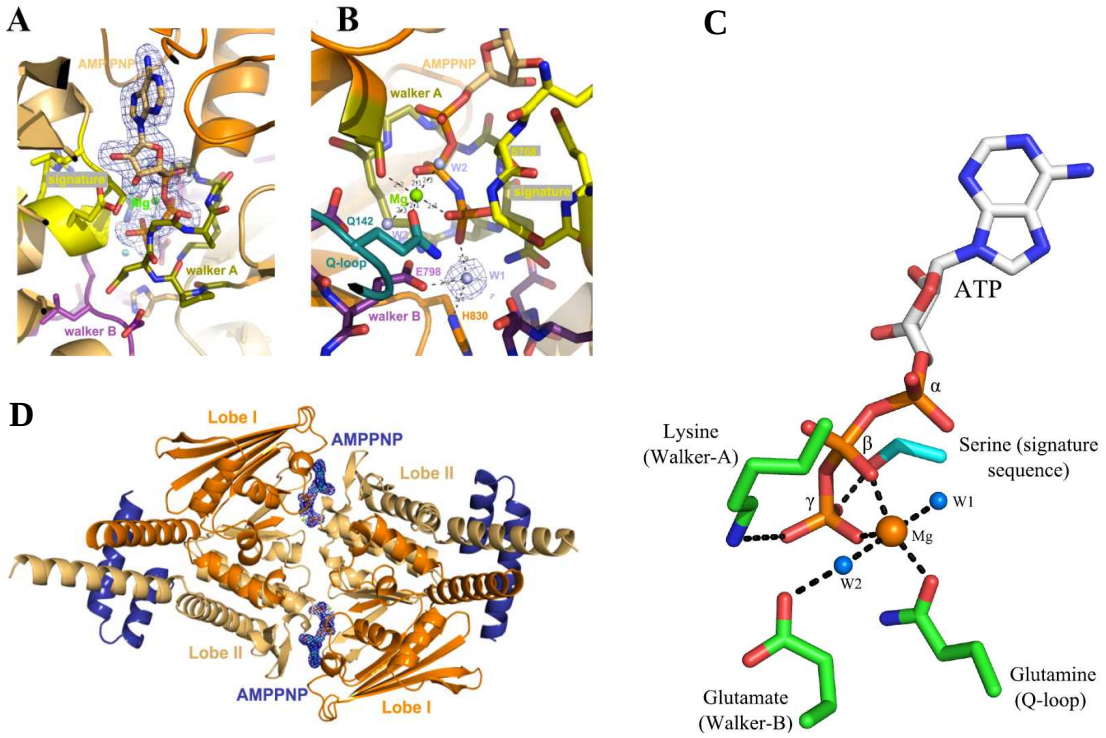


Figure 18: Close-up view of the ATP-binding site of *Pyrococcus furiosus* Rad50. A) ATP analog, AMP-PNP, is shown in density, with the conserved Walker and signature motifs labeled. B) Residues interacting with AMP-PNP are shown as sticks. Mg^{2+} ion is shown as a ball (green). C) Details of the most relevant residues involved in Mg^{2+} coordination (Q-loop and Walker-B motif) and ATP-binding (Walker-A motif and signature sequence). D) Representation of the localization of the two AMP-PNP molecules in the arrangement of Rad50 dimeric state. Figures A, B and D taken from (Mockel et al, 2011).

ATP hydrolysis is important in many processes in which energy is required to achieve protein function. The γ-ortho-phosphate of the ATP is cleaved by nucleophilic attack of a water molecule (Figure 18C), producing ADP and free inorganic phosphate. SMC proteins use ATP binding to induce head-head engagement (Hirano, 2002). Structural studies have confirmed that ATP binding induces the dimerization of NBDs (Figure 18D), a process required to form the full nucleotide-binding pocket (Hopfner et al, 2000; Lim et al, 2011; Williams et al, 2011). The ATP binding and hydrolysis cycle of SMC proteins has been dissected through site-directed mutagenesis of residues from the conserved Walker-A, Walker-B motifs and the signature sequence (Hirano et al, 2001; Hirano & Hirano, 2004; Lammens et al, 2004). The conserved lysine residue from the Walker-A motif and the signature sequence are responsible for nucleotide recognition and binding (Hanson & Whiteheart, 2005). The Walker-B motif is characterized by the presence of a glutamate residue located close to the lysine of the Walker-A motif, which catalyzes the nucleophilic

attack of a water molecule on the γ-phosphate of ATP. Mutating this conserved glutamate to glutamine in the conserved Walker-B motif, for example, disrupts the ATPase activity of SMC proteins (Lammens et al, 2004; Schwartz & Shapiro, 2011) and can result in inhibition of chromosome segregation or, more generally, interfere with stable association with DNA. This could be due to the induced "stickiness" of these mutated head domains, which trap the ATP in their binding pockets (Hirano, 2005). The effects of mutations occurring in the Walker-A motif and involving the ATP-induced protein dimerization (Koroleva et al, 2007) are not fully understood.

1.5.5. The MRN complex

As mentioned previously (§ 3.1), the MRN complex, consisting of Mre11, Rad50 and Nsb1 proteins, is involved in various aspects of DSB repair in eukaryotes (D'Amours & Jackson, 2002; Kanaar & Wyman, 2008; Wyman et al, 2009). Rad50 and Mre11 are suggested to interact with the damaged DNA, while Nbs1 (Xrs2 in yeast) associates with the MR complex, in a stoichiometry M_2:R_2 (Figure 19), upon activation of the cell-cycle checkpoint (Assenmacher & Hopfner, 2004; Dupre et al, 2006). Chromosomal aberrations can derive from abnormal regulation of MRN complex signalling pathway, which is mediated by ATM. Besides HR, the MR complex is also involved in other DNA repair pathways, like NHEJ (§ 2.3.1) and repair of interstrand cross-links.

Rad50 is a member of the SMC protein family and contains an ABC-like head domain, which can bind and hydrolyse ATP (Assenmacher & Hopfner, 2004). The head domain of Rad50 binds and hydrolyses ATP in order to drive the DNA-binding and processing activities of the MR complex (Mockel et al, 2011) (Figure 19-20). A rather flexible (van Noort et al, 2003) and extended coiled-coil domain (~50nm in length) protrudes out of the head domain and allows for intramolecular association of the NBDs on one end, while forming a hook motif on the opposite end (Figure 19). DNA is bound by the globular NBD domains with the coiled-coils on either side of them (de Jager et al, 2001; Hopfner et al, 2001; Lim et al, 2011; Williams et al, 2011; Williams et al, 2008) (Figure 7). ATP binding to the NBD stimulates engagement of the two head domains and repositioning of the coiled-coil domains (Figure 20). Extensive studies have been performed on the MRN complex over the past decade. These include structural studies on the archaeal Rad50 protein from *Pyrococcus furiosus* alone and in complex with the Mre11 nuclease (Hopfner et al, 2002; Hopfner et al, 2001; Hopfner et al, 2000; Lammens et al, 2011; Lim et al, 2011; Mockel et al, 2011; Williams et al, 2011).

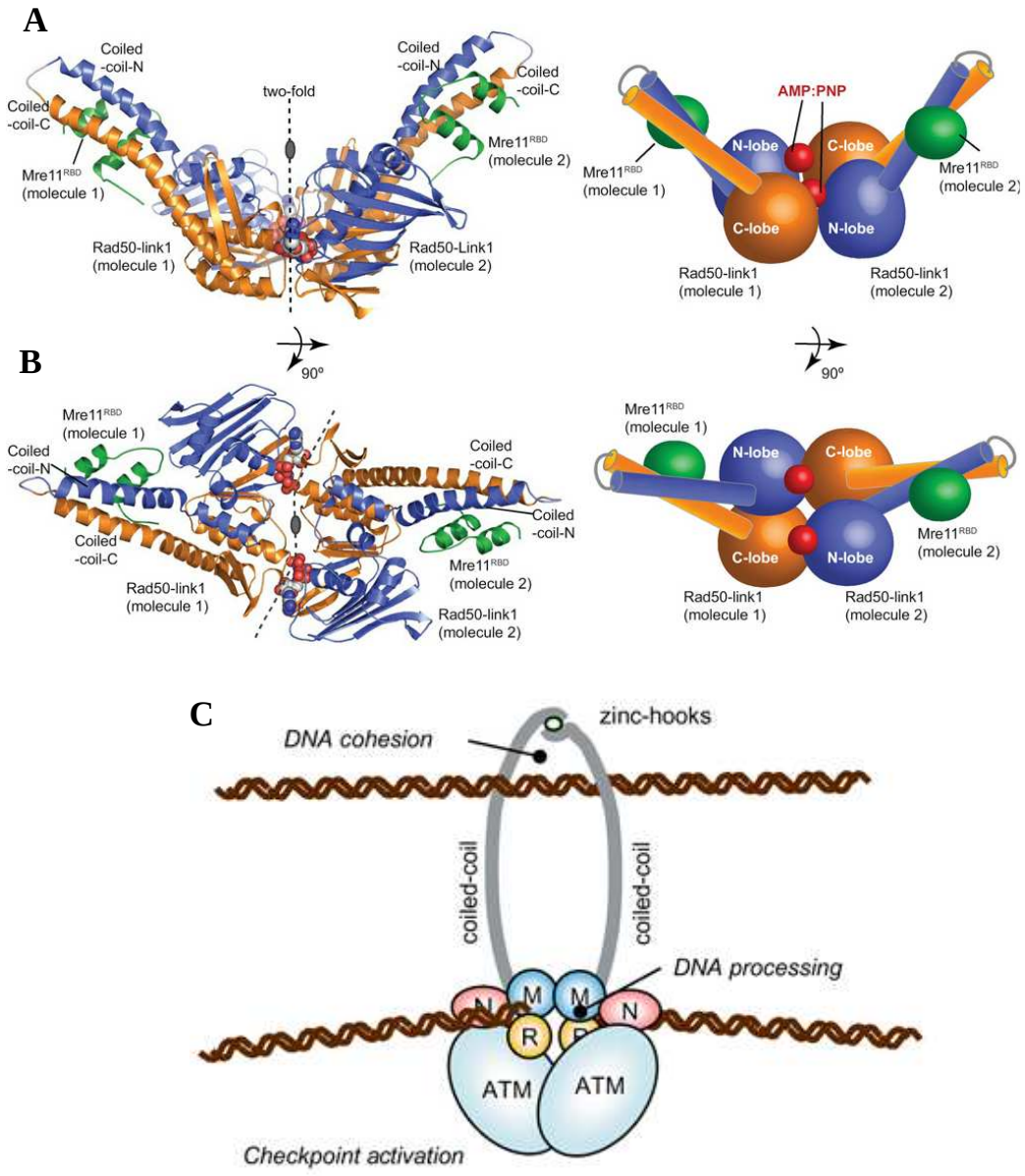


Figure 19: Dimerization process of Rad50 protein. A) Crystal structure of *P. furiosus* M_2R_2 complex. Front view is shown. B) View from the top is shown, with the Mre11 molecules indicated. C) Proposed model of interaction of the MR complex with an ATM dimer and sister chromatids. Figure taken from (Assenmacher & Hopfner, 2004).

The crystal structure of the *Thermotoga maritima* MR complex (tmMR) (Lammens et al, 2011) provided a model for the interplay of Mre11 and Rad50 in bridging dsDNA molecules (Williams et al, 2011). Entrance of ATP in the conserved binding pocket leads to a higher affinity DNA binding site. Mre11 dimers can tether and process both DNA duplexes (Figure 20) arising from either IR damage or in the case of fork collapse during DNA

replication (Williams et al, 2008). Interaction of the damaged DNA with the sister chromatid is mediated by the Rad50 hook, through binding of Zn^{2+} ions (Hopfner et al, 2002).

Recent published work allowed the proposal of an atomic level model of the ATP-dependent cycle of tmMR (Figure 20). A large transition is induced by ATP-binding, demonstrating that the tmMR is capable of adopting a closed state, even in the absence of dsDNA (Mockel et al, 2011). The proposed model (Figure 20) suggests a cooperative switch, driven by ATP, of Mre11 from the open to the closed state. The model could explain the entire mechanism of DNA tethering by the MR complex, since it has already been proven that Rad50 head domains, engaging in an ATP-dependent manner, can assemble DNA fragments (Bhaskara et al, 2007).

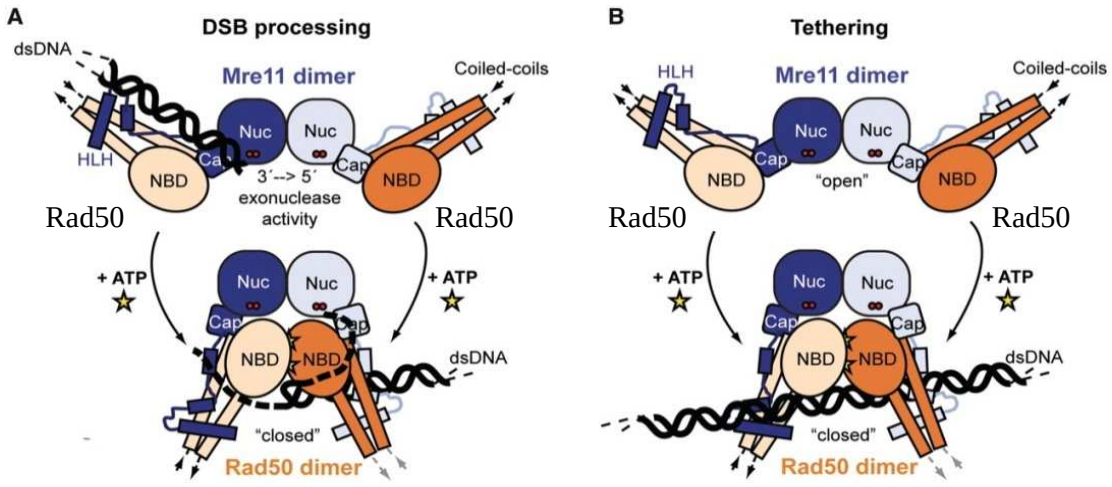


Figure 20: Hypothetical model for ATP-dependent tethering of dsDNA ends by the tmMR (Mockel et al, 2011). A) Conformational changes of the MR module could be responsible for the unwinding of dsDNA and promote endonucleolytic cleavage of ssDNA. B) The extended coiled-coil of Rad50 undergoes an extensive conformational reorientation, which could prevent intramolecular MR interactions upon ATP-dependent engagement from occurring. As a consequence, MR intermolecular interactions are the most favoured. Figure taken from (Mockel et al, 2011).

1.6. RecN: an unusual SMC-like protein

1.6.1. RecN domain organization

The bacterial RecN protein, like Rad50, is involved in DSB repair and belongs to the SMC protein family with which it shares a number of common structural and functional features (Ayora et al, 2011; Graumann & Knust, 2009; Reyes et al, 2010). Phylogenetic analysis (Figure 13) highlights how MukB, Rad50 and RecN are related both to each other and to other SMC proteins (Cobbe & Heck, 2004). A schematic model of RecN has been

proposed (Figure 21) in which the protein adopts the same architecture as other bacterial SMC proteins (Graumann & Knust, 2009).

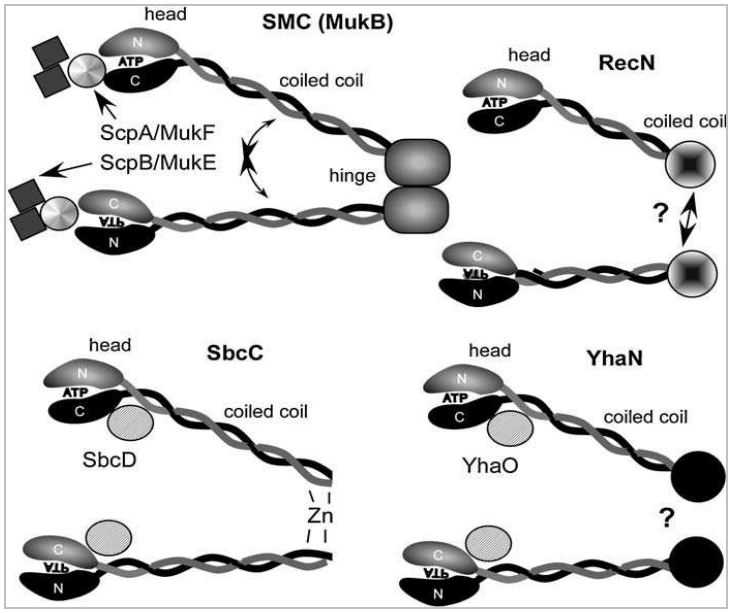


Figure 21: Schematic models of the structures of SMC and SMC-like proteins. The ***head*** domains form globular architectures, which have been shown to be responsible for ATP-binding and hydrolysis. These catalytic head domains are connected by a long coiled-coil in the case of SMC, SbcC and YhaN proteins. In RecN the length of the coiled-coil is much shorter. It is not known yet the reason for such a difference. Moreover, the mode of dimerisation of RecN is also not known (indicated as a question mark in the figure). Experimental evidence and bio-informatics studies reveal that all SMCs have interaction partners that are implicated in protein function (Graumann & Knust, 2009). An interaction partner (PNPase protein) has recently been proposed also for RecN (Cardenas et al, 2011). Figure taken from (Graumann & Knust, 2009).

Sequence alignment of the most studied bacterial RecN proteins shows that they possess highly conserved regions, which are predicted to be related to ATP hydrolysis activity (Funayama et al, 1999) (Figure 21). *D. radiodurans* RecN presents a unique N-terminal extension (Figure 22). This feature is not unusual in DR, since a number of other proteins of the recombinational repair pathway also have unusual extensions (Cox & Battista, 2005). However, it is unclear whether this extension affects the function of the protein.

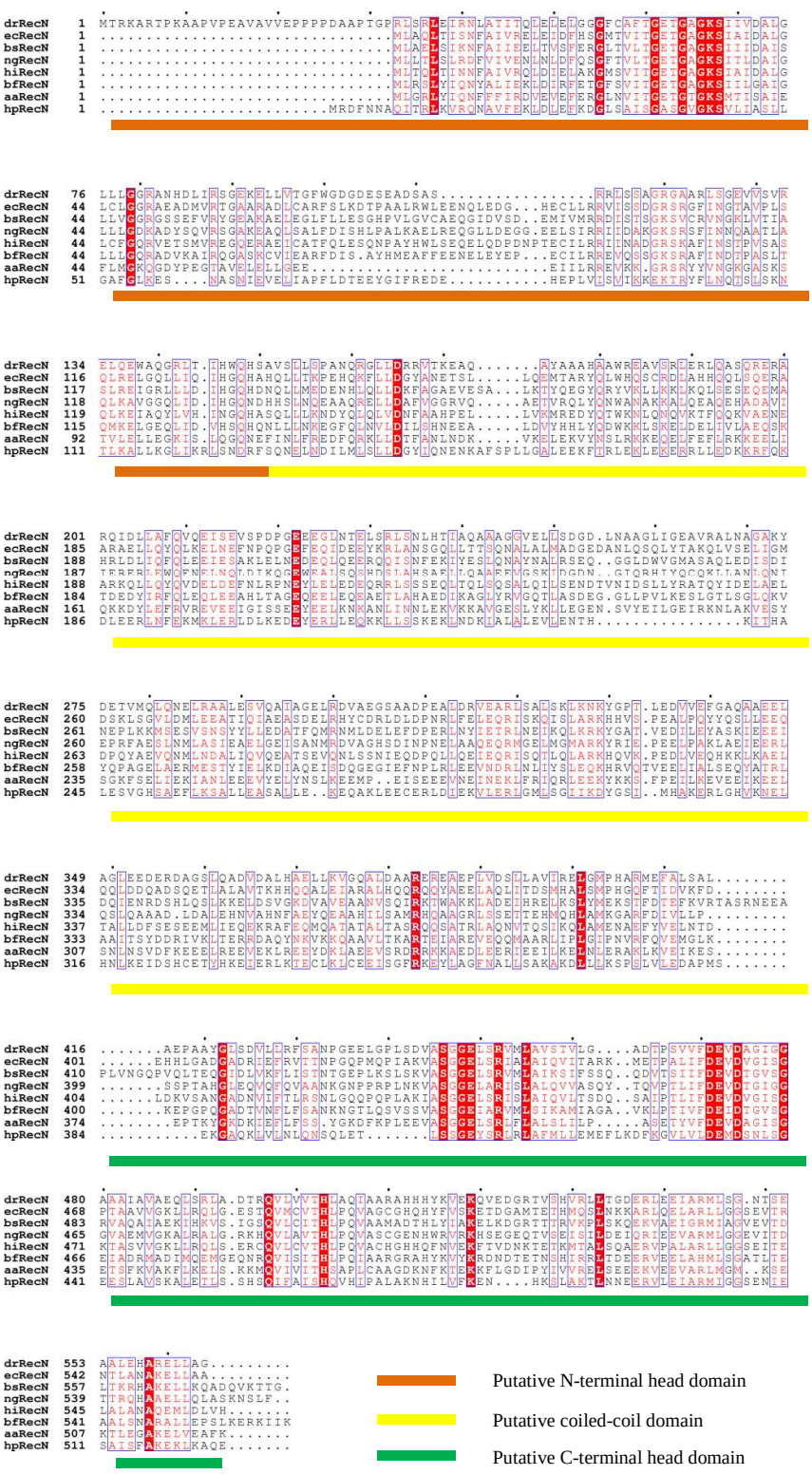


Figure 22: Amino acid sequence alignment of bacterial RecN. RecN proteins from *D. radiodurans* (dr), *E. coli* (ec), *B. subtilis* (bs), *N. gonorrhoeae* (ns), *H. influenzae* (hi), *B. fragilis* (bf), *A. aeolicus* (aa) and *H. pylori* (hp) were aligned. This and all other sequence alignments shown in this work were made using ClustalW (Goujon et al, 2010) and ESPript (Gouet et al, 1999).

The N- and C-terminal regions of SMC and SMC-like proteins, forming the NBDs and containing the major functional motifs, are particularly conserved. This is also the case for RecN (Figures 22-23).

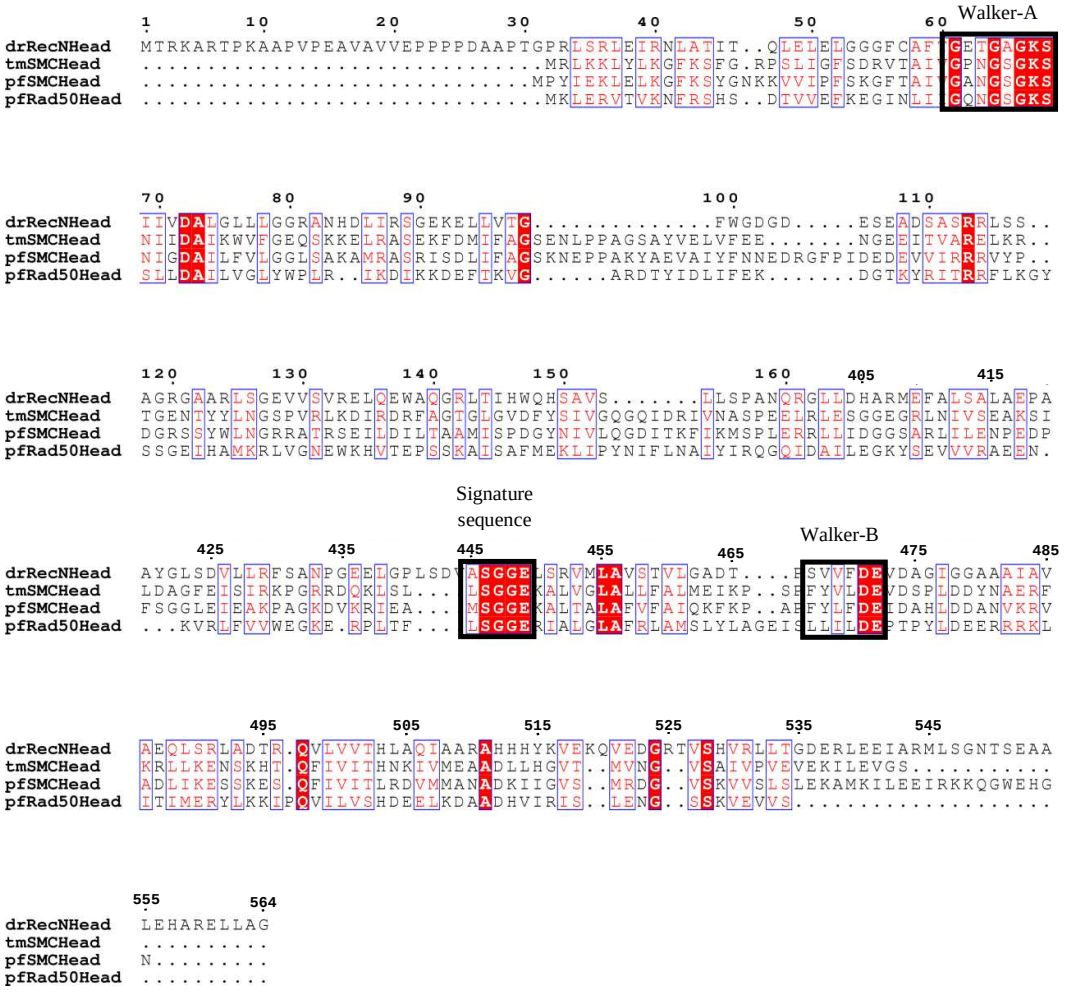


Figure 23: Amino acid sequence alignment of the putative head domain of RecN with those of *T. maritima* SMC, *P. furiosus* Rad50 and *P. furiosus* SMC.

Secondary structure prediction analysis of the RecN amino acid sequence highlighted the presence of a long α-helical central region, which likely represents a characteristic conserved structural feature of SMC-like proteins: the coiled-coil domain. This kind of architectural organization is a remarkable feature of all proteins belonging to the SMC family. It adopts a typical anti-parallel helical arrangement (de Jager et al, 2001) and has been suggested to contain local regions of flexibility (van Noort et al, 2003). In Rad50 and all bacterial and eukaryotic SMC proteins, the coiled-coil has a rather conserved length, between 900 and 1250 amino-acid residues, corresponding to a ~50 nm long structure. A dimerization

interface in the form of a hinge or hook domain (Hopfner et al, 2002; Li et al, 2010; Nasmyth et al, 2002) characterizes the region where the coiled-coil turns in order to form the anti-parallel filament. These regions drive dimer formation, which is needed to open/close the coiled-coil arms for holding DNA (Hirano, 2002). In RecN, however, the anti-parallel helical architecture and the dimerization interface, for which very little is known so far, is proposed to be much shorter than for SMC proteins (Graumann & Knust, 2009; Lowe et al, 2001) (Figure 21, 22).

1.6.2. RecN and DSB repair

HR has been presented as the most efficient pathway used by DR to restore its genome integrity after being exposed to harsh conditions. Involvement of RecN in the early steps of DSB recognition was revealed by *in vivo* studies conducted on *Bacillus subtilis* during which fluorescent labelled RecN was shown to localize to discrete foci following exposure to UV-irradiation or chemical agents causing DNA damage (Kidane et al, 2004; Sanchez & Alonso, 2005). RecN was therefore proposed to act as a sensor for DSBs.

The response of DR to IR is rather complex and involves the transcription of essential genes, a number of which encode for DNA repair proteins. However, *recN* gene is constitutively expressed in DR and its transcription is not upregulated in response to DSBs (Tanaka et al, 2004). In this way, the protein is always readily available at an early stage after induction of DNA damage. In *B. subtilis*, RecN has been shown to bind and protect 3'-ssDNA tails of duplex DNA (Sanchez & Alonso, 2005), while in DR, RecN displays preference for dsDNA (Reyes et al, 2010). Its ATP hydrolysis activity is also stimulated by addition of dsDNA. In addition, DR RecN shows cohesin-like activity (Rees et al, 2009; Reyes et al, 2010), which implies a putative role in tethering DNA ends, as suggested for Rad50 (§ 5.5).

In fact, RecN promotes dsDNA-tethering in a similar way to SMC cohesins, regardless of the type of DNA ends (blunt or overhangs). This lack of specificity suggests that RecN may carry out a structural role in tethering two DNA duplexes for subsequent repair by the HR machinery.

Objectives of the thesis

The aim of this PhD project was to gain a detailed understanding of the mechanisms underlying DSB recognition in *D. radiodurans*, which are thought to be regulated by RecN.

At the beginning of the project no structural information was available for DR RecN and very little was known in the literature about its function. The main goal was therefore to obtain structural information using X-ray crystallography, Small Angle X-ray Scattering or Electron Microscopy techniques. A second goal was the characterization of the role of RecN in DSB repair in DR, with a particular focus on the relationship between DNA-binding and ATP hydrolysis.

The achievements obtained at the end of this PhD work are presented here and have also been the subject of three manuscripts, included in Appendix-II:

I. Pellegrino S., Radzimanowski J., McSweeney S., Timmins J., Expression, purification and preliminary structural analysis of the head domain of *Deinococcus radiodurans* RecN (2012) *Acta Crystallographica section F, 68, pag. 81-84*

II. Pellegrino S., de Sanctis D., McSweeney S., Timmins J., Expression, purification and preliminary structural analysis of the coiled-coil domain of *Deinococcus radiodurans* RecN (2012) *Acta Crystallographica section F*, 68, pag. 218-221.

III. Pellegrino S., Radzimanowski J., de Sanctis D., McSweeney S., Timmins J., Structural and functional characterization of RecN: New insights into double-strand break repair (2012) *Nucleic Acids Research*, (submitted).

Chapter 2:

Materials & Methods

Summary of chapter 2

This chapter describes the principal protocols used in this thesis work: from DNA amplification to cloning, from protein expression to crystallization to ultimately finish with a section dedicated to X-ray diffraction, data processing and protein structure determination.

Résumé du chapitre 2

Ce chapitre décrit les principaux protocoles utilisés dans ce travail de thèse : de l'amplification de l’ADN au clonage, de l'expression des protéines à la cristallisation pour finalement terminer avec une section dédiée à la diffraction aux rayons X, au traitement de données et à la détermination de structure des protéines.

2.1. Gene amplification

DNA fragments codifying for the proteins of interest were amplified by Polymerase Chain Reaction (PCR), using either DR genomic DNA (method described in paragraph 1) or plasmid DNA. The full *recN* gene was amplified in order to produce the full-length protein. Later, other constructs were designed and the relevant DNA sequences were amplified for subsequent cloning into a suitable vector for protein expression (§ 2). Primers used in this study are listed in Table 4.

Construct	**Forward primer sequence**	**Reverse primer sequence**
RecN	5'−CACCATGCGCAAGGCC CGTA−3'	5'−TTAGCCAGCCAGCAACTC GC−3'
RecN*head* N-terminus	5'−CACCATGACCCGCAAG GCCCGTA−3'	5'−CACAAGCGACGTTGGATG CTTGCTCGACTCGCTGGCCT GGAGGCGCTCC−3' (Rev_1)
RecN*head* C-terminus	5'−GCCAGCGAGTCGAGCAA GCATCCAACGTCGCTTGTG CCACGAG−3' (For-2) 5'−CCAACGTCGCTTGTGCCA CGAGGCAGCGTGGACGCCC TGCACGCCG−3' (For-1)	5'−TTAGCCAGCCAGCAACTC GC−3'
RecN*cc* domain	5'−CACCCAGCGCGAGCGGG CGCGG–3'	5'−TTACACATCGGCTTGCAG GCTGCCCG−3'
RecN Coiled-coil 240-387	5'−CACCGCGCAAGCGGCGGC GGGC−3'	5'−TTAGGCCTCGCGTTCGCG GGCG−3'
RecNΔ*dd* N-terminus	5'−CACCATGACCCGCAAG GCCCGTA−3'	5'−CACAAGCGACGTTGGATG CTTGCTCGACTCGTGGAGGT TGGAAAGG−3' (Rev_1)
RecNΔ*dd* C-terminus	5'−CGCGGCGGAGTCGAGCA AGCATCCAACGTCGCTTGTG CCACGAG−5' (For-2) 5'−CCAACGTCGCTTGTGCCA CGAGGCAGCGTGCAGGCCA TCGCGG−3' (For-1)	5'−TTAGCCAGCCAGCAACTC GC−3'
RecN34	5'−CACCATGCTCTCGCGGC TGGAAATTCGCAACC−3'	5'−TTAGCCAGCCAGCAACTC GC−3'
RecNΔ47	5'−CACCATGCGCAAGGCC CGTA−3'	5'−TTATTCCACCTTGTAGTGG TGGTGG−3'

Table 4: List of the DNA oligonucleotide primers used for PCR amplification of the various RecN constructs (reported on the left) which were further characterized in this work. The vector-specific sequence (CACC) required for TOPO-cloning is indicated in red.

PCR reactions were prepared as follows, in a final volume of 25 µl:

- Primer forward: 1 µl (20 µM stock)
- Primer reverse: 1 µl (20 µM stock)
- dNTP mix: 1 µl (10 mM stock concentration)
- Buffer High Fidelity Polymerase: 2.5 µl (10X)
- High Fidelity Polymerase enzyme mix: 0.25 µl (100 units)
- Genomic DNA: 0.5 µl genomic or plasmid DNA (~100 ng/µl)
- DMSO: -/+ 1.25 µl (final concentration 5%)
- H_2O: 19.75 (18.5) µl

Reactions were carried out in a Thermocycler (Eppendorf) using the following steps:

1. 95 ºC for 5 minutes
2. 95 ºC for 1 minute
3. 50-58 ºC for 1 minute
4. 72 ºC for 1-2 minute

} 30-35 cycles

5. 72 ºC for 10 minutes
6. 4 ºC HOLD

The temperature of annealing (T_A; step 3) varied depending on the melting temperature and the length of the primers. The elongation time (step 4) changed according to the length of the sequence to be amplified since the High Fidelity (HF) polymerase can synthesise ~1000 nucleotides per minute.

Amplification of the DNA fragment codifying for RecN*head* was particularly complicated, since a peptide linker of 14 residues, with amino acid sequence ESSKHPTSLVPRGS, was designed to join the N- and C-termini of RecN. Two different PCRs were set up for amplification of the single N- and C-termini fragments using the primers listed in Table 4. A third reaction was performed using the original forward and reverse primers used for full-length RecN amplification. The N- and C-termini of the construct codifying for the RecN*head* domain contain an overlapping region, which was accounted for in the design of the primers (Figure 24). This is shown schematically in Figure 24.

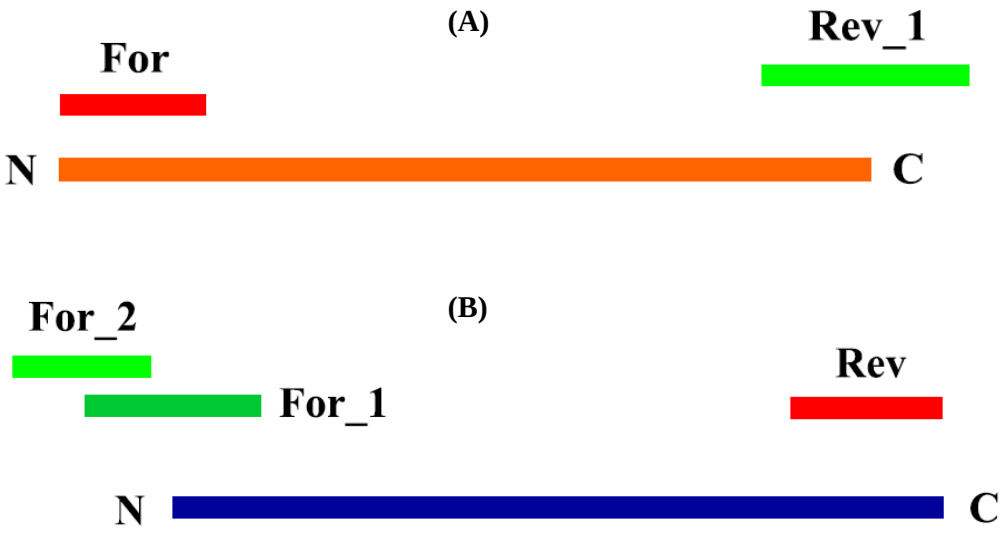


Figure 24: Schematic of the N- and C-terminal fragment amplification of RecN*head* protein. A) Schematic diagram of the N-terminal fragment for the RecN*head* domain. B) Schematic diagram of the primers designed for creation of the C-terminal region of RecN*head* construct. The N-terminal portion will have an overlapping sequence (contained in the so called Rev_1) complementary to For_2 primer.

The two overlapping forward primers, in the case of C-terminal DNA amplification, were mixed at different ratios. The "For_1" was included at a concentration 10 times lower than "For_2". The first 5 cycles of PCR (with a T_A offset at 58 °C) used the fragment specific forward primer ("For_1") for creating the first part of the overlapping region with the second forward primer. After the first cycles, the primer "For_2" will be the only forward primer available (given the higher concentration), together with the reverse primer, and the complete fragment will be amplified. The T_A was decreased to 55 °C for the last 30 PCR cycles. The two resulting DNA fragments corresponding to the N- and C-termini of RecN were purified and then mixed at equivalent concentrations for the next amplification. The original N-terminal forward primer and the C-terminal reverse primer for the full-length RecN were used at this stage. The first 5 cycles of PCR were performed in the absence of both primers, in order to stimulate annealing and extension of the overlapping region of the two fragments (For_2 with Rev_1). Once this step occurred, samples were placed on ice and supplemented with primers at the concentrations previously reported, and returned to the thermocycler to complete the remaining 30 cycles of PCR. The same principle was adapted for the amplification of the construct codifying for the deletion mutant, namely RecNΔ*dd*, which also contained the linker peptide of 14 residues.

All the PCR products were run on a 1% agarose gel (Figure 25) prepared in a solution 0.5X TBE (10x Stock: 108 g Tris base, 55 g Boric acid, 40 ml 0.5 M EDTA pH 8.0 for 1 L) containing SYBR Safe (Invitrogen). DNA bands were excised and purified using the QIAquick Gel Extraction kit (QIAGEN) and finally eluted in 30 µl of EB buffer (10mM Tris-HCl pH 8.5).

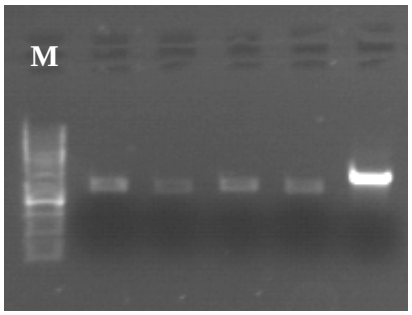


Figure 25: Purified fragments of RecN*head* domain and RecN34 constructs (on the right). Different DNA starting materials were tried (plasmid and genomic DNA) for RecN*head* and in absence or presence of 5% DMSO. M: 1kb ladder (Fermentas).

2.2. Cloning

Amplified DNA fragments coding for the target proteins were sub-cloned into a specialized vector for transformation into *E. coli* cells. The forward primers all included a characteristic nucleotide sequence, namely CACC, upstream of the gene specific sequence (Table 4). This is a feature required for directional cloning into pET151 with TOPO technology (Invitrogen). The principle of directional cloning, through a selective DNA topoisomerase, of the target fragment into the TOPO vector is summarized in the figure below (Figure 26):

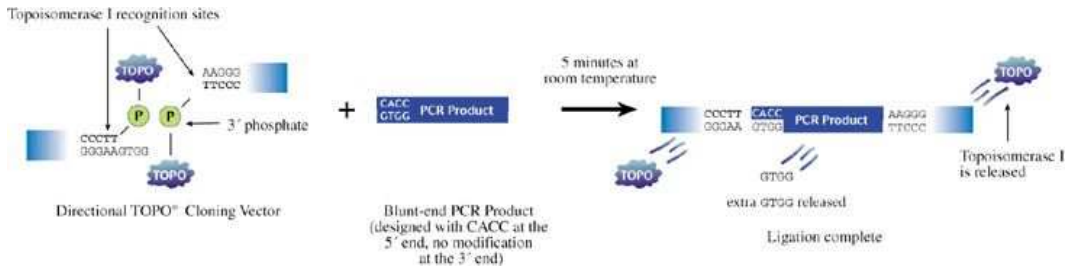


Figure 26: Schematic diagram of the directional cloning that occurs in pET-TOPO vectors.

Map of pET151-TOPO vector is shown in Figure 27. The poly-histidine tag (6xHis) together with the TEV cleavage site, are present upstream the gene insertion site. This vector contains the resistance for Ampicillin.

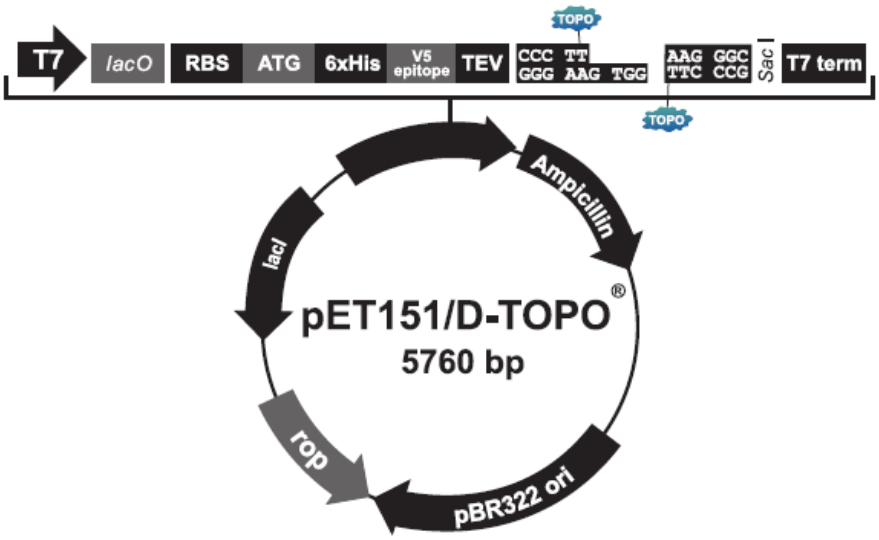


Figure 27: Schematic map of the pET151-TOPO vector used for cloning all the different constructs used during this work. The gene insertion site is indicated (TOPO), together with the poly-histidine tag (6xHis) and the TEV cleavage site (TEV). Figure taken from http://www.invitrogen.com/site/us/en/home.html

The reaction is carried out at Room Temperature (RT) for 5-30 minutes (depending on the length of the insert of interest) and then the entire reaction mix is used for the following step of transformation into *E. coli* cells.

2.3. Transformation of competent cells and control digestion of the plasmid

The products of the cloning reactions were then used for transformation of competent *E. coli* TOP10 cells (Invitrogen). For each transformation, 50 µl of cells were incubated on ice with 6 µl of the cloning reaction for 30 minutes. A heat shock of 45 seconds at 42 ºC was used for relaxing the cell membrane and allowing the vector to enter. Cells were placed 5 minutes on ice and then supplemented with 500 µl of SOC medium (2% Bacto-tryptone, 0.55% yeast extract, 10 mM NaCl, 10 mM KCl, and after autoclaving add 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose). All the cells were plated on LB-agar plates with the appropriate selective antibiotic (in this specific case, ampicillin at 0.1 mg ml^{-1}) to allow for selection of recombinant clones. The plates were incubated at 37°C for at least 12 hours, until separate colonies could be picked and tested for the presence of the gene of interest. Single colonies were grown in 5 ml of LB supplemented with 1/1000 ampicillin (stock concentration: 100 mg ml^{-1}) and then used for extraction of the plasmid containing the insert of interest. A Miniprep kit (QIAGEN) was used for extracting and purifying plasmid DNA, which was finally eluted in 30 µl of buffer EB. Digestion of a small aliquot (10 µl) with the restriction enzymes SacI

and NcoI (New England Biolab) allowed to detect whether the desired insert was present. Digestion reactions showing the presence of a DNA band of the right size (Figure 28) were sent for DNA sequencing to check the correctness of the DNA sequence

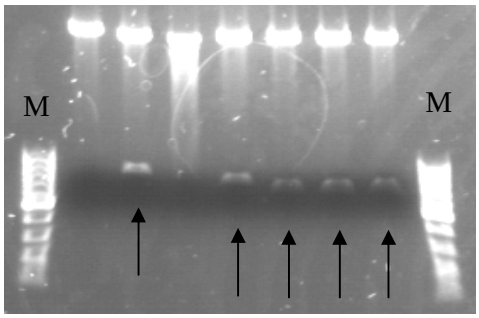


Figure 28: Digestion control for some of the colonies grown in LB+agar supplemented with the appropriate antibiotic. On the top of the gel the vector alone is visible, while the lower bands correspond to the desired fragments (black arrows). M: DNA ladder (100 bp, Fermentas).

2.4. Expression and protein purification

The constructs amplified and cloned in the pET151 vector were all transformed into *E. coli* cells, strain BL21 Star (DE3) using the protocol described in § 3. Cells were grown at 37 ºC for 2.5-3 hours, until Optical Density at 600 nm (OD_{600}) reached 0.6-0.9. At this point, isopropyl-b-D-thiogalactopyranoside (IPTG) was supplemented at a final concentration of 1 mM in order to induce protein expression.

Lysis of the proteins was achieved by mechanical force either in a cell disrupter (CONSTANT SYSTEMS Ltd.) operating at 195 MPa at 4°C or with a French press (SLM Instruments, maximum pressure achieved 1000 psi). The composition of the buffer used for lysis was 50 mM Tris-HCl pH 7.5 (RecN), 1 M NaCl, 5 mM $MgCl_2$, 5% glycerol and 5 mM β-mercaptoethanol (β-ME) for RecN (**L-1**) or 50 mM Tris-HCl pH 8, 1M NaCl, 5mM $MgCl_2$ and 5 % glycerol for RecN*head*, RecN*cc* and RecNΔ*dd* (**L-2**). The resulting lysate was centrifuged at 48,384 g at 4°C and the supernatant collected and loaded onto a 5 ml HisTrap™ HP column (GE Healthcare), previously equilibrated in buffer **A-1** containing 50 mM Tris pH 7.5 (pH 8.0 for RecN*head*, RecN*cc* and RecNΔ*dd*, named as **A-2**), 1 M NaCl (300 mM for RecN*head*, RecN*cc* and RecNΔ*dd*), 5 mM $MgCl_2$, 5mM imidazole. A wash with at least 5 column volumes of buffer A was done before starting the elution step. When the absorbance reading was stabilized, protein was eluted either with a linear gradient, from 0 to 100% over 40ml, or by steps of buffer **B-1** containing 50 mM Tris pH 7.5 (pH 8.0 for RecN*head*, RecN*cc* and RecNΔ*dd* and named as **B-2**), 1 M NaCl (300 mM for RecN*head*, RecN*cc* and RecNΔ*dd*), 5 mM $MgCl_2$ and 500 mM Imidazole. RecN was eluted in two steps: one at 4% of buffer B (20 mM imidazole) for removal of most of the contaminants, and the second at 100% of buffer B, which strips the protein of interest from the affinity-column. All

the other constructs, in contrast, were eluted using a linear gradient of imidazole. RecN*head*, RecN*cc* and RecNΔ*dd* eluted at 30-35 % of buffer B, that is approximately 150-180 mM of imidazole. Fractions containing the target protein were analyzed by SDS-PAGE (described in the next section), pooled and dialyzed overnight at 4°C in buffer **C-1** containing 50 mM Tris pH 7.5 (pH 8.0 for RecN*head*, RecN*cc* and RecNΔ*dd*, named as **C-2**), 1 M NaCl (300 mM for RecN*head*, RecN*cc* and RecNΔ*dd*) and 5 mM $MgCl_2$. At the same time His-tag cleavage was performed by addition of tobacco etch virus (TEV) protease (at a final concentration of 0.1 mg ml^{-1}), 0.5 mM EDTA and 1 mM DTT to the protein solution.

A second affinity chromatography (through a gravity-flow Ni^{2+}-column) was performed using Ni-Sepharose Fast-Flow (GE Healthcare) in order to remove the 6xHis-tag and the TEV protease from the protein solution prior to size-exclusion chromatography (SEC). Proteins were loaded onto the column and then extensively washed with buffer C, until no more protein was detected (by using the Bradford solution). A second wash was performed with a few milliliters of buffer A. Elution with buffer B was performed for stripping all the species specifically bound to the column. Samples were collected at each step and were then analyzed by SDS-PAGE.

Proteins were then concentrated to 10-15 mg ml^{-1} by ultrafiltration (Amicon Ultracel 10-30K cutoff, depending of the protein) for being injected onto a Gel Filtration (GF) column. Superdex 200 (GE Healthcare) columns (analytical or prep grade) were pre-equilibrated with buffer **D-1** containing 50 mM Tris pH 7.5 (pH 8.0 for RecN*head*, RecN*cc* and RecNΔ*dd*, named as **D-2**), 1 M NaCl (300 mM for RecN) and 5 mM $MgCl_2$. Proteins were then separated according to their molecular weight but also influenced by the shape they adopt in solution. Fractions containing the main elution peak were collected, verified by SDS-PAGE and then concentrated for further biophysical or crystallization experiments.

At the end of each step of purification, protein concentration was measured. The Nanodrop 1000 spectrophotometer (Thermo scientific) was used in the case of RecN protein: it measures the absorbance of the sample in the range of the visible-UV (750-220 nm) and calculates the concentration of the sample using the Lambert-Beer law:

$$A = \varepsilon l C \qquad \text{(Eq. 1)}$$

where A is the absorption measured at 280 nm, ε the molar absorption coefficient ($M^{-1}cm^{-1}$), l the pathlength (cm) and C the protein concentration (M). Concentration measurements for the head and coiled-coil domains were performed using the Bradford assay (Biorad). 1 µl of the sample was diluted in 999 µl of Bradford solution (red-brown colour), which is pre-diluted in

distilled H_2O (1:5). The solution turns blue in the presence of protein and the absorbance at 595 nm is then measured. A standard curve created using bovine serum albumin (BSA) was then used to convert the absorbance measurement into protein concentration.

2.5. Mutations of the conserved functional residues

Selected point mutations were introduced into full-length RecN and RecN*head* in order to disrupt ATP-binding and hydrolysis ability. Single point mutations at the Walker-A motif, namely K67A, and at the Walker-B motif, namely E472A and E472Q, were introduced using the QuickChange Site-directed Mutagenesis kit (Agilent Technologies), while double point mutations, namely K67A/E472Q and K67A/E472A, were introduced with the QuikChange Multi Site-Directed Mutagenesis Kit (Agilent Technologies). These protocols introduce the desired mutation through a one step PCR reaction. Following digestion of the methylated, non-mutated DNA plasmid with a specific restriction enzyme (DpnI), the recovered plasmid carrying the mutation can be transformed into supercompetent *E. coli* cells (XL1-Blue). The vector is then amplified, extracted and purified as previously described (§ 3) and then used for transformation into competent cells for protein expression (BL21 Star).

Mutants were then expressed and purified accordingly to the protocols described above (§ 4) for RecN and RecN*head* domain in order to compare the behavior of the different proteins in solution. Elution volumes and hydrodynamic radii were used as means for estimating changes in the properties of these proteins in solution.

2.6. SDS and Native PAGE

Purification of the proteins of interest required regular analysis of the resulting fractions in denaturing conditions through sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE). 10 to 15% gels were prepared as described in (Sambrook et al, 1989). Biorad mini-Protean II gel systems were used to run gels at 150-200 V. The solutions were prepared as follows:

- Running buffer (10X) pH 8.3: 30 g/l Tris base, 144 g/l glycine and 10 g/l SDS
- SDS sample buffer (4X) for 8 ml: 3.8 ml H_2O, 1 ml 0.5 M Tris-HCl pH 6.8, 0.8 ml glycerol, 1.6 ml 10% (w/v) SDS, 0.4 ml β-mercaptoethanol and 0.4 ml 1% (w/v) bromophenol blue.

Under denaturing conditions, the proteins are separated and migrate only accordingly to their respective sizes, since the SDS masks their overall protein charge. 10% acrylamide gels were prepared for RecN, containing, for a 5 ml final solution, 1.98 ml H_2O, 1.67 ml

Acrylamide 30%, 1.25 ml Tris 1.5M at pH 8.8, 50 µl of SDS 10%, 50 µl of APS 10% and 2 µl TEMED. 15% acrylamide gels were instead prepared for RecN*head* and RecN*cc* proteins and here water was replaced by acrylamide (2.5 ml) while keeping constant the amount of all the other reagents.

To separate proteins according to their charge, instead, 5% native TBE gels at pH 8.3 were prepared and run in non-denaturing conditions. SDS, therefore, was omitted in the preparation of the gel mix and the running buffer. Sample buffer was replaced by supplying 1 µl of glycerol 100% to the aliquot to load.

To visualise the proteins on either SDS or Native-PAGE, the gels were stained with a coomassie staining solution (containing 0.1% coomassie blue, 40% ethanol and 10% acetic acid) for 10-15 minutes and then incubated in a destaining solution (containing 5% acetic acid and 7% ethanol) until protein bands were clearly visible.

2.7. Limited proteolysis

The protein of interest (10 µg per well) was incubated with different proteases (Trypsin, Subtilisin, Proteinase K, Chymotrypsin) at a 1:5000 ratio. Samples were taken at different time points, such as 0, 1, 5, 10, 15 and 30 minutes in order to observe the degradation pattern of the target protein. Proteases cleave more easily residues present in exposed and flexible regions, like loops or linkers between domains. Samples corresponding to the different time points were then run on a 15 % SDS-PAGE gel. For N-terminal sequencing, the gel was electroblotted onto a polyvinylidene difluoride (PVDF) membrane. The membrane was stained and the bands resistant to protease cleavage were excised and sent for N-terminal sequencing analysis to identify the first 5 amino acid residues at the N-terminus of each band.

2.8. Production of seleno-methionine derivatives of RecN*head* and RecN*cc*

E. coli strain BL21 Star cells containing the plasmid (pET151-TOPO) with the gene encoding for the target protein were grown at 37ºC in a minimal medium composed as follows (for 1 liter culture): 200 ml M9 (stock 5X), 800 ml autoclaved H_2O, 2 ml $MgSO_4$ 1M (autoclaved), 0.1 ml $CaCl_2$ 1M (autoclaved), 20 ml glucose 20% (w/v) filtered. After the OD_{600} reached a value of 0.6-0.8, single amino acids were supplemented in order to inhibit the metabolic pathway for the usual synthesis of methionine. L-lysine (100 mg/l), L-phenylalanine (100 mg/l), L-isoleucine (50 mg/l), L-leucine (50 mg/l), L-valine (50 mg/l) were added to the cell culture. Finally, instead of methionines, Se-methionines (Se-Met) (60 mg/l) were added in order to be incorporated into the newly synthesized protein (Doublie,

1997). After incubation for 15 minutes in order to start Met synthesis, protein expression was induced upon addition 1mM IPTG for 16 hours at 20 ºC. Cells were then centrifuged at 7,548 g for 30 minutes, as for all the proteins characterized in this work, and then frozen at -80 ºC if not freshly used for protein purification.

Cell lysis and purification were carried out according to the protocols used for the respective native proteins. Once sufficient amounts of purified protein were obtained (~15-20 mg final), crystallization drops were set up using the same conditions as for the native proteins and then crystals were used in order to perform anomalous dispersion experimental phasing (SAD).

Composition of the media M9 (5X): 30 g Na_2HPO_4, 15 g KH_2PO_4, 5 g NH_4Cl, 2.5 g NaCl, water to 1 liter volume. The buffer is then autoclaved.

2.9. Multi-Angle Laser Light Scattering

SEC combined with Multi-Angle Laser Light Scattering (SEC-MALLS) and refractometry (RI) is a powerful method for measuring the absolute molecular mass of macromolecules. SEC experiments were performed on a Superdex 200 column (GE Healthcare) equilibrated in a buffer containing 50mM Tris-HCl pH 7.5, 5mM $MgCl_2$ and either 1M NaCl or 150mM NaCl in the case of RecN analysis. Buffer containing 50 mM Tris-HCl pH 8, 300 mM NaCl and 5 mM $MgCl_2$ was, instead, used for analysis of RecN*head*, RecN*cc* and RecNΔ*dd*. Prior to sample injection, the SEC column was calibrated using bovine serum albumin, which has a Stoke's radius of 3.4 nm. On-line MALLS detection was performed using a DAWN-EOS detector (Wyatt Technology Corp., Santa Barbara, CA) equipped with a laser emitting at 690 nm and a RI2000 detector (Schambeck SFD) for refractive index analysis. Light-scattering intensities were measured at different angles relative to the incident beam and weight averaged molar masses (Mw) were calculated using the ASTRA software (Wyatt Technology Corp., Santa Barbara, CA).

2.10. Dynamic Light Scattering

DLS measurements were performed using a fixed scattering angle Zetasizer Nano (Malvern Instruments) and quartz cuvette. The polydispersity index (pdI) was calculated as the square of the normalized standard deviation of an underlying Gaussian size distribution. All experiments were performed at 20 °C. The Stokes-Enstein equation (Nobbman et al, 2007) can be applied for conversion of the diffusion coefficient into the hydrodynamic coefficient "d_H" which ideally represents the size of a sphere with the same diffusion properties.

In the case of the analysis of hydrodynamic radii of RecN and RecN*head* proteins upon ATP addition, samples at a concentration of 80-150 μM of protein were used. Measurements were recorded after incubation for 30 minutes at 18 ºC, either in absence or in presence of 2 mM ATP.

2.11. Small Angle X-ray Scattering

SAXS experiments were all performed on ID14-3 at the ESRF (Pernot et al, 2010). Full-length RecN in high salt conditions (1M NaCl), RecN*head*, RecN*cc*, RecN*Δdd* and RecN*head*$^{E67A/E472Q}$ were measured. Data were collected at 0.931Å using a Pilatus 1M detector (Dectris, Switzerland) which is positioned 2.4 meters away from the sample capillary. The homogeneity of protein samples was previously checked by DLS. Three different protein concentrations were measured together with their corresponding buffers. Samples were flown while exposed to X-rays in order to avoid radiation damage and thus 10 frames of 10 seconds duration each were collected. Buffer subtraction and data processing were performed using *PRIMUS* (Konarev et al, 2003), which also allowed to estimate the radius of gyration (R_g) through Guinier approximation, which is reproducible among all the different measurements. Estimation of the MW was determined either from the scattering intensity at zero angle (I_0) directly from PRIMUS or from the Porod volume, using AUTOPOROD (Petoukhov et al, 2007). When the subtracted curves corresponding to the three different protein concentrations superposed well, inter-particle effects could be excluded and the data could be used for further analysis. Scattering of the buffer was measured before and after the protein sample and then subtracted from the protein in order to maximize the signal/noise ratio of the sample.

The curves corresponding to the highest and the lowest protein concentrations were then merged together and used as input file for GNOM (Svergun, 1992) for determination of the distance distribution function (*P(r)*). The output files were subsequently used for *ab initio* model reconstruction. Data were submitted to DAMMIN (Svergun, 1999) (for RecN, RecN*head*, RecN*cc* and RecNΔ*dd*) in order to perform model building cycles using the EMBL BioSAXS web server (EMBL, Hamburg). In the case of RecN*head*$^{E67A/E472Q}$, the program GASBOR (Svergun et al, 2001b) was instead used for *ab initio* model building. In all cases, several models were produced, analyzed and subsequently aligned, averaged and filtered using the DAMAVER package (Volkov & Svergun, 2003).

Comparison of experimental and theoretical scattering curves for RecN and RecN*head*$^{E67A/E472Q}$ was performed using CRYSOL (Svergun et al, 1995), which also enabled

to determine the goodness-of-fit (GOF) between the data through Chi-squared calculation (χ^2). A suitable model of the RecN*head*$^{E67A/E472Q}$ was obtained using MASSHA (Svergun et al, 2001a), which allows to move the atomic coordinates file in order to fit the experimental scattering curve. The *ab initio* envelopes of RecN, RecN*head*, RecN*cc* and RecN*head*$^{E67A/E472Q}$ were subsequently superimposed to their respective atomic models by running SUPCOMB (Svergun & Kozin, 2001), through an energy minimization calculation of the fit.

2.12. Crystallization

Initial crystallization screening for all RecN proteins was performed at 20 °C using sitting-drop vapor diffusion Greiner Crystal Quick plates. A Cartesian PixSys 4200 crystallization robot (High Throughput Crystallization Laboratory at EMBL Grenoble) was used in order to test 576 different crystallization conditions (method described in (Dimasi et al, 2007)). The following commercial screens (Hampton Research) were set-up: Crystal Screen I & II, Crystal Screen Lite, PEG/Ion, MembFac, Natrix, QuickScreen, Grid Screens (Ammonium Sulfate, Sodium Malonate, Sodium Formate, PEG 6K, PEG/LiCl, MPD) and Index Screen.

Crystals were obtained for many of the constructs used in this work, such as RecN*head*, RecN*cc*, RecNΔ*dd*. Some of the mutants produced for biochemical characterization, such as RecN*head*^{E472Q} and RecN*head*$^{K67A/E472Q}$ were also crystallized. Once initial conditions were defined, manual optimization was carried out by setting up hanging drop plates (Hampton Research). Different buffers, pH, PEG concentrations, PEG sizes, additives (from the Additive Screen, Hampton Research), and protein concentrations were assayed in order to establish the conditions that gave the best crystals.

RecN*head* and RecN*cc* were crystallized as described in paper I (Pellegrino et al, 2012b) and paper II (Pellegrino et al, 2012a) respectively, and the conditions for the native proteins were suitable also for the Se-Met substitutes.

RecNΔ*dd* was also initially crystallized using the robot at the High Throughput Crystallization (HTX) laboratory (EMBL, Grenoble). Crystals were reproducible but the best data set (4 Å resolution) was collected directly from crystals coming from the robot plates. Manual optimization did not improve the quality and, therefore, higher resolution data was not achievable.

RecN*head*^{E472Q} and RecN*head*$^{K67A/E472Q}$ mutants were also crystallized using the robot at the HTX laboratory. Crystals of RecN*head*$^{K67A/E472Q}$ were also reproducible after manual optimization of the initial conditions (chapter 5) in hanging drop plates.

2.13. X-ray diffraction data

Crystals of native and Se-Met derivatives of RecN*head* and RecN*cc* were obtained and X-ray diffraction data was collected at the European Synchrotron Radiation Facility (ESRF) protein crystallography beamlines ID14-4 (McCarthy et al, 2009), ID23-1 (Nurizzo et al, 2006) and ID23-2 (Flot et al, 2010) as described in (Pellegrino et al, 2011). Diffraction data suitable for further analysis was collected for RecN*head*, RecN*cc* and RecNΔ*dd* protein crystals at 100 K. Data were subsequently processed either using iMOSFLM (Battye et al, 2011) (RecN*head* and RecN*cc*) or XDS (Kabsch, 2010) in the case of RecNΔ*dd*. Full redundant data sets were collected for the native and substituted Se-Met proteins (RecN*head* and RecN*cc*). Single-wavelength Anomalous Dispersion (SAD) experiments were performed on Se-Met substituted protein crystals and for this purpose the energy of the X-ray beam was moved to the value corresponding to the absorption K-edge of the Selenium atoms (***peak***). Data were then processed and the intensities scaled using SCALA (Evans, 2006) where anomalous pairs were separated ($I^+ \neq I^-$), since Friedel's Law is broken in the presence of anomalous scatterers. Structure factors were then generated and the output file used either for Molecular Replacement (Mol.R.) in the case of RecNΔ*dd* or for experimental phase determination and model building (RecN*head* and RecN*cc*).

2.14. Structure determination and refinement

The RecN*head* and RecN*cc* crystal structures were solved to resolutions of 3.0 and 2.0 Å respectively, by the SAD method using Se-Met derivative crystals produced as described in Materials & Methods and in paper I and II (Pellegrino et al, 2012a; Pellegrino et al, 2012b). The Phenix suite for RecN*head* (Adams et al, 2010) and Auto-rickshaw for RecN*cc* (Panjikar et al, 2005) were used for location of heavy atoms, phase determination, density modification and initial model building (ARP-wARP was used in the case of RecN*cc*). Multiple cycles of manual model building using COOT (Emsley & Cowtan, 2004) interspersed with refinement were then performed to complete the models. Statistics of the crystallographic and refinement data are reported later in chapters 4 and 5 and also in paper I and II (Pellegrino et al, 2012a; Pellegrino et al, 2012b).

The RecN*head* crystal structure was refined using *phenix.refine* (Afonine et al, 2010), applying non-crystallographic symmetry (NCS) during the early steps of refinement in order to improve the electron density map and tightening the weighted X-ray/stereochemistry scale. Occupancies and B-factors were also refined for each atom.

The RecN*cc* crystal structure was instead refined using *Refmac5* (Murshudov et al, 1997) and *phenix.refine* as last step, using the same reflections as test sets. Translation/Libration/Screw (TLS) groups, which were treated as independent rigid bodies during the refinement steps, were determined using TLS Motion Determination (TLSMD) server (Painter & Merritt, 2006). Optimization of the stereochemistry and final round of simulated annealing resulted in reasonable statistics with good geometry (data shown in chapter 5).

The structure of RecN*Δdd* was instead solved by Molecular Replacement with the program *PHASER* (McCoy et al, 2007), which produced a single solution with good statistics RFZ (rotation function Z-score): 14.3, TFZ (translation function Z-score): 12.7 and LLG (log-likelihood-gain): 438. A truncated form (containing residues 196-238 and 306-364) of RecN*cc* and the entire RecN*head* domain structures were used as search models. Model building was carried out in *COOT* (Emsley & Cowtan, 2004) and the atomic positions and TLS parameters initially refined using *Refmac5* (Murshudov et al, 1997). Jelly-body refinement, Babinet scaling, map sharpening and the use of a weighting term which gives more weight to the stereochemistry, were used to refine the data at 4 Å resolution (Murshudov et al, 2011). For the final cycles of refinement *phenix.refine* (Afonine et al, 2010) was used instead. Tight stereochemistry weight scale, simulated annealing, secondary structure restraints and a higher resolution reference structure (RecN*head* refined at 3 Å resolution) were used as sources of restraints for improving the low-resolution refinement. B-factors were not refined due to the low resolution of the data and were instead fixed to the Wilson B-factor value (70 $Å^2$).

The final R_{work}/R_{free} values for RecN*head*, RecN*cc* and RecN*Δdd* were respectively of 21/24.9%, 20.3/25.1% and 32.5/34.7%. Final structure evaluation was performed by submitting the models to *MolProbity* server (Chen et al, 2010), which resulted in acceptable Ramachandran statistics in all three cases: the percentage of allowed/outliers is 96.78/0.3 % for RecN*head*, 98.78/0 % for RecN*cc* and 92.4/1.4 % for RecN*Δdd*. Crystal structures were finally deposited at the PDBe database with the following codes: RecN*head* (4aby), RecN*cc* (4abx), RecNΔ*dd* (4ad8).

Illustrations of protein structures were created using Pymol (Molecular Graphics System, Schrödinger, LLC). Electrostatic surface potential was calculated with PDB2PQR tool (Dolinsky et al, 2007) and then visualized using the plugin APBS developed by Michael Lerner and Heather A. Carlson.

2.15. ATPase assay

ATP hydrolysis of RecN, RecN*head* and their respective mutants was measured using the Malachite green assay (Baykov et al, 1988). Malachite green is a reagent that, together with molybdate, reacts with the ortho-phosphate released in solution upon ATP hydolysis. The assay was carried out at 37°C in buffer containing 50 mM Tris pH 7.5, 1 M NaCl and 5 mM $MgCl_2$ using 2 μM of protein. A sample with no protein was used as a blank. Reactions were started by addition of 0.5 mM ATP. 100 μl aliquots were collected at different time points: 0, 3, 5, 10, 15, 30, 45 and 60 minutes. The hydrolysis reaction was stopped by placing the tubes immediately on ice for 5 minutes. Samples were subsequently loaded on a 96-well plate and thoroughly mixed with 25 μl of malachite green reagent (Sigma). Complex formation between phosphomolybdate and malachite green was measured at 630 nm using an EPOCH microplate spectrophotometer (BioTek Instruments Inc.). The absorbance readings were converted to phosphate concentrations using a standard curve established with a series of sodium phosphate monobasic (NaH_2PO_4) solutions. The rates of phosphate release (in μmol min^{-1}) were determined from the slopes of the resulting curves. The plotted values correspond to the average of three independent measurements. Standard deviations were also calculated. Furthermore wild-type RecN and RecN*head* were tested in low salt condition. Concentrated protein (10 mg ml^{-1}) was diluted in buffer containing 50 mM Tris pH 7.5, 100mM NaCl and 5mM $MgCl_2$ and measured as previously described. 50-mer dsDNA (Eurofins MWG) at a final concentration of 0.2 μM was supplemented in order to see any difference in ATPase rate induced by the oligonucleotide.

Earlier experiments for testing ATPase activity of RecN were performed using a spectrophotometric assay, which couples ATP hydrolysis to the synthesis of pyruvate from phosphoenolpyruvate (PEP)(Kiianitsa et al, 2003).

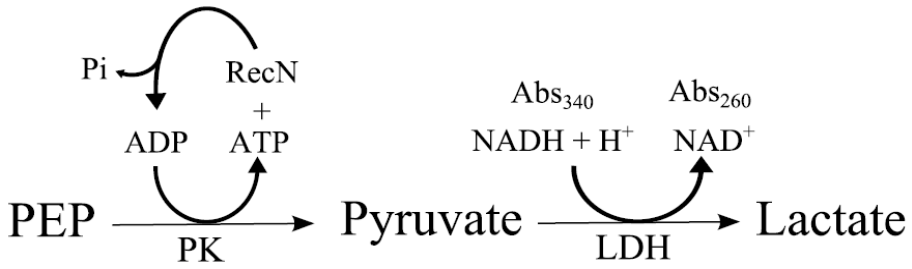


Figure 29: Schematic diagram of the reaction that regenerates ATP through the oxidation of NADH, measured as a decrease of the absorbance at 340 nm.

The assay was not used for investigation of RecN hydrolytic activity since the regeneration system (Figure 29), composed of pyruvate kinase (PK) and lactate dehydrogenase (LDH), was affected by the buffer used in the experiment. It is crucial that measurements of the blank (with no ATPase) are basically flat, with no detectable decrease in the absorbance at 340 nm. In the conditions tested, however, the blank measurement was not reliable, since a considerable slope was observed and all the other measurements were, therefore, invalidated.

2.16. Plasmid multimerization assay

DNA substrates were prepared by digesting pUC19 (Thermo Fischer Scientific Inc.) with three different restriction enzymes (Fermentas): (i) SspI produces DNA blunt ends, (ii) EcoRI creates 5'overhangs and (iii) SacI gives rise to 3'overhangs. The plasmid was digested at 37°C for 1.5 hours and then enzymes were inactivated by heating to 65°C for 20 minutes. The plasmid multimerization assay was carried out in buffer E containing 25 mM Tris-HCl pH 7.5, 1mM DTT, 17.5 mM $Mg(OAc)_2$, 40 mM KOAc, 5% (v/v) glycerol and 2.5% PEG 8K (Reyes et al, 2010). An ATP regenerating system (2.5 mM PEP and 10 units of PK / LDH) was supplemented to the sample reactions. RecN and RecN*head* samples were diluted in buffer E to reach final concentrations of 0.05, 0.1, 0.5, 1, 2 and 4 μM. Samples were mixed with the ATP regenerating system and 3 μM bp (200 ng) of DNA substrate. ATP at a concentration of 3 mM was then added and reaction mixtures were incubated at 37°C for 30 minutes. Subsequently 5X DNA ligase buffer and 0.5 μl of T4 DNA Ligase (5u/μl) was added to each reaction and the samples were incubated for an additional 30 minutes at 25°C. To stop the reaction, 80 μl of solution S (20 mM Tris-HCl pH 7.5, 20 mM EDTA and 0.5% SDS) was added. Complete enzyme inactivation was achieved by addition of 25 μg of proteinase K and incubation at 25°C for 15 minutes. Samples were subsequently purified using the QIAquick

Gel Extraction Kit (QIAGEN) and finally resuspended in 35 μl of 10 mM Tris-HCl pH 8.5. 10 μl of each sample was loaded onto a 0.5% agarose gel containing Ethidium Bromide (EtBr). Gels were run at 55 Volts for 70 minutes and DNA bands were visualized using an UV imager.

2.17. DNA-bandshift assay

Electrophoretic Mobility Shift Assays (EMSAs) were performed in order to test the DNA-binding ability of RecN. Purified proteins were diluted in buffer D (50 mM Tris-HCl pH 8, 100 mM NaCl and 5 mM $MgCl_2$) to a final concentration of 5, 20 and 50 μM. Two different DNA oligonucleotides were used, both at a final concentration of 0.2 μM: a 50-mer dsDNA and another oligonucleotide of the same length and sequence in which the thymine at position 26 is labeled with a fluorescein moiety (Eurofins MWG). Samples were incubated at 18°C for 15 minutes and then loaded onto a 5% native polyacrylamide gel prepared by mixing 30% acrylamide with the desired buffer, either Tris-Borate-EDTA (TBE), Tris-Borate (TB) or Tris-HCl at different pH values. Electrophoresis was finally performed at 4 °C and the gel was prepared and run in TB (pH 8.3) buffer for 70 minutes at 65 volts. A Typhoon 8600 Variable Mode Imager (GE Healthcare) was used to detect and quantify the fluorescently-labeled and EtBr stained DNA (unlabeled oligonucleotides). The fraction of free DNA was determined by quantifying the amount of bound and free DNA at each protein concentration. The experiments were carried out in triplicate. In the case of full-length RecN, the gels were not suitable for DNA quantification due to the significant amount of protein that remained in the wells. Attempts to improve the fraction of protein that entered the gel, using different EMSA protocols and gel types, failed most likely as a result of the very elongated shape of RecN and its tendency to form large oligomeric structures in the presence of DNA.

Chapter 3:

Biophysical characterization of full-length RecN in solution

Summary of chapter 3

RecN was cloned in a suitable vector, overexpressed and purified to homogeneity for further analysis. Characterization of RecN protein was performed taking advantage of the numerous biophysical methods available. Multi-Angle Laser Light Scattering and Small Angle X-ray Scattering were particularly useful for assessing the oligomeric state of the protein in solution and they both led to the conclusion that RecN is a dimer.

Crystallization trials, however, were not successful and therefore new constructs were designed. Bioinformatic tools and limited proteolysis allowed to identify the two principal domains, which make up RecN: a Nucleotide Binding Domain and a coiled-coil region. The study of these two constructs will be thoroughly discussed in the following chapters.

Résumé du chapitre 3

RecN a été cloné dans un vecteur approprié, surexprimé et purifié à l'homogénéité pour des analyses ultérieures. La caractérisation de RecN a été réalisée en profitant des nombreuses méthodes biophysiques disponibles. Les techniques de diffusion de lumière laser multi-angles et de diffusion de rayons X aux petits angles ont été particulièrement utiles pour évaluer l'état oligomérique de la protéine en solution, et menèrent toutes deux à la conclusion que RecN est un dimère.

Nos essais de cristallisation, cependant, n’ont pas produit de cristaux, et donc de nouvelles constructions furent conçues. Les outils bioinformatiques et la protéolyse limitée ont permis d'identifier les deux principaux domaines qui constituent RecN: un domaine de fixation du nucléotide et une région formant un faisceau d’hélices (coiled-coil). La caractérisation de ces deux constructions sera décrite en détails dans les chapitres suivants.

3.1. Cloning, expression and purification of RecN

DR RecN is a protein of 564 amino acids, with a theoretical molecular weight (MW) of 59.8 kDa and a theoretical pI of 4.83, as calculated by ProtParam (Wilkins et al, 1999). The sequence encoding DR RecN was amplified and cloned into a suitable vector for expression in *E. coli* BL21 Star cells.

Cells were resuspended in (25 ml / L culture) lysis buffer L-1 (described in Materials & Methods). The soluble fraction was loaded onto a Ni^{2+}-affinity column (Ni-chelating sepharose, GE Healthcare) in order to specifically separate proteins containing His-tags from the ensemble of other proteins contained in the cell. Buffer A-1 was used to equilibrate the Ni^{2+}-column while buffer B-1 was used for elution of the protein, given the competitive properties of the imidazole ring with the His-tag. One first wash step at 4% B, corresponding to 20mM imidazole, was necessary to remove most of the contaminants. However, a large amount of RecN (15%) (Figure 30), was also eluting at low imidazole concentration indicating that the tag was weakly bound to the column. Once the first peak eluted and the absorbance was again stabilized, a final step at 100% B-1 (containing 500 mM imidazole) was introduced to elute RecN protein, mostly in pure form (Figure 30).

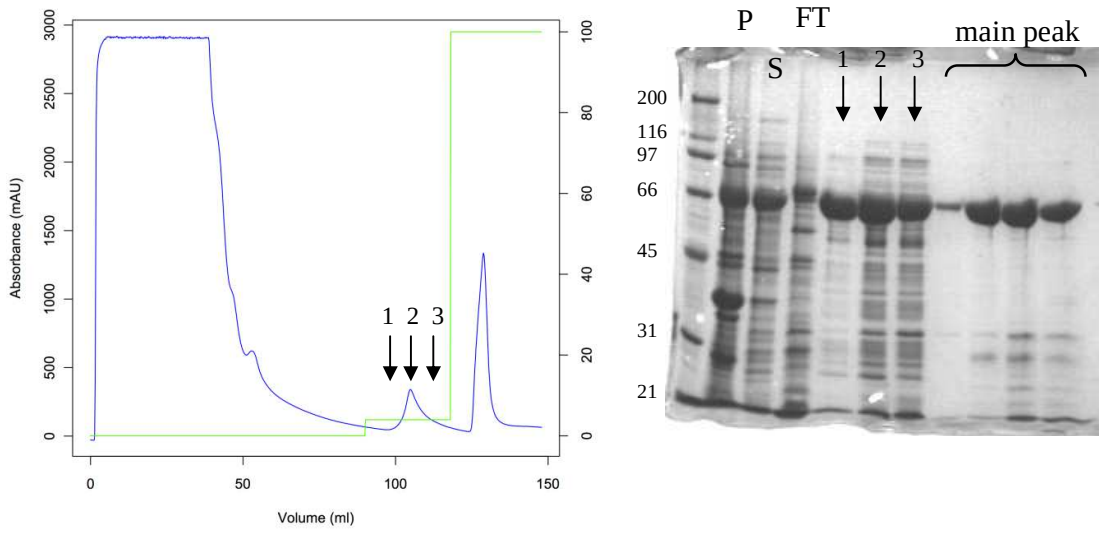


Figure 30: Affinity chromatography profile (at 280 nm) for full-length RecN protein. The first peak, eluting at 4% B-1, contained most of the contaminants but also a fraction of the target protein (as seen in the gel on the right). The peak eluting at 100% B-1, however, yielded ~15 mg of almost pure RecN for each liter of culture. The SDS-PAGE gel with the fractions corresponding to the main and secondary peaks is shown on the right. MWs are reported in kDa. P: pellet; S: soluble fraction; FT: flow-through.

RecN containing fractions were analyzed by running SDS–PAGE on the fractions corresponding to the main and also the minor peaks resulting from the chromatogram. For RecN, 10% polyacrylamide gels were prepared and then run at 200V for 45 minutes.

The pooled RecN fractions were subsequently dialyzed against buffer D-1 (described in Materials & Methods). His-tag cleavage was performed overnight at 4 ºC, at the same time as the dialysis, using the TEV protease. 0.5mM EDTA was added in order to remove traces of Ni^{2+} ions. SDS-PAGE confirmed that the tag was fully cleaved and then RecN was concentrated to a concentration of ~10 mg ml^{-1} and finally loaded onto a Superdex 200 Size Exclusion Chromatography (SEC) column (GE Healthcare), previously equilibrated in buffer D-1. RecN eluted as a single peak at 54 ml on a Superdex 200 16/60 column (GE Healthcare) (Figure 31). A low amount of contamination absorbing at 254 nm, most probably DNA, is present eluting at the void volume. Previous trials to further purify RecN using an ion exchange column (MonoQ, GE Healthcare), led to a dramatic decrease in the amount of purified protein. In the present protocol, however, the contamination is easily excluded from the pooled fraction containing RecN (Figure 31).

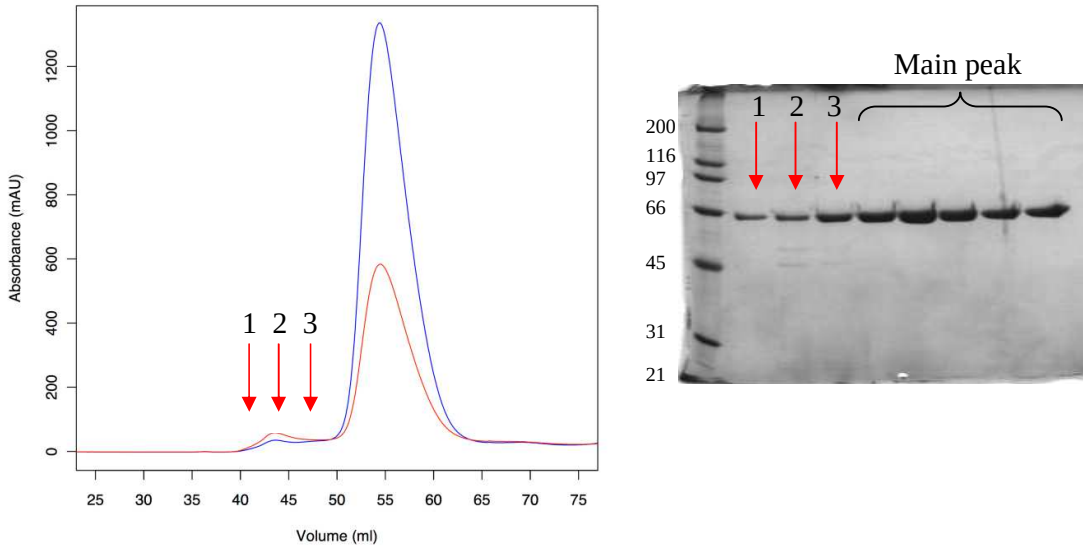


Figure 31: SEC (Superdex 200) profile of RecN in 1M NaCl condition. Protein absorbance at 280 nm is reported in blue, while nucleotide absorbance at 254 nm is in red. The protein eluted as a single, almost symmetrical peak, indicating it is largely a single species present in solution. The acrylamide gel also confirmed the purity of the protein batch. The red arrows (1, 2 and 3) are indicative of the presence of some contaminants, consisting mostly of DNA (10 µl of each fraction were loaded, while for the main peak 1-2 µl were sufficient to check the purity of the RecN elution fractions). MWs are reported in kDa.

RecN was shown to be crucial for IR sensitivity of DR, since a truncation at its C-terminal caused the decrease of the cells tolerance to irradiation (Funayama et al, 1999). We wanted to investigate at the level of the protein why the C-terminal truncation would be so relevant for the protein function. A construct lacking the last 47 amino acid residues was therefore prepared. However, we encountered some problems in purifying such a C-terminally truncated RecN. Expression tests and initial affinity chromatography steps yielded a mostly insoluble protein (Figure 32). A plausible explanation for this behavior could be that this deletion disrupts the correct folding of the globular domain. Our structural studies will reveal that this is indeed the case, as discussed later (chapter 4).

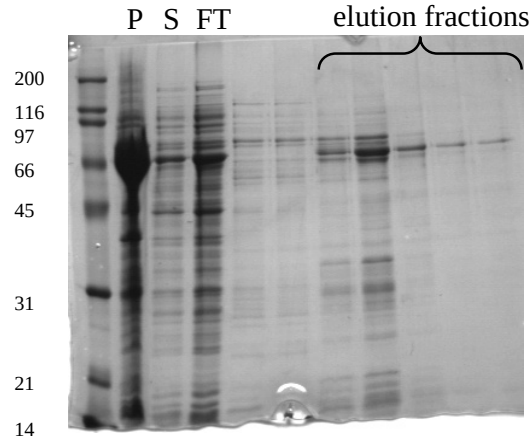


Figure 32: Affinity chromatography of RecNΔ47, which gave rise to a mostly insoluble protein (considerable amount present in the pellet). MWs are reported. P: pellet; S: soluble fraction; FT: flow-through.

3.2. Biophysical characterization of RecN

Purified RecN was used for SEC-MALLS analysis, which characterized the sample in the presence of high (1M) and low (150 mM) salt content (Figure 33). A clear shift in the elution volume from the GF column (Figure 33) highlighted the different behavior of RecN in these two conditions. RecN elutes as a heterogeneous mix of species in low salt buffer while in presence of 1 M NaCl, RecN elutes in a dimeric form as a single species with a molecular mass of 108 kDa (Figure 33).

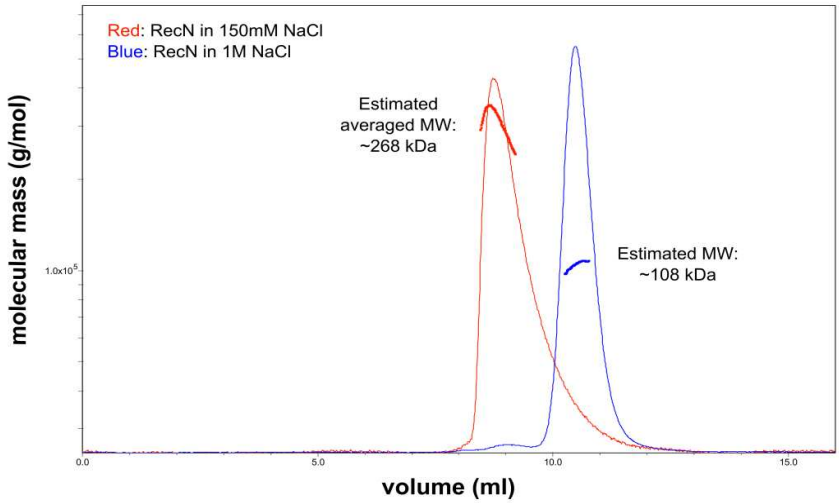


Figure 33: SEC-MALLS analysis performed on full-length RecN. Two different buffer conditions revealed that the oligomeric state of RecN is sensitive to salt concentration. A stable dimer is formed in 1 M NaCl. Averaged MW are indicated close to the respective curves and reported, in the form of a table, in paper III (Pellegrino et al, 2012 submitted).

The purified RecN in 1M NaCl was used for further biophysical analysis, since it appeared to be the most homogeneous sample. Production of a highly pure protein is fundamental in order to perform Small Angle X-ray Scattering (SAXS) studies and for obtaining protein crystals (Putnam et al, 2007). Homogeneity of our protein solution was assessed by DLS (§ Materials & Methods) prior to SAXS measurement, resulting in a polydispersity index (pdI) of 0.16, which is indicative of a highly homogeneous sample.

SAXS studies were performed on RecN (Figure 34). The data collected on full-length RecN were highly reproducible at the three different concentrations measured: 2.5, 1.25 and 0.5 mg ml^{-1}. The scattering profiles of the three measurements superimposed well, indicating that there was no inter-particle effect due to the concentration of the macromolecules in solution. The curve corresponding to the lowest concentration (red) is noisy at high *q* values, which correspond to the high resolution portion of the curve, but superpose very well at low *q*, where the Guinier region is being used for R_G estimation (Putnam et al, 2007), which was found to be ~9.5 nm. The I_0 value was in agreement with a dimeric assembly of full-length RecN and the profile of the calculated Pair distribution function (Figure 34), characterized by a long tail at higher D_{MAX} values (Putnam et al, 2007), clearly indicates that full-length RecN is elongated in solution. The shape of the P(*r*) shown in Figure 34 and also in paper III (Pellegrino et al, 2012 submitted) suggests that RecN is likely to be fairly elongated, as would be expected for an SMC-like protein (Hirano, 2002).

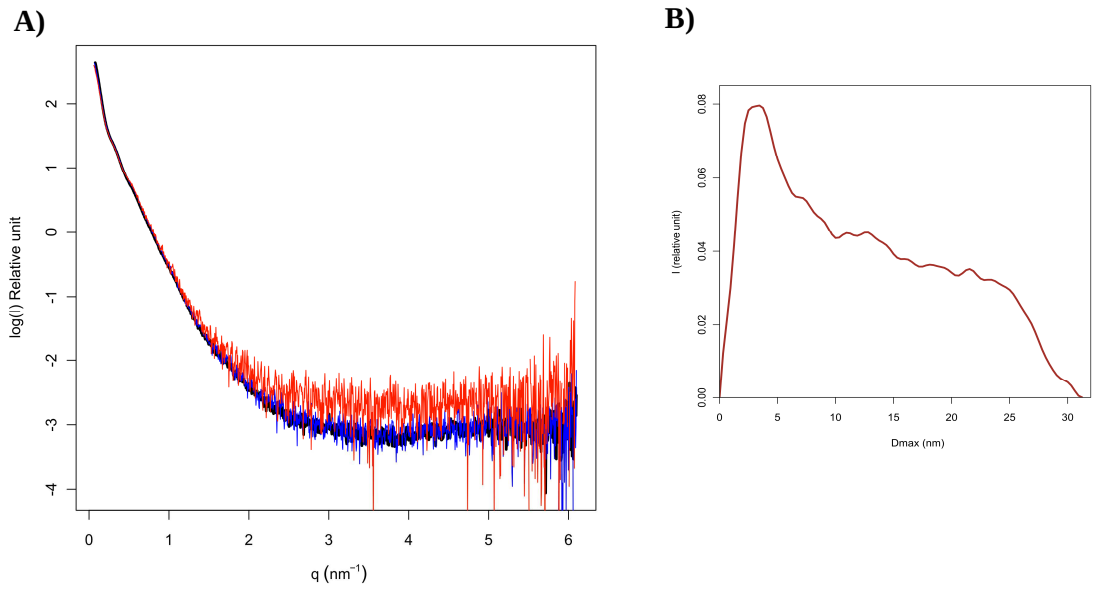


Figure 34: RecN has been proven to be a suitable sample for SAXS. A) Scattering curves were collected at three different concentrations and used for further evaluation analysis. The scattering curve corresponding to the lowest concentration adopts the same characteristic shape as the other two, but is more noisy and slightly shifted at high scattering angles. B) Pair distribution function derived from the experimental scattering curve of RecN using GNOM. The characteristic shape of this curve, with a narrow peak with an extended tail at high D_{MAX} values (reported in nm), indicates that the macromolecule in solution could be rather elongated. The R_G value was at the limit of the detector set up, so some errors need to be taken into account.

The first part of the curve at low angles was used for further analysis, since our interest focused primarily on the overall shape determination for this protein. An *ab initio* reconstruction gave rise to the envelope proposed in Figure 35, with the presence of two bulky regions at either ends. The envelope is 315 Å in length. No symmetry or prior information was used for the reconstruction calculations. This work represents the first structural study of a RecN protein.

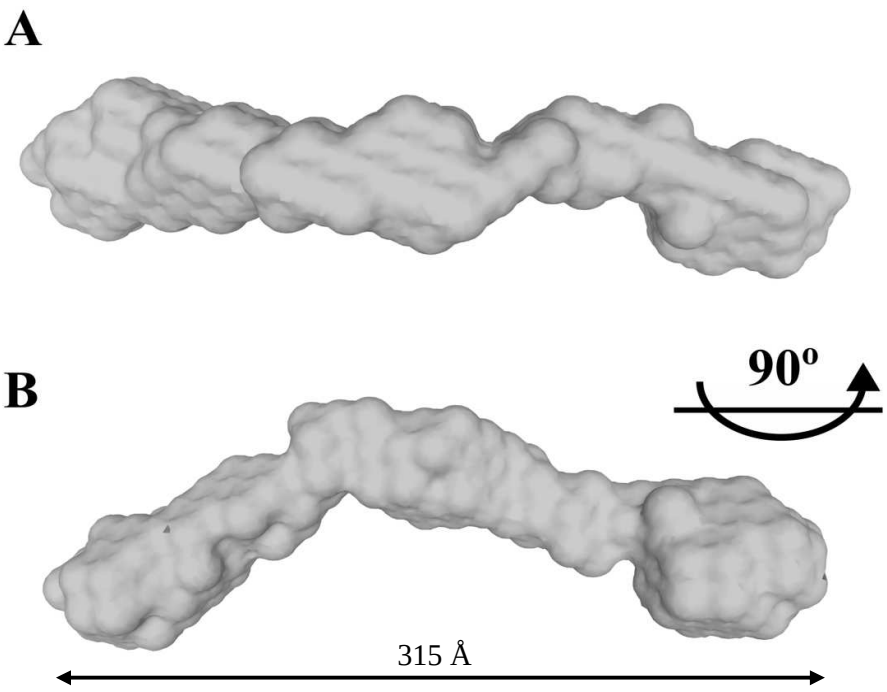


Figure 35: *Ab initio* envelope of RecN obtained by SAXS. DAMMIN was used for generating the dummy sphere model using the low *q* region of the SAXS curves, as previously described. The envelope is very elongated (315 Å long).

Full-length RecN was found to be a very challenging protein for crystallization and no crystals have so far been obtained. Protein in 1M NaCl, which appeared to be the most homogeneous and stable sample, was used for setting-up drops manually and with the crystallization robot (HTX EMBL-Grenoble). Drops were set up as described in Materials & Methods and then controlled once a week for several months but the drops did not give rise to any interesting hits: mostly precipitation was formed after addition of the precipitant solution, although some drops with promising phase separation (Figure 36) did show up after approximately 1 week in a few conditions. These, however, did not produce any crystalline material after manual optimization in hanging drop plates (Hampton Research). Drops were set up using different protein:reservoir ratios, in order to play with the final concentration of the protein.

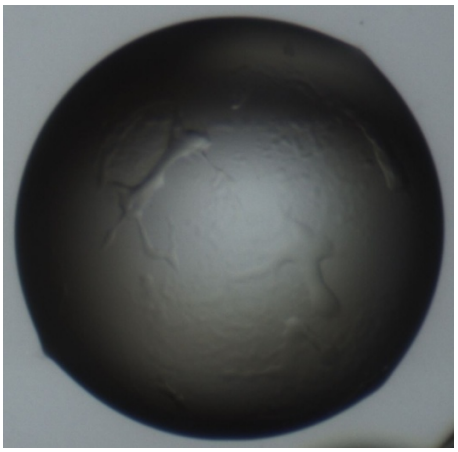


Figure 36: Example of the drop of RecN in condition 10 of MembFac screen (Hampton Research). Manual optimization did not lead to improvement. Different ratios of protein:reservoir were used but were not successful in giving rise to crystals.

3.3. Towards a high resolution structure of RecN

DR RecN was found to be fairly soluble and easy to work with compared to its *E. coli* homologue (Grove et al, 2009). Initial purification was optimized in order to stabilize the protein in solution. High salt concentration in the buffer was crucial for protein homogeneity, as seen by its low polydispersity index (DLS) or by its estimated MW by SEC-MALLS (the measure was constant through the entire elution peak) (Figure 33). SAXS analysis of RecN led to the reconstruction of an *ab initio* model (Figure 35), which provided some new insights into the overall shape of the full-length protein. RecN was characterized using various biophysical methods, each of which resulted in new findings.

However, the main goal of this first part of the project was not achieved, since crystallization studies of RecN did not succeed in producing any crystalline material. We, therefore, decided to focus on the study of single domains, since structural information of homologous proteins, such as SMC, Rad50 and MukB, was available. New constructs of RecN were therefore designed corresponding to the predicted domains. In literature, in fact, the NBD of homologous proteins have been extensively studied, together with the dimerization interface of the coiled-coil (Hopfner et al, 2002; Hopfner et al, 2000; Lammens et al, 2004; Li et al, 2010; Lowe et al, 2001).

Limited proteolysis can also aid in defining the portion or domains of the protein that are more compact and, therefore, stable than other regions that are, in contrast, more flexible and exposed to the solvent. Such an experiment was performed on full-length RecN and the results, together with secondary structure prediction (PsiPred) and the available literature, were used to define putative domain boarders (Figure 37). N-terminal sequencing (Edman, 1949) resulted in the identification of residues that likely reside in flexible or exposed regions, and therefore prone to attack by proteases (to Proteinase K in our specific case).

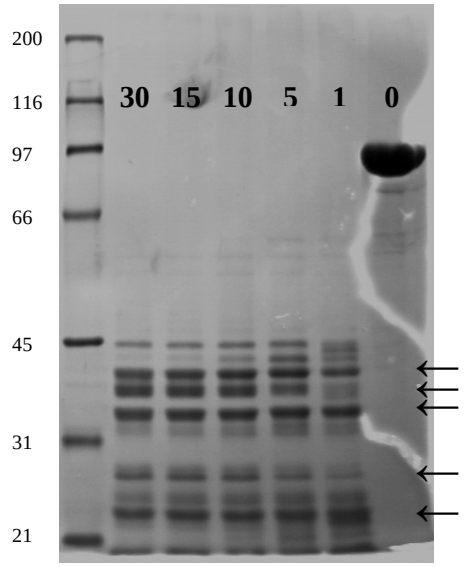


Figure 37: Limited proteolysis performed using RecN purified protein (marked as 0) and incubating it with 1/5000 ratio of Proteinase K (Sigma Aldrich). Aliquots at different time points (1-30 minutes) were collected and finally loaded onto a 12% denaturing acrylamide gel. MWs are reported in kDa.

N-terminal sequencing allowed the determination of the first five amino acid residues placed at the N-terminus of each characteristic fragment generated by proteolysis (indicated by black arrows). Once the sequence of these were known, amplification and subsequent cloning of the gene codifying for the single domains of RecN was performed as described in Materials & Methods.

Several constructs were designed (Table 5) and in the end, two of them resulted to be very successful: the globular NBD, referred to as RecN*head*, in which the N- and C-terminal domains were linked by a 14-amino acid linker and the coiled-coil region, referred to later as RecN*cc*. In the table below, a list of the constructs amplified, cloned in pET-151 TOPO and expressed in *E. coli* BL21 Star is provided (Table 5). In some cases, the cloning was unsuccessful, particularly when trying to engineer a short (8 amino acid) linker between the N- and C-terminal domains.

Constructs	**Cloning**	**Expression test**
N-term 1-196	Yes	Insoluble
N-term 20-196	Yes	Insoluble
CC 196-365	Yes	**Soluble**
CC 196-387	Yes	No expression
CC 240-365	No	
CC 240-387	Yes	**Soluble**
C-term 366-564	Yes	Insoluble
C-term 388-564	Yes	Insoluble
N+C-term 1-564 linker 8 residues	No	
N+C-term 20-564 linker 8 residues	No	
N+C term 1-564 linker 14 residues	Yes	**Soluble**
N+C-term 20-564 linker 14 residues	No	
RecN34 (N-terminal truncation)	Yes	**Soluble**

Table 5: List of the constructs designed to be cloned in pET-151 TOPO vector for expression and solubility tests in *E. coli* cells, strain BL21 Star.

Attempts with the single N- or C-terminal domains of RecN produced only insoluble protein. Cells expressing the two single domains alone were also mixed together prior to cell lysis, but this was also unsuccessful. The purpose of this experiment was to reconstitute the putative NBD within *E. coli*. Large-scale expression was therefore performed for the constructs that best behaved in the expression tests.

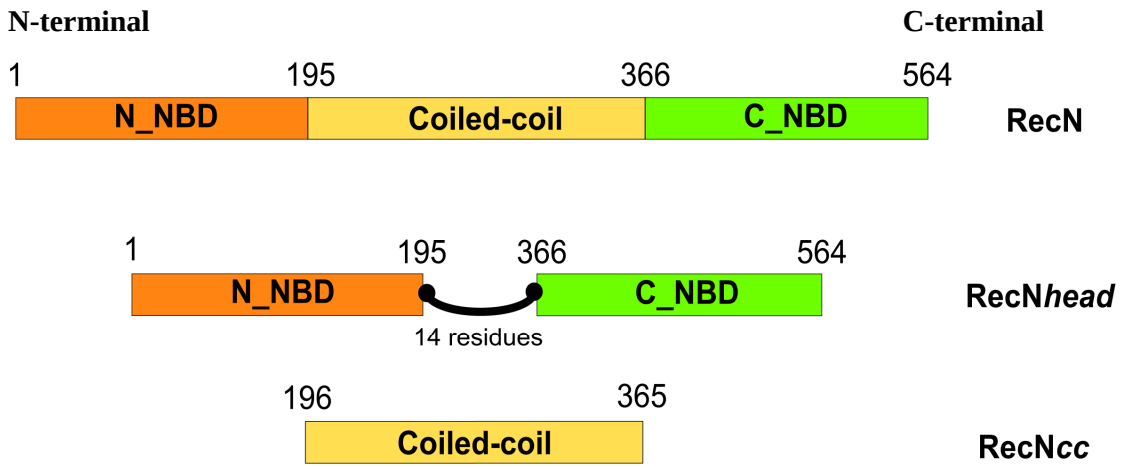


Figure 38: Schematic diagram of the constructs used for structural, biophysical and biochemical analysis of RecN. The different domains, as predicted by secondary structure prediction analysis, are colored differently: the N-terminal NBD fragment in orange, the C-terminal NBD fragment in green and the coiled-coil in yellow. The RecN*head* domain is represented with a peptide linker that fuses the N and C-termini (Pellegrino et al, 2012c).

We were particularly interested in two of them, notably the construct that codified for the N- and C-terminal domains fused by a 14-residue linker peptide (RecN*head*) and the coiled coil domain (Figure 38). The latter corresponding to residues 196-365 (RecN*cc*) was of particular interest because it represented the region that we had removed from the RecN*head* construct. Together, these two constructs would reconstitute the full-length protein. RecN*head* and RecN*cc* were amplified, cloned and expressed as described in Materials & Methods section and in (Pellegrino et al, 2012a; Pellegrino et al, 2012b).

Chapter 4:

Structural and biochemical characterization of RecNhead domain

Summary of chapter 4

The head domain of RecN is relatively conserved and similar to that of SMC or SMC-like proteins, such as Rad50 for instance. We successfully cloned, expressed and purified a construct containing the N- and C-termini of RecN fused by a short peptide linker, which we referred to as RecN*head*. We successfully crystallized this protein and collected diffraction data to a maximum resolution of 3 Å. Se-Met derivatives were needed in order to solve the phase problem by the Single-wavelength Anomalous Dispersion method. The crystal structure is presented here, together with Multi-Angle Laser Light Scattering and Small Angle X-ray Scattering studies of the protein behavior in solution. The crystal structure fitted very well with the *ab initio* model reconstructed by Small Angle X-ray Scattering and we could conclude that RecN*head* shares common structural features with homologous proteins belonging to the SMC family.

The crystal structure of RecN*head* domain represented the first high-resolution structural information known for a RecN protein.

Résumé du chapitre 4

Le domaine globulaire de RecN, nommé ‘head’, est relativement conservé et semblable à ceux des protéines de la famille SMC, telle que Rad50 par exemple. Nous avons réussi à cloner, exprimer et purifier une construction que nous appellons RecN*head* comportant les domaines N- et C-terminaux de RecN fusionnés par une courte chaine peptidique. Nous avons réussi à cristalliser cette protéine et à recueillir des clichés de diffraction à une résolution maximale de 3 Å. Des dérivés contenant des méthionines séléniées ont été nécessaires afin de résoudre le problème de phase par la méthode de dispersion anomale à une longueur d’onde unique. La structure cristalline de RecN*head* est présentée ici, accompagnée des études de RecN*head* en solution par diffusion aux petits angles (SAXS) et diffusion de lumière laser multi-angles. La structure cristalline est en accord avec le modèle *ab initio* reconstruit par SAXS et nous pouvions en conclure que RecN*head* partage de nombreuses caractéristiques structurelles avec les protéines homologues appartenant à la famille SMC.

La structure cristalline du domaine RecN*head* représentait les premières données structurelles connues à haute résolution pour une proteine RecN.

4.1. Purification of RecN*head* domain protein

The RecN*head* construct was designed to reconstitute the complete ***NBD*** of RecN (see Materials & Methods). The purification protocol of RecN*head* domain was established and resulted in a homogeneous and pure protein (as assessed by DLS, SEC-MALLS and SAXS). The amount of protein obtained for 1 litre of cell culture was approximately 75-100 mg. Buffers A-2 and B-2 used for the Ni-affinity purification step (Figure 39) and buffer C-2 for the SEC step are the following:

A) 50mM Tris-HCl pH 8
300mM NaCl
5mM $MgCl_2$
5mM imidazole pH 8

B) 50mM Tris-HCl pH 8
300mM NaCl
5mM $MgCl_2$
500mM imidazole pH 8

C) 50mM Tris-HCl pH 8
300mM NaCl
5mM $MgCl_2$

Affinity chromatography was performed as described in Materials & Methods. An imidazole gradient was used for elution of the target protein. The buffer composition is reported in Materials & Methods section.

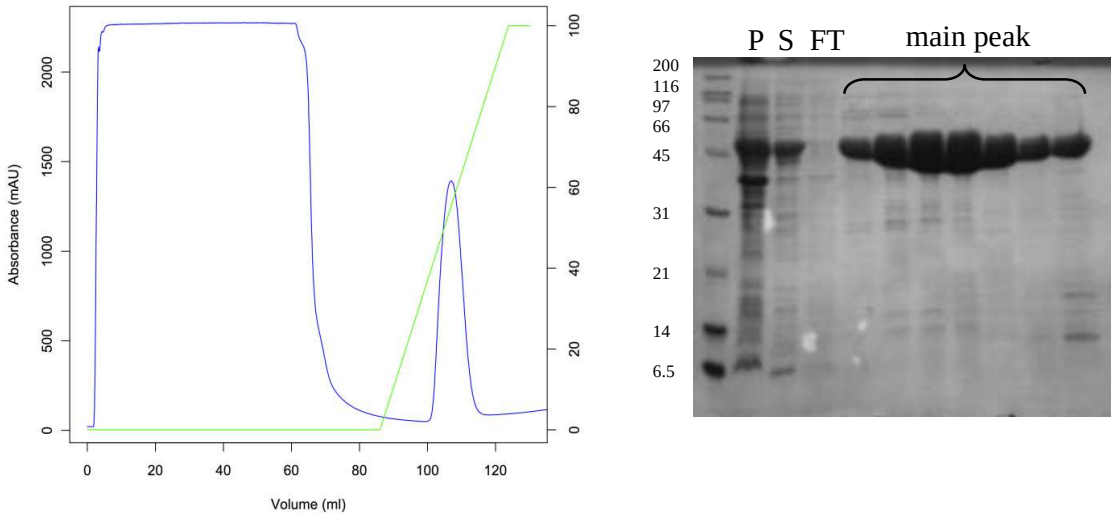


Figure 39: Metal affinity chromatography profile of the RecN*head* domain. Protein absorbance at 280 nm is shown in blue. Protein eluted at a concentration of 170 mM imidazole, corresponding to 33% of the buffer B gradient. The denaturing gel with the fractions corresponding to the main peak is shown. MWs are reported in kDa. P: pellet; S: soluble fraction; FT: flow-through.

The SEC step (Figure 40) yielded a single peak and only the purest fractions, corresponding to the center of the peak, were pooled and used for further experiments. An

aliquot of the pooled fractions was used for DLS (Zetasizer Nano, Malvern) measurements and was found to have a pdI of 0.21. The hydrodynamic radius (r_h) was estimated to be approximately 3.3 nm (Pellegrino et al, 2012b), corresponding to a MW of 55 (assuming the protein is perfectly globular).

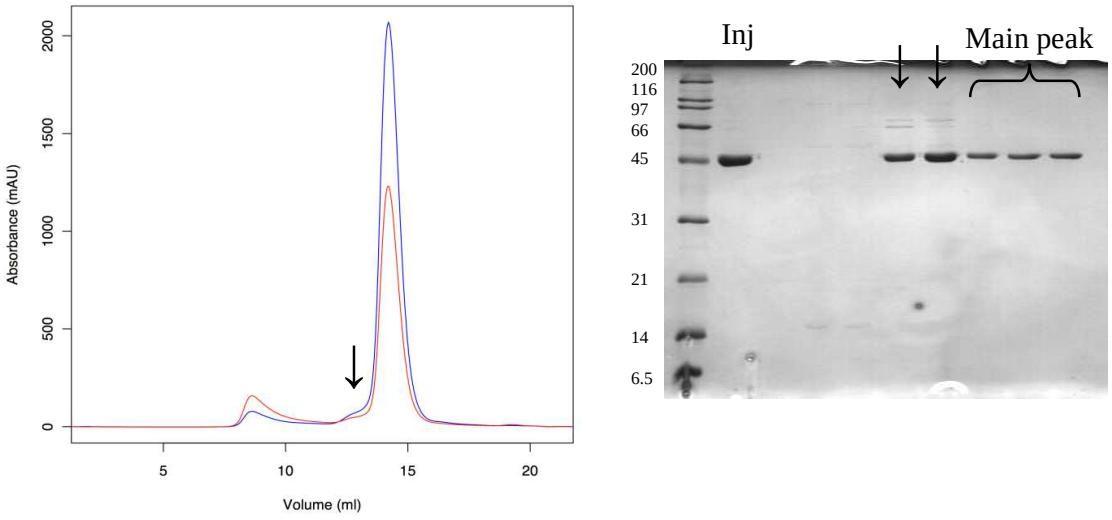


Figure 40: SEC profile of RecN*head* domain. SEC was performed using a Superdex 200 column (GE Healthcare). Protein absorbance at 280 nm is reported in blue, while nucleotide absorbance at 254 nm is in red. The presence of a little shoulder before the elution volume corresponding to the main, symmetric, peak may be due to the presence of some contaminants, as it is possible to see from the SDS-PAGE gel (indicated by arrows). MWs are reported in kDa. Inj: injection.

4.2. RecN*head* domain is a monomer in solution

SEC-MALLS experiments were performed on RecN*head* domain (Figure 41), in order to evaluate the oligomeric state of the protein. The method is described in the Materials & Methods section. The elution volume from SEC indicated that RecN*head* may form a higher MW assembly (approximately 90 kDa when comparing it to a standard curve), but the SEC-MALLS analysis clearly showed that the RecN*head* domain is a monomer in solution (Figure 41).

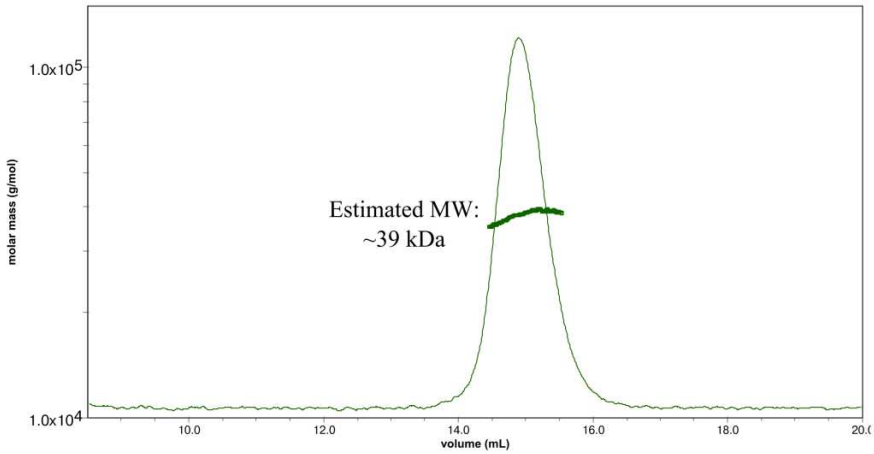


Figure 41: SEC-MALLS profile of RecN*head*. The averaged MW suggested that this chimeric construct is a monomer in solution. The refractive index has been plotted and the resulting single symmetric peak is, together with the stability of the calculated MW, index of homogeneity of the sample. Thin line: refractive index recorded; thick line: averaged MW estimation from different scattering angles.

4.3. SAXS analysis of RecN*head* domain

SAXS measurements on the wild-type RecN*head* domain were performed as described in Materials & Methods section and allowed to build an *ab initio* model (Pellegrino et al, 2012b). Measurements were performed at three different concentrations, 3.43, 2.34 and 1.12 mg ml^{-1}, to exclude any inter-particle effect (Figure 42). Processing of the data was performed as described in Materials & Methods section. GNOM was used for estimation of the D_{MAX} and the pair distribution function (P(*r*)) (Figure 42). The output file was then submitted to the model building program DAMMIN. Prior to determining the structure of RecN*head* domain, Rad50 ATP-free ATPase domain (PDB code:1II8) was used as template for superimposition with the a*b initio* envelope. The overlay suggested that RecN*head* probably adopts a similar overall fold. MW estimation was calculated from the scattering at zero angle (I_0) and resulted to be ~34 kDa. The discrepancy with the theoretical MW of monomeric RecN*head* (43 kDa) was too high, therefore a calculation with AUTOPOROD (Petoukhov et al, 2007) was performed to have a better estimation of the value. The expected MW at the end of the calculation resulted to be ~46.5 kDa, more in agreement with the theoretical value of the monomer.

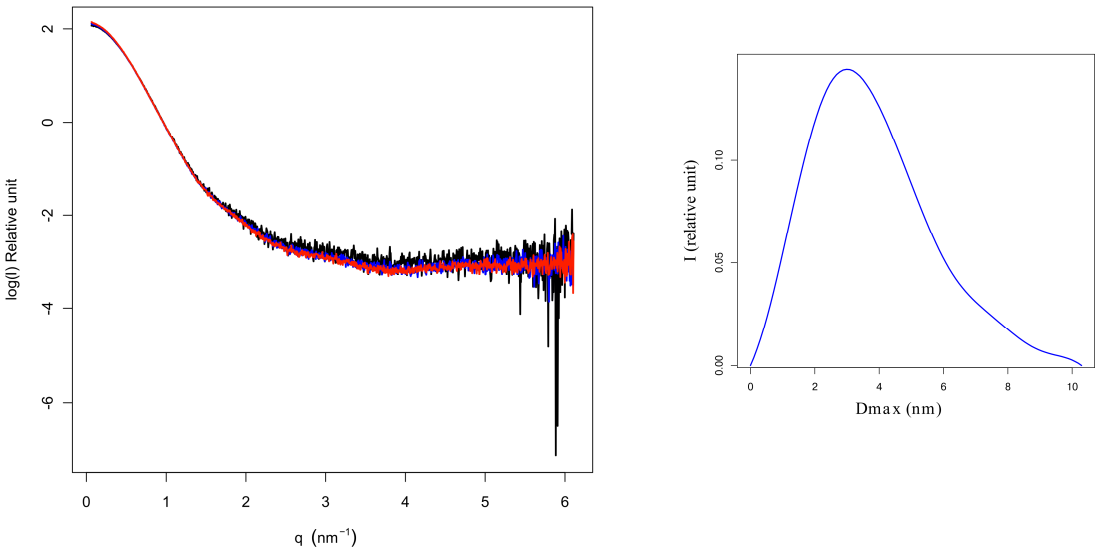


Figure 42: Superposition of the X-ray scattering curves (left) collected for RecN*head* domain at three different protein concentrations: 3.43 (red), 2.34 (blue) and 1.12 mg ml^{-1} (black). The pair distribution function (P(*r*)) is shown on the right, with a D_{MAX} value of 10.2 nm. Data were collected on ID14-3 at the ESRF (Grenoble).

Ab initio modelling was performed using DAMMIN (Svergun, 1999). The overall envelope was produced after averaging and filtering of the multiple models created, as shown in paper I (Pellegrino et al, 2012b).

4.4. Crystallization of the head domain of RecN

RecN*head* was used for crystallization experiments. Initial screening of a broad range of conditions was carried out with the crystallization robot at the HTX laboratory (Dimasi et al, 2007). Crystals were first obtained in condition 44 of the Index screen (Hampton Research) and then manually optimized by using hanging drop plates, which resulted in crystals of suitable size and diffracting at a maximum resolution of 3 Å. The final condition, which gave the best diffraction pattern, was 100mM Tris-HCl pH 7.5 and PEG3350 25%.

Diffraction experiments of RecN*head* native protein were performed on ID23-2 (Flot et al, 2010). A full data set was collected and processed with iMOSFLM (Battye et al, 2011) and the intensities were then scaled. Molecular Replacement (Mol.R.) programs, notably PHASER and MOLREP (McCoy et al, 2007; Vagin & Teplyakov, 1997), were run in order to solve the protein structure. Different search models were used, like *Pyrococcus furiosus* Rad50 (pfRad50) and *Thermotoga maritima* SMC (tmSMC) head domains and *Deinococcus radiodurans* RecF (drRecF). Poly-alanine chains of these structures were also used as models, but all trials failed to give reasonable solutions.

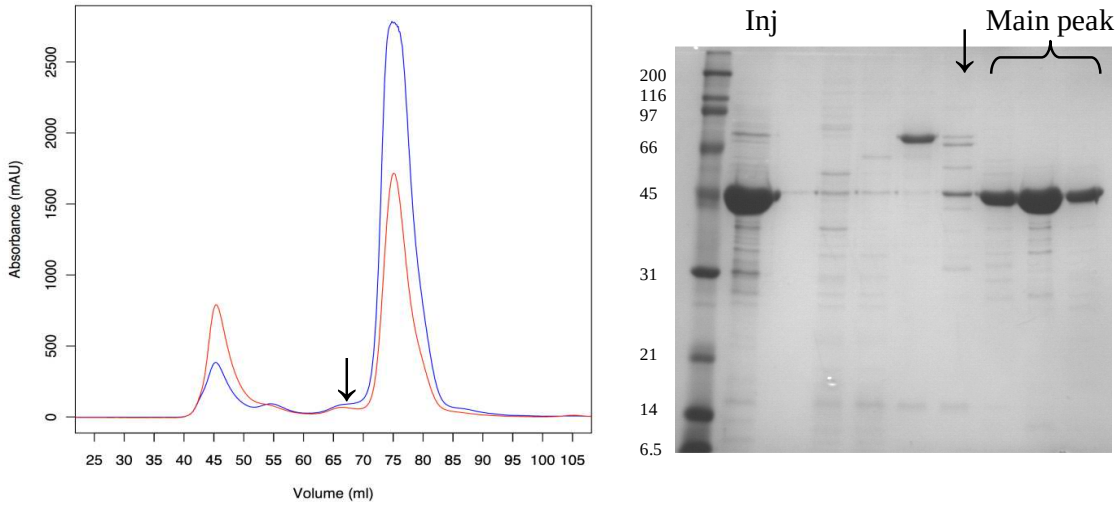


Figure 43: SEC profile of the Se-Met derivative of RecN*head* domain. Protein absorbance at 280 nm is reported in blue, while nucleotide absorbance at 254 nm is in red. The elution profile of the protein is similar to that obtained for the native protein (~74 ml). A little shoulder containing small amount of contaminants is indicated by a black arrow. SDS gel of the fractions eluting from the SEC column is shown on the right, with the mean peak's fractions indicated. MWs are reported in kDa. Inj: injection.

Since Mol.R. failed to find a good solution, Se-Met derivatives were produced (Figure 43) as described in the Materials & Methods section. Once the derivative protein was purified and crystallized, using the same protocols as for the RecN*head* native protein, the Single-wavelength Anomalous Dispersion (SAD) method was applied for calculating experimental phases. An X-ray fluorescence scan was performed to determine the wavelength corresponding to the absorption peak of the anomalous scatterers (Se atoms). A full redundant anomalous data set was collected at the peak wavelength and the processed data were submitted to Phenix *AutoSol* (Terwilliger et al, 2009) for localization of heavy atom positions and subsequent phase determination. Selenium atom positions were successfully determined (FOM Phaser: 0.278) and then used for calculation of a new set of phases which gave rise to a new electron density map (Figure 44). Phenix *AutoBuild* (Terwilliger et al, 2008) was used for model building, while refining the nascent electron density map, resulting in a reasonable starting model. Further cycles of manual building were performed using Coot (Emsley & Cowtan, 2004), employing a 2Fo-Fc and Fo-Fc (difference map) electron density maps. Refinement was performed using *phenix.refine* (Afonine et al, 2010).

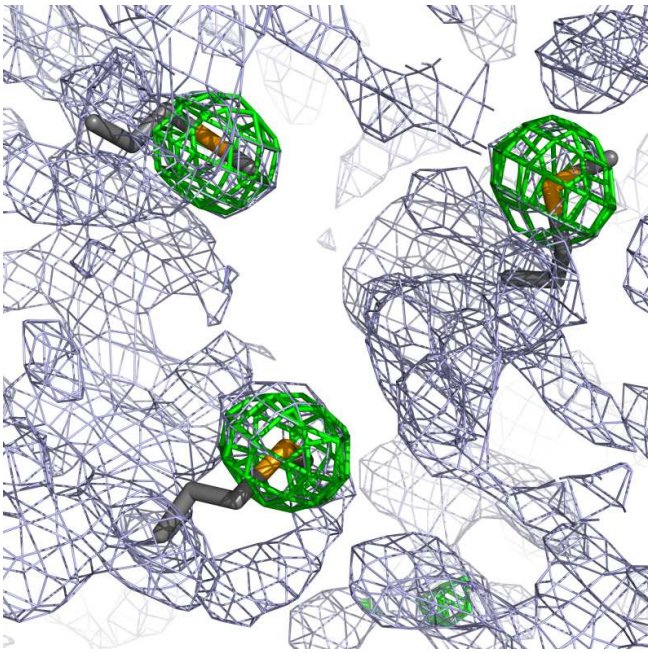


Figure 44: Fourier electron density map contoured at 1.7σ with anomalous difference density (coloured in green) contoured at 3.5σ. PHASER was run to calculate the heavy atom location and, therefore, the anomalous map of the Se atoms. This map has been superposed to the one used for initial model building (after density modification performed by RESOLVE (Terwilliger, 2003)). Se-Met residues built as part of the initial model are shown as sticks.

4.5. Structural analysis of RecN*head* domain

Although the RecN*head* construct was designed to reconstitute the complete ***NBD*** of RecN, structural analysis revealed that this construct also included some additional regions (Figure 45): the coiled-coil ***anchor motif***, the beginning of the predicted coiled-coil region and a short C-terminal helix-turn-helix (HTH) motif, which is unusual for Rad50 and SMC proteins. The coordinates of RecN*head* were submitted to DALI server (Holm et al, 2008) for 3D structural search and the best hits were found to be *Saccharomyces cerevisiae* SMC, tmSMC and pfRad50. The overall structural organization of the NBD of RecN*head* is very similar to the head domains of *S. cerevisiae* SMC (PDB: 1W1W), tmSMC (PDB: 1E69), *P. furiosus* SMC (PDB: 1XEX) and pfRad50 (PDB: 1F2T) proteins with an RMS of 1.38Å over 148 C_α, 1.57Å over 122 C_α, 1.39Å over 155 C_α and an RMS of 1.61Å over 121 C_α respectively. Sequence alignment of the NBDs of RecN with some of the above-mentioned proteins is shown in Figure 23. RecN*head* also exhibited some structural differences when compared to Rad50, notably in the architecture of its ***signature sequence***, responsible for ATP recognition, suggesting that the nucleotide binding pocket may be altered compared to

other SMC-like proteins, such as Rad50 (Hopfner et al, 2000; Hopfner & Tainer, 2003; Mockel et al, 2011; Paull et al, 2004).

The NBD of RecN*head* consists of two lobes (Figure 45-A), in which lobe I, formed primarily by the N-terminal domain of RecN, adopts a typical Rossmann fold (Rao & Rossmann, 1973), consisting of a six-stranded anti-parallel β-sheet (strands β1-β2 and β4-β7) wrapping around the Walker A containing α-helix 1 (Figure 45-A). Lobe I also contains the Walker-B motif involved, as reported in the literature (Haering et al, 2004; Hopfner et al, 2000), in the coordination of the water molecule necessary for hydrolysis of the γ-phosphate of ATP (Figure 18). Lobe II consists mostly of the C-terminal region of RecN and contains the ABC-like signature sequence, crucial for binding ATP. This C-terminal domain folds as a β-α-β sandwich in which helices α10-α12 pack against a mixed β-sheet (β3, β8, β12-β15). The N and C-terminal domains are held together by hydrogen-bonding interactions between β3 and β8 strands from the N-terminal domain and β12-β14 strands located in the C-terminal domain. The Walker-B motif is located on strand β12 while helix α10 contains the signature sequence, important for driving ATP-dependent NBD dimerization (Altenberg, 2003). A structural alignment of the NBDs of RecN and other SMC or SMC-like proteins (Figure 23) revealed that these domains only share a low overall sequence conservation (14% identity between RecN and pfRad50), but the residues responsible for ATP-binding (K67 in the Walker-A motif) and hydrolysis (D471 and E472 in the Walker-B motif) are strictly conserved and are located at similar positions. The signature sequence is also highly conserved, but unlike the Walker-A and B motifs, adopts a very different orientation, as a result of the rotation of the coiled-coil anchor motif relative to the NBD. This overall architecture gives rise to half of the functional nucleotide-binding pocket (Hopfner et al, 2000) (Figure 45-B and C).

The anchor motif is primarily composed of a three-stranded anti-parallel β-sheet (β9-β11), helix α4 and two flexible linkers connecting β8 to α4, through a short helical turn (α3), and β11 to α11. Comparison with pfRad50 head domain (Figure 45-B) reveals that the region connecting α2 to α4 including β8, which is substituted by a short helix in Rad50 (Hopfner et al, 2000), is poorly conserved. As a result the entire coiled-coil anchor motif, along with the signature sequence of RecN*head*, undergo a ~60° rotation relative to their position in pfRad50 (Figure 45-C). The anchor domain in RecN*head* is also longer than in pfRad50, being respectively 38 and 31 Å. Helix α4 plays an essential role in guiding the positioning of the helices belonging to the coiled-coil region (α5 and α10) so as to be roughly orthogonal to the

NBD (Figure 45-A). Hydrogen bonding between Y175 on helix α5 and the backbone of A419 situated in the β-sandwich of the anchor motif contributes to stabilize this conformation.

RecN*head* also has an extra HTH-like motif at its C-terminus (α14 and α15), which is present in all RecN proteins (Figure 45-A), but is entirely missing in Rad50 and partially so in SMC proteins, where only α14 is present in the case of yeast SMC (Haering et al, 2004). The function of this structural motif is unknown. Cells expressing a C-terminally truncated form of RecN (missing residues 518-564, which includes β14-15 in addition to α14-15) were reported to exhibit an increased sensitivity to irradiation and DNA-damaging agents like mitomycin C (Funayama et al, 1999). In view of the structure of RecN*head*, removal of these residues would most likely disrupt the hydrogen-bonding network between β2 and β14 strands and destabilize the β-sheet, yielding an inactive and unfolded RecN protein. Trials to express such a protein only yielded insoluble protein (Figure 32), strengthening this hypothesis.

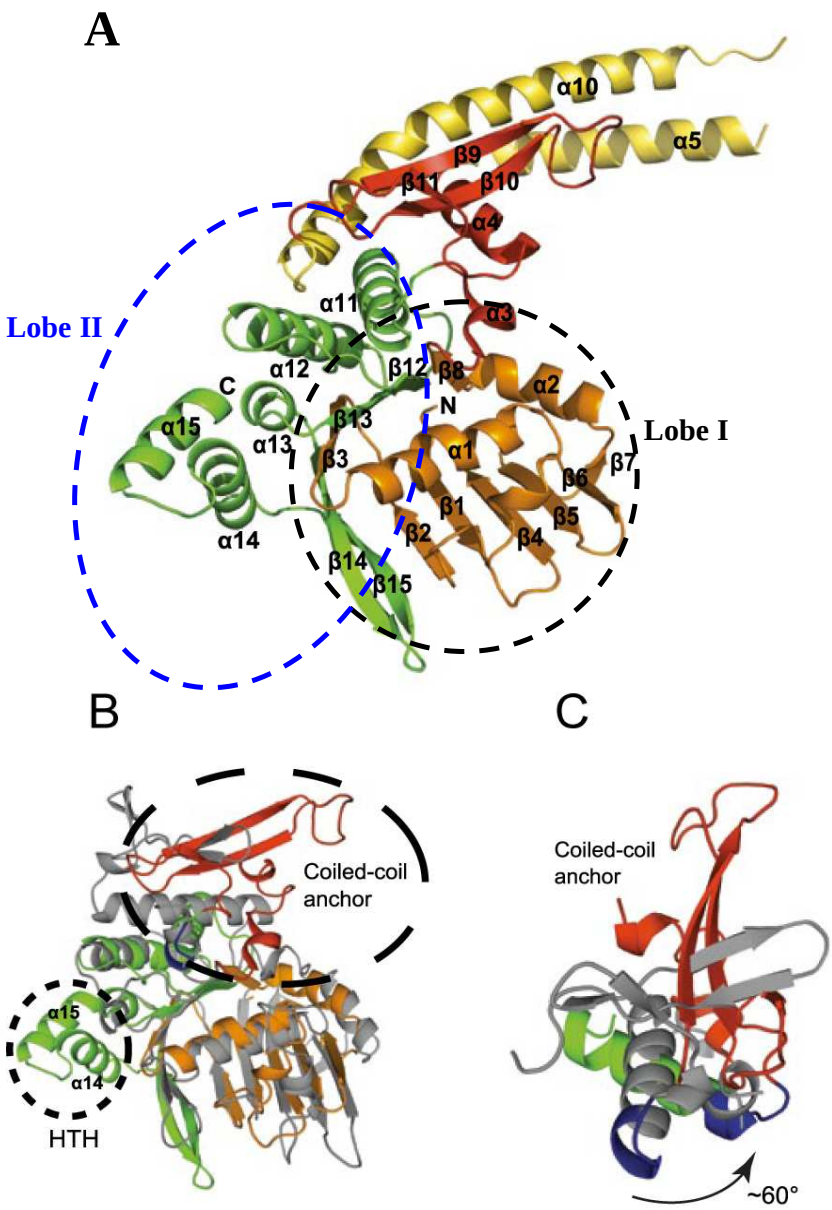


Figure 45: RecN*head* domain crystal structure. A) RecN*head* domain crystal structure adopts a Rossman fold, typical for nucleotide binding proteins, and shares structural similarity with the SMC and SMC-like proteins. B) Superposition of RecN*head* domain (colored as shown in A) with pfRad50 head domain (grey) also referred to as the ATPase domain. An HTH at the C-terminus of RecN*head*, which is not conserved in pfRad50, represented a unique feature. C) Different orientation of the RecN-Rad50 overlay, with the coiled-coil anchor motif (in red) and the signature sequence (in blue) of RecN*head* domain rotated ~60 degrees to respect to their corresponding in Rad50. The different orientation of the signature sequence, crucial for ATP-binding, might reflect differences in nucleotide binding mechanism between these two homologous proteins.

The sequence alignment shown in Figure 46 comparing RecN*head* domain sequence with the head domains of SMC and Rad50 proteins, highlighted the high level of sequence

conservation within the ATP binding/hydrolysis motifs. With the crystal structure of RecN*head* protein now available, a more detailed structural sequence alignment could be performed in order to compare the structures of RecN*head* and Rad50 (Hopfner et al, 2000). Walker-motifs and signature sequence are highly conserved (represented in black squares) and the secondary structure elements show a high level of similarity. Helix α1 containing the Walker-A motif is longer than the one carrying the same conserved motif in Rad50, but it is not fully understood if this difference could affect the functional activity of RecN. A few differences were observed, such as β8 and β11 elements of RecN*head* (Figure 46) which are not present in pfRad50.

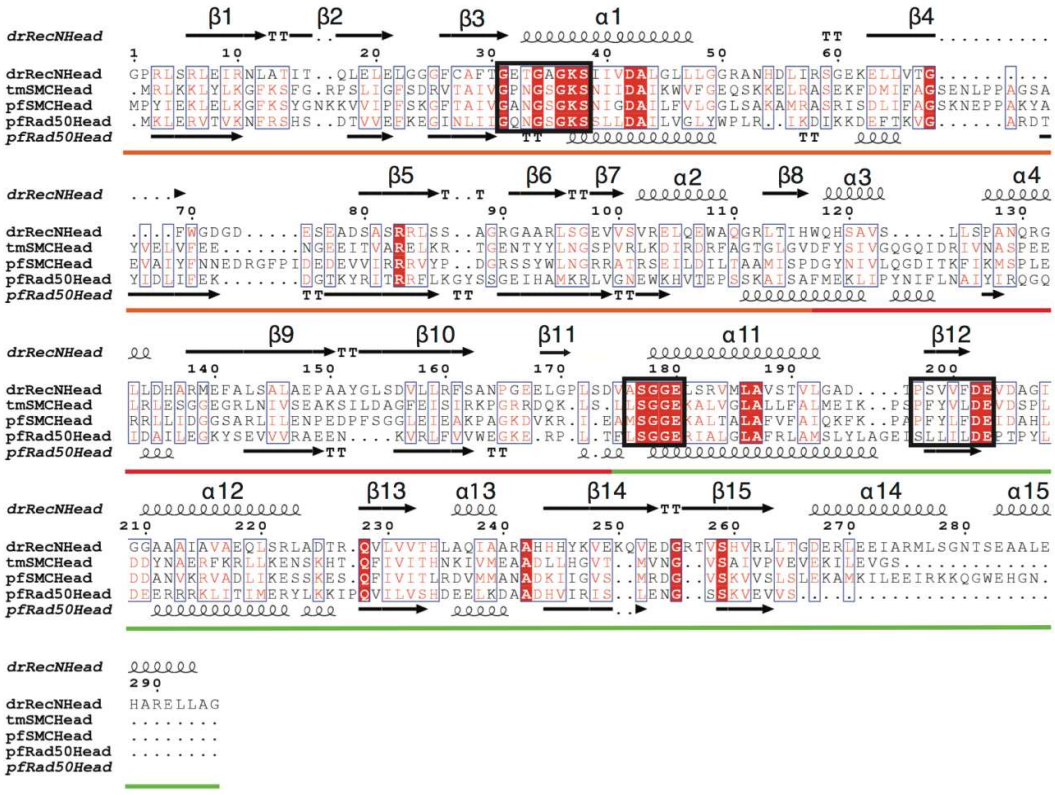


Figure 46: Sequence alignment of RecN*head* domain for comparison of structural features to homologous proteins. The three conserved motifs, Walker-A, signature sequence and Walker-B respectively (indicated by closed black boxes), are highly conserved among the different head domains used for alignment. Secondary structure elements of RecN*head* and pfRad50 head domain are presented for comparison.

RecN and consequently RecN*head* have an N-terminal extension and an HTH motif at its C-terminus (Figures 45 and 46), which are not present in other bacterial RecN proteins (Figure 22). The flexibility of the N-terminal insertion was confirmed in the RecN*head* crystal structure since there was no clear electron density for the first 31 amino acids. The role of this

extension is not understood yet. The presence of a Proline-rich region suggests that it might be involved in protein-protein interactions (Williamson, 1994).

4.6. Crystal structure vs. *ab initio* model of RecN*head*

An *ab initio* model of RecN*head* domain was built using DAMMIN (Svergun, 1999) and was presented in paper I (Pellegrino et al, 2012b). 15 models were obtained and then the electron density averaged and filtered in order to reconstruct a more consistent envelope. A picture of the superposition of the refined crystal structures with the overall shape resulting from the *ab initio* calculation is shown (Figure 47) and indicated that the overall shape adopted by RecN*head* domain in solution is consistent to that seen in the crystal structure.

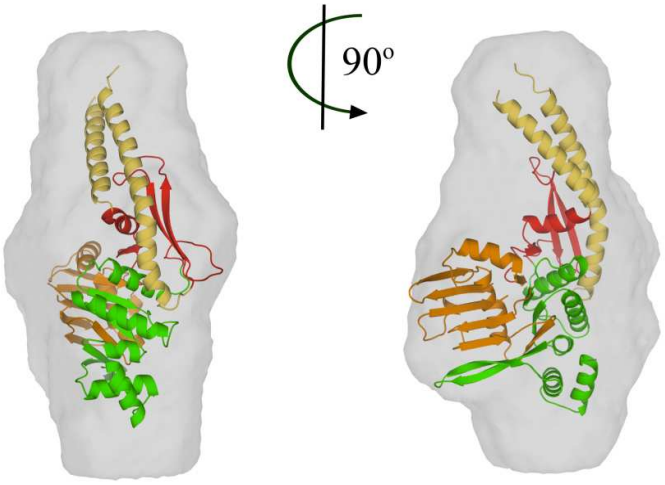


Figure 47: Overlay of the crystal structure of RecNhead with the averaged *ab initio* envelope obtained from SAXS data. Superposition was performed using SUPCOMB20 (Kozin & Svergun, 2001).

4.7. Nucleotide Binding Domain in detail

Structural analysis of RecN*head* domain highlights the presence of conserved residues in both the Walker motifs. These residues, notably K67 in the Walker-A and D471 and E472 in the Walker-B motif, are essential for ATP-binding, Mg^{2+} coordination and ATP hydrolysis respectively by mediating cleavage of γ-orthophosphate by a water molecule. Together with the signature sequence located on the same head domain molecule they give rise to half of the ATP binding pockets. Structural comparison of RecN*head* with pfRad50 head domain has shown a different positioning of the signature sequence: it is situated only 6.8Å away from Walker-B motif, while in Rad50 the same distance increases up to 13.6Å, indicating a bigger spatial separation between the two binding pockets. In the ABC protein, UvrA2 (PDB code:

2VF8), the functional and conserved motifs are placed closer, as seen in the case of RecN*head* crystal structure (as discussed in paper III (Pellegrino et al, 2012 submitted) and also shown in Figure 48).

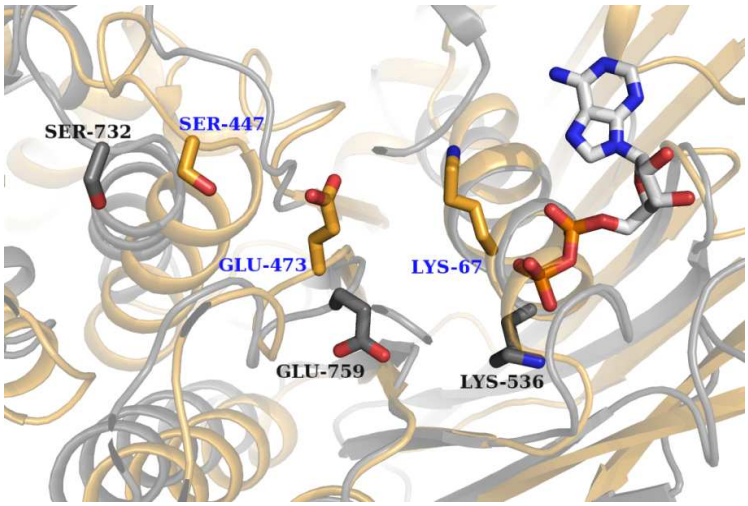


Figure 48: Representation of half of the functional NBD pocket. In the picture RecN*head* domain (grey) is superimposed to ADP-bound UvrA2 NBD2 (sand) (Timmins et al, 2009). The position of the three most conserved residues (Lys at the Walker-A, Ser in the signature sequence and Glu in the Walker-B motif) is highly conserved. ADP (crystallized together with drUvrA2 protein) is shown for clarity and represented as sticks.

As discussed in the introduction (§ 1.5.5) ATP has been shown to induce major conformational re-arrangements of the NBD, leading to head–head engagement (Hopfner et al, 2000; Mockel et al, 2011; Paull et al, 2004). 5mM nucleotides (ATP, ADP and the non-hydrolysable form AMP-PCP) were supplemented to wild-type RecN and RecN*head* proteins prior to carrying out biophysical measurements: SEC and SAXS studies for RecN and SEC-MALLS (Figure 41) and SAXS experiments for RecN*head* domain. Unfortunately, no obvious changes were observed and, therefore, no additional information was gained using these techniques. Co-crystallization experiments using RecN*head* domain and the mentioned nucleotides gave rise to nice crystals with different morphology after 8-10 days in conditions very similar to those in which the RecN*head* domain crystallized. However, these crystals diffracted only very poorly, as it will be discussed later (chapter 5).

4.8. Summary of the results on RecN*head* domain

RecN*head* domain was purified to homogeneity and successfully crystallized. Biophysical and diffraction analysis resulted in the definition of a low, obtained by SAXS, and a relatively high (3 Å) resolution model, obtained by X-ray diffraction experiments.

Analysis of the *ab initio* envelope showed that RecN*head* domain might adopt a similar structural organization to Rad50 protein (Pellegrino et al, 2012b).

Structural analysis of RecN*head* crystal structure, in addition, confirmed that RecN shares an analogous fold to SMC and Rad50 proteins, as also predicted by secondary structure prediction's tools. The crystal structure contained the well-characterized NBD, typical of ABC and ABC-like proteins, but also a portion of the nascent coiled-coil domain and an extra HTH motif at the C-terminus. The ATP-binding pocket is contained in the NBD and the functional residues of the Walker-A and Walker-B motifs are positioned similarly to those in the ABC proteins pfRad50 and drUvrA2 (Figure 48).

The crystal structure of RecN*head* also highlighted several differences within the nucleotide binding pocket which may affect ATP-binding and hydrolysis. The signature sequence, particularly, was shown to adopt a different position, closer to the Walker motifs than in Rad50 and SMC proteins. This is probably due to differences in the mechanism of recognition and binding of the nucleotide to accomplish the hydrolytic function.

RecN*head* was the first crystal structure obtained for bacterial RecN proteins and therefore could represent the starting point for the understanding of the DSB recognition pathway.

Chapter 5:

ATP-binding properties of RecNhead domain and their structural and biochemical implications

Summary of chapter 5

RecN contains a NBD and is therefore expected to bind and hydrolyze nucleotides, such as ATP. Co-crystallization experiments using different nucleotides (ATP/ADP/AMP-PNP) were performed in order to obtain the crystal structure of RecN*head* domain in complex with its cofactor, but the crystals obtained diffracted only very poorly.

Dissection of the ATP-binding mechanism was undertaken through the mutation of key residues potentially involved in this process. A double mutant with mutations in both the Walker-motifs was of particular interest since it formed a stable dimer (as assessed by Multi-Angle Laser Light Scattering and Small Angle X-ray Scattering studies). The scattering curve allowed to build a model illustrating how two RecN*head* domains interact upon ATP-binding, providing the first low resolution model of dimeric RecN*head*.

Résumé du chapitre 5

RecN contient un domaine de fixation de nucléotide, appelé NBD, et par conséquent, il devrait lier et hydrolyser les nucléotides, tel que l'ATP. Des expériences de co-cristallisation utilisant différents nucléotides (ATP / ADP / AMP-PNP) ont été réalisées afin d'obtenir la structure cristalline du domaine RecN*head* en complexe avec son cofacteur, mais les cristaux obtenus ne diffractaient que faiblement.

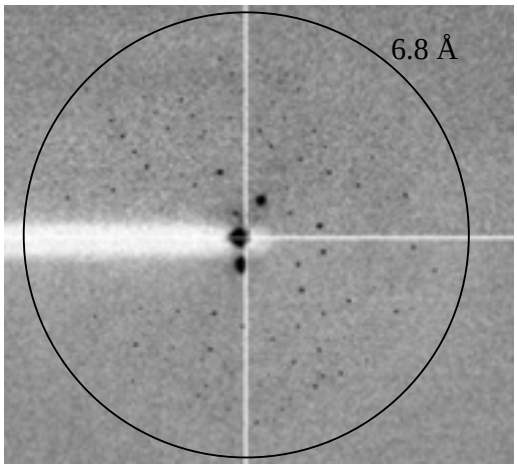
La dissection du mécanisme de liaison de l'ATP a été entreprise par la mutation des résidus clés potentiellement impliqués dans ce processus. Un double mutant avec des mutations dans les deux motifs Walker était tout particulièrement intéressant car il formait un dimère stable (évalué par diffusion aux petits angles et diffusion de lumière laser multi-angles. La courbe de diffusion obtenu par SAXS permit de construire un modèle illustrant la manière dont les deux domaines RecN*head* interagissent lors de la fixation de l'ATP, offrant le premier modèle basse résolution d'un dimère de RecN*head*.

5.1. Co-crystallization trials of RecN*head* domain with nucleotides

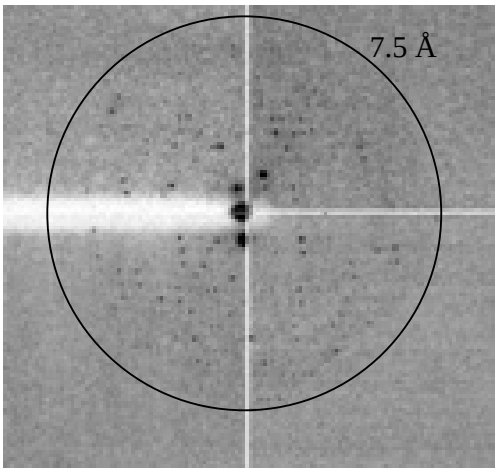
ATP is known to induce a major conformational re-arrangement of the NBD, leading to head-head engagement (Hopfner et al, 2000; Mockel et al, 2011; Paull et al, 2004). RecN*head* was analysed by SEC, SEC-MALLS and SAXS in the presence of a large excess of different nucleotides (ATP/ADP/AMP-PCP) but no significant changes were observed. The crystal structure of RecN*head* domain does not contain a bound nucleotide. Co-crystallization experiments were therefore performed using RecN*head* protein at 9.6 mg ml^{-1} and nucleotides (ATP, ADP or AMP-PCP) at a final concentration of 5mM. Crystals of different morphology (Figure 49) appeared after 8-10 days in conditions containing 100mM Tris-HCl pH 8-8.5 and 22.5-27.5% PEG 6000. Crystals were tested at Room Temperature (RT) on ID14-1 at the ESRF (Wakatsuki et al, 1998) using the humidity control device (HC1) developed jointly by ESRF and EMBL (Russi et al, 2011). The crystals showed very poor diffraction. With the use of the humidifier control device it was not possible to improve the resolution limit beyond the 6.8 Å (Table 6) obtained at 99.5 % of humidity. These crystals belonged to space group I_4 with cell parameters a,b = 128.4 Å, c = 314.3 Å. Unfortunately it was not possible to extrapolate any useful structural information, since the experiment was carried out at RT and the crystals suffered from severe radiation damage already after just a few images.

Humidity (%)	**Space group**	**Resolution (Å)**	**Cell parameters (Å, °)**
99.5	I4	6.8	a,b=128.4, c=314.3
98.5	C2	7.5	a=182.6, b=181.2, c=182.3 β=120.6
97.5	Failed	8.3	
96.5	C2	8	a=184.9, b=306.5, c=191.2 β=85.2
95.5	C2	10	a=309.4, b=179.7, c=181.9 β=89.0
94.5		20	
87.5		>20	

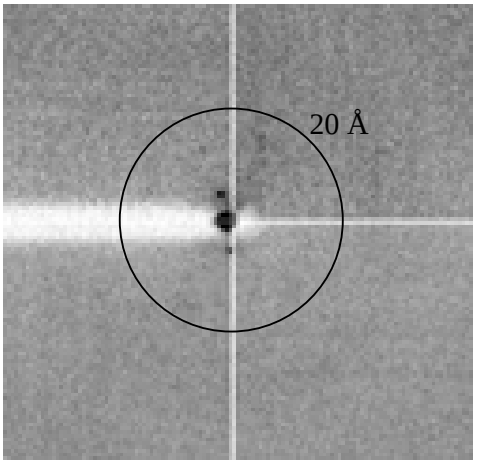
Table 6: Summary of the integration of the two images collected for the crystal tested at different humidity conditions. Experiments were performed on ID14-1 (ESRF) at RT using the humidifier device HC1.



Diffraction pattern at 99.5% humidity

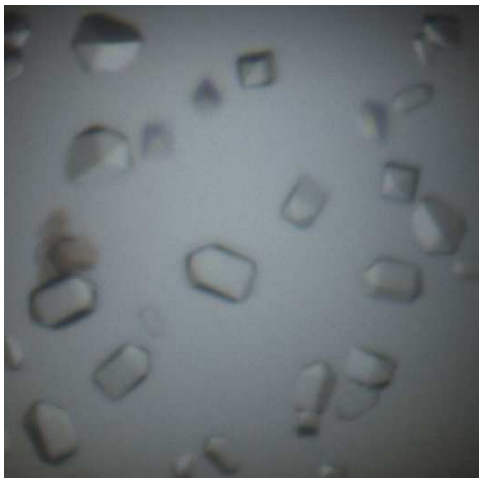


Diffraction pattern at 98.5% humidity



Diffraction pattern at 87.5% humidity

B)



Picture of crystals of RecNhead domain co-crystallized with nucleotide

Figure 49: Images collected at different humidity concentrations (the values are indicated). Circles indicating the resolution limit are also shown for clarity. B) Crystals of RecN*head* domain in presence of nucleotides. Morphology of the crystals is completely different from what was seen for the RecN*head* apo-protein (Pellegrino et al, 2012c).

5.2. Mutations of the conserved functional residues

Characterization of the ATP-binding site was of particular interest for the understanding of the role of ATP hydrolysis in the putative DNA end-joining activity of RecN (Reyes et al, 2010). Single and double amino acids substitutions were introduced into the Walker A and B motifs: K67A, E472Q, E472A, K67A/E472Q, K67A/E472A and D471A. The mutant proteins were successfully expressed and purified as described in Materials & Methods.

5.3. Purification and preliminary analysis of RecN*head*$^{K67A/E472Q}$

The double mutant, RecN*head*$^{K67A/E472Q}$, in which the Walker A lysine was mutated to alanine and the Walker B glutamate was mutated to glutamine, was purified to homogeneity. The first affinity chromatography step produced a huge amount of protein (~200 mg of protein from 1 liter of cell culture). Subsequently, a SEC column was run to further purify the protein. The mutant protein eluted as a single peak, but its elution volume (12.5 ml) was considerably smaller than the RecN*head* domain (14.1 ml). Interestingly, during the SEC step, the main protein peak exhibited an unusually high 254 / 280nm absorption ratio, indicating that a molecule absorbing at 254 nm may be bound to RecN*head*$^{K67A/E472Q}$. The ratio was maintained even after performing an ion exchange chromatography prior to injection on the SEC column, confirming that the species present is tightly bound to RecN*head*$^{K67A/E472Q}$. We speculate that it may be an ADP or ATP molecule that could have co-purified with RecN*head*$^{K67A/E472Q}$ (Pellegrino et al, 2012 submitted). The protocol used for purification of RecN*head*$^{K67A/E472Q}$ involved the same two-steps procedure used for RecN*head* domain. Affinity chromatography (Figure 50), followed by thorough dialysis and the inclusion of a SEC as final step (Figure 51), yielded a highly pure protein suitable for biophysical analysis like SEC-MALLS (Figure 52) and SAXS.

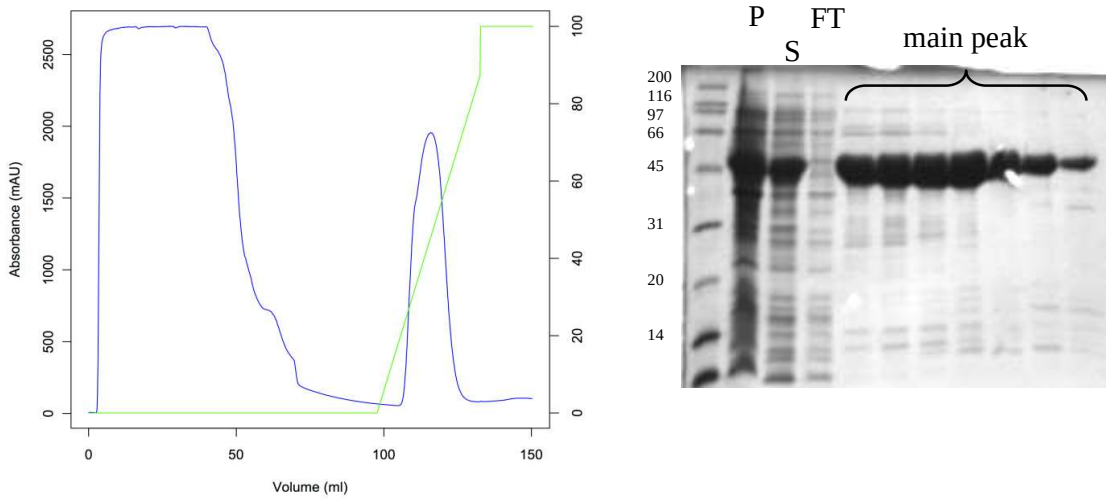


Figure 50: Affinity chromatography profile of RecN*head*$^{K67A/E472Q}$. Protein absorbance at 280 nm is shown in blue. The protein eluted at 35 % of buffer B-2, corresponding to approximately 170mM imidazole. SDS-PAGE gel with the main peak fractions is presented on the right. MWs are reported in kDa. P: pellet; S: soluble fraction; FT: flow-through.

Buffer C-2, as for RecN*head*, represented the optimal buffer for elution and allowed also to compare the behavior of this mutant with the wild-type protein.

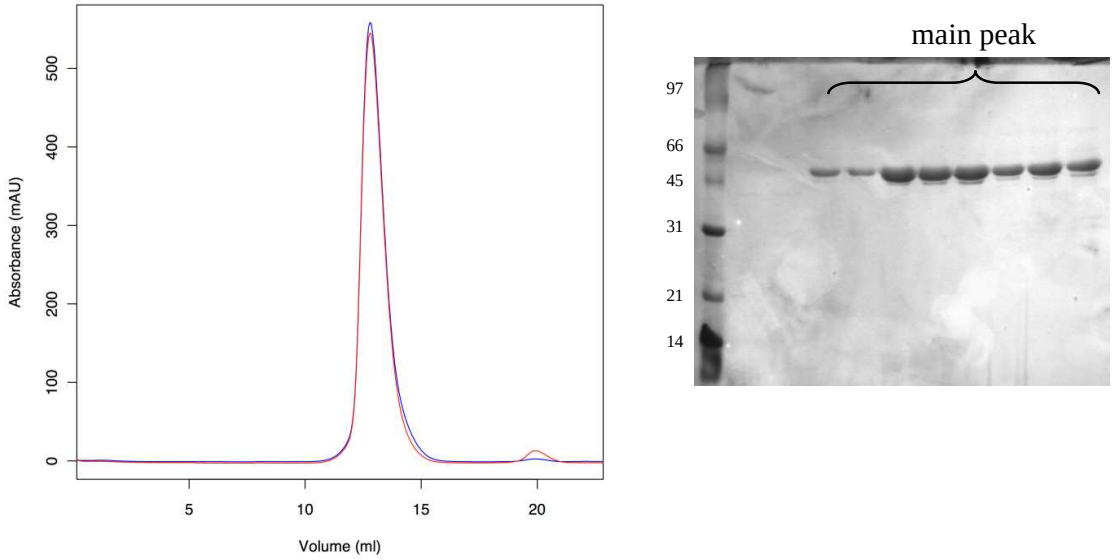


Figure 51: SEC chromatography profile of RecN*head*$^{K67A/E472Q}$. The ratio between the absorbance at 280 nm (blue), corresponding to protein, and at 254nm (red), corresponding to nucleotide absorption, is approximately 1. This result suggests the presence of tightly bound nucleotide to the protein (Pellegrino et al, 2012 submitted). SDS-PAGE gel with the main peak fractions is presented on the right. MWs are reported in kDa.

5.4. SEC-MALLS confirmed the dimeric state of RecN*head*$^{K67A/E472Q}$

Comparison of the SEC-MALLS profiles of RecN*head* and RecN*head*$^{K67A/E472Q}$ constructs demonstrated that the mutant is double the size of the native protein, suggesting that the specific mutations induced a structural rearrangement that favored a dimeric assembly (Figure 52).

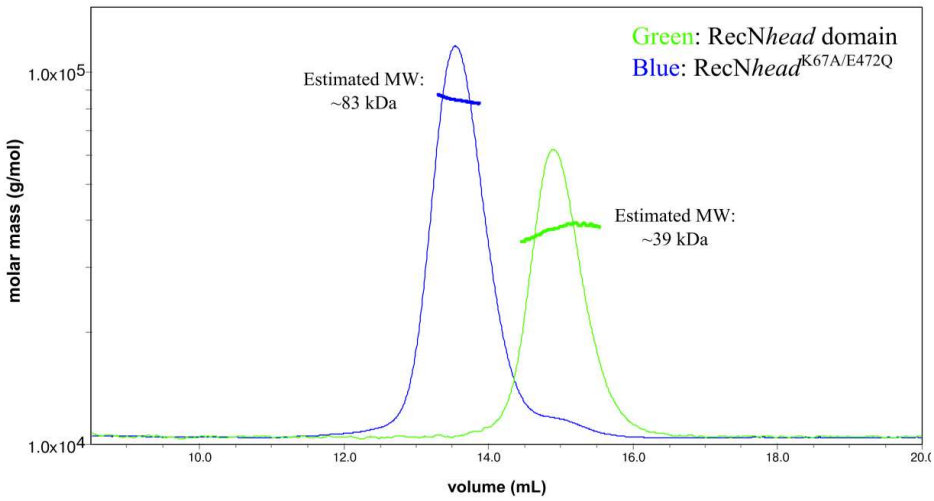


Figure 52: SEC-MALLS profiles of RecN*head* domain and RecN*head*$^{K67A/E472Q}$. Comparison of the two curves evidenced that the double mutant is a stable dimer in solution. Experiments were conducted in the same buffer conditions: 50mM Tris-HCl pH 8, 300 mM NaCl and 5 mM $MgCl_2$. Thin line: refractive index recorded; thick line: averaged MW estimation from different scattering angles.

5.5. SAXS studies of RecN*head*$^{K67A/E472Q}$

SAXS experiments were performed on ID14-3 at the ESRF and scattering curves were collected on RecN*head*$^{K67A/E472Q}$ (Figure 53). SAXS data could provide us with more information regarding the overall envelope adopted by RecN*head* domain when nucleotide enters and binds to the ATP-binding site.

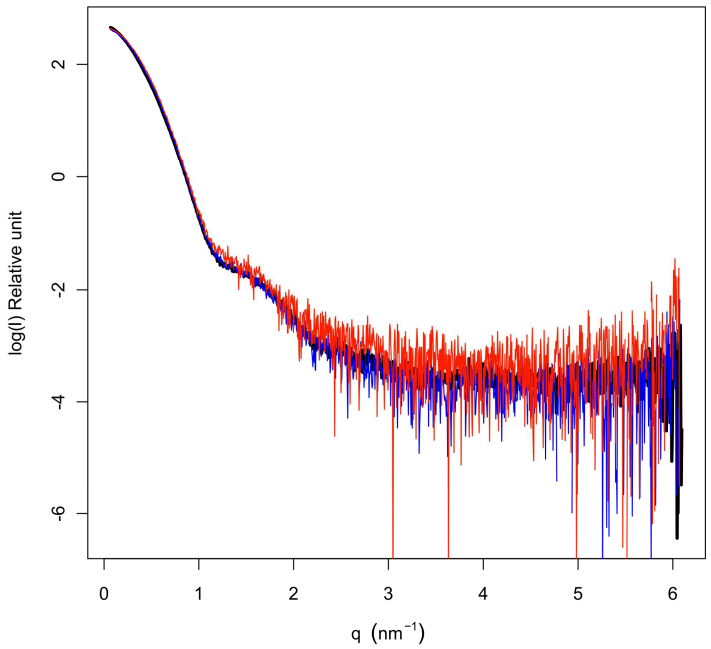


Figure 53: SAXS measurements were performed on RecN*head*$^{K67A/E472Q}$ protein. Three different concentrations (3.5, 1.7 and 0.8 mg ml^{-1}) were measured in order to check for any inter-particle effect. The curves corresponding to the lowest (red) and highest (black) protein concentrations were merged and used for *ab initio* model reconstruction by using GASBOR (Svergun et al, 2001b).

The scattering curves (Figure 53) obtained from three different protein concentrations (3.5, 1.7 and 0.8 mg ml^{-1}) were used to successfully reconstruct an *ab initio* envelope (Figure 54). The Guinier approximation estimated the radius of gyration (R_g) to be 3.45 nm and the molecular weight (~70 kDa) derived from the scattering at zero angle (I_0) was in agreement with the size of a dimer. Taking advantage of the available knowledge regarding ATP-binding sites (Hopfner & Tainer, 2003) and our overall envelope, we manually built a model of the putative RecN*head* dimer using our recently obtained crystal structure (Figure 45). The calculated theoretical scattering curve derived from this model was superimposed on the experimental SAXS curve. The excellent fit with a χ^2 value of 1.56 over the entire curve indicates that this model describes very well the behavior of the protein in solution (Figure 55).

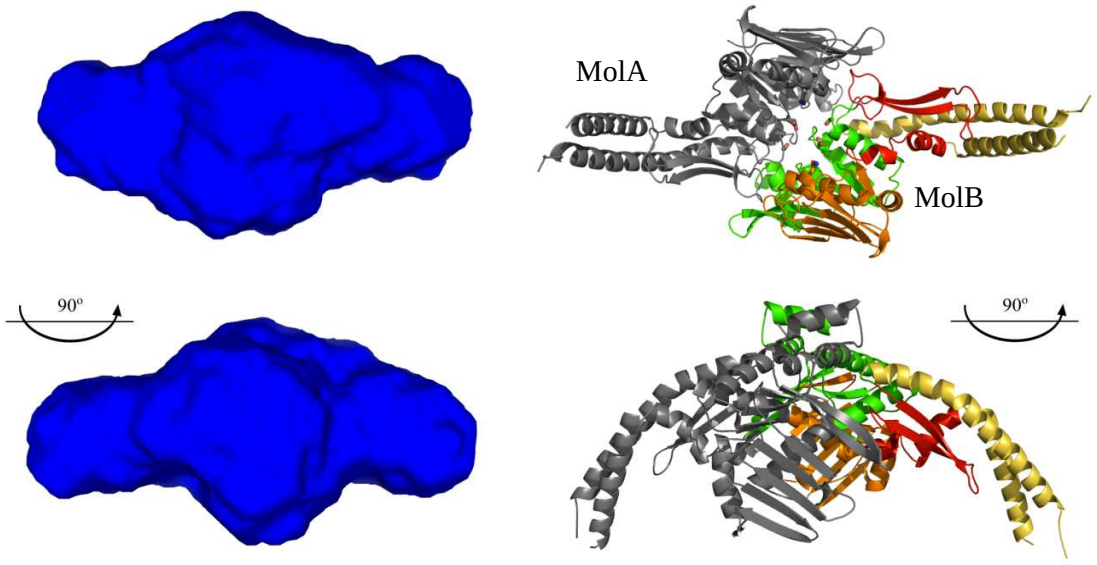


Figure 54: SAXS envelope (on the left) and model built using the RecN*head* domain crystal structure (on the right) are shown. All pictures are shown 90 degrees apart in order to view the orientation of the coiled-coil helices, which point out in the same direction. The envelope is the averaged results of several runs of *ab initio* modeling using DAMMIN (Svergun, 1999). In the model built form the RecN*head* crystal structure one monomer is colored accordingly to the colors used in chapter 4, while the other monomer is colored in grey for clarity.

In the model proposed in Figure 54, the coiled-coil domains are located on either side of the two RecN*head* monomers (MolA and MolB), that come together in a head-to-tail orientation to form two functional ATP-binding pockets as observed previously for dimerization of NBDs (Hopfner et al, 2000; Lammens et al, 2004). Each RecN*head* monomer contains all the conserved motifs necessary for the formation of the functional ATP binding pocket. As in Rad50 and other ABC proteins, the Walker motifs of one RecN*head*$^{K67A/E472Q}$ molecule and the signature sequence of a second molecule come together to form a functional nucleotide binding pocket (Hopfner et al, 2000; Timmins et al, 2009). In the crystal structure of RecN*head*, the signature sequence is located only 9 Å away from the Walker-B motif, while this distance increases up to ~20Å for SMC or Rad50 head domains, suggesting that the two ATP binding pockets at the NBD dimerization interface would be much closer together in the case of RecN. The arrangement of the Walker-A, -B and signature sequence motifs in RecN are more similar to that found in the C-terminal NBD2 of UvrA2 (RMSD: 1.54 Å over 116 C_α) in its ADP-bound form (Figure 47).

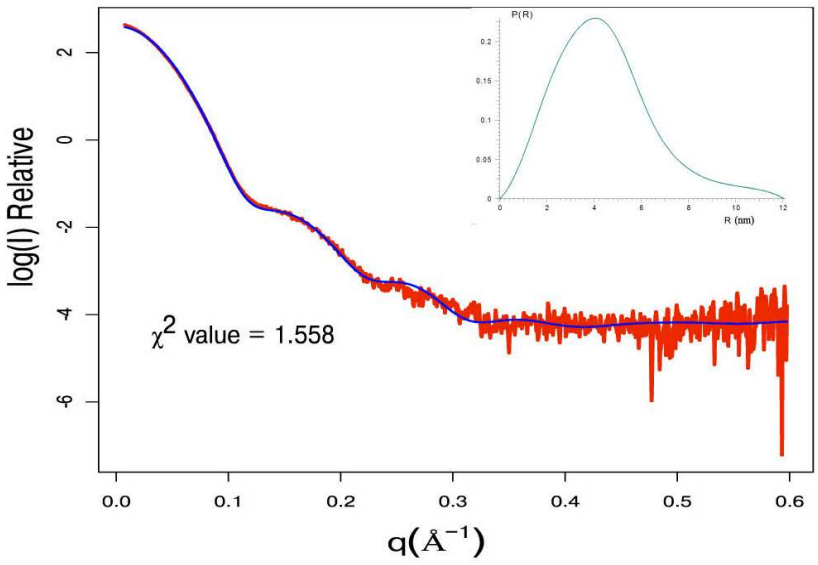


Figure 55: Overlay of the experimental SAXS curve (in red), created by merging highest and lowest concentrations scattering curves, with the theoretical curve (in blue) calculated from the manually built model of RecN*head*$^{K67A/E472Q}$. The dimeric model of RecN*head* double mutant approaches very well the behavior of this construct in solution. Goodness-of-fit (χ^2) is reported.

The *ab initio* envelope resulting from this SAXS study was used as a template for superposition with the built model, proposed in Paper III (Pellegrino et al, 2012 submitted). The coiled-coil α-helices are pointing out at the same direction, suggesting a model similar to what has been proposed for Rad50 while interacting with Mre11 (Mockel et al, 2011). The finding that RecN head-head engagement shares a number of conserved features with Rad50, suggests that the molecular mechanisms underlying DSB recognition might also be conserved.

5.6. Crystallization trials on RecN*head* domain mutants

To clarify the contribution of the signature motif rearrangement on the nucleotide binding properties of RecN (Pellegrino et al, 2012 submitted), atomic resolution data was desirable. Crystallization trials on RecN*head*$^{K67A/E472Q}$ were carried out in parallel to the SAXS study. Crystals were obtained from the crystallization robot screening (Figure 56). Initial conditions were nº 20 of the PEG/Ion Screen, containing 0.2 M,magnesium formate pH 5.9 and 20 %w/v PEG 3350. Crystals appeared also in the condition nº 22 of the Crystal Screen I, containing 0.2 M sodium acetate trihydrate, 0.1 M Tris-HCl pH 8.5 and 30 %w/v PEG 4000 (Hampton Research). Despite the fact that crystals were obtained and were also reproducible, diffraction spots were not visible on the processed image.

Instead, spectroscopy was used to compare these crystals to those of the native protein (paragraph 5.7).

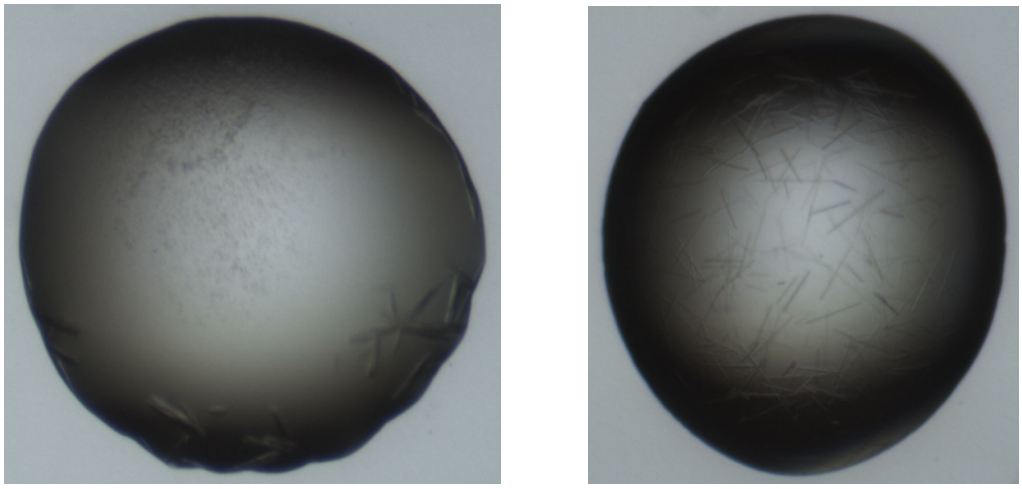


Figure 56: Crystals of RecN*head*$^{K67A/E472Q}$ were obtained after initial screening using the crystallization robot at the EMBL-Grenoble (Dimasi et al, 2007). The conditions at which these appeared were nº 20 of the PEG/Ion Screen (left) and nº 36 of the Crystal Screen I (right) (Hampton Research). Crystals were tested at ID23 and ID29 (ESRF) but diffraction was very poor.

Crystallization trials were also set up for the RecN*head*^{E472Q} mutant and crystals diffracting to 3.5 Å were obtained (Figure 57). The purpose of this experiment was to improve our understanding, at an atomic level, of the mechanism by which ATP is hydrolyzed and, in addition, to understand why the activity of this mutant is so much higher than that of the wild-type protein (chapter 8). Crystals belonged to space group $P2_12_12_1$ with cell parameters a = 63.7 Å, b = 121.8 Å, c = 161.7 Å. An almost complete data set was collected (~87 % completeness) because the crystal suffered from radiation damage. Mol.R. was performed using *PHASER* (McCoy et al, 2007) and a reasonable solution was found with a likelihood gain (LLG) of 1508. Analysis of the resulting electron density map at the ATP-binding pocket site revealed no difference in the orientation of the Walker-A and –B conserved residues, notably K67 and E472. As with the wild-type protein, these crystals were obtained in the absence of nucleotide and most likely, further information would be gained by co-crystallizing the mutant protein with ATP or AMP-PNP (non-hydrolyzable form).

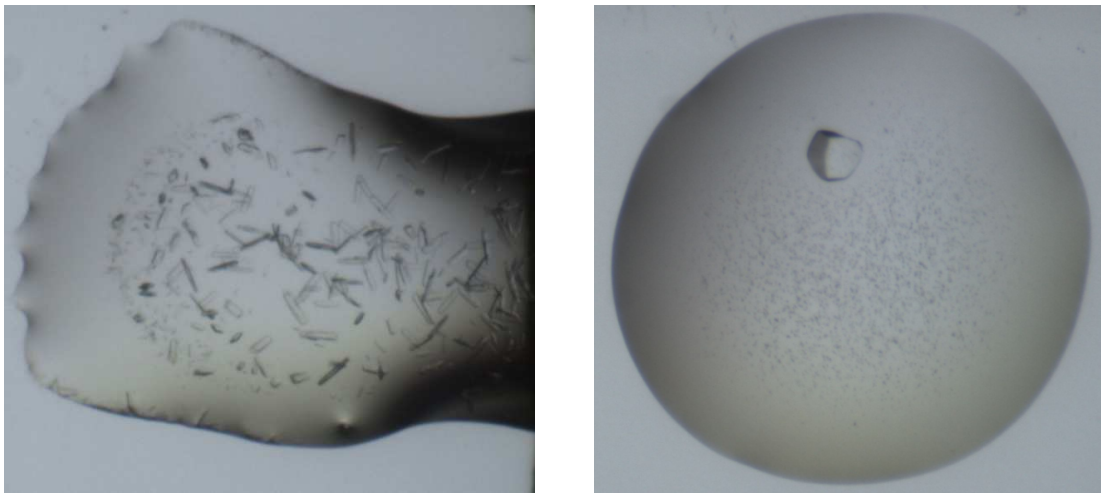


Figure 57: Crystal of RecN*head*^{E472Q} construct. On the left: crystals obtained in drops using the crystallization robot at the HTX laboratory (EMBL-Grenoble) and grown in condition nº 21 of the PEG-Ion screen (Hampton Research). These crystals diffracted to 3.5Å on ID29 at the ESRF. On the right: crystal obtained after 3 weeks equilibration. The condition in which this crystal appeared was the nº 29 of the Index Screen (Hampton Research).

5.7. Spectroscopic measurements in crystallo

Steady-state fluorescence emission spectra were carried out on ID29-S at the Cryo-Bench at the ESRF (Royant et al, 2007). Spectra were recorded on crystals of RecN*head* and RecN*head*$^{K67A/E472Q}$ domains. Since X-ray diffraction data were not available for RecN*head*$^{K67A/E472Q}$, the aim of these measurements was to further characterize the unidentified, bound ligand. Spectra were recorded on each of the two crystals over the full UV-visible wavelength range and then compared.

At wavelengths approaching those where protein (280 nm) and nucleotides (254 nm) absorb, the spectra of the two crystals present slightly different features (Figure 58). This result confirmed, first of all, that the crystals tested were indeed protein crystals of RecN*head*$^{K67A/E472Q}$. However, unfortunately these spectra were not sufficient to allow us to discriminate between ADP and ATP, given the similar chemical properties of the two nucleotides.

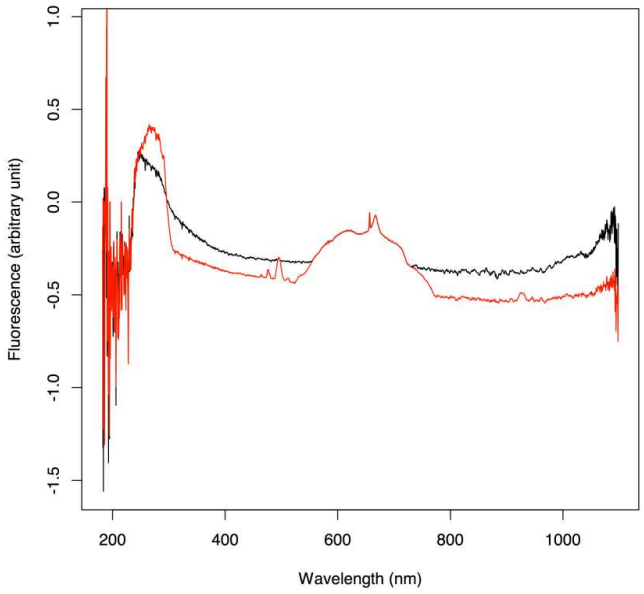


Figure 58: Comparison of spectra taken from crystals of wild-type RecN*head* domain (in black) and crystals of RecN*head*$^{K67A/E472Q}$ (in red), in which a nucleotide is suggested to be trapped from the early purification steps. Data were recorded at the Cryo-Bench at ID29 (ESRF).

5.8. Summary of the structural changes of RecN*head* induced by ATP

A double mutant of RecN*head* domain, namely RecN*head*$^{K67A/E472Q}$, provided us with a model of the possible dimeric organization that occurs upon ATP-binding. An *ab initio* envelope reconstructed from SAXS data allowed to define the probable nucleotide-bound form of RecN*head*. In this model the coiled-coil domains are oriented in a similar way and point out in the same direction from the globular region. The dimeric assembly proposed in this work and built considering the position of the conserved functional residues fitted very well with the experimental scattering curve (Figure 55). However, the low-resolution nature of SAXS did not allow us to determine if local conformational re-arrangements took place in the ATP-binding pocket of RecN*head* in preparation to receive ATP. Crystals of RecN*head*$^{K67A/E472Q}$ were obtained but unfortunately no diffraction was visible. A comparative spectroscopic study provided a spectrum in the UV-Vis range, which at least excluded the possibility that we were working with salt crystals. Comparison with wild-type RecN*head* crystals highlighted a difference in the spectra, especially in the region between 250-300 nm (Figure 58).

More efforts need to be made to improve the quality of the RecN*head*$^{K67A/E472Q}$ crystals in order to better understand the ATP-binding and hydrolysis mechanisms.

Biochemical experiments could also provide new insights into the effects of single amino acid substitutions on the ATP release.

Chapter 6:

Structural investigation of the coiled-coil domain of RecN

Summary of chapter 6

RecN also possesses a central coiled-coil domain, interconnecting the N- and C-termini. In RecN, this domain is particularly short compared to the other members of the SMC family, such as SMCs, Rad50 or MukB. We successfully expressed and purified two different constructs corresponding to residues 240-387 and 196-365 (RecN*cc*). Both of them were characterized by Multi-Angle Laser Light Scattering, which highlighted the different oligomerization state of these two constructs. While the former was a monomer in solution, the latter adopted a dimeric state in solution (as seen also by Small-Angle X-ray Scattering analysis) suggesting that this region could be responsible for RecN dimerization. RecN*cc* was also successfully crystallized and the structure solved by Single-wavelength Anomalous Dispersion (using Se-Met derivatives) at a maximum resolution of 2 Å. The dimerization interface is formed through a network of van der Walls interactions.

A deletion mutant (RecNΔ*dd*), lacking this dimerization interface, was also expressed, purified and characterized by SEC-MALLS, SAXS and X-ray diffraction. In the crystal structure (resolution 4 Å) RecNΔ*dd* is a monomer, confirming our previous observations.

Résumé du chapitre 6

RecN est également constitué d'un domaine central formant un faisceau d'hélices (coiled-coil), connectant les régions N- et C-terminales. Pour RecN ce domaine est particulièrement court par rapport à ceux des autres membres de la famille SMC, tels que SMC, Rad50 ou MukB. Nous avons réussi à exprimer et purifier deux constructions différentes, correspondant aux résidus 240-387 et 196-365 (RecN*cc*). Toutes deux ont été caractérisées par la diffusion de lumière laser multi-angles (MALLS), qui met en évidence des états d'oligomérisation différents pour les deux protéines. Alors que le premier est un monomère en solution, le deuxième adopte un état dimérique en solution (confirmé par l'analyse de diffusion aux petits angles, SAXS), suggérant que cette région pourrait être responsable de la dimérisation de RecN. RecN*cc* a également été cristallisée et sa structure résolue par dispersion anomale à une longueur d'onde (en utilisant une protéine séléniée) à une résolution maximale de 2 Å. L'interface de dimérisation est formée par un réseau d'interactions Van der Walls.

Un mutant de délétion (RecN*Δdd*), manquant cette interface de dimérisation, a également été exprimé, purifié et caractérisé par MALLS, SAXS et diffraction aux rayons X. Dans la structure cristalline (résolution 4 Å), RecN*Δdd* est un monomère, confirmant ainsi nos précédentes observations.

6.1. Expression and purification of coiled-coil domain of RecN

Different constructs (Table 5) were designed for the coiled-coil domain but most of our attention focused on the one including residues 196 to 365 (RecN*cc*). Another construct, corresponding to residues 240-387 of RecN, was also successfully expressed and was characterized by SEC-MALLS studies (§ 6.3).

The DNA codifying for RecN*cc* protein was amplified by PCR, cloned in pET151-TOPO vector and transformed into *E. coli* cells BL21* as for the other constructs. RecN*cc* was subsequently purified according to the protocol described in Materials & Methods section and in (Pellegrino et al, 2012a). A two-step purification (Figures 59 and 60) was performed using the same buffer solutions as for RecN*head* protein purification and yielded pure protein, suitable for further structural (X-ray crystallography and SAXS) and biophysical analysis by DLS and SEC-MALLS (Figure 60).

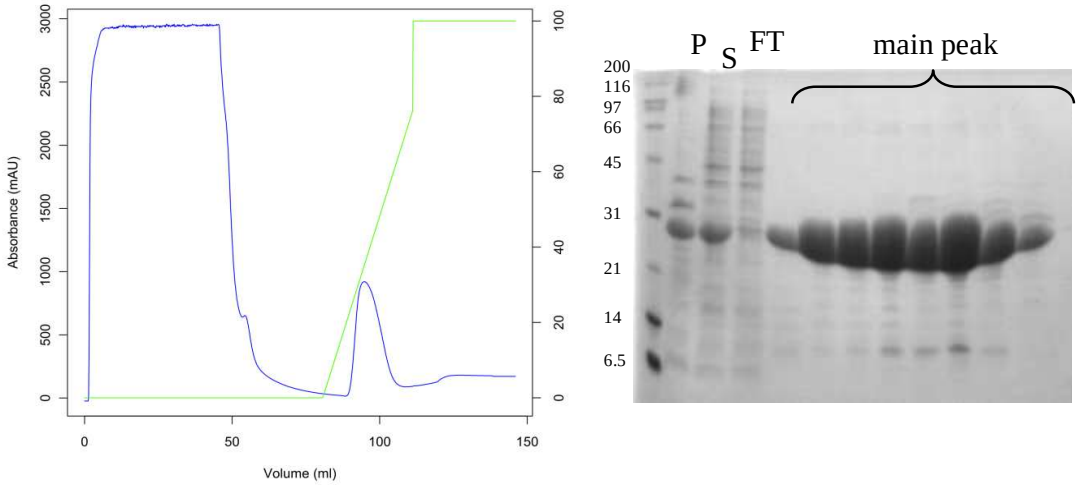


Figure 59: Affinity chromatography profile of RecN*cc*. Protein absorbance at 280 nm is shown in blue. The main peak eluted at 32 % of buffer B-2, which corresponds to approximately 165 mM imidazole. 10µl of fractions corresponding to the main peak were loaded onto a 15% SDS-PAGE gel in order to check the purity of the fractions. The gel with the main peak fractions is presented on the right. MWs are reported in kDa. P: pellet; S: soluble fraction; FT: flow-through.

SEC was performed using buffer C-2 (Materials & Methods) and the RecN*cc* eluted as a single peak from the Superdex 200 column, which was used for comparison with the other constructs used in this work. Protein purity was then checked by SDS-PAGE: there are two bands of lower MW (Figure 60) under the band corresponding to RecN*cc*. This could be due to minor degradation of the construct target.

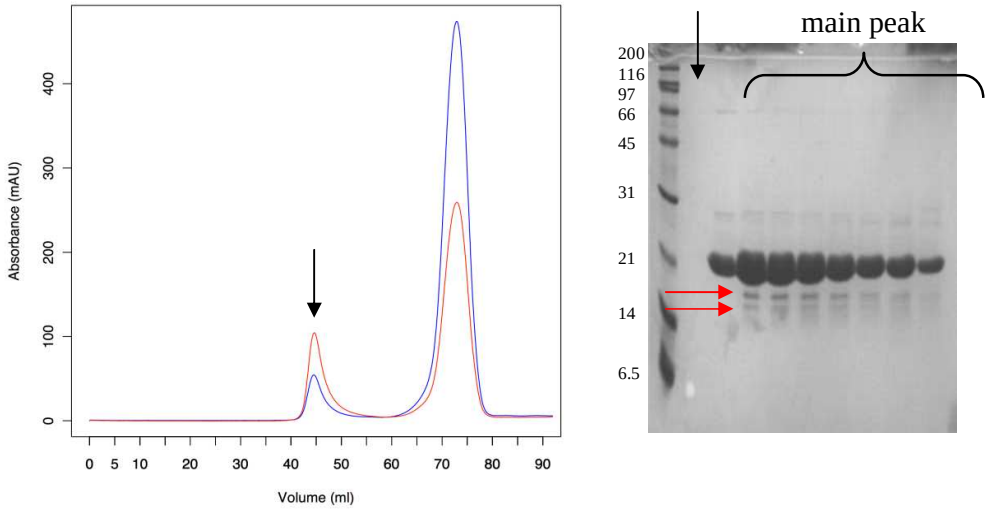


Figure 60: SEC of RecN*cc*. The protein eluted as a single peak, with a small tail before the main symmetrical peak. Protein absorbance at 280 nm is reported in blue, while nucleotide absorbance at 254 nm is in red. A small amount of degradation is visible (red arrows). The protein fractions were checked by DLS and the ones with a low pdI(0.18) were pooled for further analysis. SDS-PAGE gel with the main peak fractions is presented on the right. MWs are reported in kDa.

6.2. SEC-MALLS measurements on RecN*cc* domain

SEC-MALLS measurements (Figure 61) were performed on purified RecN*cc* and resulted in an averaged molecular mass of ~32 kDa, which clearly corresponded to a dimeric state of the coiled-coil domain (theoretical MW of the monomer is 18 kDa as calculated by ProtParam). This finding was the first clear indication of the possible location of the dimerization interface in RecN, as also suggested by a previously published paper (Graumann & Knust, 2009). The RecN dimerization process is therefore most likely achieved through the coiled-coil region. Information is available for the dimerization interfaces of homologous proteins, such as Rad50, SMC and MukB, but there is not a common architecture for these. The question was now, how does the dimerization process take place in RecN?

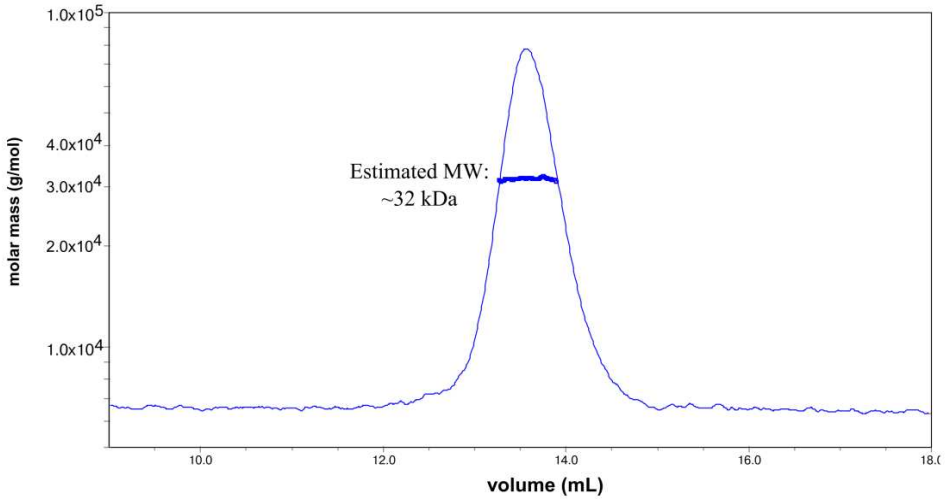


Figure 61: SEC-MALLS measurements performed on RecN*cc* for determination of the absolute MW. Experiments displayed that RecN*cc* is a stable dimer in solution, supporting what is suggested in the literature (Graumann & Knust, 2009). Thin line: refractive index recorded; thick line: averaged MW estimation from different scattering angles. The averaged MW is reported.

6.3. Investigation of the overall shape of RecN*cc* by SAXS

SAXS measurements on RecN*cc* were performed on ID14-3 at the ESRF, using the experimental set-up described in (Pellegrino et al, 2012a). Three different protein concentrations were measured and inter-particle effects were not detectable when comparing the three scattering curves (Figure 62). Data were then truncated for *ab initio* model reconstruction, since there were different features appearing at *q* values higher than 3 $Å^{-1}$. The intensity at zero angle (I_0) allowed the estimation of the MW of RecN*cc*, which resulted to be ~37 kDa, corresponding to the size of a dimer. AUTOPOROD (Petoukhov et al, 2007) was also run for estimation of the ***Porod*** volume and the MW, a method that is, unlike the I_0, independent of concentration measurements and their associated errors. The expected MW calculated by Porod volume resulted to be ~40 kDa, which is in agreement with the value calculated by I_0.

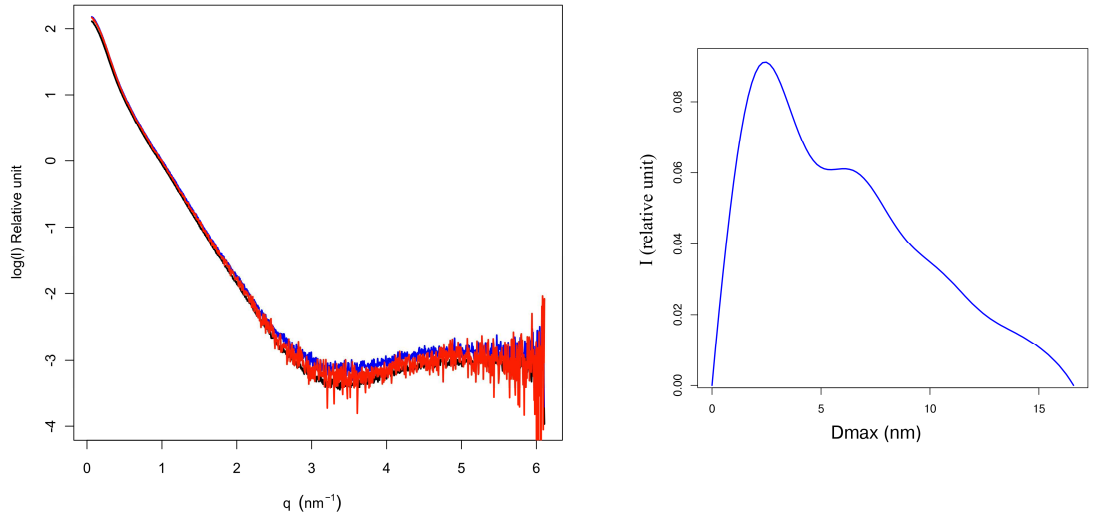


Figure 62: Overlay of the three scattering curves (left) corresponding to three different RecN*cc* concentrations: 6.0 (black), 2.9 (blue) and 1.5 mg ml^{-1} (red). The curves superpose very well at low *q* angles values. The data was truncated at *q* of 3Å^{-1} for following *ab initio* modelling. The pair distribution function (P(r)) is shown on the right and the long tail is characteristic for elongated macromolecules.

The resulting *ab initio* envelope was presented in (Pellegrino et al, 2012a). The characteristic profile of the P(r) was indicative of the presence of a rather elongated molecule in solution and the model building confirmed this hypothesis. Further details will be provided by the crystal structure of RecN*cc*.

6.4. Crystallization of the RecN*cc* and structure determination

RecN*cc* was expressed, purified and crystallized as described in Materials & Methods section and in (Pellegrino et al, 2012a). RecN*cc* was characterized in solution (SEC-MALLS and SAXS) and by X-ray diffraction experiments. Optimization of the crystals was carried out in hanging drop plates (Hampton Research, Aliso Viejo, CA) and drops were set-up mixing 1 µl of protein at either 14.82 or 7.7 mg ml^{-1} and 1 µl of reservoir. Manual optimization of the initial conditions was achieved using the Additive Screen (Hampton Research, Aliso Viejo, CA): 1,2,3-heptantriol gave rise to better diffraction quality and final concentration of 3% produced crystals that diffracted to 2.04 Å. To solve the crystal structure of RecN*cc* experimental phases were determined by the SAD method using a Se-Met derivative crystal, since Mol.R. failed due to the lack of prior structural information. Data were integrated and processed with *iMOSFLM* (Battye et al, 2011) and then submitted to Auto-Rickshaw (Panjikar et al, 2005). Heavy atoms (one Se for each chain) were localized and a new set of

phases calculated using *SHELX* (Sheldrick, 2010). An initial chain tracing was performed using *SHELXE*. The nascent map was then run through *DM* for density modification (Figure 63) in order to improve the quality through several cycles. Extensive model building was performed using *ARP-wARP*, which resulted in an almost complete model of the 4 molecules of the asymmetric unit (over 75 % of the main chain was built). Several cycles of manual model building using *COOT* (Emsley et al, 2010) and refinement using *REFMAC5* (Murshudov et al, 2011) allowed completion of the crystal structure with good final statistics (Pellegrino et al, 2012 submitted).

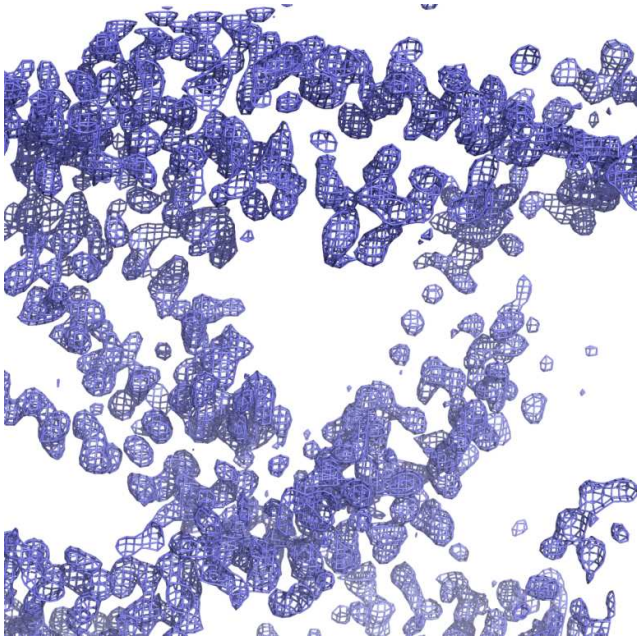


Figure 63: Detail of the Fourier electron density map for RecN*cc* contoured at 1.5 σ after density modification (*DM*). The α-helical arrangement is clearly visible and the backbone can be traced: 75% of the model was initially built from ARP/wARP {Langer, 2008 #1312}.

6.4.1. The coiled-coil domain showed a new interaction interface

The central region of RecN proteins is predicted to be a coiled-coil and to be involved in dimer formation (Graumann & Knust, 2009). This domain is conserved in length but shows very low sequence identity (Figure 22). The crystal structure of RecN*cc* is composed of 6 α-helices (named α5-α10) of varying lengths (ranging from 17 to 29 residues long), which follow an anti-parallel arrangement. The RecN*cc* was found to form a stable dimer with a clear two-fold symmetry (Figure 64). The dimer interface is formed by residues located on helices α6-α8 interacting with their equivalent residues from the second molecule through an extended network of van der Waals interactions. The buried interface covers an

area of ~1400 Å^2 and involves 34 residues altogether. The anti-parallel arrangement of each RecN*cc* monomer is favored by dipole and hydrogen-bonding interactions (Figure 64, magnification): in particular the highly conserved E222 and R319 form a salt bridge that stabilizes the packing of helix α6 against helix α9. As a result, the overall structure of RecN*cc* appears to be very rigid. Several loops located between helices α5-α6, α8-α9 and α9-α10 confer some degrees of freedom to the entire rod-like structure, as displayed by the analysis of the B-factor distribution per residue (see § 6.5 and paper III (Pellegrino et al, 2012 submitted)).

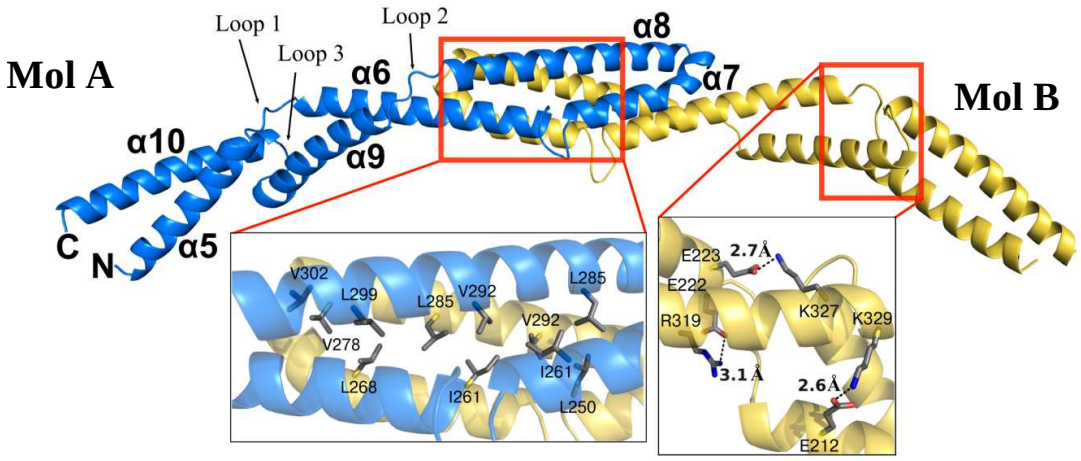


Figure 64: RecN*cc* is a dimer. Extensive inter-molecular interactions between helices α6-α7-α8 contribute to the formation of the dimerization interface. In total 34 residues (magnification below on the left) are responsible for creating an extended hydrophobic buried area (~1400 Å^2) which holds together the RecN*cc* monomers. In the other magnification (on the right) a detail of the intra-molecular polar interactions is shown. We suggested these interactions are involved in the process of folding of RecN*cc*.

The crystal structure of the coiled-coil domain presents an anti-parallel α-helical arrangement (Figure 64 and 65), similar to the *E.coli* MukB (PDB code: 3IBP) (Li et al, 2010), with an RMS of 1.33Å over 61 Cα. The main difference was found to be the mode of dimerization: in MukB, like in SMC proteins, the monomers interact through an extended (2930 Å^2) hinge domain mediated by dipole interactions (Griese & Hopfner, 2011; Li et al, 2010). In the case of Rad50, the dimerization process takes place through a different kind of interface and, precisely, a Zn-hook domain (Hopfner et al, 2002). The entire process is stimulated by the presence of a pocket, created by four cysteines, which allows for a Zn^{2+} ion to bind.

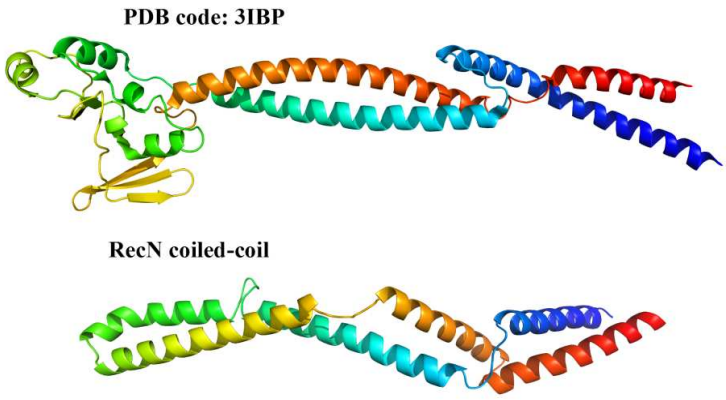


Figure 65: Comparison of ecMukB (Li et al, 2010) (on the top) and RecN*cc* domain. The structural organization between the two proteins is very conserved and it is characterized by the presence of an anti-parallel arrangement of the coiled-coil region. The hinge domain, which is involved in SMC and SMC-like proteins dimer formation and, in a few cases, in DNA-binding, is totally missing in RecN. The two structures are coloured in rainbow in order to distinguish the anti-parallel arrangement.

In conclusion the crystal structure of RecN*cc* revealed that the coiled-coil regions of SMC and SMC-like proteins share an anti-parallel helical arrangement, but differ in their mode of dimerization. RecN*cc* utilizes a novel means of dimerization involving an extended hydrophobic interface that has not been previously reported in the literature for similar interaction interfaces of coiled-coil domains (Griese & Hopfner, 2011; Griese et al, 2010; Hopfner et al, 2002; Li et al, 2010).

6.4.2. Are loops in between the coiled-coil helices involved in flexibility?

In view of the particular structural organization adopted by RecN*cc*, the coiled-coil domain appears to be a mostly rigid, rod-like structure. In SMC proteins, however, flexibility of this region is expected to be crucial in order to allow the "closure" of two NBDs for the formation of an active ATP-binding pocket. In RecN, in contrast, the coiled-coil does not allow the NBDs at either ends of the coiled-coil to engage and therefore form the typical ABC-like pocket. Limited flexibility is nevertheless possible and the main candidates for coiled-coil movements are the loops connecting the α-helices, particularly those connecting α6 to α7 and α8 to α9 (Figure 66). A graph reporting the B-factors of the two chains, A and B, constituting the RecN*cc* dimer, was created using *BAVERAGE* (Dodson, 1991) and showed a clear increase in thermal motion of the Cα of each chain in between the helices.

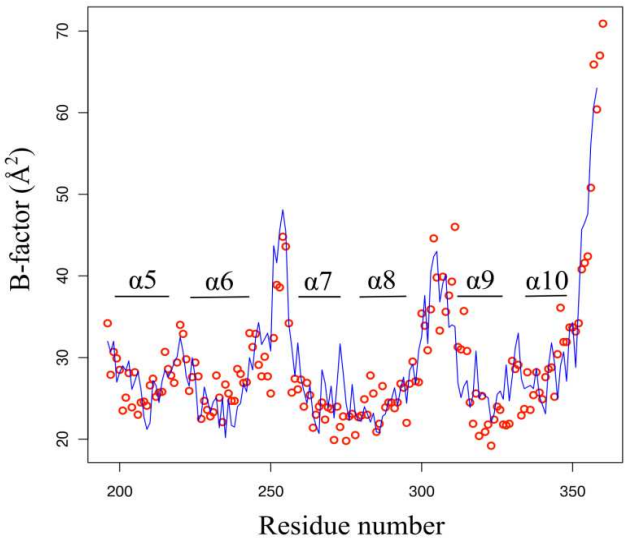


Figure 66: The trend of B-factors in MolA and MolB of RecN*cc* is represented. BAVERAGE was used to extrapolate B-factor statistics from the PDB file, while R was used to build the graph (Gentleman & Ihaka, 2000). The values were lower in correspondence of the secondary structure elements, while increasing in correspondence of the loops, which are more flexible. Red circles: MolA. Blue lines: MolB.

6.5. SAXS vs crystal structure of RecN*cc*

SAXS experiments were performed on RecN*cc* and described in paper II (Pellegrino et al, 2012a). An *ab initio* model of RecN*cc* was built using DAMMIN (Svergun, 1999) and the crystal structure was subsequently overlayed with the overall envelope of the coiled-coil domain (Figure 67) using SUPCOMB (Petoukhov et al, 2007).

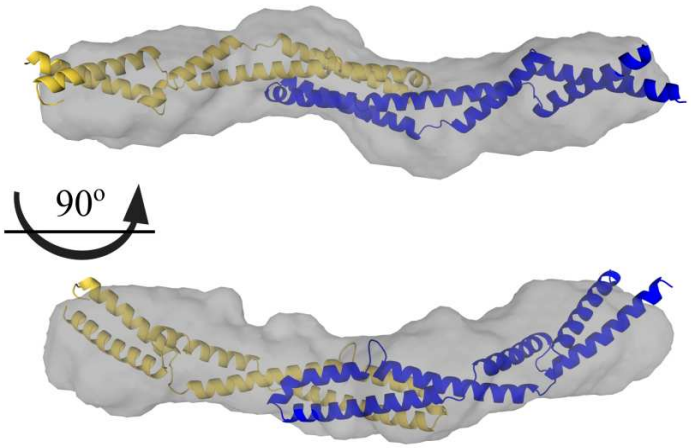


Figure 67: Superposition of RecN*cc* domain crystal structure with the *ab initio* model obtained by SAXS. Model is averaged after several runs of *ab initio* modeling. Superposition was performed using SUPCOMB20 (Kozin & Svergun, 2001).

The overlay (Figure 67) showed a close similarity between the low- and high-resolution structures indicating that the coiled-coil domain of RecN most likely adopts the same architecture both in solution and in crystals.

6.6. Characterization of a shorter coiled-coil construct

The gene codifying for a shorter coiled-coil construct (residues 240-387) was amplified by PCR and subsequently cloned into pET151-TOPO vector for expression in *E. coli*, strain BL21 Star cells. Purified RecN coiled-coil 240-387 construct was obtained following the two steps purification described for RecN*cc* (Pellegrino et al, 2012a). The Immobilized Metal ion Affinity Chromatography (IMAC) step was similar to RecN*cc* purification, with protein eluting at approximately 34 % of the imidazole gradient. Dialysis and TEV cleavage allowed removal of the imidazole and the 6xHis-tag from the protein and the resulting sample was further purified by SEC. It typically eluted around ~81 ml (Figure 68), which is later than with RecN*cc* (73.8 ml). The elution peak also exhibited a significant shoulder, indicating that the sample was not apparently very homogeneous, despite the pdI value of 0.18 suggested that the sample was homogeneous.

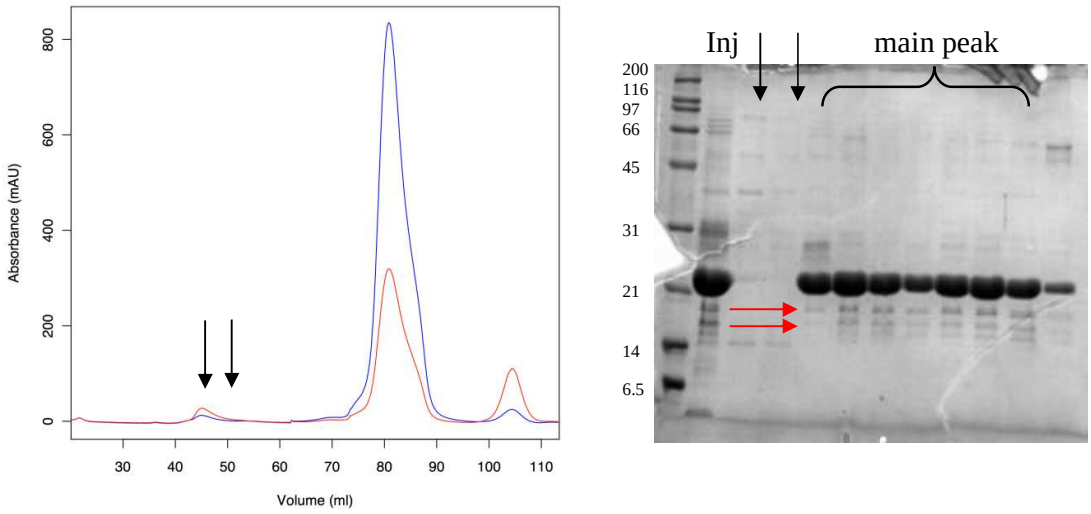


Figure 68: SEC of the shorter coiled-coil construct (res. 240-387). The main peak is less symmetric than RecN*cc*. Protein absorbance at 280 nm is reported in blue, while nucleotide absorbance at 254 nm is in red. The SDS-PAGE, however, showed that the purity of the protein was suitable for further biophysical experiments. SDS-PAGE gel with the main peak fractions is presented on the right. MWs are reported in kDa. Inj: injection.

SEC-MALLS analysis of the shorter coiled-coil construct (240-387) indicated that it behaved differently from RecN*cc* in solution and estimated the molecular mass to be 14.7

kDa, which was more consistent with a monomer in solution (Figure 69). The theoretical MW for this construct is 15.3 kDa, as calculated by ProtParam (Wilkins et al, 1999). A possible explanation for this behavior could be that the correct anti-parallel folding of the protein is being disrupted. In this shorter construct, α5 and part of α6 are missing and as a result a number of stabilizing intramolecular interactions, in particular the salt bridges between E212 and K329 and E222 and R319, are lost. The disruption of the correct interaction pattern may interfere with the proper anti-parallel fold and consequently lead to inappropriate formation of the dimerization interface.

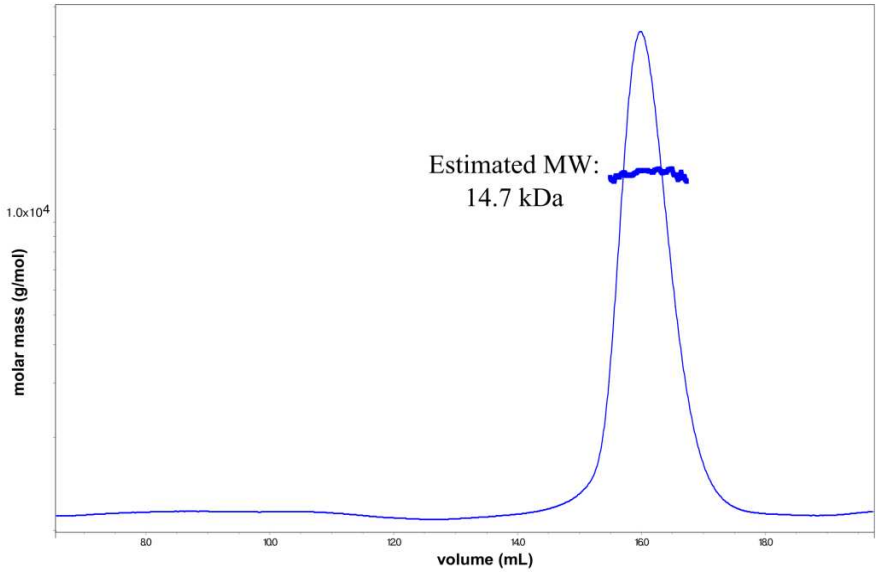


Figure 69: SEC-MALLS measurement of the coiled-coil domain of RecN, residues range 240-387. This construct elutes as a monomer, indicating that the dimerization interface is no longer functional. Thin line: refractive index recorded; thick line: averaged MW estimation from different scattering angles.

6.7. Deletion of the hydrophobic interface favors a monomeric arrangement of RecN*cc*

In order to characterize the role of this newly identified dimerization interface in the oligomerization of full-length RecN, a construct in which the dimerization domain (residues 237-291) was deleted was designed giving rise to a deletion mutant (RecN*Δdd*) (Figure 70). This construct was also cloned into pET151-TOPO after PCR amplification of the gene encoding for this construct. The primers used for PCR are listed following Table 5 (see Materials & Methods section).

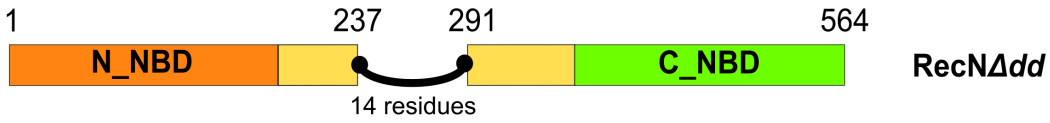


Figure 70: Schematic representation of the deletion mutant of RecN. Conserved motifs and domains are colored according to the scheme proposed for the other constructs designed as single domains of RecN.

6.7.1. Purification of RecNΔ*dd*

RecN*Δdd* was expressed and purified according to the protocol used for RecN*head* (Pellegrino et al, 2012b). The same peptide linker (amino-acid sequence ESSKHPTSLVPRGS) used for the design of the RecN*head* was used to fuse the N- and C-terminal domains. A two-step purification (IMAC and SEC) was established in order to obtain a pure protein sample (Figure 71 and 72) in order to perform SEC-MALLS (Figure 73), DLS and SAXS experiments.

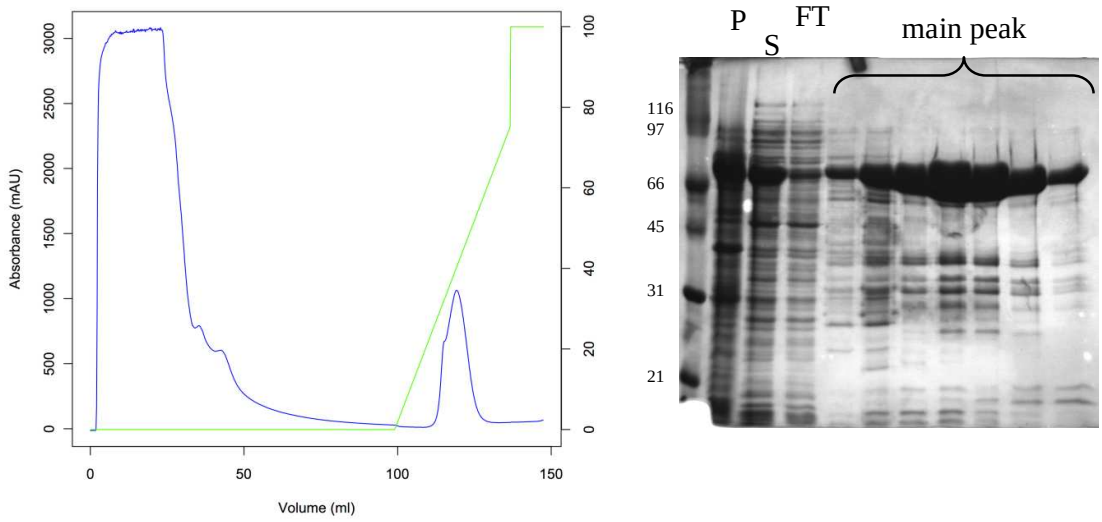


Figure 71: Affinity chromatography profile of RecN*Δdd*. A gradient of buffer B (50mM Tris-HCl pH 8, 300mM NaCl, 5mM $MgCl_2$) was used for protein elution from a Ni-NTA column. The main peak was eluting at 40 % of buffer B, corresponding to 205 mM imidazole. Protein absorbance at 280 nm is shown in blue. SDS-PAGE gel with the main peak fractions is presented on the right. MWs are reported in kDa. P: pellet; S: soluble fraction; FT: flow-through.

SEC step was performed on a Superdex 200 column (GE Healthcare) in order to compare the resulting elution volume with the other purified constructs. RecN*Δdd* in solution adopts an intermediate behavior in between the RecN*head* and the full-length RecN. The single symmetric peak resulting from SEC elutes at 12.5 ml, in between RecN (10.8 ml) and

RecN*head* (14 ml). DLS was performed to check the homogeneity of the sample. The pdI was found to be 0.073 indicating that the sample was homogeneous and suitable for further analysis by SAXS.

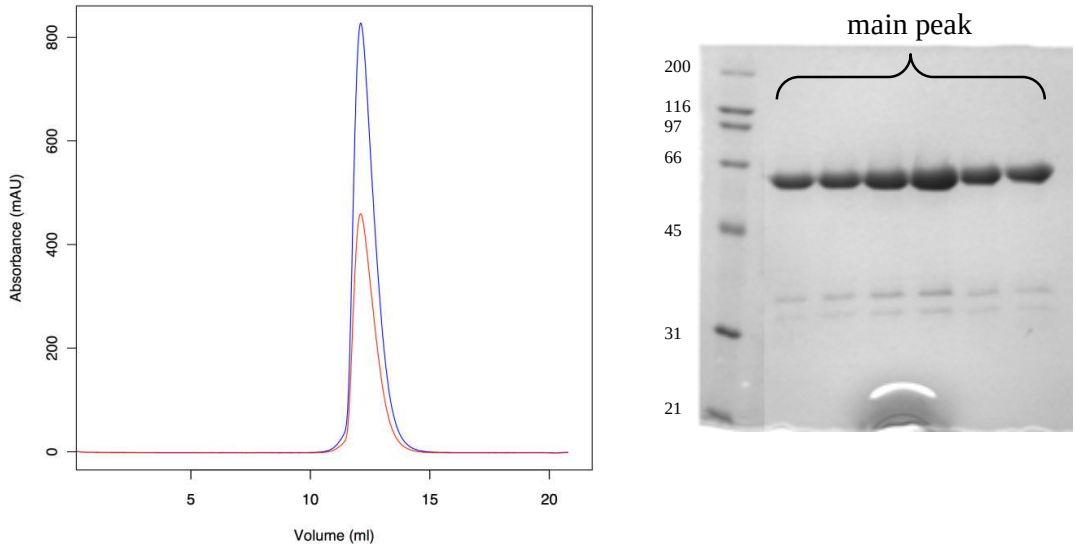


Figure 72: SEC profile of RecN*Δdd* in the same buffer condition as for RecN*head* domain protein. Protein absorbance at 280 nm is reported in blue, while nucleotide absorbance at 254 nm is in red. SDS-PAGE gel with the main peak fractions is presented on the right. MWs are reported in kDa.

6.7.2. RecNΔ*dd* elutes as a monomer from a GF column

SEC-MALLS analysis (Figure 73) revealed that this construct elutes as a monomer in solution, confirming our finding that RecN dimerization is induced by interactions between helices α6-α7-α8 of the coiled-coil domain.

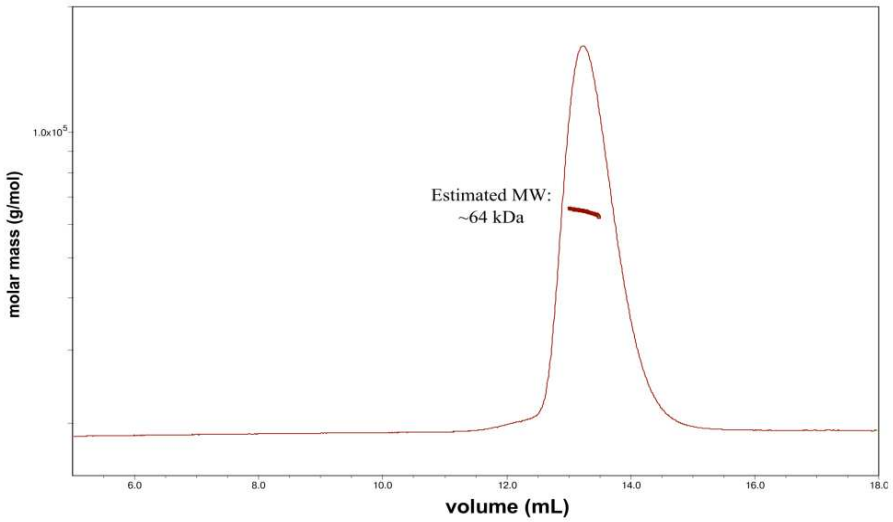


Figure 73: SEC-MALLS analysis of RecNΔ*dd* protein. The deletion mutant construct is a monomer in solution and eluted as a single peak. Thin line: refractive index recorded; thick line: averaged MW estimation from different scattering angles. The MW is reported in kDa.

6.7.3. Low resolution study of RecNΔ*dd* by SAXS

SAXS measurements were performed on RecNΔ*dd*. Data collected on ID14-3 (ESRF) showed a discrepancy at low angles between the highest and lowest concentration scattering curves (Figure 74), indicating some interparticle effects. This also made data analysis more complicated.

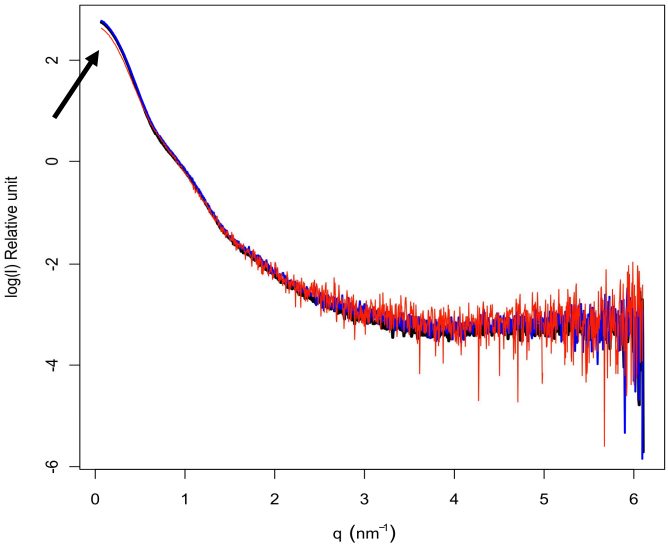


Figure 74: SAXS measurements carried out on RecN*Δdd* purified protein (described earlier). Three different concentrations (6.3, 3.0 and 1.4 mg ml^{-1}) were measured to check for any inter-particle effect. The three curves almost superpose, but the lowest (in red) concentration showed a different behaviour at very low angles (indicated by the black arrow), crucial for determination of the protein envelope (Putnam et al, 2007). The highest concentration is represented in black, the middle in blue.

The three scattering curves (sample concentrations: 6.3, 3.0 and 1.4 mg ml^{-1}) match quite well, except at low *q* values, which is very important for *ab initio* shape determination. An averaged model is shown in Figure 75. In addition the MW estimated by I_0 and the one estimated by Porod volume were very different: in the first case the value was closer to the size of a monome*r*, while the second value indicated a dimeric assembly, as calculated by AUTOPOROD (Petoukhov et al, 2007).

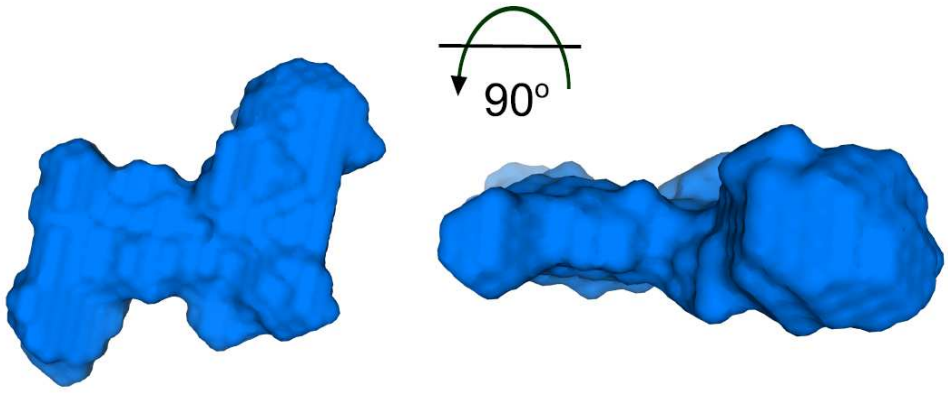


Figure 75: *Ab initio* model of RecNΔ*dd* resulting from envelope reconstruction using DAMMIN. The views, 90º apart, are represented. The envelope does not look like an elongated structure, as we speculated RecNΔ*dd* to be.

6.7.4. Crystallization and structural analysis of RecNΔ*dd*

Crystals of RecNΔ*dd* were obtained (Figure 76) by using the crystallization robot (HTX laboratory, EMBL Grenoble) and a full data set at a maximum resolution of 4 Å was collected. The crystal used for data collection was fished from the robot plates and flash-frozen directly on the beamline (ID23-2) adding 25 % glycerol to the reservoir solution. The crystal structure was solved as described in paper III (Pellegrino et al, 2012 submitted) by Mol.R. using RecN*head* and a modified coordinates file of RecN*cc* containing only residues 196-238 and 306-364 as search models.

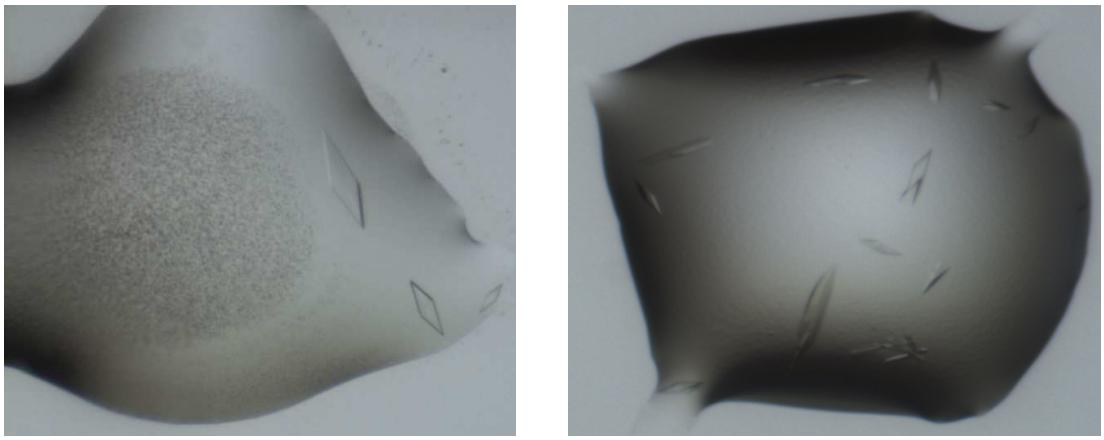


Figure 76: crystals were obtained in several conditions, with the best conditions being nº 18 of the Crystal Screen Lite (left) and nº 7 of the PEG-Ion Screen (Hampton Research). Crystals were tested on ID23-2 (ESRF) and diffracted to a maximum resolution of 4 Å.

The crystal structure of RecN*Δdd* solved at 4 Å resolution provided valuable information concerning the organization of the coiled-coil in the full-length RecN. The electron density map allowed to extend helices α5 and α10 (Figure 77), belonging to the coiled-coil domain, that are only partially present in the RecN*head* structure (Figure 45). These two very long and kinked helices make up the beginning of the coiled-coil domain (Figure 77), and confer extreme rigidity to the coiled-coil protrusions. Short loops between the helices contribute to the possible quaternary rearrangement, by providing small degrees of freedom to the entire RecN architecture.

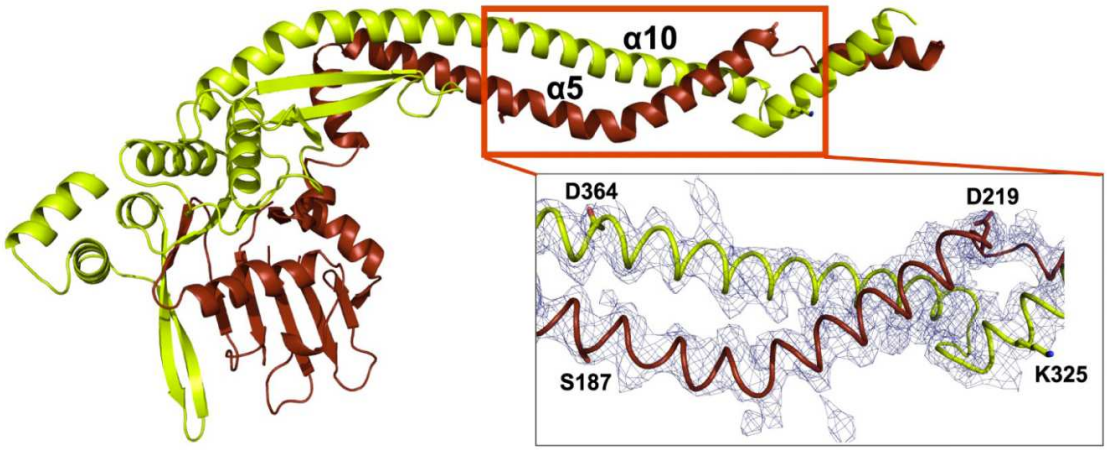


Figure 77: Crystal structure of RecNΔ*dd*. This construct contains an overlapping region with RecN*head* and RecN*cc* crystal structures. The beginning of the coiled-coil is formed by two long helices that protrude out from the globular domain (containing the NBD). In helix α5, a kink allows the helix to bend in order to wrap around α10 in an anti-parallel fashion. It is possible now to have a better understanding of the arrangement of the coiled-coil domain. In the magnification a detailed view of the electron density map contoured at 1.5 σ is shown.

6.8. Summary of the results on RecN*cc* and RecNΔ*dd*

The RecN*cc* crystal structure represents the first complete structure of a coiled-coil from an SMC or SMC-like protein. This is most likely related to the fact that the coiled-coil region in RecN is considerably shorter than the ones present in homologous proteins, such as Rad50, SMC and MukB.

RecN*cc* and RecNΔ*dd* were designed and purified to homogeneity as described in Materials & Methods section and in (Pellegrino et al, 2012a; Pellegrino et al, 2012 submitted). The coiled-coil domain was crystallized as a dimeric anti-parallel assembly (Figure 64), a finding that was also confirmed by analysis of the protein in solution (Figure 67). Dimerization is driven by the formation of a large hydrophobic area between two RecN*cc* monomers.

Removal of this interacting interface (to form the RecNΔ*dd* protein) favored the monomeric form instead of the dimeric assembly, as confirmed by the SEC-MALLS analysis (Figure 73) and the crystal structure solved by Mol.R. (Figure 77). Due to difficulties in data processing, the SAXS experiments performed on RecNΔ*dd* did not provide a reliable *ab initio* model of the protein in solution. It is indeed not clear whether disruption of the dimeric interface, leaving most of the coiled-coil intact, might lead to artefacts that produce oligomers

More accurate SAXS data collection and processing might provide more useful information regarding the structure of RecNΔ*dd* in solution.

Chapter 7:

A quasi-atomic model of RecN provides new insights into DSB repair

Summary of chapter 7

A quasi-atomic model of full-length RecN is proposed in this work, based on the three crystal structures obtained for RecN*head*, RecN*cc* and RecNΔ*dd*. RecN is very elongated, as speculated by previous studies. The constructed model fits very well with the *ab initio* model reconstructed from the Small Angle X-ray Scattering data, indicating that this model reliably describes the behavior of the protein in solution.

Résumé du chapitre 7

Un modèle quasi-atomique de la protéine RecN entière est proposé dans ce travail, à partir des trois structures cristallines de RecN*head*, RecN*cc* et RecN*Δdd*. RecN est très allongée, comme prédit par les études antérieures. Le modèle obtenu est en parfait accord avec le modèle reconstruit *ab initio* à partir des données de diffusion des rayons X aux petits angles. Ce modèle décrit donc correctement le comportement de la protéine en solution.

7.1. Building a reliable model of RecN

Detailed information regarding the structure of full-length RecN could now be derived from the analysis of the crystal structures of RecN*head*, RecN*cc* and RecNΔ*dd* (Figures 45, 64 and 77). A model of the entire protein (Figure 78) could be assembled using these three overlapping fragments. The structure of RecN*Δdd* was used as a template for superposition of the RecN*head* and RecN*cc* structures. The final RecN monomer is 190 Å long, with the coiled-coil that protrudes out of the NBD (Figure 78).

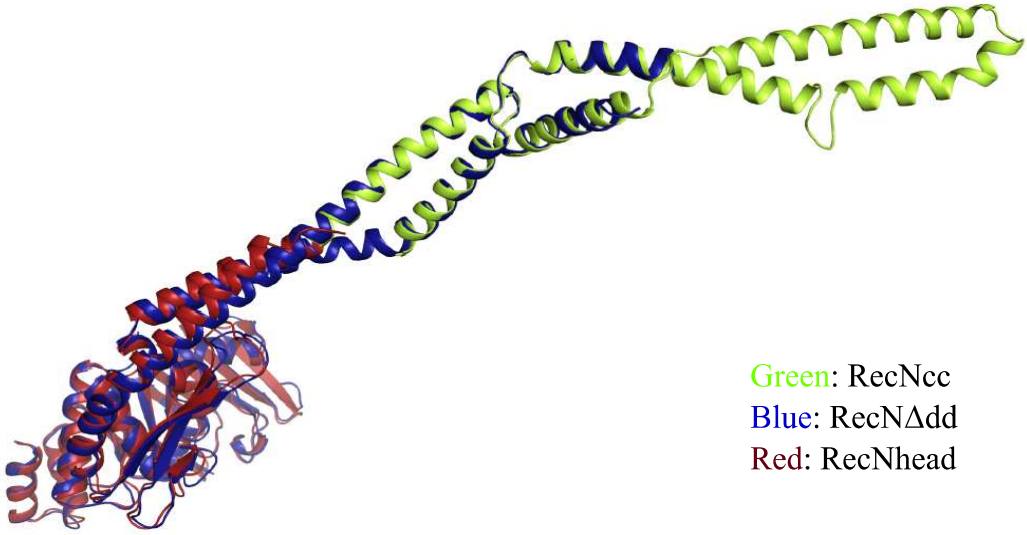


Figure 78: Monomeric state of RecN. This model was built by superimposing the three crystal structures obtained in this work and colored differently for clarity. The overlapping region in RecNΔ*dd* allowed to superimpose well the RecN*head* and RecN*cc* crystal structures and have a full picture of the very elongated full-length RecN.

A complete model of dimeric RecN (Figure 79) was produced using the model of monomeric RecN (Figure 78) and the dimeric interface of RecN*cc* (Figure 65). In this model, the two head domains are located ~300 Å apart at the two extremities of the dimer and are twisted 180˚ apart (Figure 79). Such an elongated structure of RecN is supported by literature (Graumann & Knust, 2009) and by all our biophysical experiments performed so far on RecN (chapter 3).

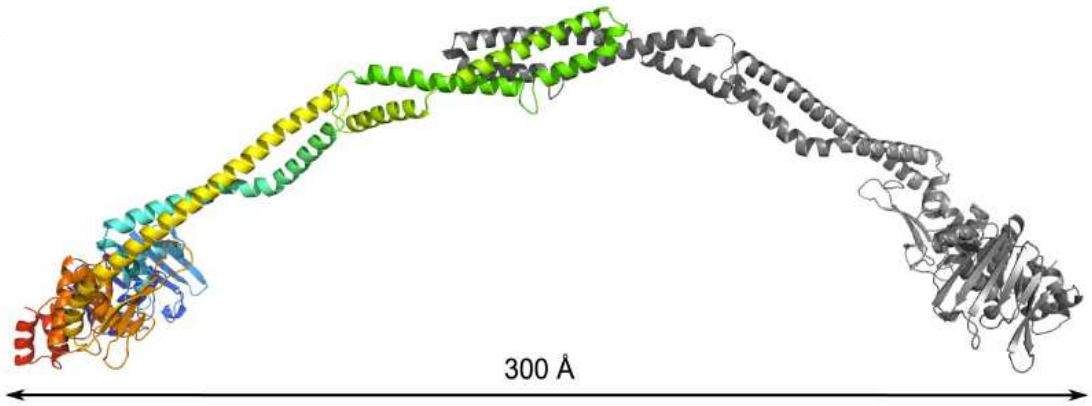


Figure 79: Dimeric assembly of DR RecN. One monomer is coloured in grey and the second in rainbow colours, for clarity. The distance between the two NBD is ~300 Å, confirming the D_{MAX} value of the *ab initio* envelope determined by SAXS (315 Å, § 3.2).

This model could then be compared to the SEC-MALLS and SAXS data previously performed (§ 3.3) that indicated that RecN adopts a dimeric assembly in solution. The experimental SAXS curve could now be fitted to the theoretical curve derived from the assembled, quasi-atomic model, as described in (Pellegrino et al, 2012 submitted). The fit was very good, with a χ^2 value of 3.44 over the whole scattering curve (Figure 80). The good fit of the two curves is indicative that the model describes well the structural organization of RecN in solution.

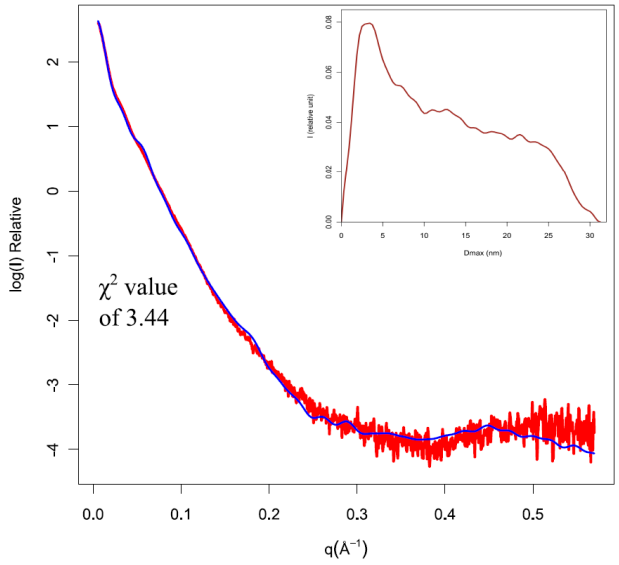


Figure 80: Overlay of the experimental (in red) and theoretical (in blue) scattering curve of RecN. The theoretical curve was derived from the quasi-atomic model proposed in this chapter. The goodness-of-fit is reported as χ^2 value.

The dimeric model of RecN is shown in Figure 80. A rather elongated structure (approximately 300 Å) confirmed the hypothesis found in the literature (Graumann & Knust, 2009) and was in agreement with the experiments performed so far on DR RecN (chapter 3).

Overlay of the quasi-atomic model of RecN with the *ab initio* envelope (Figure 81) obtained by shape reconstruction from the SAXS data shows that these two models are in very good agreement. Movements within the more flexible regions, such as the loops interconnecting the α-helices of the coiled-coil domain or the linker between the NBD and the anchor motif, could allow the NBD to move a little with respect to the entire structure.

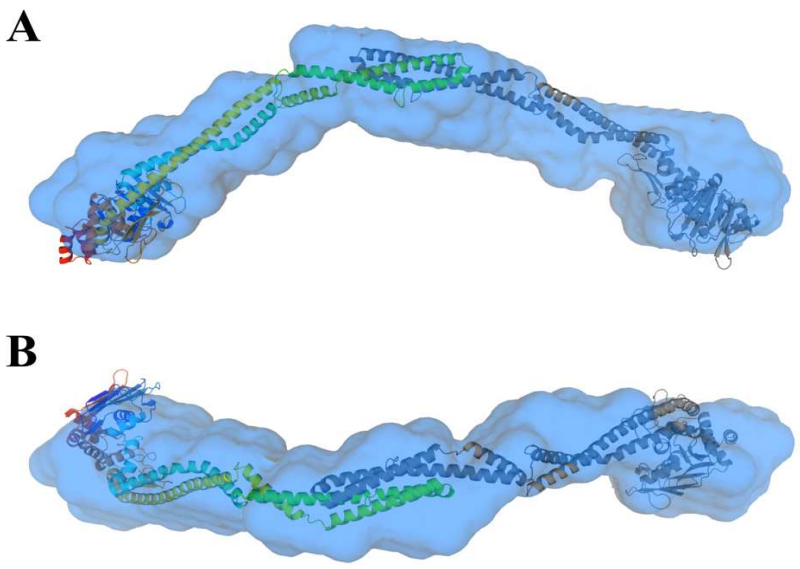


Figure 81: The proposed model of RecN reflects the behavior of the protein in solution. *Ab initio* model obtained after SAXS measurements (Figure 34) was superimposed to the quasi-atomic model of RecN built from the crystal structures of its constitutive domains. The two head domains fits with the bulky regions at each side of the *ab initio* envelope.

7.2. ATP-induced structural re-organization

DLS measurements were performed on the different full-length RecN mutants and compared with the data obtained for the wild-type protein. The full-length RecN displayed the same hydrodynamic radius (r_h) both in presence and absence of 2mM ATP. The Walker A and B mutants, in contrast, showed a different behavior than the wild-type and their hydrodynamic radii increased upon addition of ATP. Moreover, in the absence of ATP, these mutants also showed a smaller hydrodynamic radius compared to wild-type RecN (Table 7). At present, it is not clear what this is due to. No changes in the hydrodynamic radius of

RecN*Δdd* were observed upon ATP addition, in conformity with the wild type RecN and RecN*head* domain.

RecN	**DLS (w/o ATP) (r_h)**	**DLS (with ATP) (r_h)**
$RecN^{WT}$	10.22 nm (0.156)	10.26 nm (0.137)
$RecN^{E472Q}$	7.5 nm (0.31)	9.5 nm (0.22)
$RecN^{K67A/E472Q}$	7.5 nm (0.29)	9.5 nm (0.44)
$RecN^{K67A/E472A}$	n/a	n/a
$RecN^{K67A}$	7.6 nm (0.3)	8.5 nm (0.23)
$RecN^{E472A}$	8.40 (0.33)	12.56 (0.2)
$RecN^{D471A}$	7.45 (0.14)	18.01 (0.25)
RecN*head*^{WT}	3.2 (0.25)	3.2 (0.1)
RecN*head*$^{K67A/E472Q}$	4.0 (0.13)	4.2 (0.13)
RecNΔdd	4.66 (0.07)	4.60 (0.22)

Table 7: List of the hydrodynamic radii prior and after addition of 2 mM ATP. Poly-dispersity value (pdI) estimated by DLS is reported in parentheses.

7.3. Summary of RecN modeling

The quasi-atomic model of RecN, built from the three single crystal structures of its constitutive domains, revealed that RecN is an elongated, dimeric protein with its NBDs located at each of its extremities, almost 300 Å apart. This model is consistent (Figure 81) with our SAXS data obtained in solution on a sample in high salt (1 M NaCl) and shown initially in Figure 35. The dimeric assembly of RecN provided the first complete model for an SMC-like protein and is of interest also for the understanding of the role of RecN. In a model in which the two NBDs are located at either ends of the dimer it is unclear how such an assembly might interact with DNA to trigger DSB repair.

The model we proposed is very elongated and displays limited flexibility located principally at the level of the loops connecting the different helices of the coiled-coil domain. This could be a reason why crystallization of full-length was not successful. However, our quasi-atomic model supported by our SAXS study represents a very good approximation of the architecture of RecN. Further SAXS studies in the presence of ATP/ADP/AMP-PNP

could provide additional information regarding the conformational re-arrangements that occur upon nucleotide-binding.

Chapter 8:

Biochemical activities of RecN, RecNhead and their respective mutants

Summary of chapter 8

Biochemical studies were performed on full-length RecN and RecN*head*, since this last construct was designed so as to conserve the integrity of the NBD. RecN was shown to have both ATPase and DNA end-joining activities, confirming previously work on RecN and homologous proteins, such as Rad50. DNA-binding activity was also assayed and a ladder-like band-shift pattern was observed suggesting that RecN might form oligomers along the dsDNA. In contrast, RecN*head* domain displayed strongly reduced activity for both activity assays and a very different DNA binding pattern.

Mutants of RecN and RecN*head* domain that had previously been expressed and purified were also characterized biochemically. ATP hydrolysis, DNA end-joining and DNA band-shift assays were performed and the results of the two different constructs compared.

Résumé du chapitre 8

Des études biochimiques ont été effectuées sur RecN entiere et RecN*head*, puisque cette construction avait été conçue afin de conserver l'intégrité du domaine de fixation de nucleotide. Nous avons démontré que RecN possède à la fois une activité ATPase et une activité de liaison d'ADN confirmant ainsi des travaux précédant sur RecN et des protéines homologues, telle que Rad50. L'interaction de RecN avec de l'ADN double brin a également été testée, produisant plusieurs bandes de tailles variables (telle une échelle) sur le gel, ce qui suggère que RecN pourrait former des oligomères le long de l'ADN double brin. Le domaine RecN*head* quant à lui affichait des activités ATPase et liaison de l'ADN fortement réduites et un mode d'interaction avec l'ADN double brin très diffèrent.

Des mutants de RecN et RecN*head* qui avait précédemment été exprimés et purifiés ont également été caractérisés au niveau biochimique. L'hydrolyse de l'ATP, la capacité à joindre deux extrémités d'ADN et des tests d'interaction avec l'ADN ont été effectués et les résultats des deux constructions furent comparés.

8.1. ATP hydrolysis activity is disrupted in RecN*head* domain

A weak ATPase activity was detected for full-length RecN and it was shown to be in the range of what is reported in the literature (Pellegrino et al, 2012 submitted). The Malachite Green (Baykov et al, 1988) method was used for quantifying ATP hydrolysis and a detailed protocol is described in the section of Materials and Methods.

RecN*head* domain was designed in order to reconstitute the active NBD, typical of SMC and SMC-like proteins (Hopfner et al, 2000; Lammens et al, 2011; Lowe et al, 2001) and, we therefore expected the ATPase activity to be retained in such a construct. The RecN*head* domain, however, showed a severely reduced activity (Pellegrino et al, 2012 submitted), suggesting the coiled-coil region is essential for the *in vitro* ATPase activity (Figure 82 and Table 8).

Mutagenesis is a powerful tool for dissecting the ATP-hydrolysis mechanism and its influence on the structural rearrangement of SMC-like proteins. The activities of Walker A and B mutants were also measured and compared to the activities of their respective wild-type proteins. For both RecN and RecN*head*, single and double mutations of K67 (Walker-A) and E472 (Walker-B) resulted in a severely reduced ATPase activity. Exception comes from the E472Q mutation that leads to a 2-fold increase in the ATPase rate of RecN and a dramatic increase (approximately 25-fold) in the activity of RecN*head* domain (Figure 82). Such a mutation is expected to mimic a transition state and therefore facilitate head-head engagement, but block ATP hydrolysis. Previous studies have indeed reported that this mutation leads to impaired hydrolysis activity of SMC proteins (Lammens et al, 2004). In RecN, however, mutant E472Q displays an increased rate of ATP hydrolysis (from ~2-fold in RecN to ~25-fold for RecN*head* domain). This may be an additional indication that the nucleotide-binding pocket of RecN is slightly different from those of other SMC and SMC-like proteins.

Furthermore we tested the ATPase activity of RecN and RecN*head* in the presence of an excess of dsDNA substrate (50-mer). An increased activity was observed for RecN*head* (Figure 82), while in the case of full-length RecN, its ATPase activity was unchanged (Table 8). Low salt conditions were used in order to stimulate DNA binding, which would otherwise have been inhibited if the high salt conditions, required for RecN solubility, had been maintained.

	RecN	**RecN*head***
Wild-type	351.7 ± 19.8	49.7 ± 14
K67A	101.1 ± 13.6	27.0 ± 7.6
E472A	60.4 ± 20.9	50.5 ± 1.8
E472Q	760.0 ± 32.4	1206.3 ± 101.9
K67A / E472A	22.9 ± 5.0	20.3 ± 2.9
K67A / E472Q	59.8 ± 12.2	4.4 ± 2.5
D471A	349.3 ± 9.84	n/a
Wild-type in low salt buffer	258.7 ± 24.5	53.9 ± 6.65
WT + dsDNA 50mer	157.2 ± 13.8	197.7 ± 33.2

Table 8: ATP hydrolysis rates of RecN and RecN*head* are displayed as (nmol P_i released / min). Activities of mutants of both protein constructs are also shown.

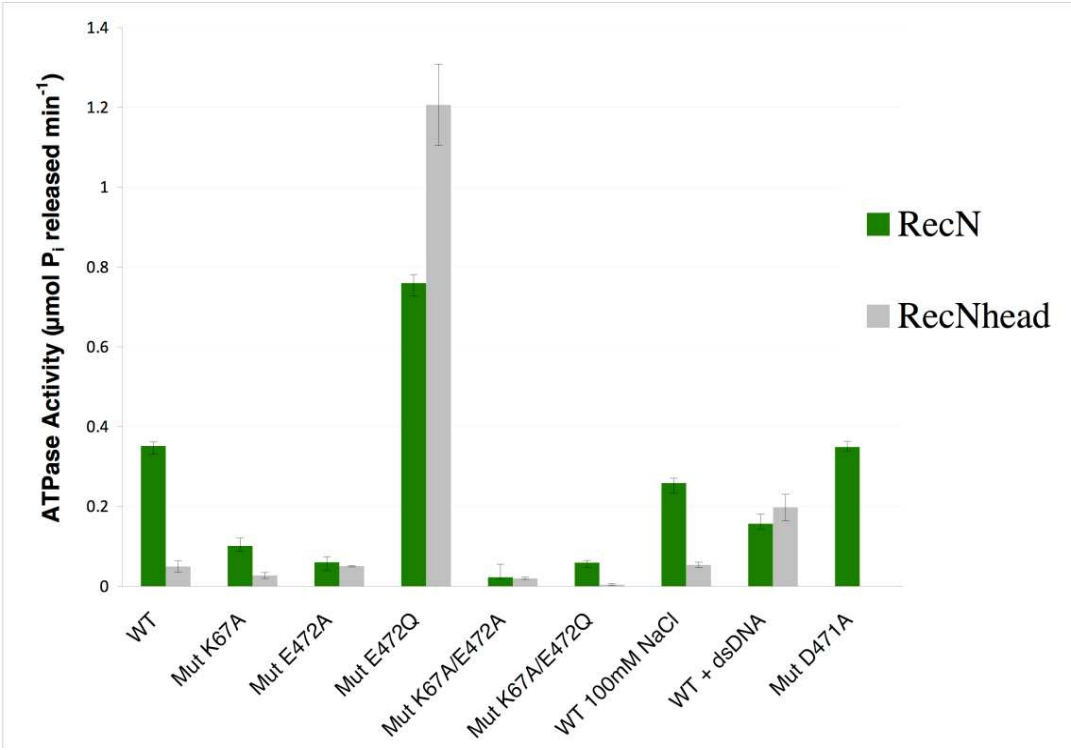


Figure 82: ATP hydrolysis activity of RecN, RecN*head* domain and their respective mutants. The RecN*head* domain showed reduced activity compared to RecN. The mutants (K67A, E472A, K67A/E472Q, K67A/E472A) of both construct displayed reduced or abolished activity, as reported in the previous literature. The mutant E472Q, in contrast, showed an increased ATPase rate where, in the literature, is reported to inhibit the activity.

8.2. DNA-end joining activity

Previous work has shed light on the possible functions of RecN in cells and, particularly, in DNA damage recognition. DNA end-joining activity has recently been reported for DR RecN (Reyes et al, 2010). In this study, RecN was shown to promote the tethering of two DNA ends for subsequent ligation by a DNA ligase. The protocol was fully described in (Reyes et al, 2010) and, adapted for our specific case, in (Pellegrino et al, 2012 submitted).

RecN displayed DNA end-joining activity (Figure 83), while as in the case of the ATPase activity, RecN*head* was largely inactive (Figure 84) probably as a result of the loss of the coiled-coil domain. RecN mutants showed reduced activity and a behavior that was largely reminiscent of that of RecN in the presence of the non-hydrolysable analogue of ATP, ATPγS (Reyes et al, 2010).

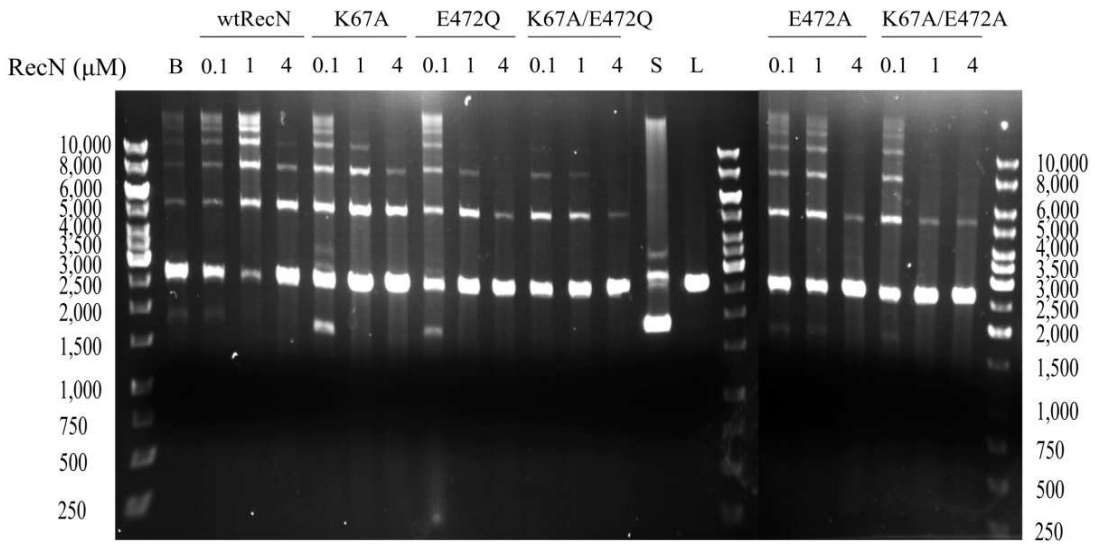


Figure 83: DNA end-joining activity of RecN protein and its mutants. pUC19 plasmid was linearized using the restriction enzyme SspI, which creates blunt DNA ends, and used as substrate for performing the *in vitro* assay (3 μM bp). Three concentrations (0.1, 1, 4 μM) are reported, but a broader range of concentrations were tested in order to have a better understanding of the effect of concentration on the DNA end-joining activity. MWs are reported in kDa. B: blank; S: supercoiled; L: linearized.

Experiments performed on Walker-A and B mutants of RecN*head* revealed that mutated RecN*head* domains showed restored ability to induce DNA end-ligation (Figure 84), but the activity appeared to be less regulated and yielded very high-molecular weight DNA species (Pellegrino et al, 2012 submitted). We speculate that these mutations affect the release of the nucleotide from the ATP-binding pocket and thus favor ATP-dependent dimerization, which seems to be essential for this DNA-end-joining activity.

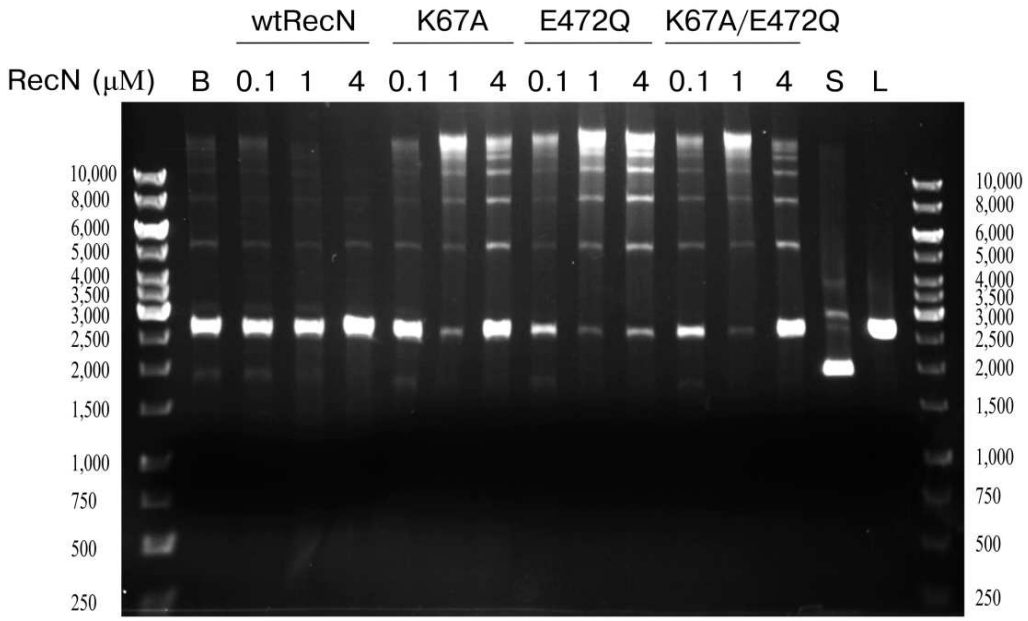


Figure 84: DNA end-joining activity of RecN*head* domain and its mutants. pUC19 plasmid was linearized using the restriction enzyme SspI, which creates blunt DNA ends, and used as substrate for performing the *in vitro* assay (3 μM bp). Three concentrations (0.1, 1, 4 μM) are reported, but a broader range of concentrations were tested. The mutants displayed restored activity compared to the wild-type RecN*head*, suggesting that the mechanism could not be related to ATP hydrolysis activity. MWs are reported in kDa. B: blank; S: supercoiled; L: linearized.

RecN was previously shown (Reyes et al, 2010) to have no preference for the dsDNA substrate. This finding was also confirmed by our DNA end-joining experiments performed on RecN using 5'-overhang (created by EcoRI, Fermentas) and 3'-overhang dsDNA (created by SacI, Fermentas). Figure 85 represents a DNA end-joining assay performed on RecN using the 3'-overhang dsDNA.

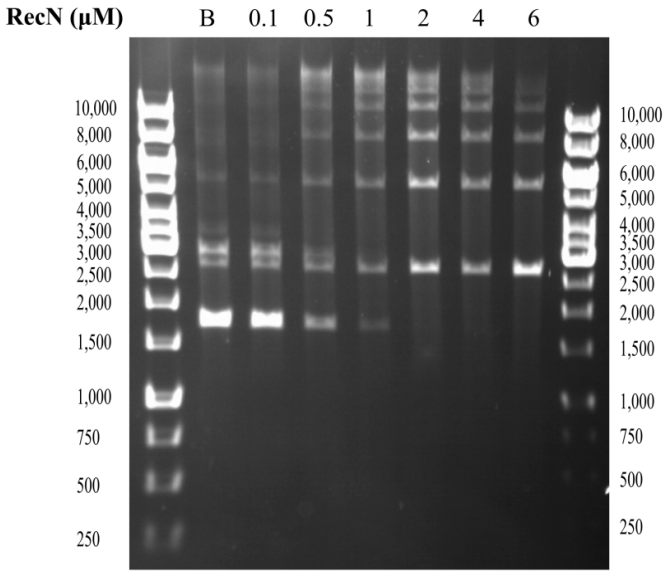


Figure 85: DNA end-joining assay performed on RecN wild-type protein, using 3'-overhanged dsDNA previously digested with SacI. The pattern of DNA end-joining activity is similar to what seen previously using dsDNA with blunt ends. MWs are reported in kDa. B: blank.

8.3. Band-shift assay for detection of RecN binding to DNA

Our in vitro studies (§ 8.2) had previously highlighted the capacity of RecN to stimulate intermolecular ligation of DNA linear molecules (Figure 83). In order to further characterize RecN's ability to interact with DNA we performed Electrophoretic Mobility Shift Assay (EMSA) as described in Materials & Methods. The results are presented here and in (Pellegrino et al, 2012 submitted). 50-mer dsDNA was used and 5 % non-denaturing acrylamide gels were run in order to detect possible shifts of the DNA. Detection of DNA bandshifts with full-length RecN proved to be very difficult since the protein alone did not enter such gels. This is most likely a consequence of its elongated shape, which prevents the protein from running smoothly into the gel matrix. Different native gel systems (acrylamide, agarose and mixed acrylamide/agarose) and buffers (TBE pH 8.3 and Tris-HCl pH 8.0) were tried in order to improve the protocol, but at best, only a small fraction of the protein entered the gels. In these conditions it was impossible to carry out a quantitative study of RecN binding to DNA.

RecN*head* domain, in contrast to full-length RecN, is more globular (chapter 5) and entered the native gels without the problems experienced for RecN protein. In the conditions tested, notably in 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 5 mM $MgCl_2$, wild-type

RecN*head* domain displayed a clear capacity to bind dsDNA (Figure 86). The presence of ATP did not significantly affect DNA binding.

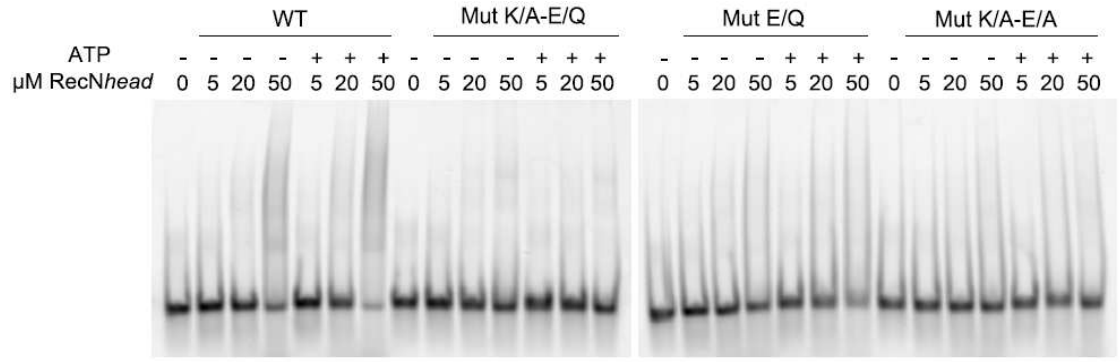


Figure 86: EMSA performed on RecN*head* domain. A DNA smear is visible for the wild-type protein while for the mutants did not show a detectable binding. Experiments were carried out in presence or absence of 1 mM ATP and using a 50-mer dsDNA with a fluorescein inserted on the middle.

Full-length RecN was also found to interact with dsDNA. However, more than one form was favored and a ladder-like pattern was generated (Figure 86). This result suggested that RecN could form oligomers upon DNA interaction, leading to the formation of different forms of DNA:protein complexes. The mechanism underlying the DNA-binding of RecN, however, remains an open question and data obtained so far do not allow us to clearly define how the process takes place. Our results, nonetheless, suggest that RecN binds to dsDNA non-specifically (Figure 87) most likely by sensing DNA ends rather than a particular sequence and shows a tendency to polymerize on the DNA. This ability to form protein filaments along the DNA seems to be regulated by ATP, since the presence of ATP in the EMSA assay interfered with the formation of a ladder-like pattern.

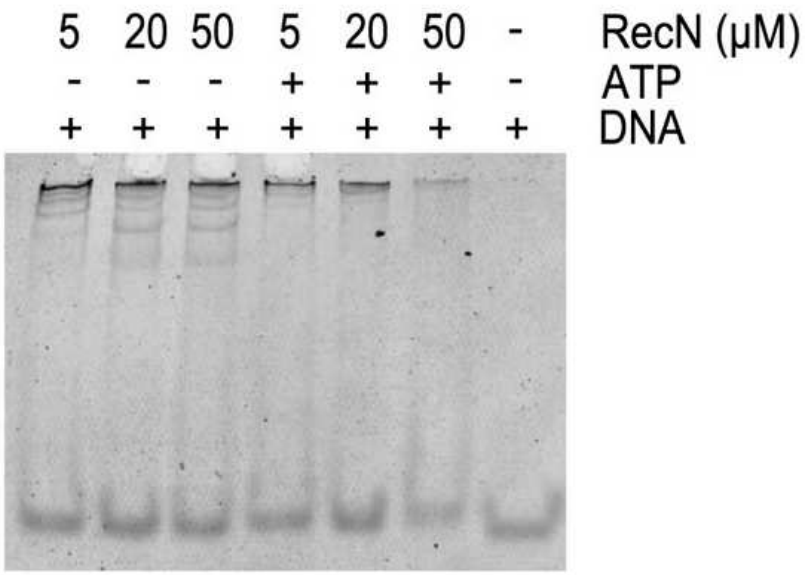


Figure 87: EMSA on the full-length RecN protein. The characteristic pattern is indicative of the presence of multiple forms of protein:DNA assemblies, which ultimately lead to a ladder-like profile on the native gel. In presence of ATP the effect seems to be reduced, suggesting that in presence of nucleotide RecN might have a different mechanism of interaction with DNA.

8.4. Summary of the biochemical characterization of RecN and RecN*head*

Our biochemical data confirmed what was proposed in the literature for proteins homologous to RecN. The ATP hydrolytic activity is weak (approximately 350 nmol ATP hydrolyzed min^{-1}) as reported for ABC and ABC-like proteins (Hopfner et al, 2000). DNA end-joining activity of DR RecN was also confirmed (Reyes et al, 2010). RecN*head* domain was also tested and its behavior was very different compared to full-length RecN. RecN*head* displayed a strongly reduced ATPase activity and no DNA end-joining activity.

Extensive study of the different mutants produced both for RecN and RecN*head* proteins led to the conclusion that the mutated residues were crucial for ATP-binding and hydrolysis. Only the E472Q mutant was found to behave in an unexpected way: this mutant displayed a remarkably increased ATP hydrolysis rate, instead of a reduced activity as reported in the literature (Lammens et al, 2004). The reason for this is not clear, since in SMC and SMC-like proteins this mutation, that mimics a *transition state*, reduces considerably the ATPase activity.

The DNA end-joining assay was reproduced (Reyes et al, 2010) and the results were very similar. RecN*head* domain was also assayed and it was found to be deficient in DNA end-joining activity. The coiled-coil domain thus appears to be needed for RecN's role in

stimulating the oligomerization of the DNA substrate. Mutants were also tested both for RecN and RecN*head* proteins: in the first case (Figure 84) the results were comparable to that seen when ATPγS was added instead of ATP (Reyes et al, 2010). Monomers and dimers were the most favored species especially at high protein:DNA ratios. However for the RecN*head* mutants, the presence of the altered ATP-binding site allowed an increased head-head engagement and this stimulated the production of higher order MW DNA bands.

In conclusion the mutations introduced in the functionally conserved Walker-motifs of RecN led to a decrease of the activity compared to the wild-type protein (Figure 83). Removal of the coiled-coil domain caused the loss of the regulation of DNA end-joining activity, as demonstrated by the marked decrease of the multimerization pattern seen by RecN*head* domain. On the other hand, mutations introduced in RecN*head* protein led to increased formation of higher order DNA bands, stimulated by more favorable head-head engagements.

In order to further dissect the mechanisms of DNA binding and end-joining by RecN, it would be particularly interesting to prepare constructs carrying mutations in the residues actively involved in DNA-binding, once identified. The use of FRET technology could also help in understanding the preference of RecN for DNA substrates, labeling different ends with different dyes and then measuring the fluorescence transferred between the two labels. Crystal structures of RecN or RecN*head* domain in complex with DNA would of course be of major relevance for a more detailed understanding of the interactions occurring during DSB recognition and repair.

Chapter 9:

Discussion

Summary of chapter 9

In this chapter an extensive discussion of the results obtained during this work is presented. The model of the full-length RecN built from the crystal structures of RecN*head*, RecN*cc* and RecNΔ*dd* and the Small Angle X-ray Scattering envelope is discussed.

Biochemical results are also discussed and a model of the mechanism underlying the Double Strand Break recognition step in DNA repair is proposed. Furthermore, future experiments are suggested in order to gain a deeper understanding of the role of RecN in this essential cellular process of DNA repair.

Résumé du chapitre 9

Dans ce chapitre, une discussion approfondie des résultats obtenus au cours de ce travail est présentée. Le modèle de la protéine RecN entière construite à partir des structures cristallines de RecN*head*, RecN*cc* et RecNΔ*dd* et de l'enveloppe obtenue par diffusion des rayons X aux petits angles est discuté.

Les données biochimiques sont également discutées et un modèle du mécanisme de reconnaissance des cassures double brin lors de la réparation de l'ADN est proposé. Par ailleurs, des expériences complémentaires sont proposées afin d'approfondir notre compréhension du rôle de RecN dans cet important processus cellulaire qu'est la réparation de l'ADN.

9.1. Structural characterization of RecN

The crystal structures obtained in this work revealed that RecN consists of two major domains: a globular region containing the NBD, which we referred to as RecN*head* (Figure 45) and a coiled-coil region, which we referred to as RecN*cc* (Figure 64). The NBD is typical of ABC-like proteins, such as SMC or Rad50, and is involved in binding and hydrolysis of tri-phosphate nucleotides, particularly ATP. Comparative structural analysis of the head domain of RecN showed that the fold of the ATP-binding pocket was conserved. However, there were small differences in the orientation of the active site residues involved in binding and hydrolysis, even though the location of the Walker A and B motifs was relatively conserved (Figure 45). The signature sequence displayed a different orientation in RecN compared to Rad50 (Figure 45-B and C) with a ~60 degree rotation, which also greatly affected the positioning of the anchor motif (Figure 45-C) and its associated coiled-coil domain. The crystal structure of RecN*head* confirmed our previous low-resolution studies performed by SAXS (Figure 47), suggesting that the behavior in solution is very close to what was seen in the crystal.

The crystal structure of RecN*cc*, on the other hand, was found to form a stable ATP-independent dimer, through extensive hydrophobic interactions between the α6-α7-α8 helices of the two monomers (Figure 64). The coiled-coil in SMC proteins is thought to be very flexible and allow the NBDs of the two units constituting the functional dimer to interact upon ATP-binding. In RecN, in contrast, the extended dimerization region and the intrinsic architecture (Figure 64) of the coiled-coil domain confer rigidity to this rod-like structure. It is therefore very unlikely that two RecN monomers from a given dimer interact with each other via their head domains. This finding is also consistent with our SAXS data, which resulted in the reconstruction of an envelope that fitted well with the crystal structure obtained (Figure 67). Comparison of RecN*cc* with other crystal structures of SMC or SMC-like dimerization domains was discussed thoroughly in (Pellegrino et al, 2012 submitted).

The crystal structure of a deletion mutant of RecN (Figure 77), which we referred to as RecNΔ*dd*, allowed to better define the orientation and the organization of the coiled-coil, especially in the region which was missing in the crystal structures of RecN*head* and RecN*cc*. The resolution of the data (4 Å) was sufficient to trace with good approximation the main chain of the protein and several cycles of refinement allowed to describe the kink present on helix α5 needed to induce the torsion of the coiled-coil in order to form the characteristic anti-parallel architecture.

In addition to these crystallographic models, a low resolution SAXS study of a double mutant of RecN*head* domain, named RecN*head*$^{K67A/E472Q}$, also provided a model of the large-scale conformational rearrangements of the RecN*head* domain upon nucleotide binding (Figure 54). The envelope clearly showed the orientation and positioning of helices α5-α10, which correspond to the beginning of the coiled-coil domain. The helices pointed out in the same direction suggesting that the two NBDs interact in a head-to-tail orientation in order to reconstitute the two nucleotide binding pockets at their interface. However, the resolution of the envelope was not high enough to gain a full understanding of the local structural re-arrangements occurring at the level of the ATP-binding pockets. Higher resolution data, through improvement of the initial crystallization conditions (Figure 56), would provide a better understanding of the detailed molecular mechanism underlying nucleotide binding and hydrolysis by RecN.

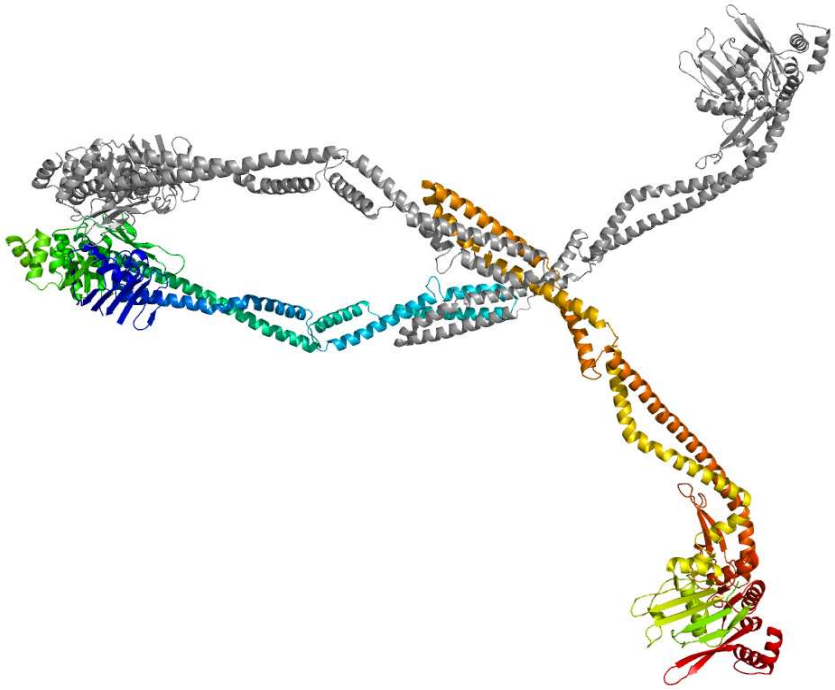


Figure 88: Possible model for the tetrameric form of DR RecN. ATP stimulates the interaction of two NBDs, as seen from the SAXS data on RecN*head*$^{K67A/E472Q}$; the other two NBDs, in contrast, are free to interact with the NBDs from nearby molecules or, alternatively, may induce dissociation of the first two head domains, upon ATP hydrolysis. This suggests a dynamic role of RecN in DSB repair.

All our structural information could be brought together in order to define a quasi-atomic model of the monomeric, dimeric and also tetrameric arrangements of RecN. RecN monomer and dimer were discussed thoroughly in this work and also in paper III (Pellegrino

et al, 2012 submitted), and led to the definition of a 300 Å long structure, which we proposed could bind DNA via each of its head domains, with the coiled-coil region acting as a spacer between the two DNA fragments. A tetrameric arrangement (Figure 88), however, could provide a possible explanation for the dynamics of DSB recognition and repair initiation by RecN.

9.2. ATP hydrolysis: a regulatory mechanism?

Full-length RecN was found to possess both ATPase (Figure 82) and DNA-end joining activities (Figure 83). Its weak ATPase rate was consistent with what was previously published for ABC-like proteins such as Rad50 (Holland & Blight, 1999; Hopfner et al, 2000). ATP binding and hydrolysis is crucial for modulation of the putative activity; therefore the ATP binding pockets have evolved to avoid uncontrolled ATP hydrolysis (Pellegrino et al, 2012 submitted).

Biochemical characterization of RecN*head* protein revealed that its ATPase activity was strongly reduced compared to full-length RecN, indicating that head-head engagement is very inefficient in the absence of the coiled-coil domain (Figure 82 and Table 8). This may result from the increased degree of freedom of the NBD in RecN*head* compared to full-length RecN, which would allow the head domain to adopt alternative conformations that may not be functional.

Mutations were introduced into both RecN and RecN*head* proteins in order to dissect the ATP binding and hydrolysis processes. Most of the mutants displayed very little or completely abolished ATPase activity, confirming that the mutated residues are actively involved in hydrolysis (Figure 82). The E472Q mutant, however, did not behave as previously reported in the literature (Hirano & Hirano, 2004; Lammens et al, 2004). This is a common mutation introduced into Walker-B motifs, which is suggested to favor head-head engagement and mimic a transition state, leading to the inhibition of the activity. In the biochemical studies we performed, the ATP hydrolysis rate measured for this particular mutant showed a remarkable increase. The unexpected trend of the E472Q mutant may result from the increased head-head engagement together with a local rearrangement of the nucleotide-binding pocket allowing for a neighboring residue (D471 would be a good candidate for such a role) to compensate for the E/Q mutation. Crystallographic data collected on crystals of this mutant did not reveal any remarkable differences in the ATP-binding pocket. Improvement of the diffraction quality (3.5 Å was the maximum resolution achieved) and co-crystallization

with different nucleotides might lead to a better in-depth view of the mechanism underlying such an increased hydrolytic activity.

The tight regulation of the ATPase activity of RecN may be disrupted by the introduced mutation. In DR, *recN* gene is constitutively expressed (Tanaka et al, 2004) and this could represent one of the reasons for keeping the basal hydrolytic activity of RecN low. In the presence of DNA the ATPase activity increases, by a yet unknown mechanism (Figure 82) (Pellegrino et al, 2012 submitted; Reyes et al, 2010; Sanchez & Alonso, 2005). DR engineering with the cited mutations might be used for *in vivo* studies in order to see whether these substitutions occurring at the Walker-motifs influence the entire DSB recognition mechanism, thus leading to different DR phenotypes.

9.3. DNA end-joining activity of RecN

DNA end-joining activity has been reported for human cohesins, yeast Rad50 and DR RecN proteins (Losada & Hirano, 2001; Reyes et al, 2010; Tomkinson et al, 2005). Activity was found to be independent of ATP hydrolysis (Tomkinson et al, 2005) and the studies performed on RecN and RecN*head* domain confirmed this finding (Figures 83 and 84). RecN did not display a particular preference for DNA ends (both overhang (Figure 85) and blunt ends were processed) suggesting that the protein might recognize topological features of DNA rather than specific sequences. We would therefore expect that residues in the putative DNA-binding site would interact preferentially with the DNA backbone. The results obtained from the DNA end-joining assay (chapter 8) showed that the DNA oligomerization pattern resulting from the activity of the RecN mutants was similar to experiments performed on the wild-type but in the presence of ATPγS, a non-hydrolyzable ATP analogue. In these experimental conditions monomers and dimers were the most favored species (Reyes et al, 2010). These findings suggest that the DNA end-joining is also promoted in the absence of an efficient ATP hydrolysis activity.

The mutated RecN constructs might therefore be unable to release the DNA and be recycled so as to promote the ligation of additional DNA molecules into high molecular weight forms. Wild-type RecN*head* domain, in contrast to its mutants, displayed a characteristic DNA-end joining pattern in which plasmid DNA monomers were the only species produced, suggesting that the head domain alone was unable to promote inter-molecular DNA ligation, while at the same time inhibiting re-circularization of the plasmid DNA (Figure 84). The presence of the coiled-coil therefore appears to play a pivotal role in the capacity of RecN to stimulate DNA end-joining activity (discussed in § 8.2).

These results indicate that head-head engagement, which is ineffective in RecN*head* but favored in RecN*head* mutants, is essential for the DNA end-joining activity of RecN. This finding is supported by the SAXS study of RecN*head*$^{K67A/E472Q}$ in which head-head engagement was found to be induced by introducing the double mutation K67A / E472Q.

9.4. DNA binding ability of RecN

DNA binding studies showed that RecN*head* was able to interact weakly with a dsDNA oligonucleotide (50-mer) (Figure 84), even though the RecN*head* protein was found to be deficient in DNA end-joining activity. This suggested that engagement of the head domains and a functional ATPase activity are not necessarily required for DNA binding. The DNA-binding ability of RecN*head* domain was also not improved after addition of ATP (Figure 86), confirming the hypothesis mentioned above. In addition to this finding, no detectable DNA binding was observed for the ATPase deficient mutants of RecN*head*, which means that single head domains interact more tightly with DNA than ATP-dependent dimers. A crystal structure of RecN*head* domain in complex with DNA would be a priority in order to understand the driving mechanism of DSB recognition.

The DNA binding pattern of full-length RecN, in contrast, reflected a controlled assembly on the DNA (Figure 85). Binding to DNA was observed as a ladder-like pattern and not as smear as in the case of RecN*head* domain, indicating that the coiled-coil may here again be playing a central role in regulating specific protein:DNA interactions. Here again, a crystal structure of RecN in complex with DNA would be of great benefit in order to gain further insight into the process of protein:DNA interaction. SAXS coupled to SANS experiments would also provide low-resolution information regarding the architecture and overall shape of protein:DNA complexes. SANS would be particularly useful to identify the region of RecN that interacts with DNA, by taking advantage of the difference in scattering of DNA compared to protein (method known as contrast variation).

Calculation of the electrostatic surface potential of RecN revealed that its surface is largely dominated by negatively charged residues, especially concentrated in the coiled-coil region (Figure 89). In contrast, a few patches of positive charges are detectable and all of them are located on the globular region, suggesting that the head domain represents a good candidate as a potential DNA binding domain (Figure 90). This would suggest a difference in the mechanism compared to SMC proteins, for which the hinge domain, involved in protein dimerization, has also been reported to be involved in DNA binding (Griese & Hopfner, 2011; Griese et al, 2010).

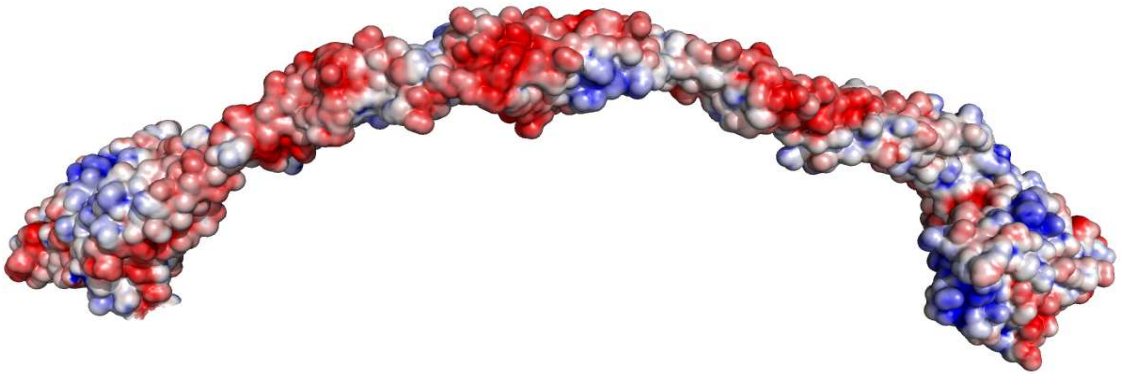


Figure 89: Electrostatic surface representation of dimeric RecN protein. The coiled-coil domain is mainly negatively charged, so it is unlikely to be the putative DNA binding site. This contrasts with SMC hinge domains, for which DNA-binding activity has been reported (Griese & Hopfner, 2011; Griese et al, 2010). Positive charges are represented in blue while negative are represented in red.

This positively charged patch consists of three arginine residues (R81, R120 and R133), two of which are highly conserved within the RecN protein family (Figure 22).

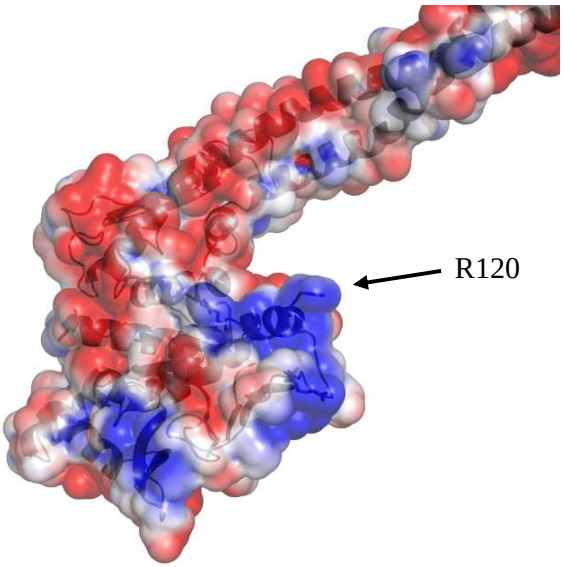


Figure 90: Positively charged patch on the surface of the globular head domain of full-length RecN. The R120 residue that points out to the solvent is shown as sticks and indicated by a black arrow. Positive charges are represented in blue while negative are represented in red.

These residues are respectively located on β7, α2 and the loop connecting α1 to β4 of the NBD (Figure 45) and all of them point out towards the solvent area and are therefore ideally positioned to interact with the negatively charged DNA backbone (Pellegrino et al, 2012 submitted) (Figure 91). Identification of the DNA-binding site (Figure 90 and 91) might

be dissected through specific mutations, but prior to this the DNA-binding conditions need to be improved in order to obtain reliable band-shifts.

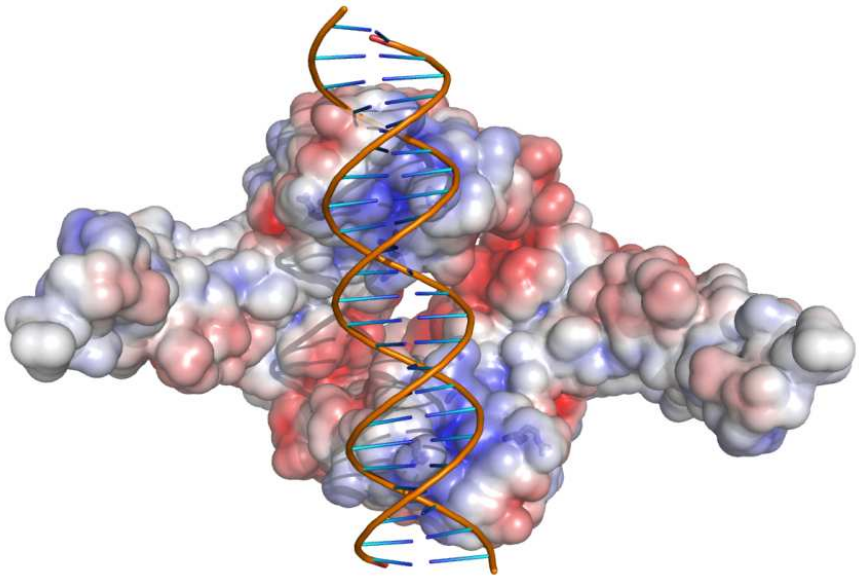


Figure 91: Model of DNA binding proposed on the basis of the results obtained in this work. DNA could find place at the top of the dimeric structure of RecN*head*, where the two R120 point out in the minor groove of DNA. Positive charges are represented in blue while negative are represented in red. DNA is represented as a cartoon.

Finally, we speculate that RecN*head* domain, through ATP-induced dimerization, might regulate the interaction of RecN with a molecule of double stranded DNA (Pellegrino et al, 2012 submitted) (Figure 91). The R120 belonging to each RecN*head* monomers nicely fits into two consecutive grooves of the DNA structure (Figure 92).

Alternatively, dimeric RecN*head* protein might bind simultaneously to two different dsDNA molecules (Figure 92A), which would then face each other adopting a yet unknown orientation.

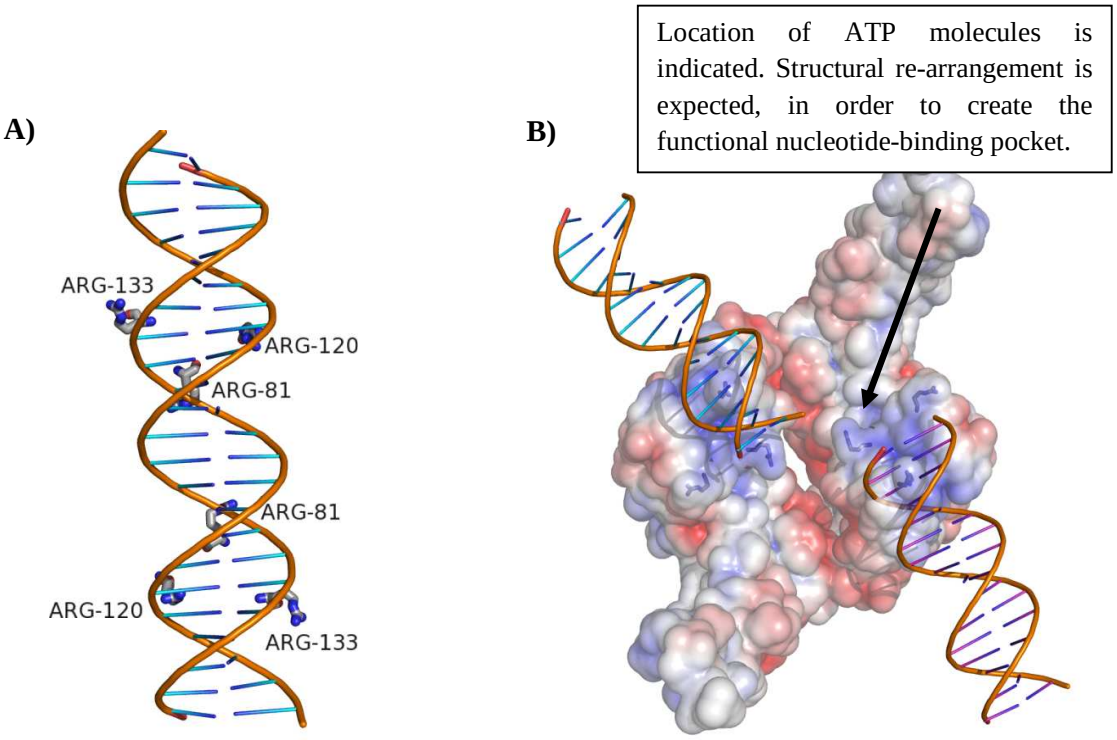


Figure 92: Representation of DNA interaction with RecN*head* domain. A) Detail of the conserved arginine residues, namely R81, R120 and R133, putatively involved into DNA-binding. B) Alternative model representing the RecN*head*$^{K67A/E472Q}$, in which nucleotide has been suggested to be trapped into the ATP-binding cleft (black arrow). Structural re-arrangement of the pocket is expected in view of the highly negative character of the pocket (from the top view of the cleft). Positive charges are represented in blue while negative are represented in red. DNA is represented as a cartoon.

9.5. DSB repair initiation

The mechanism of action of RecN is still only poorly understood, but in view of our structural and biochemical results we can speculate on how RecN may function in DSB recognition and repair (Figure 93). We propose that DSBs in the DNA are recognized by a single head domain of a RecN dimer. The second monomer would then interact with the homologous strand. In this way the coiled-coil domain acts as a spacer, maintaining the homologous strands far apart, before the DNA repair machinery intervenes. Binding of the single NBDs to DNA would then favor nucleotide (ATP) binding, leading to the formation of a dimer of RecN*head* domains between two dimeric RecN assemblies. Eventually, this would promote polymerization of RecN along the DNA.

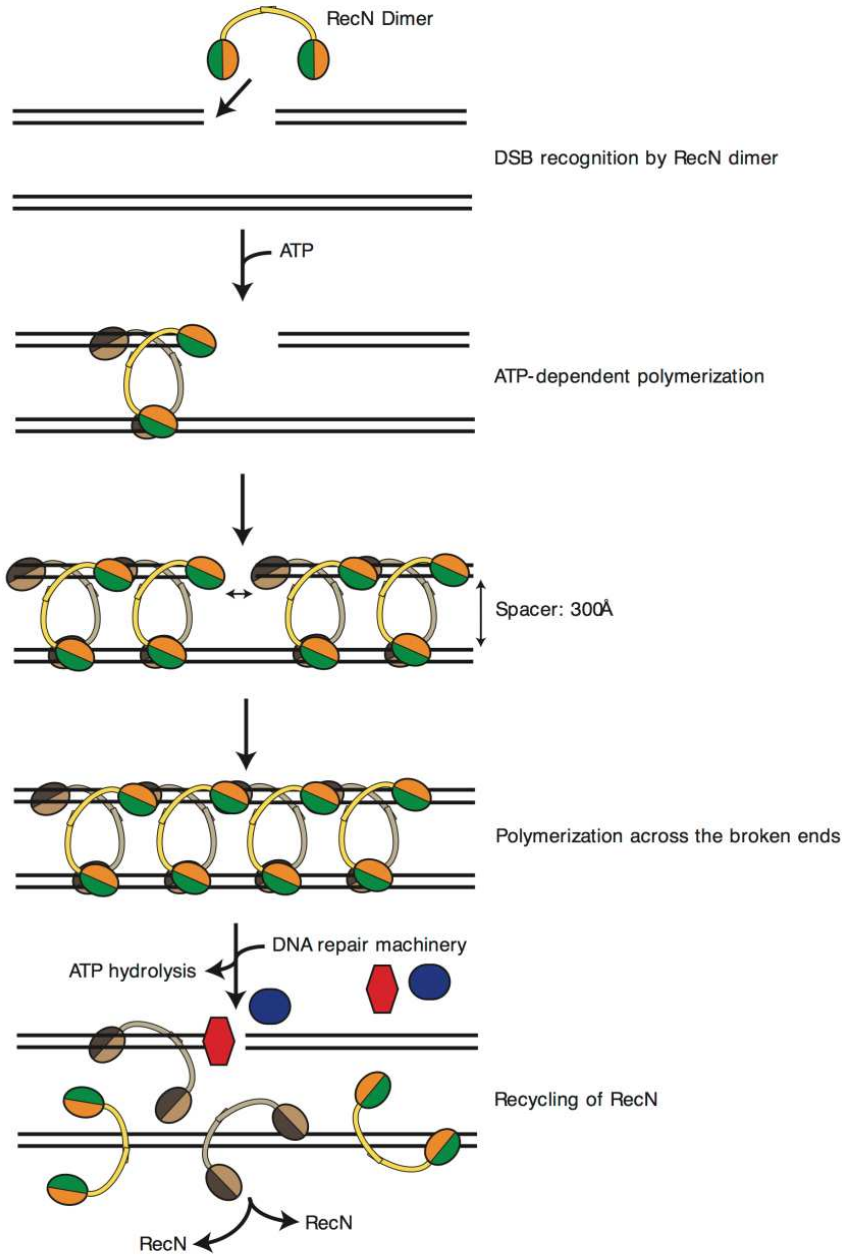


Figure 93: Model for RecN DSB recognition and HR initiation. RecN was speculated to bind both to damaged and "healthy" homologous DNA and thus induce the recruitment of the repair machinery on specific sites of DSB, as supported by *in vivo* studies on *B. subtilis*, where RecN was recruited to distinct DSB repair foci. Once RecN binds to the damaged DNA, the presence of ATP allow the formation of tetramers of RecN, which then polymerises along the DNA. This probably induces a variation in the DNA conformation which recruits the DNA repair machinery to restore the integrity of the DNA.

Once RecN has assembled on either sides of the DSB, the terminal NBDs associated with the DNA ends may also engage so as to tether the two broken ends (Figure 93). The arrival of the

DNA repair machinery would then stimulate ATP hydrolysis and disassembly of RecN from the DNA to be recycled for a new round of repair. This model is similar to the one proposed for cohesin proteins (Hirano, 2006) but the chain-like structural organization of RecN revealed by our quasi-atomic structure of RecN is the first experimental evidence supporting such a model. Electron microscopy (EM) studies might represent a suitable tool for investigation of the polymerization state of RecN induced by the presence of ATP and DNA substrate.

9.6. Conclusion

The results obtained during this work were discussed thoroughly in the three manuscripts (Pellegrino et al, 2012a; Pellegrino et al, 2012 submitted; Pellegrino et al, 2012b) and enabled us to draw a schematic model of DSB recognition in *Deinococcus radiodurans*, in which we speculate on the putative function of RecN in DSB repair. The crystal structures of RecN*head*, RecN*cc* and RecNΔ*dd* were solved to resolutions of 3.0, 2.0 and 4.0 Å respectively and together provided a quasi-atomic resolution model of the full RecN protein. This model represents the first complete model of a member of the SMC protein family. This model of RecN along with all our crystallographic data were supported by solution scattering studies, which indicated that the structures seen in our crystals were maintained in solution. Moreover, in cases when protein crystals were not obtained, SAXS data were very valuable and provided further insight into the shape and architecture of the assemblies. This was notably the case for the study of the dimeric, nucleotide-bound RecN*head*$^{K67A/E472Q}$ mutant.

Our structural work together with our in-depth biochemical characterization of RecN's *in vitro* activities have greatly contributed to a better understanding of the role and mode of action of RecN in DSB repair. But more broadly, our work has also provided valuable experimental evidence in support of a model for the function of SMC and SMC-like proteins in the numerous cellular processes involving DNA tethering and cohesion activities.

References

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* **66:** 213-221

Afonine PV, Mustyakimov M, Grosse-Kunstleve RW, Moriarty NW, Langan P, Adams PD (2010) Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 11)**:** 1153-1163

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Watson J (2008) *Molecular biology of the cell*: Garland Science, Taylor & Francis Group.

Altenberg GA (2003) The engine of ABC proteins. *News in Physiological Sciences* **18:** 191-195

Amundsen SK, Smith GR (2003) Interchangeable parts of the Escherichia coli recombination machinery. *Cell* **112**(6)**:** 741-744

Anderson AW, Nordan HC, Cain RF, Parrish G, Duggan D (1956) Studies on a Radio-Resistant Micrococcus .1. Isolation, Morphology, Cultural Characteristics, and Resistance to Gamma Radiation. *Food Technology* **10**(12)**:** 575-578

Assenmacher N, Hopfner KP (2004) MRE11/RAD50/NBS1: complex activities. *Chromosoma* **113**(4)**:** 157-166

Aubert E, Lecomte C (2007) Illustrated Fourier transforms for crystallography. *Applied crystallography* **40**(6)**:** 1153-1165

Ayora S, Carrasco B, Cardenas PP, Cesar CE, Canas C, Yadav T, Marchisone C, Alonso JC (2011) Double-strand break repair in bacteria: a view from Bacillus subtilis. *FEMS Microbiol Rev*

Battista JR (1997) Against all odds: The survival strategies of Deinococcus radiodurans. *Annual Review of Microbiology* **51:** 203-224

Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67**(Pt 4)**:** 271-281

Bauermeister A, Moeller R, Reitz G, Sommer S, Rettberg P (2011) Effect of Relative Humidity on Deinococcus radiodurans' Resistance to Prolonged Desiccation, Heat, Ionizing, Germicidal, and Environmentally Relevant UV Radiation. *Microb Ecol* **61**(3)**:** 715-722

Baykov AA, Evtushenko OA, Avaeva SM (1988) A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Anal Biochem* **171**(2)**:** 266-270

Bentchikou E, Servant P, Coste G, Sommer S (2010) A Major Role of the RecFOR Pathway in DNA Double-Strand-Break Repair through ESDSA in Deinococcus radiodurans. *Plos Genet* **6**(1)

Best B (2009) Nuclear DNA damage as a direct cause of aging. *Rejuvenation Research* (12)**:** 199–208

Bhaskara V, Dupre A, Lengsfeld B, Hopkins BB, Chan A, Lee JH, Zhang X, Gautier J, Zakian V, Paull TT (2007) Rad50 adenylate kinase activity regulates DNA tethering by Mre11/Rad50 complexes. *Mol Cell* **25**(5)**:** 647-661

Blasius M, Sommer S, Hubscher U (2008) Deinococcus radiodurans: What belongs to the survival kit? *Critical Reviews in Biochemistry and Molecular Biology* **43**(3)**:** 221-238

Blow D (ed) (2002) *Outline of Crystallography for Biologist*: Oxford University Press

Brim H, McFarlan SC, Fredrickson JK, Minton KW, Zhai M, Wackett LP, Daly MJ (2000) Engineering Deinococcus radiodurans for metal remediation in radioactive mixed waste environments. *Nat Biotechnol* **18**(1)**:** 85-90

Burma S, Chen BP, Chen DJ (2006) Role of non-homologous end joining (NHEJ) in maintaining genomic integrity. *DNA Repair (Amst)* **5**(9-10)**:** 1042-1048

Burmeister WP (2000) Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallogr D* **56:** 328-341

Cardenas PP, Carzaniga T, Zangrossi S, Briani F, Garcia-Tirado E, Deho G, Alonso JC (2011) Polynucleotide phosphorylase exonuclease and polymerase activities on single-stranded DNA ends are modulated by RecN, SsbA and RecA proteins. *Nucleic Acids Res* **39**(21)**:** 9250-9261

Carvin D, Islam S, Sternberg M, Blundell T (2001) The preparation of heavy-atom derivatives of protein crystals for use in multiple isomorphous replacement and anomalous scattering. In *International tables for Crystallography: "Crystallography of Biological macromolecules"* E. RMaA (ed), Vol. F, pp 247- 262. Dordrecht: Kluwer Academic

Chayen NE, Saridakis E (2008) Protein crystallization: from purified protein to diffraction-quality crystal. *Nat Methods* **5**(2)**:** 147-153

Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 1)**:** 12-21

Chow KH, Courcelle J (2007) RecBCD and RecJ/RecQ initiate DNA degradation on distinct substrates in UV-irradiated Escherichia coli. *Radiat Res* **168**(4)**:** 499-506

Ciccia A, Elledge SJ (2010) The DNA damage response: making it safe to play with knives. *Mol Cell* **40**(2)**:** 179-204

Cobbe N, Heck MM (2004) The evolution of SMC proteins: phylogenetic analysis and structural implications. *Mol Biol Evol* **21**(2)**:** 332-347

Coppens P (2006) The structure factor. In *International Tables for Crystallography* Vol. B, pp 10-24.

Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**(Pt 9)**:** 1002-1011

Cowtan K (2010) Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 4)**:** 470-478

Cowtan KD, Zhang KY (1999) Density modification for macromolecular phase improvement. *Prog Biophys Mol Biol* **72**(3)**:** 245-270

Cox MM (2007) Motoring along with the bacterial RecA protein. *Nat Rev Mol Cell Biol* **8**(2)**:** 127-138

Cox MM, Battista JR (2005) Deinococcus radiodurans - The consummate survivor. *Nature Reviews Microbiology* **3**(11)**:** 882-892

Cox MM, Keck JL, Battista JR (2010) Rising from the Ashes: DNA Repair in Deinococcus radiodurans. *Plos Genet* **6**(1)

D'Amours D, Jackson S (2002) The Mre11 complex: at the crossroads of dna repair and checkpoint signalling. *Nat Rev Mol Cell Biol*(3(5))**:** 317-327

Daly MJ, Minton KW (1995) Interchromosomal recombination in the extremely radioresistant bacterium Deinococcus radiodurans. *J Bacteriol* **177**(19)**:** 5495-5505

Daly MJ, Minton KW (1997) Recombination between a resident plasmid and the chromosome following irradiation of the radioresistant bacterium Deinococcus radiodurans. *Gene* **187**(2)**:** 225-229

Davidson AL, Dassa E, Orelle C, Chen J (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol Mol Biol Rev* **72**(2)**:** 317-364, table of contents

De Bont R, van Larebeke N (2004) Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **19**(3)**:** 169-185

de Jager M, van Noort J, van Gent DC, Dekker C, Kanaar R, Wyman C (2001) Human Rad50/Mre11 is a flexible complex that can tether DNA ends. *Molecular Cell* **8**(5)**:** 1129-1135

Debye P (1947) Molecular-weight determination by light scattering. *J Phys Colloid Chem* **51**(1)**:** 18-32

Della M, Palmbos PL, Tseng HM, Tonkin LM, Daley JM, Topper LM, Pitcher RS, Tomkinson AE, Wilson TE, Doherty AJ (2004) Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science* **306**(5696)**:** 683-685

Dillingham MS, Kowalczykowski SC (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol Mol Biol Rev* **72**(4)**:** 642-671, Table of Contents

Dimasi N, Flot D, Dupeux F, Marquez JA (2007) Expression, crystallization and X-ray data collection from microcrystals of the extracellular domain of the human inhibitory receptor expressed on myeloid cells IREM-1. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **63**(Pt 3)**:** 204-208

Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research* **35:** W522-W525

Doublie S (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol* **276:** 523-530

Dupre A, Boyer-Chatenet L, Gautier J (2006) Two-step activation of ATM by DNA and the Mre11-Rad50-Nbs1 complex. *Nat Struct Mol Biol* **13**(5)**:** 451-457

Edman P (1949) A method for the determination of amino acid sequence in peptides. *Arch Biochem* **22**(3)**:** 475

Elder FR, Gurewitsch AM, Langmuir RV, Pollock HC (1947) Radiation from Electrons in a Synchrotron. *Phys Rev* **71**(11)**:** 829-830

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1)**:** 2126-2132

Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D* **66:** 486-501

Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**(Pt 1)**:** 72-82

Evans P, McCoy A (2008) An introduction to molecular replacement. *Acta Crystallogr D Biol Crystallogr* **64**(Pt 1)**:** 1-10

Feeney KM, Wasson CW, Parish JL (2010) Cohesin: a regulator of genome integrity and gene expression. *Biochem J* **428**(2)**:** 147-161

Flot D, Mairs T, Giraud T, Guijarro M, Lesourd M, Rey V, van Brussel D, Morawe C, Borel C, Hignette O, Chavanne J, Nurizzo D, McSweeney S, Mitchell E (2010) The ID23-2 structural biology microfocus beamline at the ESRF. *Journal of Synchrotron Radiation* **17:** 107-118

Friedel G (1913) The crystalline symmetries that can reveal the diffraction of Rontgen rays. *Cr Hebd Acad Sci* **157:** 1533-1536

Funayama T, Narumi I, Kikuchi M, Kitayama S, Watanabe H, Yamamoto K (1999) Identification and disruption analysis of the recN gene in the extremely radioresistant bacterium Deinococcus radiodurans. *Mutation Research-DNA Repair* **435**(2)**:** 151-161

Fuss JO, Cooper PK (2006) DNA repair: dynamic defenders against cancer and aging. *PLoS Biol* **4**(6)**:** e203

Garman EF, Nave C (2009) Radiation damage in protein crystals examined under various conditions by different methods. *Journal of Synchrotron Radiation* **16:** 129-132

Garman EF, Owen RL (2006) Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **62**(Pt 1)**:** 32-47

Gentleman R, Ihaka R (2000) Lexical Scope and Statistical Computing. *Journal of Computational and Graphical Statistics* **9:** 491-508

Glatter O, Kratky O (1982) *Small angle X-ray scattering*.

Goodsell DS (2001) The molecular perspective: ultraviolet light and pyrimidine dimers. *Stem Cells* **19**(4)**:** 348-349

Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**(4)**:** 305-308

Goujon M, McWilliam H, Li WZ, Valentin F, Squizzato S, Paern J, Lopez R (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research* **38:** W695-W699

Graumann PL, Knust T (2009) Dynamics of the bacterial SMC complex and SMC-like proteins involved in DNA repair. *Chromosome Res* **17**(2)**:** 265-275

Griese JJ, Hopfner KP (2011) Structure and DNA-binding activity of the Pyrococcus furiosus SMC protein hinge domain. *Proteins* **79**(2)**:** 558-568

Griese JJ, Witte G, Hopfner KP (2010) Structure and DNA binding activity of the mouse condensin hinge domain highlight common and diverse features of SMC proteins. *Nucleic Acids Res* **38**(10)**:** 3454-3465

Grove JI, Wood SR, Briggs GS, Oldham NJ, Lloyd RG (2009) A soluble RecN homologue provides means for biochemical and genetic analysis of DNA double-strand break repair in Escherichia coli. *DNA Repair* **8**(12)**:** 1434-1443

Gruber S, Haering CH, Nasmyth K (2003) Chromosomal cohesin forms a ring. *Cell* **112**(6)**:** 765-777

Haering CH, Schoffnegger D, Nishino T, Helmhart W, Nasmyth K, Lowe J (2004) Structure and stability of cohesin's Smc1-kleisin interaction. *Mol Cell* **15**(6)**:** 951-964

Hanson PI, Whiteheart SW (2005) AAA+ proteins: Have engine, will work. *Nature Reviews Molecular Cell Biology* **6**(7)**:** 519-529

Harper JW, Elledge SJ (2007) The DNA damage response: ten years after. *Mol Cell* **28**(5)**:** 739-745

Hauf S, Roitinger E, Koch B, Dittrich CM, Mechtler K, Peters JM (2005) Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. *PLoS Biol* **3**(3)**:** e69

Helleday T, Petermann E, Lundin C, Hodgson B, Sharma RA (2008) DNA repair pathways as targets for cancer therapy. *Nature Reviews Cancer*(8)**:** 193-204

Helliwell JR (1992) *Macromolecular crystallography with synchrotron radiation*: Cambridge University press.

Heyer WD, Ehmsen KT, Liu J (2010) Regulation of homologous recombination in eukaryotes. *Annu Rev Genet* **44:** 113-139

Hirano M, Anderson DE, Erickson HP, Hirano T (2001) Bimodal activation of SMC ATPase by intra- and inter-molecular interactions. *EMBO J* **20**(12)**:** 3238-3250

Hirano M, Hirano T (1998) ATP-dependent aggregation of single-stranded DNA by a bacterial SMC homodimer. *EMBO J* **17**(23)**:** 7139-7148

Hirano M, Hirano T (2002) Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. *EMBO J* **21**(21)**:** 5733-5744

Hirano M, Hirano T (2004) Positive and negative regulation of SMC-DNA interactions by ATP and accessory proteins. *EMBO J* **23**(13)**:** 2664-2673

Hirano T (2002) The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion and repair. *Gene Dev* **16**(4)**:** 399-414

Hirano T (2005) SMC proteins and chromosome mechanics: from bacteria to humans. *Philosophical Transactions of the Royal Society B-Biological Sciences* **360**(1455)**:** 507-514

Hirano T. (2006) At the heart of the chromosome: SMC proteins in action. *Nature Reviews Molecular Cell Biology*, Vol. 7, pp. 311-322.

Hoeijmakers JH (2001) Genome maintenance mechanisms for preventing cancer. *Nature* **411**(6835)**:** 366-374

Hoeijmakers JH (2009) DNA damage, aging, and cancer. *N Engl J Med* **361**(15)**:** 1475-1485

Holland IB, Blight MA (1999) ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *J Mol Biol* **293**(2)**:** 381-399

Hollenstein K, Dawson RJ, Locher KP (2007) Structure and mechanism of ABC transporter proteins. *Curr Opin Struct Biol* **17**(4)**:** 412-418

Holliday R (1964) A mechanism for gene conversion in fungi. *Genetic Research* **5:** 282-304

Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics* **24**(23)**:** 2780-2781

Hopfner KP, Craig L, Moncalian G, Zinkel RA, Usui T, Owen BA, Karcher A, Henderson B, Bodmer JL, McMurray CT, Carney JP, Petrini JH, Tainer JA (2002) The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. *Nature* **418**(6897)**:** 562-566

Hopfner KP, Karcher A, Craig L, Woo TT, Carney JP, Tainer JA (2001) Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. *Cell* **105**(4)**:** 473-485

Hopfner KP, Karcher A, Shin DS, Craig L, Arthur LM, Carney JP, Tainer JA (2000) Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* **101**(7)**:** 789-800

Hopfner KP, Tainer JA (2003) Rad50/SMC proteins and ABC transporters: unifying concepts from high-resolution structures. *Current Opinion in Structural Biology* **13**(2)**:** 249-255

Huyghe P (1998) Conan the bacterium. *Sciences-New York* **38**(4)**:** 16-19

Inoue J, Honda M, Ikawa S, Shibata T, Mikawa T (2008) The process of displacing the single-stranded DNA-binding protein from single-stranded DNA by RecO and RecR proteins. *Nucleic Acids Res* **36**(1)**:** 94-109

Jackson AL, Chen R, Loeb LA (1998) Induction of microsatellite instability by oxidative DNA damage. *Proc Natl Acad Sci U S A* **95**(21)**:** 12468-12473

Jones PM, George AM (2004) The ABC transporter structure and mechanism: perspectives on recent research. *Cell Mol Life Sci* **61**(6)**:** 682-699

Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 2)**:** 133-144

Kanaar R, Wyman C (2008) DNA repair by the MRN complex: break it to make it. *Cell* **135**(1)**:** 14-16

Kashchiev D (2000) *Nucleation: basic theory with applications*: Oxford

Kidane D, Sanchez H, Alonso JC, Graumann PL (2004) Visualization of DNA double-strand break repair in live bacteria reveals dynamic recruitment of Bacillus subtilis RecF, RecO and RecN proteins to distinct sites on the nucleoids. *Mol Microbiol* **52**(6)**:** 1627-1639

Kiianitsa K, Solinger JA, Heyer WD (2003) NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities. *Anal Biochem* **321**(2)**:** 266-271

Kim JI, Cox MM (2002) The RecA proteins of Deinococcus radiodurans and Escherichia coli promote DNA strand exchange via inverse pathways. *Proc Natl Acad Sci U S A* **99**(12)**:** 7917-7921

Kinoshita E, van der Linden E, Sanchez H, Wyman C (2009) RAD50, an SMC family member with multiple roles in DNA break repair: how does ATP affect function? *Chromosome Res* **17**(2)**:** 277-288

Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* **36:** 1277-1282

Koroleva O, Makharashvili N, Courcelle CT, Courcelle J, Korolev S (2007) Structural conservation of RecF and Rad50: implications for DNA recognition and RecF function. *EMBO J* **26**(3)**:** 867-877

Kozin MB, Svergun DI (2001) Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* **34:** 33-41

Kratky O, Porod G (1949) Diffuse small-angle scattering of X-rays in colloid systems. *J Colloid Sci* **4**(1)**:** 35-70

Lammens A, Schele A, Hopfner KP (2004) Structural biochemistry of ATP-driven dimerization and DNA-stimulated activation of SMC ATPases. *Curr Biol* **14**(19)**:** 1778-1782

Lammens K, Bemeleit DJ, Mockel C, Clausing E, Schele A, Hartung S, Schiller CB, Lucas M, Angermuller C, Soding J, Strasser K, Hopfner KP (2011) The Mre11:Rad50 Structure Shows an ATP-Dependent Molecular Clamp in DNA Double-Strand Break Repair. *Cell* **145**(1)**:** 54-66

Langer G, Cohen SX, Lamzin VS, Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**(7)**:** 1171-1179

Lawrence CW, Borden A, Banerjee SK, LeClerc JE (1990) Mutation frequency and spectrum resulting from a single abasic site in a single-stranded vector. *Nucleic Acids Res* **18**(8)**:** 2153-2157

Li G, Xiang Y, Zhang Y, Wang DC (2001) A simple and efficient innovation of the vapor-diffusion method for controlling nucleation and growth of large protein crystals. *J Appl Cryst* **34:** 388-391

Li YY, Schoeffler AJ, Berger JM, Oakley MG (2010) The Crystal Structure of the Hinge Domain of the Escherichia coli Structural Maintenance of Chromosomes Protein MukB. *Journal of Molecular Biology* **395**(1)**:** 11-19

Lim HS, Kim JS, Park YB, Gwon GH, Cho Y (2011) Crystal structure of the Mre11-Rad50-ATP gamma S complex: understanding the interplay between Mre11 and Rad50. *Gene Dev* **25**(10)**:** 1091-1104

Lisby M, Barlow JH, Burgess RC, Rothstein R (2004) Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. *Cell* **118**(6)**:** 699-713

Liu L, Hawkins DM, Ghosh S, Young SS (2003) Robust singular value decomposition analysis of microarray data. *Proc Natl Acad Sci U S A* **100**(23)**:** 13167-13172

Longhese MP, Bonetti D, Manfrini N, Clerici M (2010) Mechanisms and regulation of DNA end resection. *EMBO J* **29**(17)**:** 2864-2874

Losada A, Hirano T (2001) Intermolecular DNA interactions stimulated by the cohesin complex in vitro: implications for sister chromatid cohesion. *Curr Biol* **11**(4)**:** 268-272

Lowe J, Cordell SC, van den Ent F (2001) Crystal structure of the SMC head domain: An ABC ATPase with 900 residues antiparallel coiled-coil inserted. *Journal of Molecular Biology* **306**(1)**:** 25-35

Lusetti SL, Cox MM (2002) The bacterial REcA protein and the recombinational DNA repair of stalled replication forks. *Annu Rev Biochem* **71:** 71-100

Mattimore V, Battista JR (1996) Radioresistance of Deinococcus radiodurans: Functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. *Journal of Bacteriology* **178**(3)**:** 633-637

McCarthy AA, Brockhauser S, Nurizzo D, Theveneau P, Mairs T, Spruce D, Guijarro M, Lesourd M, Ravelli RBG, McSweeney S (2009) A decade of user operation on the macromolecular crystallography MAD beamline ID14-4 at the ESRF. *Journal of Synchrotron Radiation* **16:** 803-812

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**(Pt 4)**:** 658-674

Mockel C, Lammens K, Schele A, Hopfner KP (2011) ATP driven structural changes of the bacterial Mre11:Rad50 catalytic head complex. *Nucleic Acids Res*

Moore JK, Haber JE (1996) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Mol Cell Biol* **16**(5)**:** 2164-2173

Moreno-Herrero F, de Jager M, Dekker NH, Kanaar R, Wyman C, Dekker C (2005) Mesoscale conformational changes in the DNA-repair complex Rad50/Mre11/Nbs1 upon binding DNA. *Nature* **437**(7057)**:** 440-443

Morimatsu K, Kowalczykowski SC (2003) RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. *Mol Cell* **11**(5)**:** 1337-1347

Murayama Y, Kurokawa Y, Mayanagi K, Iwasaki H (2008) Formation and branch migration of Holliday junctions mediated by eukaryotic recombinases. *Nature* **451**(7181)**:** 1018-1021

Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**(Pt 4)**:** 355-367

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **53:** 240-255

Nakamura J, Swenberg JA (1999) Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues. *Cancer Res* **59**(11)**:** 2522-2526

Nasmyth K, Haering CH, Lowe J, Hochwagen A (2002) Molecular architecture of SMC proteins and the yeast cohesin complex. *Molecular Cell* **9**(4)**:** 773-788

Navaza J (1994) Amore - an Automated Package for Molecular Replacement. *Acta Crystallographica Section A* **50:** 157-163

Niki H, Imamura R, Kitaoka M, Yamanaka K, Ogura T, Hiraga S (1992) E.coli MukB protein involved in chromosome partition forms a homodimer with a rod-and-hinge structure having DNA binding and ATP/GTP binding activities. *EMBO J* **11**(13)**:** 5101-5109

Nobbman U, Connah M, Fish B, Varley P, Gee C, Mulot S, Chen J, Zhou L, Lu Y, Sheng F, Yi J, Harding SE (2007) Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies. *Biotechnology and Genetic Engineering Reviews* **24:** 117-128

Nurizzo D, Mairs T, Guijarro M, Rey V, Meyer J, Fajardo P, Chavanne J, Biasci JC, McSweeney S, Mitchell E (2006) The ID23-1 structural biology beamline at the ESRF. *Journal of Synchrotron Radiation* **13:** 227-238

Ordoukhanian P, Taylor JS (2000) Caged single and double strand breaks. *Bioconjug Chem* **11**(1)**:** 94-103

Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D* **62:** 439-450

Panjikar S, Parthasarathy V, Lamzin VS, Weiss MS, Tucker PA (2005) Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr D* **61:** 449-457

Paull TT, Moncalian G, Lengsfeld B, Bhaskara V, Hopfner KP, Karcher A, Alden E, Tainer JA (2004) The Rad50 signature motif: Essential to ATP binding and biological function. *Journal of Molecular Biology* **335**(4)**:** 937-951

Pellegrino S, de Sanctis D, McSweeney S, Timmins J (2011) Expression, purification and preliminary structural analysis of the coiled-coil domain of *Deinococcus radiodurans* RecN. *Acta Crystallogr F*

Pellegrino S, de Sanctis D, McSweeney S, Timmins J (2012a) Expression, purification and preliminary structural analysis of the coiled-coil domain of Deinococcus radiodurans RecN. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **68**(Pt 2)**:** 218-221

Pellegrino S, Radzimanowski J, de Sanctis D, McSweeney S, Timmins J (2012 submitted) Structural and functional characterization of RecN: New insights into double-strand break repair *Nucleic Acids Res*

Pellegrino S, Radzimanowski J, McSweeney S, Timmins J (2012b) Expression, purification and preliminary structural analysis of the head domain of Deinococcus radiodurans RecN. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **68**(Pt 1)**:** 81-84

Pellegrino S, Radzimanowski J, McSweeney S, Timmins J (2012c) Expression, purification and preliminary structural analysis of the head domain of *Deinococcus radiodurans* RecN. *Acta Crystallogr F* **68:** 81-84

Pernot P, Theveneau P, Giraud T, Fernandes RN, Nurizzo D, Spruce D, Surr J, McSweeney S, Round A, Felisaz F, Foedinger L, Gobbo A, Huet J, Villard C, Cipriani F (2010) New beamline dedicated to solution scattering from biological macromolecules at the ESRF. *Xiv International Conference on Small-Angle Scattering (Sas09)* **247**

Petoukhov MV, Konarev PV, Kikhney AG, Svergun DI (2007) ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. *J Appl Crystallogr* **40:** S223-S228

Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* **40**(3)**:** 191-285

Radman M, Slade D, Lindner AB, Paul G (2009) Recombination and Replication in DNA Repair of Heavily Irradiated Deinococcus radiodurans. *Cell* **136**(6)**:** 1044-1055

Rao ST, Rossmann MG (1973) Comparison of Super-Secondary Structures in Proteins. *Journal of Molecular Biology* **76**(2)**:** 241-&

Ravelli RBG, McSweeney SM (2000) The 'fingerprint' that X-rays can leave on structures. *Struct Fold Des* **8**(3)**:** 315-328

Rees DC, Johnson E, Lewinson O (2009) ABC transporters: the power to change. *Nat Rev Mol Cell Biol* **10**(3)**:** 218-227

Reyes ED, Patidar PL, Uranga LA, Bortoletto AS, Lusetti SL (2010) RecN Is a Cohesin-like Protein That Stimulates Intermolecular DNA Interactions in Vitro. *Journal of Biological Chemistry* **285(**22)**:** 16521-16529

Rocha EP, Cornet E, Michel B (2005) Comparative and evolutionary analysis of the bacterial homologous recombination systems. *Plos Genet* **1**(2)**:** e15

Royant A, Carpentier P, Ohana J, McGeehan J, Paetzold B, Noirclerc-Savoye M, Vernede X, Adam V, Bourgeois D (2007) Advances in spectroscopic methods for biological crystals. 1. Fluorescence lifetime measurements. *J Appl Crystallogr* **40:** 1105-1112

Russi S, Juers DH, Sanchez-Weatherby J, Pellegrini E, Mossou E, Forsyth VT, Huet J, Gobbo A, Felisaz F, Moya R, McSweeney SM, Cusack S, Cipriani F, Bowler MW (2011) Inducing phase changes in crystals of macromolecules: status and perspectives for controlled crystal dehydration. *J Struct Biol* **175**(2)**:** 236-243

Saikrishnan K, Powell B, Cook NJ, Webb MR, Wigley DB (2009) Mechanistic basis of 5'-3' translocation in SF1B helicases. *Cell* **137**(5)**:** 849-859

Sakai A, Cox MM (2009) RecFOR and RecOR as Distinct RecA Loading Pathways. *Journal of Biological Chemistry* **284**(5)**:** 3264-3272

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, Vol. 1, 2, 3: Cold Spring Harbor Laboratory Press.

Sanchez H, Alonso JC (2005) Bacillus subtilis RecN binds and protects 3 '-single-stranded DNA extensions in the presence of ATP. *Nucleic Acids Research* **33**(7)**:** 2343-2350

Savir Y, Tlusty T (2010) RecA-Mediated Homology Search as a Nearly Optimal Signal Detection System. *Molecular Cell* **40**(3)**:** 388-396

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

Schwartz MA, Shapiro L (2011) An SMC ATPase mutant disrupts chromosome segregation in Caulobacter. *Mol Microbiol*

Sharples GJ, Ingleston SM, Lloyd RG (1999) Holliday junction processing in bacteria: Insights from the evolutionary conservation of RuVABC, RecG, and RusA. *Journal of Bacteriology* **181**(18)**:** 5543-5550

Sheldrick GM (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 4)**:** 479-485

Shintomi K, Hirano T (2007) How are cohesin rings opened and closed? *Trends in Biochemical Sciences* **32**(4)**:** 154-157

Shuman S, Glickman MS (2007) Bacterial DNA repair by non-homologous end joining. *Nat Rev Microbiol* **5**(11)**:** 852-861

Singleton MR, Dillingham MS, Gaudier M, Kowalczykowski SC, Wigley DB (2004) Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* **432**(7014)**:** 187-193

Sinha RP, Hader DP (2002) UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci* **1**(4)**:** 225-236

Slade D, Radman M (2011) Oxidative stress resistance in Deinococcus radiodurans. *Microbiol Mol Biol Rev* **75**(1)**:** 133-191

Spies M, Kowalczykowski S (2004) Homologous recombination by RecBCD and RecF pathways. In *Bacterial Chromosomes*, pp 389-403. ASM Press

Svendsen JM, Harper JW (2010) GEN1/Yen1 and the SLX4 complex: Solutions to the problem of Holliday junction resolution. *Genes Dev* **24**(6)**:** 521-536

Svergun D, Barberato C, Koch MHJ (1995) CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* **28:** 768-773

Svergun DI (1992) Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J Appl Crystallogr* **25:** 495-503

Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing (vol 76, pg 2879, 1999). *Biophysical Journal* **77**(5)**:** 2896-2896

Svergun DI, Konarev PV, Petoukhov MV (2001a) MASSHA - a graphics system for rigid-body modelling of macromolecular complexes against solution scattering data. *J Appl Crystallogr* **34:** 527-532

Svergun DI, Kozin MB (2001) Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* **34:** 33-41

Svergun DI, Petoukhov MV, Koch MH (2001b) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* **80**(6)**:** 2946-2953

Tanaka M, Earl AM, Howell HA, Park MJ, Eisen JA, Peterson SN, Battista JR (2004) Analysis of Deinococcus radiodurans's transcriptional response to ionizing radiation and desiccation reveals novel proteins that contribute to extreme radioresistance. *Genetics* **168**(1)**:** 21-33

Taylor AF, Smith GR (1999) Regulation of homologous recombination: Chi inactivates RecBCD enzyme by disassembly of the three subunits. *Gene Dev* **13**(7)**:** 890-900

Taylor G (2003) The phase problem. *Acta Crystallogr D* **59:** 1881-1890

Taylor GL (2010) Introduction to phasing. *Acta Crystallogr D* **66:** 325-338

Terwilliger TC (2000) Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr* **56**(Pt 8)**:** 965-972

Terwilliger TC (2003) Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr D* **59:** 38-44

Terwilliger TC, Adams PD, Read RJ, Mccoy AJ, Moriarty NW, Grosse-Kunstleve RW, Afonine PV, Zwart PH, Hung LW (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D* **65:** 582-601

Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW, Read RJ, Adams PD (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D* **64:** 61-69

Timmins J, Gordon E, Caria S, Leonard G, Acajjaoui S, Kuo MS, Monchois V, McSweeney S (2009) Structural and mutational analyses of Deinococcus radiodurans UvrA2 provide insight into DNA binding and damage recognition by UvrAs. *Structure* **17**(4)**:** 547-558

Tomkinson AE, Chen L, Trujillo KM, Van Komen S, Roh DH, Krejci L, Lewis LK, Resnick MA, Sung P (2005) Effect of amino acid substitutions in the Rad50 ATP binding domain on DNA double strand break repair in yeast. *Journal of Biological Chemistry* **280**(4)**:** 2620-2627

Vagin A, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* **30:** 1022-1025

van Noort J, van der Heijden T, de Jager M, Wyman C, Kanaar R, Dekker C (2003) The coiled-coil of the human Rad50 DNA repair protein contains specific segments of increased flexibility. *P Natl Acad Sci USA* **100**(13)**:** 7581-7586

Volkov VV, Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* **36:** 860-864

Wakatsuki S, Belrhali H, Mitchell EP, Burmeister WP, McSweeney SM, Kahn R, Bourgeois D, Yao M, Tomizaki T, Theveneau P (1998) ID14 'Quadriga', a beamline for protein crystallography at the ESRF. *Journal of Synchrotron Radiation* **5:** 215-221

Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**(8)**:** 945-951

Wang H, Perrault AR, Takeda Y, Qin W, Iliakis G (2003) Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic Acids Res* **31**(18)**:** 5377-5388

Wang J, Julin DA (2004) DNA helicase activity of the RecD protein from Deinococcus radiodurans. *J Biol Chem* **279**(50)**:** 52024-52032

Watrin E, Peters JM (2006) Cohesin and DNA damage repair. *Exp Cell Res* **312**(14)**:** 2687-2693

Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**(4356)**:** 737-738

White O, Eisen JA, Heidelberg JF, Hickey EK, Peterson JD, Dodson RJ, Haft DH, Gwinn ML, Nelson WC, Richardson DL, Moffat KS, Qin HY, Jiang LX, Pamphile W, Crosby M, Shen M, Vamathevan JJ, Lam P, McDonald L, Utterback T, Zalewski C, Makarova KS, Aravind L, Daly MJ, Minton KW, Fleischmann RD, Ketchum KA, Nelson KE, Salzberg S, Smith HO, Venter JC, Fraser CM (1999) Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1. *Science* **286**(5444)**:** 1571-1577

Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **112:** 531-552

Williams GJ, Williams RS, Williams JS, Moncalian G, Arvai AS, Limbo O, Guenther G, SilDas S, Hammel M, Russell P, Tainer JA (2011) ABC ATPase signature helices in Rad50 link nucleotide state to Mre11 interface for DNA repair. *Nature Structural & Molecular Biology* **18**(4)**:** 423-U454

Williams RS, Moncalian G, Williams JS, Yamada Y, Limbo O, Shin DS, Groocock LM, Cahill D, Hitomi C, Guenther G, Moiani D, Carney JP, Russell P, Tainer JA (2008) Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. *Cell* **135**(1)**:** 97-109

Williamson MP (1994) The Structure and Function of Proline-Rich Regions in Proteins. *Biochemical Journal* **297:** 249-260

Wong RW (2010) An update on cohesin function as a 'molecular glue' on chromosomes and spindles. *Cell Cycle* **9**(9)**:** 1754-1758

Wyman C, Kinoshita E, van der Linden E, Sanchez H (2009) RAD50, an SMC family member with multiple roles in DNA break repair: how does ATP affect function? *Chromosome Res* **17**(2)**:** 277-288

Wyman C, Ristic D, Kanaar R (2004) Homologous recombination-mediated double-strand break repair. *DNA Repair* **3**(8-9)**:** 827-833

Xie JM, Wang L, Wu N, Brock A, Spraggon G, Schultz PG (2004) The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat Biotechnol* **22**(10)**:** 1297-1301

Yabuki M, Fujii MM, Maizels N (2005) The MRE11-RAD50-NBS1 complex accelerates somatic hypermutation and gene conversion of immunoglobulin variable regions. *Nat Immunol* **6**(7)**:** 730-736

Zahradka K, Slade D, Bailone A, Sommer S, Averbeck D, Petranovic M, Lindner AB, Radman M (2006) Reassembly of shattered chromosomes in Deinococcus radiodurans. *Nature* **443**(7111)**:** 569-573

Zimmerman JM, Battista RJ (2005) A ring-like nucleoid is not necessary for radioresistance in the Deinococcaceae. *BMC Microbiology* **5**(17)

Appendix-I:

Application of synchrotron radiation to Structural Biology

Summary of Appendix-I

A basic presentation of synchrotron radiation and how this is routinely used for Structural Biology purposes is presented in this chapter. The protein crystallization process is also discussed and the phase diagram is described.

In order to solve a protein structure, it is necessary to solve the phase problem, since the diffraction experiment performed on a protein crystal only provides information about the amplitudes of the diffracted photons and not about the phases. There are three methods for solving such a problem: Molecular Replacement, Isomorphous Replacement and Anomalous Dispersion.

Small Angle X-ray Scattering (SAXS) is nowadays extensively used as a technique for overall envelope reconstruction and visualization of conformational rearrangements induced by specific cofactors or substrates. Basic theory of SAXS is presented in this chapter.

Résumé de l'annexe

Ce chapitre présente brièvement le rayonnement synchrotron et comment cela est couramment utilisé en Biologie Structurale. Le processus de cristallisation des protéines est également discuté et le diagramme de phase est décrit.

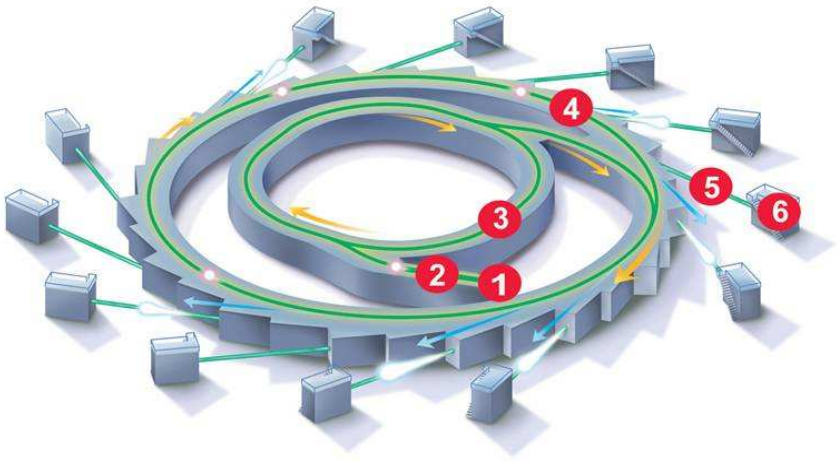
Afin de résoudre la structure cristalline d’une protéine, il est nécessaire de résoudre le problème de phase, puisque seules les amplitudes des photons diffractés et non leurs phases peuvent être extraites des expériences de diffraction effectuée sur un cristal de protéine. Il existe trois méthodes pour résoudre un tel problème: le remplacement moléculaire, le remplacement isomorphe et la dispersion anomale.

La diffusion des rayons X aux petits angles est aujourd'hui largement utilisée comme technique pour la reconstruction d’enveloppe globale et la visualisation de changements conformationnels induits par des cofacteurs ou des substrats. La théorie de base du SAXS est présentée dans ce chapitre.

1. Generation of synchrotron radiation

In the last two decades synchrotron radiation (SR) has become very important in Structural Biology.

A synchrotron is a cyclic particle accelerator in which charged particles travel very close to the speed of light (Helliwell, 1992). When high-energy particles, including electrons, are in rapid motion and are forced to travel in a curved path by a magnetic field, SR is produced. To produce SR (Figure A1), electrons are injected through a linear accelerator (Linac) into a small ring, called a booster synchrotron. Here the electrons are accelerated to nearly the speed of light and, then, injected into the storage ring. The synchrotron light is produced at each bending event of the electrons: magnetic elements, which induce a curvature in the trajectory, are placed along the path in the storage ring. The synchrotron light is produced tangentially to the direction of the travelling electrons (Figure A1) (Elder et al, 1947).



1. Electron Gun: produces electrons.
2. Linear Accelerator (LinAc): accelerates electrons to near the speed of light.
3. Booster Ring: further increases the energy of the electrons.
4. Storage Ring: maintains the electron energy and produces radiation.
5. Beamline: set up of radiation for experiments.
6. Experimental Hutch: User area.

Figure A1: A general view of how SR is generated. All the steps necessary to produce the synchrotron light are described. Figure taken from http://shelx.uni-ac.gwdg.de/~tg/teaching/seminars.

1.1. Properties of synchrotron radiation in relation to Structural Biology

Synchrotron radiation can be used for a broad variety of experiments, from Physics to Chemistry and, ultimately, Biology. The main properties of SR which are of importance for Structural Biology experiments are:

- High brilliance.
- Tunable over a wide range of different wavelengths, usually in the range of 5 to 20 keV. The shorter the wavelength of the beam, the lower the absorption of the protein crystal will be (Helliwell, 1992).

Protein crystals are difficult to study due to their intrinsic properties. First of all, scattering from such crystals is weak due to the constitutive light atoms, mostly carbon (C), oxygen (O), nitrogen (N) and hydrogen (H). Large solvent regions, together with the flexibility and disorder that can occur in the crystals, can be factors that reduce the scattering into the Bragg reflections. The high brightness of the SR beam allows a collimated beam to hit a small cross-section area of the crystal, maximizing the signal to noise ratio (Helliwell, 1992). This can improve the resolution of the data obtained, even in protein crystals, as long as the ***radiation damage*** is taken into account. The use of high doses of IR can lead to the creation of photoelectrons, which result when high-energy photons are absorbed by light atoms (Garman & Nave, 2009). The principal consequence of such an event is the creation of secondary electrons, which cause further excitation and ionization inside the protein crystal. This process ultimately creates free radicals (mainly from water molecules) that may lead to decarboxylation, disulphide bond breakage, cleavage of C-S bonds in methionines, in addition to the crystal lifetime decay (Burmeister, 2000; Ravelli & McSweeney, 2000). In order to avoid the above-mentioned serious problems, ***cryo-cooling*** of protein crystals has been introduced (Garman & Owen, 2006). The use of short wavelength SR X-rays reduces also the amount of absorbed X-rays and, by consequence, the damage to the crystal.

2. Protein crystallization

Crystalline material is an ordered and repeated array of atoms or molecules. In Structural Biology single crystals of proteins can be used to perform X-ray diffraction experiments, which ultimately lead to the determination of the atomic structure of the protein.

Supersaturation is the driving force in the crystallization process: it is only when this situation is achieved that crystalline nuclei are produced (Figure A2) (Kashchiev, 2000). However, too much supersaturation can lead to precipitation and formation of aggregates,

therefore finding the best nucleation condition is a crucial step. Once crystal nuclei have been formed, the system moves, in principle, to the metastable zone (Figure A2). Here, no more nuclei are produced and crystal growth is stimulated.

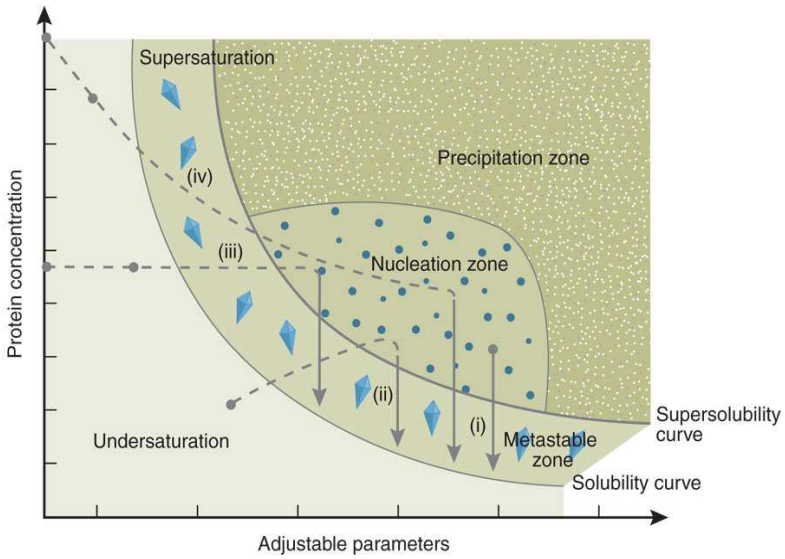


Figure A2: The crystallization process can be explained by the following phase diagram. During the crystallization process the protein moves towards supersaturation. Crystal nuclei are only produced in the nucleation zone. Finding the right condition, which maximizes the relation between protein concentration and other adjustable parameters are essential for the crystal nuclei to move into the metastable zone. Here, protein crystals can grow, with no more nuclei formation. Figure taken from (Chayen & Saridakis, 2008)

Vapor diffusion is the most common technique used in protein crystallization. Two methods are prevalently used: hanging and sitting drops. In both cases (Figure A3) a droplet containing the protein solution is mixed with the same volume of a reservoir solution, which contains a precipitant solution. The resulting droplet is then left to equilibrate against a much larger volume of the reservoir solution. This induces the transfer of water from the drop to the reservoir. This will increase the level of saturation of the protein sample in the drop thus, (hopefully) triggering the production of single crystals. The quality of the crystals can then be defined by:

- Size, crystal habit or shape
- Mosaicity: crystals are made up of “building” blocks (unit cells), which are angularly misaligned, deviating from the ideal behavior. Mosaicity is also affected by the crystal-freezing step.

- Resolution limit: depends mostly on the amount of solvent in the crystal and the flexibility of the protein in arrangement of the unit cells.
- Radiation lifetime (Helliwell, 1992).

Like all crystals, those of protein molecules act as diffraction gratings for X-rays with the analysis of the resulting diffraction pattern allowing the determination of the structure of the protein molecule contained in the crystals.

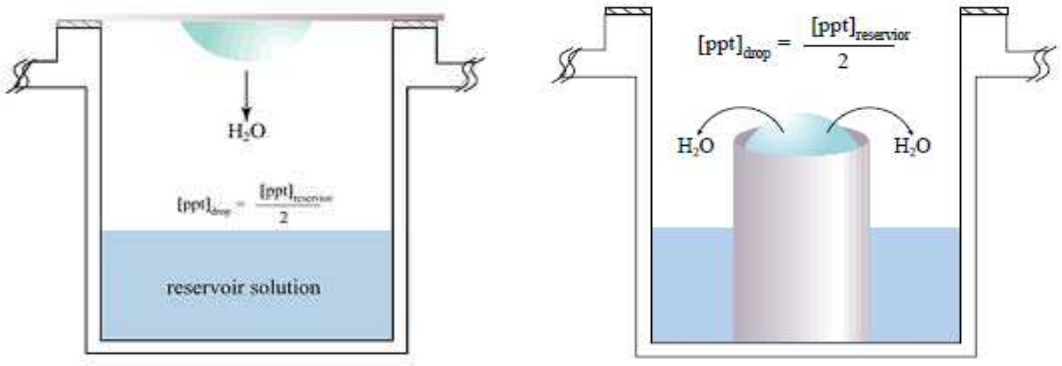


Figure A3: Hanging and sitting drops. Both protocols are used for protein crystallogenesis and ttake advantage of vapor diffusion technique (Li et al, 2001). Figures taken from http://hamptonresearch.com.

3. X-ray diffraction and application to protein crystallography

X-rays were discovered by Wilhelm von Rontgen in the late 19th century. X-rays are a form of electromagnetic radiation with a wavelength in the range of 0.01 to 10 nm, which corresponds to an energy range between 120 eV and 120 keV. They are called “soft X-rays” when they have a wavelength of more than 2Å and interact strongly with air (absorption phenomenon) and water. X-rays with a shorter wavelength are much more penetrating and, especially when in the range of 0.5-1.6Å, can be suitable for protein crystallography.

In order to determine the crystal structure of a protein we need to perform an experiment of X-ray diffraction. Interaction of incident waves with the crystalline sample results in the generation of a diffracted beam (Figure A4), only if the scattered rays of each repeating unit are all in phase. If we consider that a crystal comprises layers of lattice points, acting as mirrors, we can hypothesize that X-rays will be scattered by the first plane, while other rays will pass through and will be scattered by the second layer and so on. The two mirrors will scatter in phase only if the difference in path length of the beams reflected from the various layers is an integer number of wavelength (nλ) (Figure A4). It is thus possible to

define the condition for diffraction to occur from any given plane, as defined by the Bragg's law:

$$n\lambda = 2d\sin\theta \qquad \text{(Eq. A1)}$$

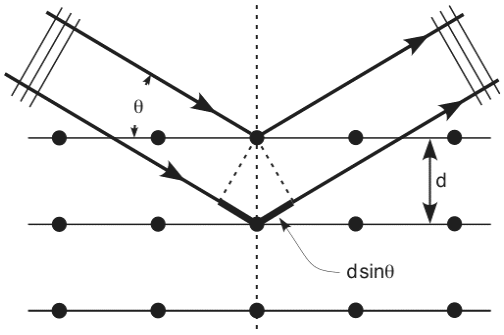


Figure A4: Bragg's reflections generated by a three-dimensional lattice. Figure taken from http://photonicswiki.org

Scattering of incident radiation from atoms in a crystal can be mathematically described by the ***structure factor*** (Coppens, 2006):

$$F_{hkl} = \sum_{N} f_i \exp\left[2\pi i\left(hx_i + ky_i + lz_i\right)\right]\exp\left[\frac{-B_i \sin^2\theta}{\lambda^2}\right] \qquad \text{(Eq. A2)}$$

where f_i is the atomic scattering factor of the *i*th atom, B is the Debye-Waller factor, accounting for the thermal motion of the atoms and x_i, y_i and z_i are the positional coordinates of the *i*th atom. The structure factor (F_{hkl}) contains information about the unit cell contents and about the amplitude $\left|F_{hkl}\right|$ and phase of the scattered wave (Blow, 2002).

Structure factors are related to the electron density function ρ(*x*), by applying a Fourier transform (Aubert & Lecomte, 2007):

$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl} F_{hkl} \exp\left[2\pi i\left(hx + ky + lz\right)\right] \qquad \text{(Eq. A3)}$$

An electron density map can thus be reconstructed and the crystal structure "solved" once in possession of both amplitude and phase information for each structure factor.

3.1. The phase problem

The photons diffracted from a protein crystal are recorded on a detector, which is able to measure the intensity of each diffracted beam (I_{hkl}). To solve a crystal structure we need

information about both the amplitudes and the phases of the diffracted waves in order to reconstruct the exact position of the scattering atoms. The amplitudes ($|F_{hkl}|$) are known, since they can be derived directly from the measured intensities $|I_{hkl}|$:

$$|F_{hkl}| \propto \sqrt{I_{hkl}} \qquad \text{(Eq. A4)}$$

In contrast, phase information is lost and we need to derive this for each structure factor in order to reconstruct the molecular structure. In Structural Biology the so-called ***phase problem*** is generally solved in one of three ways:

- Molecular Replacement
- Single/multiple isomorphous replacement
- Anomalous dispersion

3.1.1. Molecular Replacement

Molecular Replacement (Mol.R.) is a phasing technique, which takes advantage of the prior knowledge of a known crystal structure (model), suggested to share a relatively high degree of similarity with an unknown crystal structure (target). The Mol.R. method is based on the correlation of the Patterson map generated from the model crystal structure with the Patterson function derived from the data collected from the target structure. The ***Patterson function*** (Equation 5) produces an inter-atomic vector map.

$$P(U,V,W) = \frac{1}{V_{EZ}} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{-} \left|F(hkl)\right|^2 \cdot \exp\left[-2\pi i(hU + kV + lW)\right] \qquad \text{(Eq. A5)}$$

The crystal structure of the target molecule can be solved when the Patterson map for the model crystal structure superposes well with the map generated from the target structure. Two functions are crucial for the superposition of the two Patterson maps: the ***rotation function***, which determines the orientation of the model in the new unit cell (target), and the ***translation function***, which searches for the right translation, with respect to the origin of the new unit cell, of the correctly oriented model (Taylor, 2010). Once the best model has been found, Mol.R. programs, like Phaser (McCoy et al, 2007), AMoRe (Navaza, 1994), MolRep (Vagin & Teplyakov, 1997) for instance, try several possible orientations and positions in order to match the observed diffraction (experiment) with the predicted diffraction (model) (Figure A5). Phases are then taken from the rotated and translated model structure and used in

combination with the observed structure factor amplitudes for the calculation of an initial electron density map (Evans & McCoy, 2008).

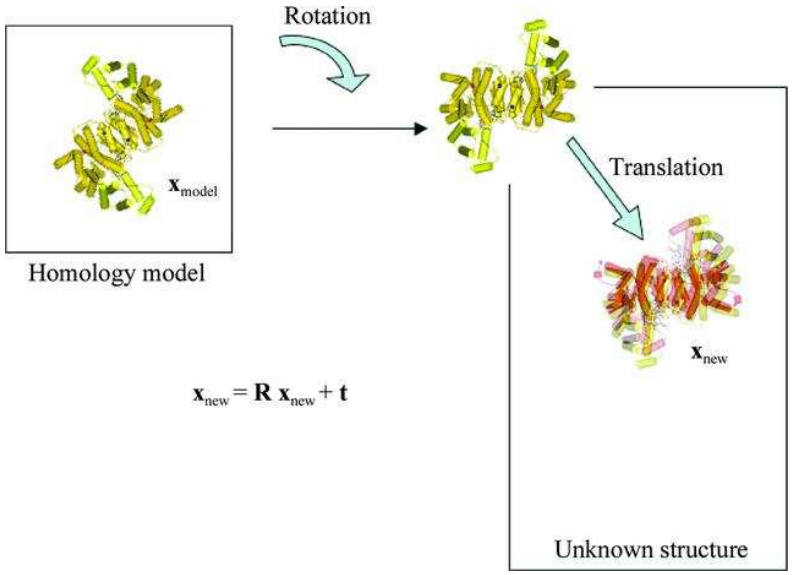


Figure A5: Process of how Mol.R. works for placing a similar crystal structure in the right position in order to have a good set of phases for a target crystal structure. Figure taken from (Taylor, 2010).

3.1.2. Isomorphous replacement

The isomorphous replacement method is based on the comparison of different data from two crystals, one that carries one or more heavy atoms (called "*derivative*"), which therefore scatter X-rays more than light atoms, and one crystal that does not contain such heavy atom scatterers (called "*native*"). Heavy atoms can be incorporated chemically (i.e. by iodination of phenylalanine residues (Xie et al, 2004)), by soaking the crystals in a solution containing the heavy atoms (Carvin et al, 2001), or by genetic engineering (i.e. seleno-methionine incorporation (Doublie, 1997)). However, these incorporations can cause changes in the crystal properties meaning that native and derivative crystals are no longer isomorphous. Such non-isomorphism can lead to inaccuracy of the phase determination.

The Harker construction (Figure A6) provides the phases for the native protein structure. When one data set for the native and one for the derivative form are collected, the Patterson map allows determining the heavy atom positions. This means that both amplitudes and phases for the heavy atom are determined. Since the structure factor of the heavy atom derivative ($|F_{PH}|$) is the vector sum of the structure factor of the heavy atom alone ($|F_H|$)

and the structure factor of the native protein crystal ($|F_P|$), the possible phases can be drawn geometrically, using the mentioned Harker construction (Figure A6).

Both Single (SIR) / Multiple (MIR) Isomorphous Replacement can be used as methods for phase determination (Figure A6-A). SIR requires only one derivative in addition to the native crystal but results in a phase ambiguity, in which two equivalent solutions are possible (Figure A6-A) (Taylor, 2010). In the case of MIR the phase ambiguity is, in principle, resolved (Figure A6-B). However, more than one derivative crystal is needed.

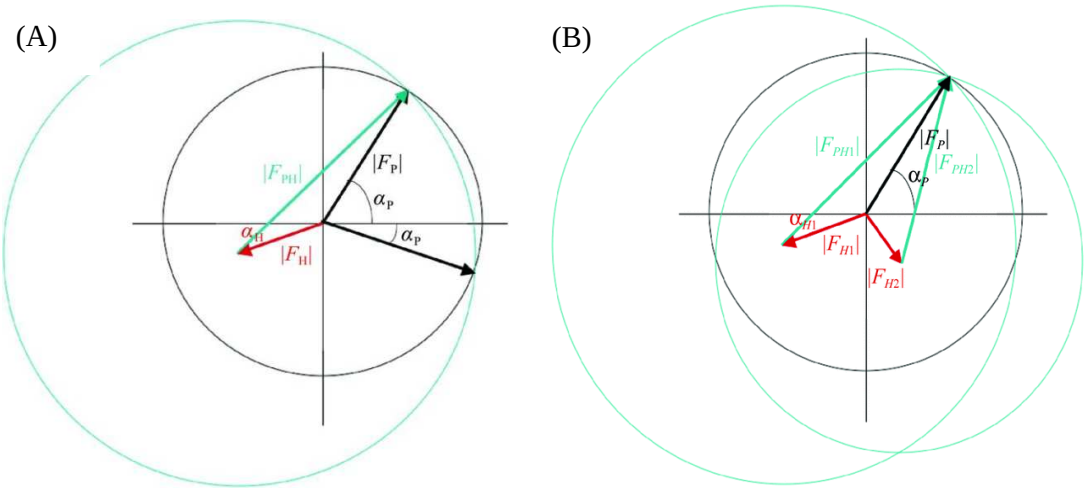


Figure A6: Harker construction for isomorphous replacement case. A) Harker diagram for Single Isomorphous Replacement (SIR) where two possible phases can be drawn. B) Harker construction for Multiple Isomorphous Replacement method (MIR). Figures taken from (Taylor, 2003).

3.1.3. Anomalous dispersion

When we perform a diffraction experiment on a protein crystal, we normally assume that, for reflections related by a centre of inversion, Friedel's law is true (Friedel, 1913):

$$|F_{hkl}| = |F_{\overline{hkl}}| \quad \text{and} \quad \varphi_{hkl} = -\varphi_{\overline{hkl}} \qquad \text{(Eq. A6-A7)}$$

However, if there is an anomalous scatterer (a heavy atom for instance present in the crystal), then this is no longer the case and $|F_{hkl}| \neq |F_{\overline{hkl}}|$ (Figure A7).

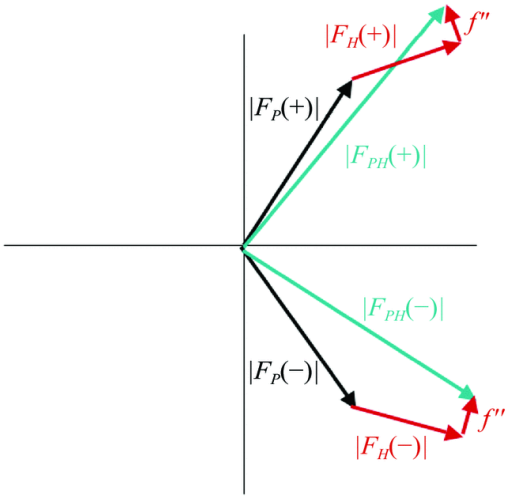


Figure A7: Representation of the breakdown of Friedel's law when an anomalous scatterer is present in a crystal. Figure taken from (Taylor, 2010).

One of the most widely used techniques for phase determination is the anomalous dispersion method, which takes advantage of structure factor differences when anomalous scatterers are present in a crystal. Anomalous scattering takes place when incident photon energy is closed to the transition energy of an atom, bringing it to an excited state (Taylor, 2003). In this particular case, the radiated energy has a different phase and the intensity of scattering is also reduced. Modified methionine residues, in which the S atom is substituted by a Se (heavier than the sulphur), are largely used for anomalous experimental phasing (Doublie, 1997).

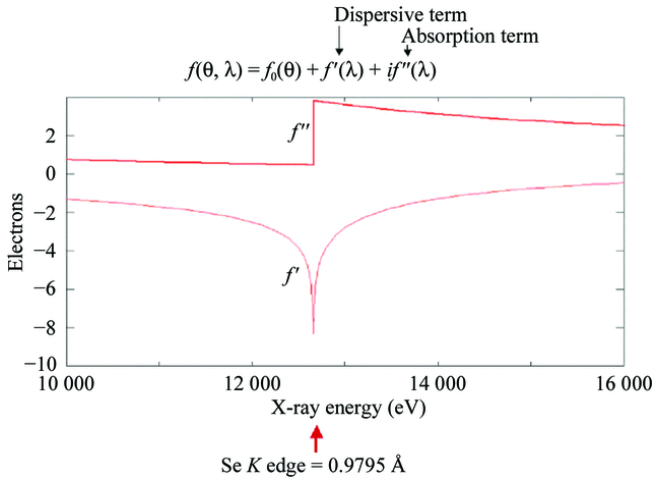
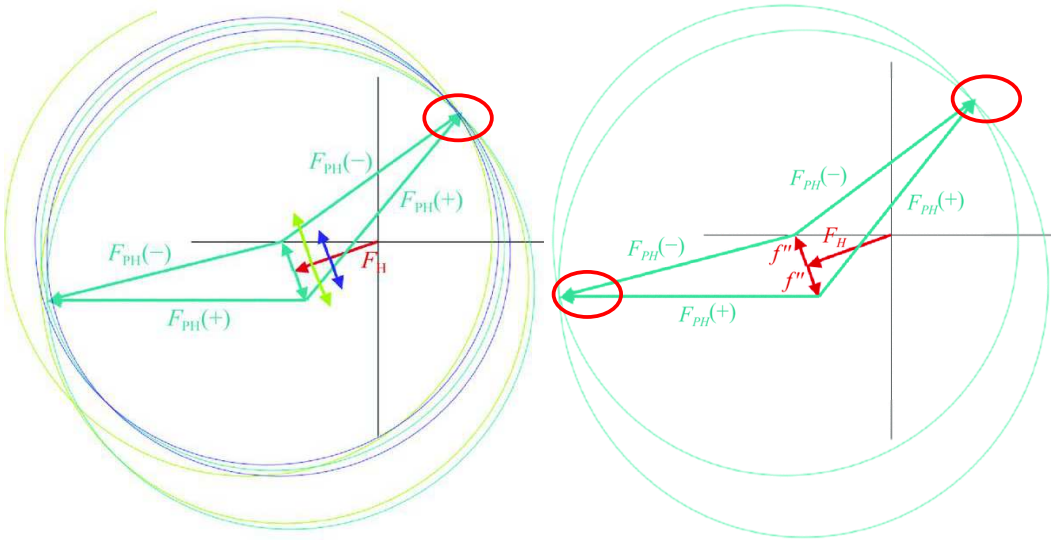


Figure A8: Representation of the theoretical anomalous scattering of Selenium atoms around the K absorption edge. Absorption and dispersive terms are indicated. Figure taken from (Taylor, 2010).

When taking into account anomalous scattering, the atomic scattering vector is defined as:

$$f(\theta,\lambda) = f_0(\theta) + f'(\lambda) + if''(\lambda) \qquad \text{(Eq. A8)}$$

where f_0 is the atomic scattering vector, $f'(\lambda)$ is the factor representing the amount of normal scattering reduced at λ and $f''(\lambda)$ is the amount of 90º out-to-phase scattering at λ (Blow, 2002) (Figure A8). The advent of tunable beamlines have made the anomalous scattering technique more powerful and suitable for the analysis of a broad range of heavy atoms, avoiding also the need for several crystals and thus overcoming non-isomorphism problems. Anomalous dispersion can overcome the non-isomorphism problem. Data are collected from a single crystal at several, typically three, wavelengths (Multi-wavelength Anomalous Dispersion, MAD (Figure A9-A)) or only one (Single-wavelength Anomalous Dispersion, SAD (Figure A9-B)).



(A) **(B)**

Figure A9: Harker constructions for anomalous dispersion case. A) Harker construction for Multi-wavelength Anomalous Dispersion (MAD). Three different data sets are collected from the same crystals, at three different wavelengths. The signal extrapolated from the three data sets allows unambiguous determination of the phase angle. B) Harker construction for Single-wavelength Anomalous Dispersion (SAD). Possible phases are indicated with red circles. Figures taken from (Taylor, 2010).

3.1.3.1. MAD phasing: principles

In order to maximize the absorption and dispersive effects (Equation A8), three data sets are collected at three chosen wavelengths: one is collected at the absorption (f'') peak

(λ_1) corresponding to the absorption K-edge of the anomalous scatterer; one at the inflection point (λ_2) where the dispersive term *f'* reaches its minimum; the last one is collected at a remote wavelength (λ_3 and/or λ_4) in order to maximize the dispersive difference to λ_2 (Taylor, 2010) (Figure A10). Figure A9-A represents the Harker construction for a MAD experiment which yield a unique solution for the phase.

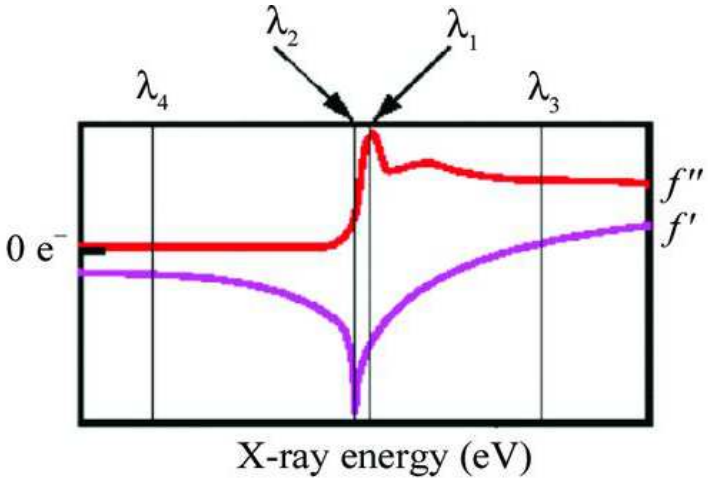


Figure A10: MAD phasing. Experimentally measured absorption edge for an anomalous scatterer, showing the wavelengths at which data are collected for maximization of the anomalous signal. Figure taken from (Taylor, 2010).

Radiation damage, however, needs to be taken into account, since fully redundant and, in the case of MAD, more than one complete data set needs to be collected.

3.1.3.2. SAD phasing: principles

SAD experimental phasing is becoming more and more powerful, since only one data set may be sufficient to solve the phase problem. However, compared to MAD, in which data are collected at three different wavelengths from the same crystal (in correspondence of the absorption peak, inflection point and high/low energy remote), phase ambiguity needs to be resolved (Figure A9-B). This issue is overcome by employing density modification tools, like DM (Cowtan, 2010; Cowtan & Zhang, 1999), RESOLVE (Terwilliger, 2000) etc., which use solvent flattening, histogram matching and NCS-averaging for modifying an initial electron density map to generate a new set of phases. These are then combined with the experimental phases to calculate a new electron density map, possibly suitable for model

building programs, such as ARP/wARP (Langer et al, 2008), BUCCANEER (Cowtan, 2006) or Phenix_AutoBuild (Terwilliger et al, 2008).

4. Small Angle X-ray Scattering

Small Angle X-ray Scattering (SAXS) is a technique widely used to obtain low-resolution information on protein structures. Scattering of the sample is measured at very low angles (0.1-10°) resulting in the determination of the overall protein shape and size, gain of information regarding protein folding/unfolding/aggregation and finally conformational changes (Putnam et al, 2007). In a SAXS experiment, what is measured is the difference in average electron density of the solute $\rho\left(\vec{r}\right)$ (protein molecules in our case) and the surrounding liquid phase ρ_s (the buffer in which the protein is resuspended):

$$\Delta\rho\left(\vec{r}\right) = \rho\left(\vec{r}\right) - \rho_s \qquad \text{(Eq. A9)}$$

In SAXS the scattering curve I(q), once the buffer has been subtracted, is isotropic (the scattering is symmetric) due to the randomly-oriented distribution of the molecules in solution. I(q) is measured as a function of the momentum transfer:

$$q = \frac{(4\pi \sin \theta)}{\lambda} \qquad \text{(Eq. A10)}$$

where 2θ is the scattering angle and λ the wavelength of the incident X-rays. The scattering curve of a mono-disperse sample, essential conditions in SAXS, can be derived from the P(r) (pair distribution) function:

$$I(q) = 4\pi \int_0^{D_{max}} P(r) \frac{\sin(qr)}{qr} dr \qquad \text{(Eq. A11)}$$

where D_{max} is the maximum dimension of the scattering particle. The pair-distribution function P(r) in Small Angle Scattering is considered to correspond to the Patterson function in X-ray crystallography (Figure A11).

4.1. The Pair Distribution function

The pair distribution function provides direct information about the distances between electrons in the scattering particles in the sample (Putnam et al, 2007) (Figure A11). The Pair distribution function can be used for calculation of radius of gyration (R_G) and the scattering intensity at zero angle (I_0) and does not only take into account the region of the direct beam, as the Guinier approximation does. Calculation of molecular weight is much more accurate if the P(r) is taken rather than the Guinier region, especially where the latter is affected by partial inter-particle effects. I_0 is evaluated from the lowest q value and is represented as the square of the number of electrons in the scatterer and, therefore, is unaffected by particle shape (Putnam et al, 2007).

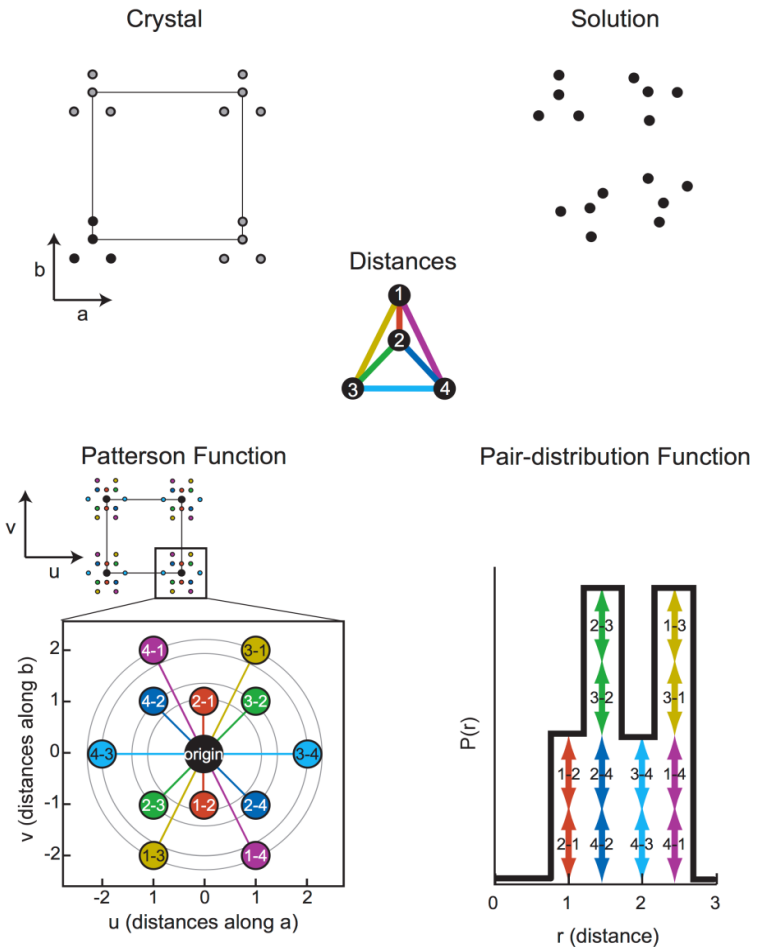


Figure A11: Comparison of the Patterson autocorrelation and the pair-distribution autocorrelation functions. A theoretical two-dimensional molecule of four atoms is placed in an arbitrary two-dimensional crystal in solution. The Patterson function contains cross peaks for every interatomic distance in the crystal and these crosspeaks in the u, v plane, are indicated by circles and retain directional information about their positions in the crystal. The

pair-distribution function, on the other hand, resolves distances but not directions within each scattering unit. Thus, all equivalent distances in the four-atom molecule add-up. Figure taken from (Putnam et al, 2007).

Globular macromolecules have a P(r) function with a single peak, while elongated macromolecules have a longer tail at large r and can have multiple peaks (Figure A12). The maximum length in the particle, D_{MAX}, is the position where the P(r) function returns to zero at large values of r.

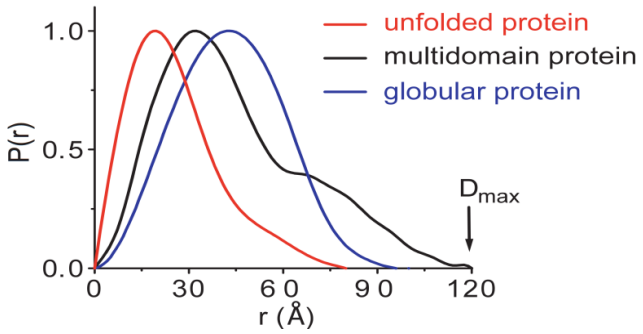


Figure A12: Pair distribution function profiles of proteins with different quaternary structures. Figure taken from (Putnam et al, 2007).

Disagreements for values of R_G and I_0 calculated from the P(r) function and from the Guinier plot can indicate small amounts of aggregation that primarily affect the low resolution data and the accuracy of the Guinier plot. Guinier approximation uses the Small Angle *q* values for estimation of the R_G, following the formula in Table A1 (Putnam et al, 2007).

SAXS experiment can also provide clues regarding the folding state of the sample (Figure A13). Extrapolation of data at higher *q* value provides information concerning the molecular shape. For folded molecules, the intensity of scattering can be described as follows (Porod's law, (Glatter & Kratky, 1982)):

$$I(q) \propto q^{-4} \qquad \text{(Eq. A12)}$$

This is directly calculated from the scattering curve and plotted as $q^2I(q)$ as a function of *q*, named as Kratky plot (Kratky & Porod, 1949).

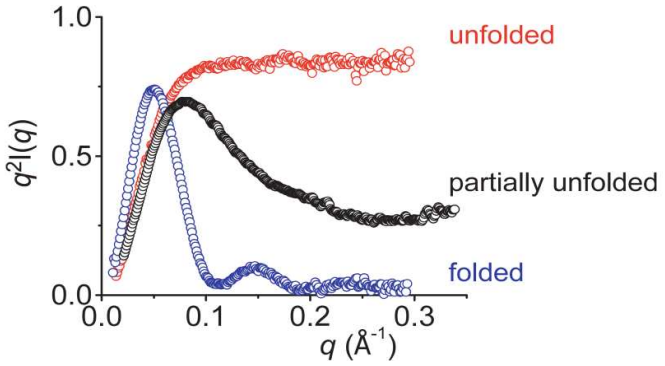


Figure A13: Kratky plot for protein in folded, partially folded and unfolded state. Figure taken from (Putnam et al, 2007).

The Kratky plot identifies unfolded samples. Globular macromolecules follow Porod's law and have bell-shaped curves (Figure A13). Extended molecules, such as unfolded peptides, lack this peak and have a plateau or are slightly increasing in the larger q range. In the presence of elongated proteins, the Debye approximation is more appropriate (Debye, 1947):

Parameter	**Formula**	**Range of Data used and variable definitions**
Radius of Gyration (R_G): Guinier approximation	$\ln[I(q)] = \ln[I(0)] - \frac{q^2 R_G^2}{3}$	$qRg<1.3$ in case of globular, $qRg<0.8$ in case of elongated molecules
Radius of Gyration (R_G): Debye approximation	$I(q) = \frac{2I(0)}{q^4 R_G^4}\left(q^2 R_G^2 - 1 + e^{-q^2 R_G^2}\right)$	$qRg<1.4$ for elongated molecules

Formulas for elongated or flexibly linked linear macromolecules

Radius of Gyration of cross-section (R_{XC})	$\ln[qI(q)] = \ln[qI(0)] - \frac{q^2 R_{XC}^2}{2}$	Intermediate q values
Length (L)	$L = \left(12\left(R_G^2 - R_{XC}^2\right)\right)^{\frac{1}{2}}$	See R_G and R_{XC}

Table A1: Basic formulas for Small Angle Scattering. Radius of gyration R_G is calculated considering the formulas above, which take into account different approximations related to the particular shape of the macromolecules.

SAXS data are really useful for the determination of the so-called *ab initio* envelope, which can be extrapolated from the scattering curve of a macromolecule in solution. Tools like DAMMIN, GASBOR, DAMMIF (Svergun, 1999; Svergun et al, 2001b) allow to determine the overall shape of a target molecule, through mathematical algorithms that randomly place "dummy" spheres, or residues, in a D_{MAX} defined volume and, finally, refine them against the scattering curve.

Comparing crystallographic models to Small Angle Scattering curves is very useful for understanding the influences of the crystal lattice on the atomic structure. The crystal structure is likely the lowest energy state under crystallization conditions; however, it is not necessarily the lowest energy state in solution. Many studies have suggested that the effects of the crystal lattice do not alter the folding of domains, but rather influence the conformations adopted by flexible termini or linkers between domains (Putnam et al, 2007). SAXS tools are readily available for comparing experimental curves obtained from protein in solution with a theoretical scattering curve derived from a model built starting from a crystallography structure (CRYSOL, (Svergun et al, 1995)). Formulation of the χ^2 value (Eq. A13) has been introduced for evaluation of the goodness of fit:

$$\chi^2 = \frac{1}{N_P - 1}\sum_i \left[\frac{I(q_i)_{\exp} - cI(q_i)_{calc}}{\sigma(q_i)}\right]^2 \qquad \text{(Eq. A13)}$$

Appendix-II:

collection of manuscripts

ABBREVIATIONS

ABC: ATP-Binding Cassette
ADP: Adenosine 5'-Diphosphate
ATM: Ataxia Telangiectasia Mutated
ATP: Adenosine 5'-Triphosphate
ATPase: ATP hydrolase
BER: Base Excision Repair
bp: base pair
kDa: kilo Dalton
DDR: DNA Damage Response
DLS: Dynamic Light Scattering
DMSO: Dimethyl sulfoxide
DNA: Deoxyribonucleic Acid
dNTPs: Deoxyribonucleoside triphosphates
dsDNA: double strand DNA
ssDNA: single strand DNA
DR: *Deinococcus radiodurans*
DSB: Double Strand Break
DTT: 1,4-Dithio-DL-threitol
EB: Elution Buffer
EDTA: Ethylene diamine tetra-acetic acid
EMSA: Electrophoretic Mobility Shift Assay
ESDSA: Extended Synthesis DNA Strand Annealing
EtBr: Ethidium Bromide
FRET: Fluorescence Resonance Energy Transfer
mg: milligram
µg: microgram
GF: Gel Filtration
h: hour
HF: High Fidelity
His: Histidine

HR: Homologous Recombination

IMAC: Immobilized Metal ion Affinity Chromatography

IPTG: Isopropyl-b-D-thiogalactopyranoside

IR: Ionizing Radiation

LB: Lysogenic Broth

LDH: lactate dehydrogenase

LLG: Log Likelihood Gain

M: Molar

mM: Millimolar

μM: Micromolar

nM: Nanomolar

pmol: Picomoles

MAD: Multi-wavelength Anomalous Dispersion

SEC-MALLS: Size Exclusion Chromatography coupled to Multi-Angle Laser Light Scattering

min : minutes

ml: Millilitre

μl: Microlitre

nm: Nanometre

MMR: Mismatch Repair

Mol.R: Molecular Replacement

MR: Mre11-Rad50

MRN: Mre11-Rad50-Nbs1

MRX: Mre11-Rad50-Xrs2

MW: Molecular Weight

NBD: Nucleotide Binding Domain

NER: Nucleotide Excision Repair

NHEJ: Non Homologous End-Joining

NTA: Nitrilotriacetic acid

OD: Optical Density

PCR: Polymerase Chain Reaction

pdI: polydispersity index

PEG: Polyethylene glycol

PEP: Phosphoenolpyruvate

PK: Pyruvate kinase

rpm: revolutions per minute

Rec: Recombinase

RFZ: Rotation Function Z-score

RT: Room Temperature

SAD: Single-wavelength Anomalous Dispersion

SAXS: Small Angle X-ray Scattering

SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SEC: Size Exclusion Chromatography

Se-Met: Selenium Methionine

SMC: Structural Maintenance of Chromosomes

SSB: Single Strand DNA-Binding protein

N-terminus: Amino-terminus

C-terminus: Carboxyl-terminus

TEV: Tobacco Etch Virus

TFZ : Translation Function Z-score

TLS : Translation/Libration/Screw

TOPO : DNA Topoisomerase

UV: Ultra-violet

List of Tables & Figures

- **Tables:**

- **Figures:**

Acknowledgements