

Dissertation zur Erlangung des Doktorgrades  
der Fakultät für Chemie und Pharmazie  
der Ludwig-Maximilians-Universität München

**Structural and biochemical studies on the  
structural maintenance of chromosomes protein  
from *Pyrococcus furiosus***



Alfred Lammens

aus

Würzburg

München, 2007

## **Erklärung**

Diese Dissertation wurde im Sinne von § 13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 von Herrn Prof. Dr. Karl-Peter Hopfner betreut.

## **Ehrenwörtliche Versicherung**

Diese Dissertation wurde selbstständig, ohne unerlaubte Hilfsmittel erarbeitet.

München, am 08.03.2007

Alfred Lammens

Dissertation eingereicht am 08.03.2007

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

2. Gutachter: Herr Prof. Dr. Patrick Cramer

Mündliche Prüfung am 07.05.2007

## Publications:

Lammens, A., Schele, A., Hopfner, K.-P.

**Structural Biochemistry of ATP-Driven Dimerization and DNA-Stimulated Activation of SMC ATPases.**

Curr Biol. 2004 Oct 5;14(19):1778-82.

Assenmacher, N., Wenig, K., Lammens, A., Hopfner, K.-P.

**Structural Basis for Transcription Coupled Repair: the N Terminus of Mfd Resembles UvrB with Degenerate ATPase Motifs**

J Mol Biol. 2006 Jan 27;355(4):675-83.

Lammens, A., Hopfner, K.-P.

**Structural Basis for Adenylate Kinase Activity in ABC ATPases**

submitted

Alt, A., Lammens, K., Lammens, A., Pieck, J.C., Chiocchini, C., Hopfner, K.-P., Carell, T.

**Structural investigation of the replicative bypass of a cisplatin DNA lesion by polymerase  $\eta$**

submitted

## Presentations

### Talk:

Gene Center Annual Retreat 2004

14<sup>th</sup> May 2004, Wildbad-Kreuth, Germany

9th Biennial Meeting of the DGDR

13<sup>th</sup> September, Hamburg, Germany

### Poster:

Mechanisms of Genomic Integrity Workshop

21<sup>st</sup>-24<sup>th</sup> June 2004, Galway, Ireland

Murnau Conference on Structural Biology of Molecular Recognition

15<sup>th</sup>-17<sup>th</sup> September 2005, Murnau, Germany

14. Jahrestagung der Deutschen Gesellschaft für Kristallographie

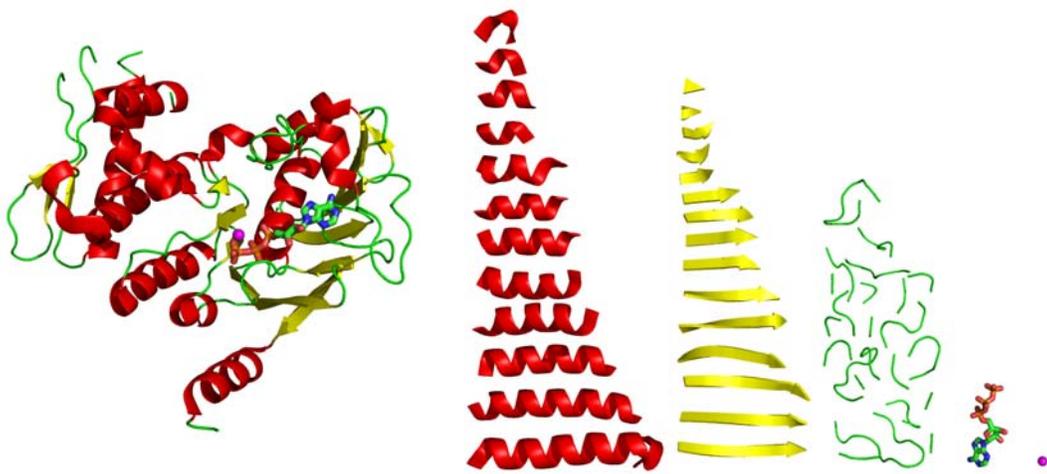
3<sup>rd</sup>-6<sup>th</sup> April 2006, Freiburg, Germany

8th International School on the Crystallography of Biological Macromolecules

21<sup>st</sup>-25<sup>th</sup> May 2006, Como, Italy

9th Biennial Meeting of the DGDR

12<sup>th</sup> -15<sup>th</sup> September, Hamburg, Germany



*The whole is more than the sum of its parts.*

Aristotle

Image inspired by Ursus Wehrli.

## Table of contents

Index of Figures .....	ii
Index of Tables.....	ii
1 Summary .....	2
2 Introduction .....	2
2.1 The ABC ATPase family .....	2
2.2 The SMC protein family .....	2
2.2.1 Architecture of SMC proteins.....	2
2.2.2 The SMC 1/3 cohesin complex.....	2
2.2.3 The SMC 2/4 condensin complex.....	2
2.2.4 The SMC 5/6 DNA repair complex .....	2
2.2.5 The bacterial SMC / Kleisin complex.....	2
2.2.6 The bacterial MukB/E/F complex.....	2
2.3 Structure determination by X-ray crystallography.....	2
2.3.1 Physical and mathematical background.....	2
2.3.2 Single- and multi-wavelength anomalous dispersion .....	2
2.3.3 Molecular replacement .....	2
3 Objectives.....	2
4 Materials.....	2
4.1 Chemicals.....	2
4.2 Enzymes, standards, Kits and chromatographic material .....	2
4.3 Oligonucleotides.....	2
4.4 <i>E.coli</i> strains and Plasmids.....	2

---

5	Methods	2
5.1	Cloning	2
5.2	Expression and purification	2
5.3	Analytical size exclusion chromatography	2
5.4	ATPase activity assay	2
5.5	ATP binding assay	2
5.6	Reverse adenylate kinase activity assay	2
5.7	Crystallization	2
5.8	Data collection and processing	2
5.9	Structure solution and refinement	2
6	Results	2
6.1	Cloning and expression	2
6.1.1	<i>Pyrococcus furiosus</i> SMC	2
6.1.2	<i>Pyrococcus furiosus</i> ScpA	2
6.1.3	<i>Pyrococcus furiosus</i> ScpB	2
6.2	Biochemical studies	2
6.2.1	Analytical size exclusion chromatography	2
6.2.2	ATPase activity assay	2
6.2.3	ATP binding assay	2
6.2.4	Reverse adenylate kinase activity	2
6.3	Crystallization, structure solution and refinement	2
6.3.1	Apo wild type SMCcd	2
6.3.2	ATP bound E1098Q SMCcd	2
6.3.3	ADP-AlF <sub>4</sub> <sup>-</sup> and ADP-BeF <sub>3</sub> <sup>-</sup> bound to SMCcd	2

---

6.3.4 SMCcd in complex with AP5A .....	2
6.4 Structural analysis .....	2
6.4.1 Overall structure of SMCcd .....	2
6.4.2 Structure of the ATP bound dimer .....	2
6.4.3 The active site .....	2
6.4.4 Structural comparison between monomeric and dimeric SMCcd .....	2
6.4.5 The conserved arginine finger .....	2
6.4.6 Structures with bound transition state analogues .....	2
6.4.7 Structure of SMCcd in complex with AP5A .....	2
7 Discussion .....	2
8 References .....	2
9 Abbreviations .....	2
Appendix .....	II
Acknowledgements .....	II
Curriculum vitae .....	II

## Index of Figures

Figure 1: Structures of ABC ATPases.....	2
Figure 2: Conserved ABC ATPase motifs.....	2
Figure 3: Mechanism of ABC ATPases.....	2
Figure 4: Model of the adenylate kinase activity of CFTR.....	2
Figure 5: Architecture of SMC proteins. ....	2
Figure 6: The SMC hinge domain.....	2
Figure 7: Subunit composition of SMC protein complexes.....	2
Figure 8: Model of the cohesin cycle in yeast.....	2
Figure 9: Working model of the condensin action.....	2
Figure 10: Scattering factors as function of X-ray energy.....	2
Figure 11: The atomic scattering factor. ....	2
Figure 12: Expression and purification of full length SMC.....	2
Figure 13: Coiled-coil prediction for the full-length <i>Pfu</i> SMC protein.....	2
Figure 14: Expression and purification of SMCcd.....	2
Figure 15: Expression and purification of His-ScpA.....	2
Figure 16: Thrombin cleavage of His-ScpA and complex formation with SMCcd.....	2
Figure 17: Expression and purification of His-ScpB.....	2
Figure 18: Size exclusion profiles of SMCcd.....	2
Figure 19: Size exclusion profiles of SMCcd and ScpA.....	2
Figure 20: ATPase activity assay of wt and mutant SMCcd.....	2
Figure 21: ATPase activity of SMCcd in presence of inhibitors.....	2
Figure 22: ATP binding assays of wild type and mutant SMCcd.....	2

---

Figure 23: Reverse adenylate kinase activity of SMCcd. ....	2
Figure 24: Images of the structure solution steps of SMCcd apo form. ....	2
Figure 25: Crystals, diffraction and density images of ATP bound E1098Q SMCcd. ....	2
Figure 26: Electron densities around ADP- $\text{AlF}_4^-$ and ADP- $\text{BeF}_3^-$ .....	2
Figure 27: Crystals, diffraction and structure images of AP5A-SMCcd. ....	2
Figure 28: Ribbon presentation of the SMCcd crystal structure .....	2
Figure 29: Two orthogonal views of dimeric SMCcd. ....	2
Figure 30: Conservation and ABC ATPase motifs mapped on the surface of SMCcd. ....	2
Figure 31: Stereo view of ATP bound to the active center. ....	2
Figure 32: Overlay of nucleotide free and ATP bound SMCcd.....	2
Figure 33: Surface of dimeric SMCcd with mapped electrostatic potential and R-loop. .	2
Figure 34: Detail view of the R-loop and alignment of the arginine finger.....	2
Figure 35: Superimposed active site of SMCcd bound to ATP, ADP- $\text{AlF}_4^-$ and ADP- $\text{BeF}_3^-$ .....	2
Figure 36: Binding of AP5A to SMCcd.....	2
Figure 37: Comparison of ATP and AP5A bound structures of SMCcd. ....	2
Figure 38: Comparison of AP5A bound SMCcd to CFTR NBD1 and ABCE1. ....	2
Figure 39: ATPase cycle of SMC proteins with functional mutations. ....	2
Figure 40: Model for the establishment of sister chromatid cohesion. ....	2
Figure 41: Unifying model for the ATPase and adenylate kinase activity. ....	2
Figure 42: Structure based sequence alignment of SMC proteins. ....	II

---

## Index of Tables

Table 1: Components of eukaryotic SMC complexes.....	II
Table 2: Crystallographic table of SMCcd and ATP-SMCcd.....	II
Table 3: Crystallographic table of ADP-AlF <sub>4</sub> <sup>-</sup> and ADP-BeF <sub>3</sub> <sup>-</sup> bound SMCcd.....	II
Table 4: Crystallographic table of AP5A-SMCcd .....	II

# 1 Summary

Correct propagation of the genetic information to the daughter cells is a central step in life. Structural maintenance of chromosomes (SMC) proteins support this process as key players in higher-order chromosome architecture in all organism. They form the core of several multi-protein complexes that are involved in sister chromatid cohesion, chromosome condensation and DNA repair. For their function ATP binding and hydrolysis is essential.

In this work structural and biochemical studies were performed to understand the mechanism how SMC proteins fulfill their biological role. The crystal structure of the ATPase domain from *Pyrococcus furiosus* SMC (SMCcd) was determined in the nucleotide free form, in the presence of ATP and two transition state analogues. ATP mediates the dimerization of two SMC ATPase domains by binding to opposing Walker A and signature motifs, indicating that ATP can directly assemble SMC protein rings. The dimerization of the head domains could be proved biochemically. Comparison of the four obtained SMCcd structures revealed only minor structural changes during the ATPase cycle indicating that no “power stroke” is involved in the function of SMC proteins.

The largest conformational changes occur at a surface loop that is ordered upon ATP binding. Structural and mutagenesis data identified an arginine residue in this loop that is conserved among all SMC proteins. This arginine finger links a putative DNA interaction site on the surface of the SMC protein to the bound ATP and stimulates its hydrolysis. The importance of the connection between DNA interaction and ATP hydrolysis for the biological function of SMC proteins could be proved together with a collaborating group.

SMCcd was also used to gain insights into the mechanism of the adenylate kinase activity of ABC ATPases. The crystal structure of SMCcd in complex with the adenylate kinase inhibitor AP5A suggests that ATP binds to the P-loop while AMP binds to the Q-loop glutamine. Furthermore, the adenylate kinase reaction probably occurs in the engaged dimer and requires the signature motif. These findings indicate

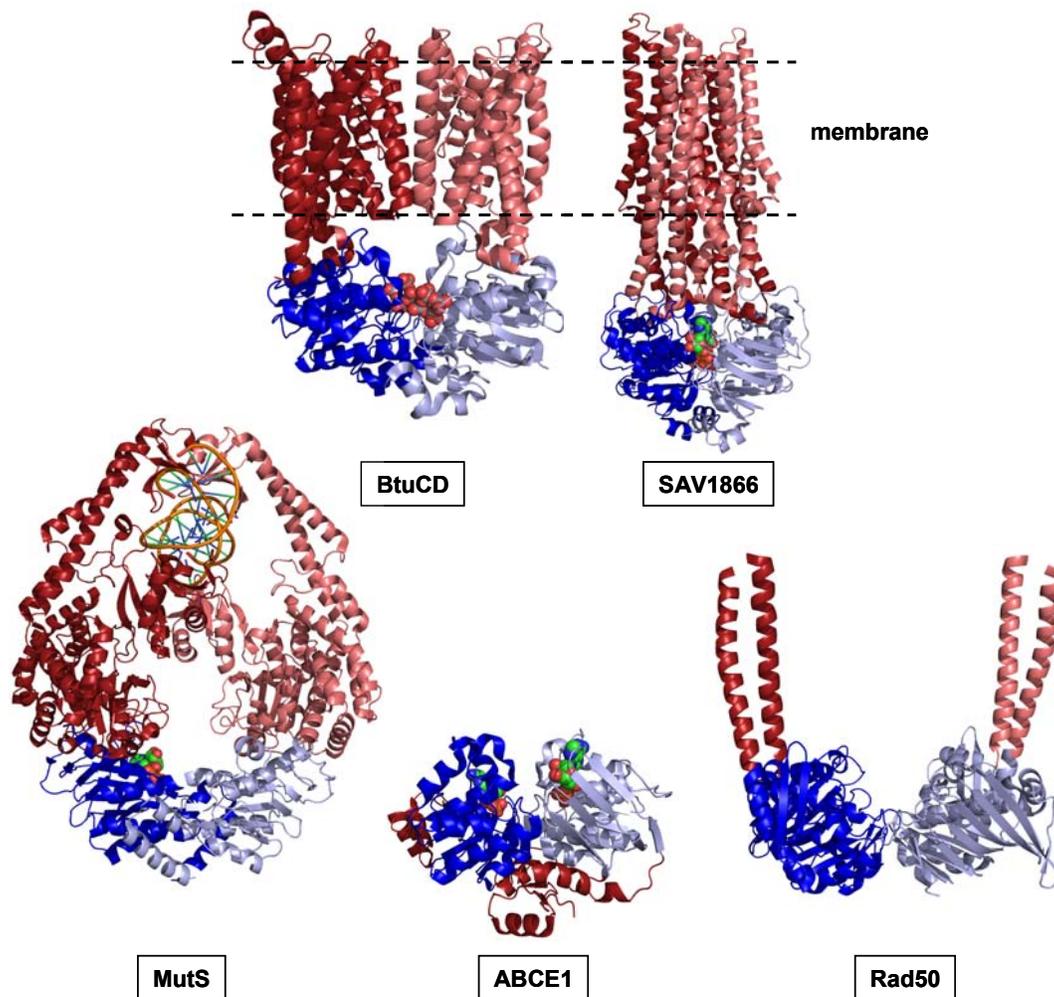
that ATP hydrolysis and adenylate kinase activity of ABC ATPases are achieved by the same set of conserved motifs enabling both reactions in one protein. Different ABC ATPases developed mechanisms to favor one or the other reaction for their biological function.

Taken together this thesis revealed important insights into the mechanism of SMC proteins. The obtained results mark powerful starting points for further experiments on the biological function of SMC proteins. Mutational analysis of the arginine finger discovered in this work was already used to dissect the establishment of sister chromatid cohesion *in vivo*. Furthermore, it could be shown that the ability to catalyze the adenylate kinase reaction is probably an intrinsic ability of all ABC ATPases.

## 2 Introduction

### 2.1 The ABC ATPase family

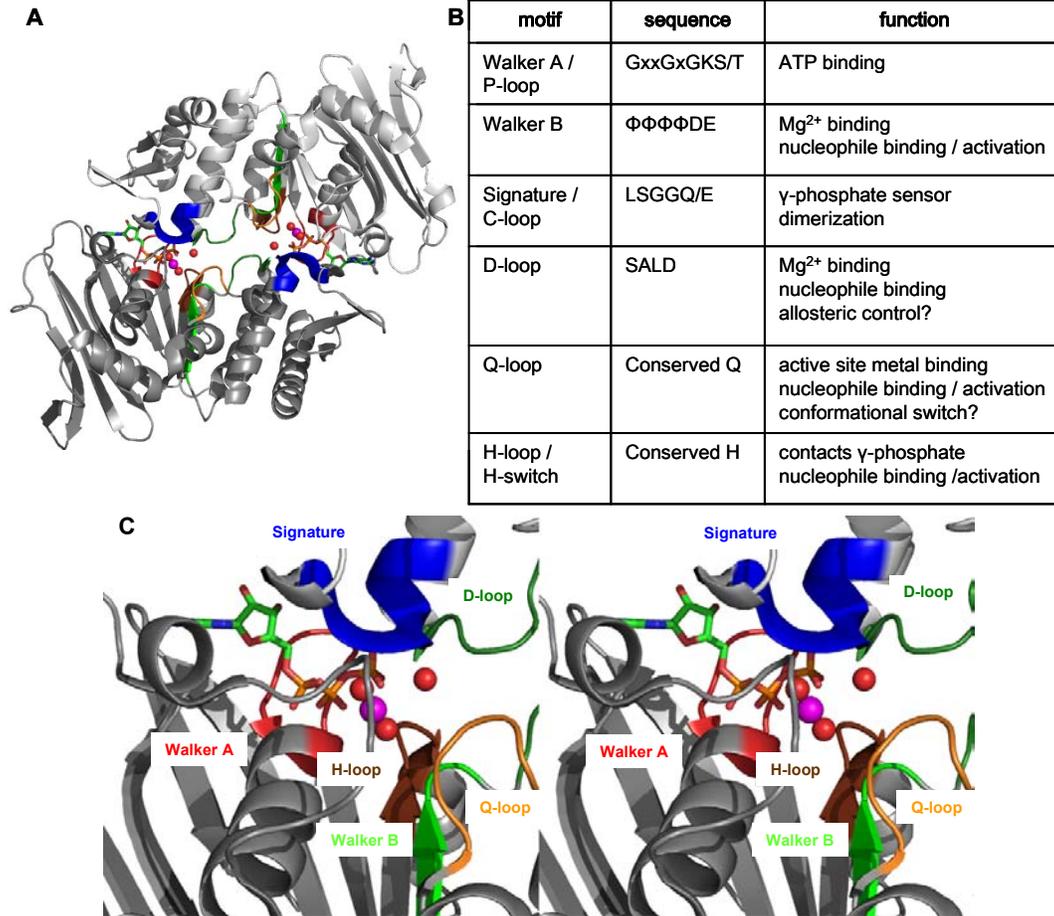
ABC-type ATPases are ubiquitous in all kingdoms of life and are characterized by the conserved ATP-binding cassette (ABC). Originally this type of nucleotide binding domain was described in transmembrane proteins using the energy of ATP hydrolysis to transport various substrates from ions to proteins across the membrane. The histidine permease was the first representative of this family described in 1982. Other notable examples are the P-glycoproteins (MDR-1 and MDR-2) which are over-expressed in tumor cells leading to resistance of the cells against chemotherapy., the vitamin B12 transporter from *E.coli* and the yeast mitochondrial ABC transporter Mdl1 involved in peptide transport (Bradley et al., 1988; Higgins et al., 1982; Locher et al., 2002; Young et al., 2001). ABC transporters share a conserved central architecture of two transmembrane domains (TMD) and two nucleotide binding domains (NBD) to which additional functional domains might be added. The transporters can be encoded by one gene for all four subdomains, two domains can be fused (TMD-NBD, TMD-TMD or NBD-NBD) or all subdomains are encoded by a different gene. ABC transporters can function as importers or exporters. Importers need an additional substrate binding protein that allocates the substrate to be functional (Oswald et al., 2006). The TMDs form the translocation tunnel and are very diverse in sequence thereby affording the broad range of transported substrates. The NBDs or ABC domains contain the conserved motifs for ATP binding and hydrolysis. Beside the high sequence homology the NBDs from different ABC ATPase proteins show structural similarity as seen in various high resolution structures obtained over the last years (for examples see Fig. 1).



**Figure 1: Structures of ABC ATPases.**

Ribbon representation of high resolution structures of ABC proteins obtained by X-ray crystallography. The two NBDs are colored in dark and light blue, additional domains in dark and light red. Nucleotides are depicted as spheres and DNA bound to MutS as sticks. The upper part shows the structures of BtuCD (PDB ID: 1L7V) and of a bacterial multidrug transporter (PDB ID: 2HYD). The predicted position of the membrane is indicated with dashed lines. The bottom row shows structures of ABC proteins not involved in cross membrane transport: the DNA mismatch repair enzyme MutS with bound DNA (PDB ID: 1E3M), in the middle ABCE1 (PDB ID: 1YQT) and Rad50 (PDB ID: 1II8 / monomer structure arranged as dimer for better comparison).

To obtain a functional active site two NBDs have to dimerize “head-to-tail“ upon binding of ATP. This brings the ABC ATPase motifs of two NBDs together, forming two active sites each binding one ATP (Fig.1 and 2) (Hopfner et al., 2000; Obmolova et al., 2000; Smith et al., 2002).

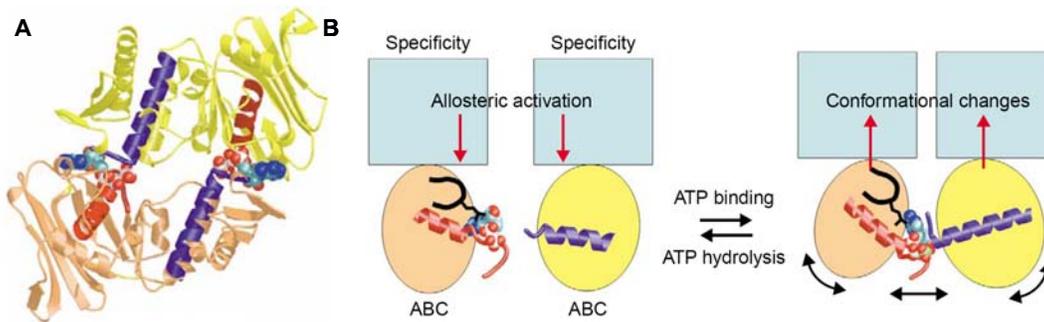


**Figure 2: Conserved ABC ATPase motifs.**

A) Ribbon representation of the Rad50 dimer with bound AMP-PNP as color coded sticks, the Mg<sup>2+</sup> (magenta), the putative nucleophilic water and the 2 waters coordinated by the Mg<sup>2+</sup> as spheres (PDB ID 1F2U). Functional motifs are colored as in C. B) Table listing the functional ABC ATPase motifs with their consensus sequence and the proposed function. X stands for any and Φ represents a hydrophobic amino acid. C) Stereo view of one active site of dimeric Rad50 with annotated motifs in the same representation as in A.

Among the functional motifs the Walker A motif or P-loop is the major ATP binding site. The Walker B motif is involved in ATP hydrolysis (Walker et al., 1982). The third highly conserved motif is the C-loop or signature motif. This loop with the consensus sequence LSGGQ/E is located distant from the Walker motifs in the three dimensional structure of the NBD. In the biological relevant dimer, however, it forms a composite active center with the Walker motifs of the second NBD (Fig. 2). By formation of specific hydrogen bonds of the serine and glycine residues to the γ phosphate, the function of this motif can be described as a specific γ phosphate sensor (Hopfner et al., 2000). The P-loop and signature motif are located at the N-termini of two helices and

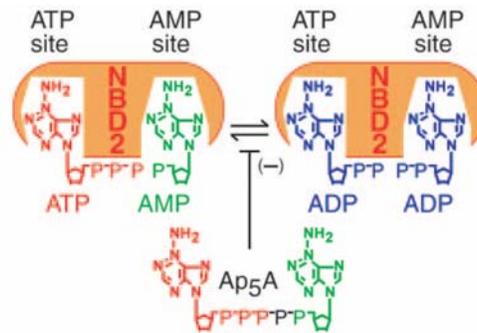
compensate the increased negative charge during ATP hydrolysis (Fig. 3). The D- and Q-loop are involved in binding the active site ion and activating the attacking water molecule. This nucleophile is positioned and activated together with the H-loop form the opposing NBD, therefore hydrolysis can only take place in an engaged dimer.



**Figure 3: Mechanism of ABC ATPases.**

A) The ribbon model of ATP-bound Rad50 that the two NBDs dimerize head to tail and form two composite active sites. The P-loop of one domain and the signature motif of the other compensate the phosphate charge being located at positively charged N-termini of the highlighted helices (red and blue). B) ATP binding to the interface of two NBDs of ABC ATPases leads to rearrangement of the P-loop and signature helices (same color as A), causing domain engagement and conformational changes are conducted to the substrate binding domains (Hopfner and Tainer, 2003).

A second catalytic activity besides the ATP hydrolysis was described for the cystic fibrosis transmembrane conductance regulator (CFTR). The CFTR protein, whose loss of function leads to the cystic fibrosis disease, is structurally related to the ABC transporters and serves as an anion channel (Riordan et al., 1989). In the presence of ATP the isolated NBD2 of CFTR displays ATPase activity but the addition of AMP to the reaction leads to formation of ADP in an adenylate kinase reaction ( $\text{ATP} + \text{AMP} \rightleftharpoons \text{ADP} + \text{ADP}$ ) (Fig.4). This adenylate kinase activity was suggested to regulate the channel opening for the passive flux of the ions instead of control gating via the highly energy consuming ATPase reaction. (Gross et al., 2006; Randak and Welsh, 2003).



**Figure 4: Model of the adenylate kinase activity of CFTR.**

This illustration shows a cartoon of the CFTR-NBD2 with hypothetical ATP- and AMP-binding sites. The indicated reversible adenylate kinase reaction is inhibited by AP5A, a bisubstrate inhibitor (Randak and Welsh, 2005).

The ABC-type ATPase domain is also found in a variety of proteins not involved in transport across membranes. For example in the ABCE1 protein (RLI / HP68) which is one of the most conserved proteins among species. This protein shows a unique arrangement of the ABC domains. It comprises two NBDs on a single polypeptide chain linked by a hinge domain and contains two additional 4Fe4S cluster (Fig. 1). The fundamental cellular function of this protein which makes it essential in all types of organism is not known so far, but it is supposed to play a role in translation (Bisbal et al., 1995; Karcher et al., 2005; Zimmerman et al., 2002).

Another interesting non-transporter ABC protein is the DNA mismatch repair enzyme MutS (Fig. 1). MutS forms homodimers in prokaryotes and heterodimers in eukaryotes and is responsible for the recognition of DNA mismatches. After sensing the mismatch MutS activates MutL in an ATP dependent manner while MutL is stimulating the endonuclease activity of MutH. (Hopfner and Tainer, 2000; Jun et al., 2006; Lamers et al., 2000; Obmolova et al., 2000; Sixma, 2001).

UvrA, a protein involved in nucleotide excision repair, also contains ATPase domains of the ABC type. Although the actual mechanisms of damage recognition, signaling and repair conducted by the UvrABC machinery is not completely dissected so far, ATP binding to UvrA and subsequent dimerization are essential for the repair process (Husain et al., 1986; Oh et al., 1989; Van Houten et al., 2005).

The central protein of the Rad50-Mre11 complex which is involved in double strand break repair is also a member of the ABC ATPase family. Rad50 is characterized by an

ABC ATPase domain from which a long coiled-coil protrudes that may serve as a linker between different DNA strands. This protein was the first case where the question of the dimerization mode of ABC ATPases could be answered by a high resolution structure. The structure of the NBD from *Pyrococcus furiosus* Rad50 with bound ATP revealed the biological relevant “head to tail” dimeric state (Hopfner et al., 2000).

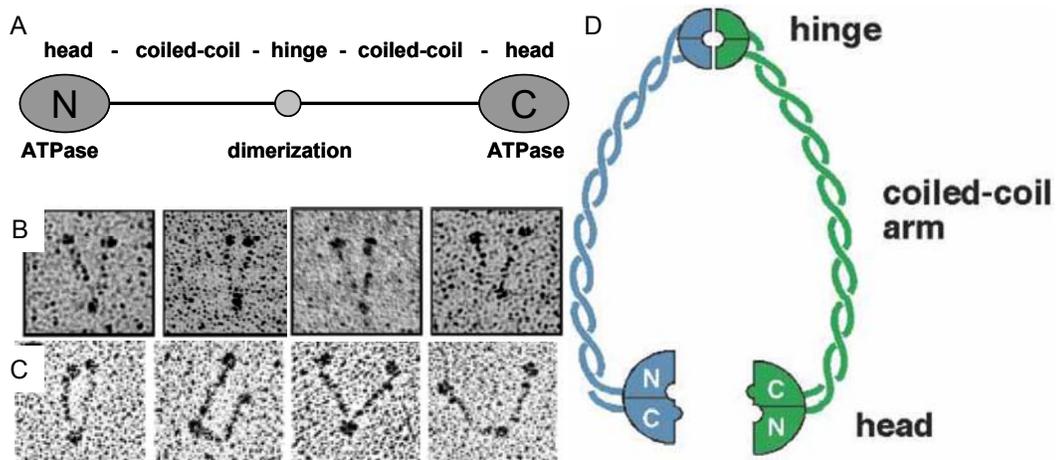
An important subfamily of the ABC ATPases is the structural maintenance of chromosomes (SMC) protein family, containing proteins structurally related to Rad50 which play an essential role in genome integrity.

## 2.2 The SMC protein family

Cell division is one of the most crucial processes for living cells. The genetic information has to be handed over faithfully to the daughter cells during division. Key players involved in this process are protein complexes with SMC proteins as central component. SMC1 was the first protein of the SMC family discovered in 1985 in a mutant yeast strain with impaired maintenance of artificial centromeric minichromosomes (Larionov et al., 1985). Derived from this function the gene was named Stability of Mini Chromosomes. After the discovery of other proteins sharing sequence and structural homology but other functions the name was generalized to Structural Maintenance of Chromosomes. Members of the SMC protein family are ubiquitous in all kingdoms of life emphasizing their importance for genome integrity. In eukaryotes six SMC proteins are known so far, which form the core of different complexes involved in chromosome condensation (SMC2+4), sister chromatid cohesion (SMC1+3) and DNA repair (SMC5+6). Bacterial and archaeal genomes have only one gene for a SMC protein and so far not much is known about the different complexes they might form. According to the mostly accepted model SMC proteins associate with other proteins, forming large proteinous rings that entrap the DNA one or several times to fulfill their function.

## 2.2.1 Architecture of SMC proteins

Members of the structural maintenance of chromosomes (SMC) family share a characteristic domain structure. The 1000 to 1300 residues long polypeptide chain folds into an elongated three domain protein. The N- and C-terminal parts are forming a globular head domain, the middle part folds into a globular hinge domain and a long coiled-coil joins both domains.



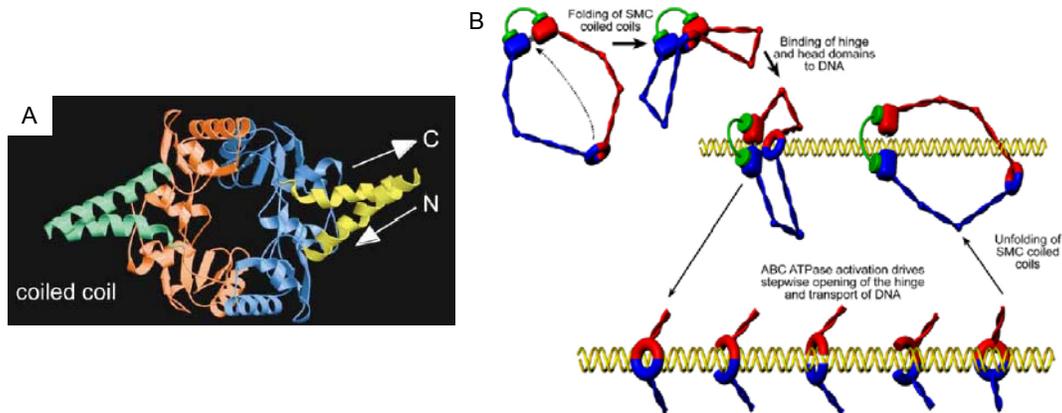
**Figure 5: Architecture of SMC proteins.**

A) Domain organization of SMC proteins. B) Electro micrographs of *S. cerevisiae* SMC1/3 dimer (Haering et al., 2002). C) Electro micrographs of *B. subtilis* SMC homodimers (Melby et al., 1998). D) Model of a SMC dimer (Losada and Hirano, 2005).

SMC proteins belong to the ABC group of ATPases with all typical motifs located at their head domain formed by the N- and C-terminal parts of the protein. Structural analysis of the *Thermotoga maritima* SMC head domain revealed high structural similarity to other members of the ABC ATPase family such as Rad50 catalytic domain, MukB N-terminal domain and histidine permease ATP binding domain (Lowe et al., 2001). To form a catalytically active ATPase site, it has been proposed that two head domains have to dimerize upon ATP binding to form a composite active center as shown for Rad50 (Hopfner et al., 2000). This idea is supported by different biochemical and structural analysis on SMC proteins over the last years. For the activity of SMC proteins ATP binding and hydrolysis capacity is essential (Hirano and Hirano, 1998; Kimura and Hirano, 2000).

The coiled-coil domain protruding from the head domain consists of two anti parallel helices and is  $\sim 500$  Å long. This remarkable structural feature allows SMC proteins to span over a distance equivalent to  $\sim 150$  bp dsDNA. Besides this more structural property at least the coiled-coil of SMC1 has a functional role in DNA damage response. SMC1 contains two serine residues (957 and 966 in human) in a non-helical region of the coiled-coil close to the head domain. It was shown that these residues are phosphorylated *in vitro* and *in vivo* by the checkpoint kinase ATM after ionizing radiation and ATM independently after other DNA damaging events. The phosphorylation of SMC1 in response to DNA damages and S-phase checkpoint activation underlines the central role of SMC proteins in genome maintenance (Kim et al., 2002; Kitagawa et al., 2004; Wakeman et al., 2004; Yazdi et al., 2002).

At the end of the coiled-coil the two helices fold back on each other, forming the hinge domain. Two SMC proteins associate via this hinge, forming a V-shaped heterodimer in eukaryotes and homo dimer in prokaryotes as seen in electron micrographs (Fig. 5B). The crystal structure of the *T. maritima* hinge domain reveals a dimer with coiled-coils protruding from the same site of the molecule which is consistent with a V-shaped molecule (Fig. 6A)(Haering et al., 2002).



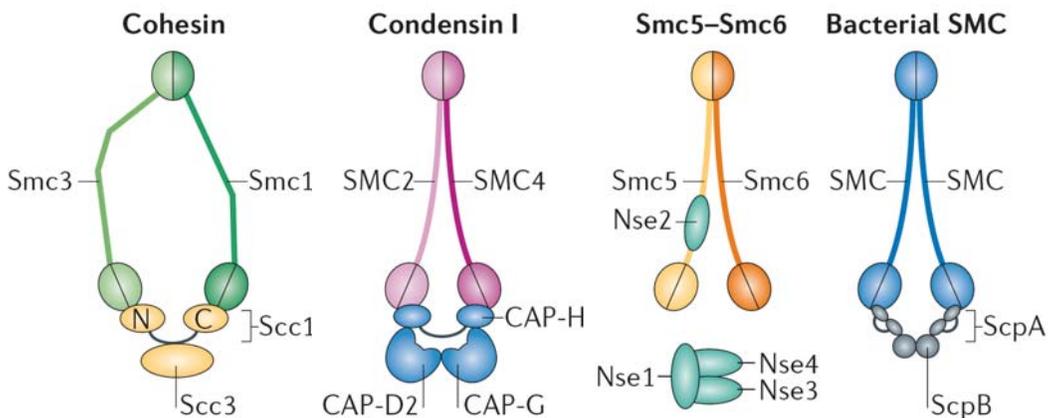
**Figure 6: The SMC hinge domain.**

A) Crystal structure of the dimeric *T. maritima* hinge domain (Haering et al., 2002). B) Model for the entry of DNA into the SMC1/3 complex via the hinge domain (Gruber et al., 2006).

The importance of the dimerization via the hinge domain is emphasized by mutational analyzes in which mutation impeding dimerization of *B. subtilis* SMC proteins abolish their interaction with DNA (Hirano and Hirano, 2002). Furthermore, it was observed

that the hinge domain of *B. subtilis* is able to interact with DNA, thereby stimulating the ATPase activity (Hirano and Hirano, 2006). Additionally, it was shown that SMC 1 and 3 can no longer associate with DNA, when connected via an artificial hinge. Based on these findings a model was proposed where the DNA enters the SMC ring through the hinge domain by stepwise opening of the hinge (Fig. 6B) (Gruber et al., 2006).

Different complexes between SMC proteins and various non-SMC proteins result in different cellular functions. An overview of the complexes is shown in figure 7. The relationship between the names for the proteins in different eukaryotes is given in table 1 of the appendix.



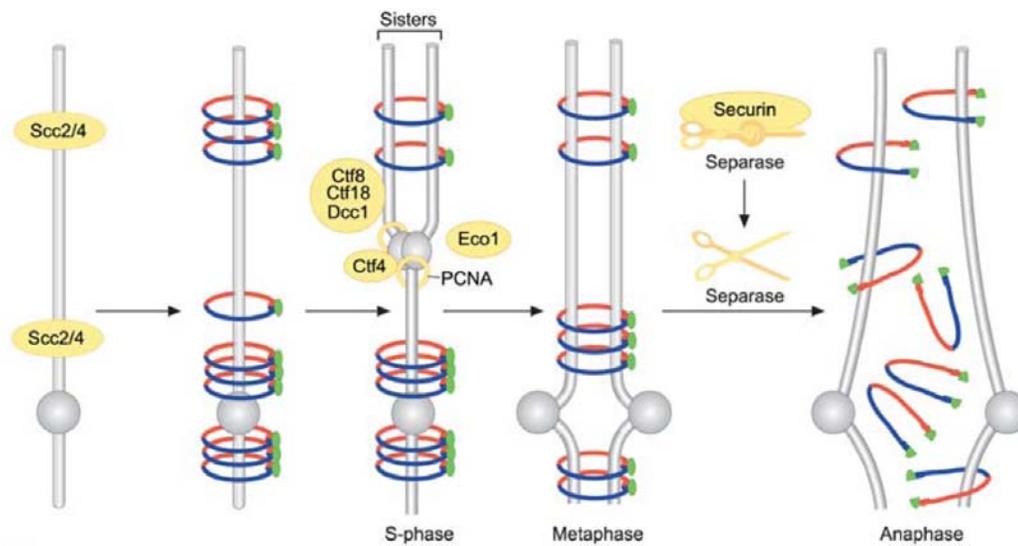
**Figure 7: Subunit composition of SMC protein complexes.**

Architecture and composition of the different SMC protein complexes described in sections 2.2.2 to 2.2.5 (Losada and Hirano, 2005).

## 2.2.2 The SMC 1/3 cohesin complex

The best characterized SMC protein complex is the cohesin complex consisting of SMC1 and SMC3 and the non-SMC proteins SCC1 and SCC3. This complex holds the sister chromatids together till the onset of anaphase (Michaelis et al., 1997). For proper sister chromatid cohesion the SMC1/SMC3/Sccl/Sccl3 complex has to be loaded in late G1 phase before DNA replication starts. The replication fork has then to slide either through the ring structure of the cohesion complex or the ring is somehow opened in front of the replication fork and is held in close proximity to be re-associated after passage of the replication machinery (Carson and Christman, 2001; Lengronne et al., 2006). For the association of the cohesion complex with chromosomes ATP

hydrolysis is absolutely essential and it was proposed that the entry door to the ring is perhaps the hinge region (Arumugam et al., 2003; Gruber et al., 2006; Weitzer et al., 2003). The binding and loading of the preassembled cohesin complex is mediated by the Scc2/Scc4 protein complex that is associated to distinct regions of the chromosomes (Ciosk et al., 2000; Gillespie and Hirano, 2004; Takahashi et al., 2004). Additionally, the Eco1 and some other proteins are essential for establishing sister chromatid cohesion during S phase and couple cohesion to PCNA (proliferating cell nuclear antigen) dependent DNA replication (Skibbens et al., 1999; Toth et al., 1999).



**Figure 8: Model of the cohesin cycle in yeast.**

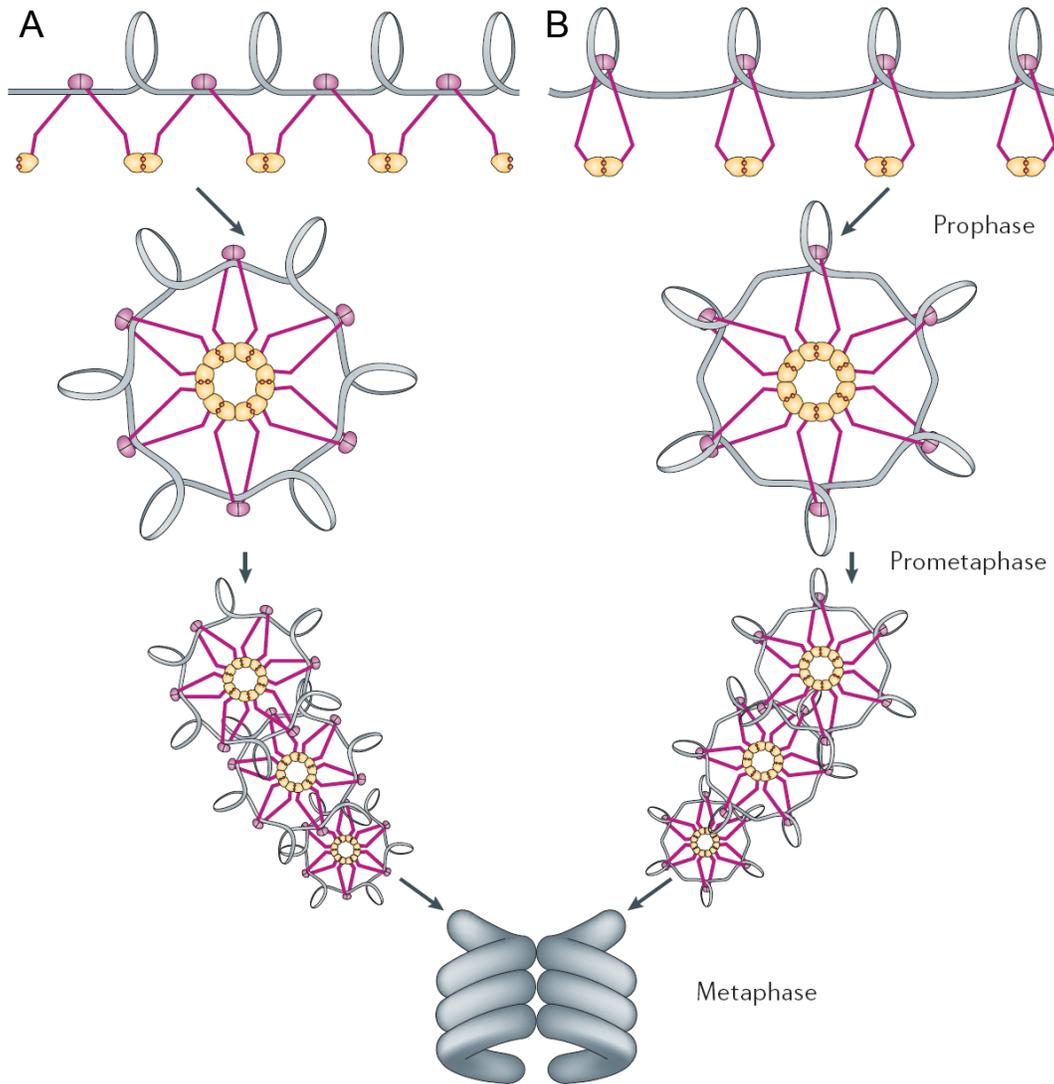
Cohesin (SMC in blue and red, non-SMC subunits in green) is loaded onto the chromosomes with the help of Scc2/Scc4 and sister chromatid cohesion is established by interaction of Eco1 with the replication fork. At anaphase onset cleavage of Scc1 by separase leads to ring opening and sister chromatid segregation (Nasmyth and Haering, 2005).

From the area of loading the cohesion rings are relocated by the replication forks to places of convergent transcription (Lengronne et al., 2004). If the chromosomes are properly aligned on the mitotic spindle and no DNA damages are encountered the APC/C (anaphase promoting complex/cyclosome) marks securin for degradation by ubiquitylation. A protease called separase is no longer blocked by its inhibitor securin and cleaves the Scc1 subunit of the cohesin complex. The cohesin ring is open and the sister chromatides can be segregated to opposing poles of the cell (Fig. 8) (Uhlmann, 2001; Uhlmann et al., 1999; Uhlmann et al., 2000). During meiosis Scc1 is replaced by Rec8, that is cleaved in the first division only distal to crossing over events, to allow

segregation of homologues while the proximal complexes are cleaved only in the second meiotic division to allow sister segregation (Klein et al., 1999; Watanabe and Nurse, 1999).

### 2.2.3 The SMC 2/4 condensin complex

The core of the condensin complex is formed by SMC2 and SMC4. Together with three non-SMC proteins called CAP-D2, CAP-G and CAP-H in condensin I and CAP-D3, CAP-G2 and CAP-H2 in condensin II two different subtypes of condensin complexes are formed. The architecture of the functional complexes is similar to that of cohesion but their function is different. Condensin is responsible for the compaction of the chromosomes before cell division to obtain shorter and mechanically more stable chromosomes that can be handled by the cell (Hirano et al., 1997; Hirano and Mitchison, 1994; Steffensen et al., 2001). The condensin complexes are loaded onto the chromosomes during early mitosis where condensin II initiates the early stage of condensation and condensin I is recruited to the chromosomes after nuclear envelope break down (Gerlich et al., 2006; Hirano et al., 1997; Hirota et al., 2004; Kireeva et al., 2004; Ono et al., 2003). How the compaction of the chromosomes is achieved in the cells is not clear so far. *In vitro* experiments could prove the introduction of positive supercoils into closed circular DNA in the presence of condensin complex, ATP and Topoisomerase I. This reaction is a good candidate for the condensation of DNA *in vivo*. The kind of interaction of the complexes with the chromosomes leading to superhelical torsion is however unclear. Condensin might form rings like cohesin, entrapping different regions of one chromatid, or it might form filaments that loop the DNA (Hagstrom et al., 2002; Hirano, 2006; Kimura and Hirano, 1997; Kimura and Hirano, 2000).



**Figure 9: Working model of the condensin action.**

After binding of condensin (magenta: coiled-coil / hinge and yellow: head) to DNA intermolecular head-head interactions can lead to nucleoprotein filaments in which super helical tension is trapped (A) or chiral loops might be produced by intramolecular head-head engagement (B). In a next step protein-protein interaction could form rosette like structures that then stack to form the chromatin fiber (Hirano, 2006).

A second role of condensin besides of the compaction of DNA for cell division seems to be the dosage compensation. The investigation of the dosage compensation complex in *C.elegans* revealed a high degree of homology to the composition of the condensin complex (Chuang et al., 1994; Lieb et al., 1998; Lieb et al., 1996). Dosage compensation via this complex might be achieved through compaction of the DNA

regions of the X chromosome with genes that are intended to be silenced (Martinez-Balbas et al., 1995).

#### 2.2.4 The SMC 5/6 DNA repair complex

The third SMC complex in eukaryotes was found in a genetic screen of a radiation-sensitive *S. pombe* mutant (Nasim and Smith, 1975; Phipps et al., 1985). Rad18 / SMC6, the complementing gene product in this screen, dimerizes with SMC5 and forms together with the four non-SMC subunits Nse1-4 the functional complex. The whole complex is assembled by the binding of Nse2 to the coiled-coil of SMC5 and recruitment of a complex of Nse1, Nse3 and Nse4 to Nse2. This architecture is strikingly different from the two other complexes. Furthermore, two non-SMC proteins involved in the complex show functional motifs in their primary structure. Nse1 contains a RING finger motif conserved in E3 ubiquitin ligases and Nse2 a zinc finger motif characteristic for SUMO ligases (Fujioka et al., 2002; McDonald et al., 2003). The importance of these activities for the complex function is underlined by the fact that mutations in the Nse2 RING finger impairing sumoylation of the other subunits and especially decreased level of SMC6 sumoylation lead to radiation sensitive but viable cells. Complete loss of any of the proteins involved in this complex leads to accumulation of DNA damages and abnormal structures in the cell over a few rounds of divisions and finally to cell death (Zhao and Blobel, 2005). An additional role of the SMC5/6 complex in segregation of repetitive chromosome regions was reported (Torres-Rosell et al., 2005). However, the distinct function of the SMC5/6 complex in DNA repair is not known so far but its involvement in the maintenance of genome integrity is definite.

#### 2.2.5 The bacterial SMC / Kleisin complex

In bacteria and also archaea the core of the SMC complex is formed by a homodimer of SMC proteins. The non-SMC proteins of the functional complex are named ScpA and ScpB for “segregation and condensation protein” and at least the gene for ScpA can be

found in all bacteria. Deletion of either SMC, ScpA or ScpB in *B.subtilis* leads to disorganization of nucleoids, changes in superhelicity of plasmids and nucleoid partitioning effects and are finally lethal (Britton et al., 1998; Graumann, 2000; Graumann, 2001; Mascarenhas et al., 2002). The stoichiometry of the complex is not clear so far. ScpA is able to interact directly with the head domains of the SMC core whereas ScpB binds to ScpA. The subcomplexes themselves can multimerize so that the biological composition of the SMC-kleisin complex still has to be determined (Dervyn et al., 2004; Hirano and Hirano, 2004). Due to the fact that bacterial SMC proteins form homodimers especially *B.subtilis* proteins were used to investigate the mechanism of SMC proteins. Using this model it was found that dimerization of the hinge domain is essential for interaction with DNA (Haering et al., 2002; Hirano and Hirano, 2002). Furthermore, the activity of SMC protein complexes to bind and restructure DNA is depending on ATP binding and hydrolysis (Hirano and Hirano, 1998). The ATPase activity itself is stimulated by interaction of DNA with the hinge region and is influenced by the kleisins (Hirano and Hirano, 2004; Hirano and Hirano, 2006).

### 2.2.6 The bacterial MukB/E/F complex

MukB is the homologue of the SMC proteins in  $\gamma$ -proteobacteria (Melby et al., 1998; Niki et al., 1991). Although having no sequence similarities the structure and function is similar to that of SMC proteins. It also forms V-shaped homodimers connected by their hinge domains and is responsible for DNA condensation in complex with the non-SMC proteins MukE and MukF (Niki et al., 1992; Yamanaka et al., 1996; Yamazoe et al., 1999). Knockout of any of the three proteins leads to inability to grow above 30°C, diffuse nucleoids and high incidence of anucleate, elongated or filamentous cells (Sunako et al., 2001).

## 2.3 Structure determination by X-ray crystallography

The function of a protein is directly related to its three dimensional structure. To obtain insights into the three dimensional structure of a protein and thereby being able to link

its structure to its function, three different methods are available today. One method is electron microscopy with relatively low resolution around 10 down to so far  $\sim 4$  Å, but it is applicable for big proteins and complexes. A second method is nuclear magnetic resonance spectroscopy (NMR) which is able to determine the atomic structure of small proteins in solution (usually around 30 kDa up to 50 kDa). The third technique is the structure determination by X-ray diffraction using protein crystals. This method can be used to obtain high resolution structures without size limitations. The latter method was used in this work and will be briefly described in the following section. For more detailed information several textbooks are available (e.g. (Blow, 2002; Drenth, 1999; McPherson, 2001)).

### 2.3.1 Physical and mathematical background

Two distinct points can only be recognized as individuals if the wavelength of the electromagnetic radiation used for observation is in the range of the distance of the two (Abbe limit). In the case of proteins this means that the wavelength has to be in the range of 1 Å corresponding to the typical bond length of e.g. 1.497 Å for a C $\alpha$ -C $\beta$  bond (Engh, 1991). Radiation with a wavelength between 0.1 and 10 Å is called X-ray radiation. For structure determination this high energy radiation is produced with rotating copper anodes or at synchrotrons. The interaction between X-rays and matter is very weak and so the prerequisite for structure determination using X-rays is the growth of protein crystals. In the crystal proteins are periodically repeated in a three dimensional lattice resulting in enough scattering power to be able to detect a diffraction signal. X-rays interact with the electrons of the proteins and are emitted in a sphere. The interference of the waves emitted from all electrons in the crystal lead to a distinct diffraction pattern with only a few regions of positive interference. These regions of positive interference are recorded as spots on a detector with their intensity and location. The reflections are related via the Bragg's law (equation 1) to a set of parallel lattice planes in the reciprocal space and are classified by the Miller indices h, k and l.

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

The scattered waves positively interfere and lead to detectable spots if the distance  $d$  and the angle  $\theta$  between the plane and the incident beam fulfill this condition with  $n$  is an integer and  $\lambda$  the wavelength.

The intensity of a reflection  $(hkl)$  is proportional to the square of the structure amplitude  $|F_{hkl}|$ . The structure factor  $F$  mathematically describes the scattering of the incident beam by the crystal (equation 2).

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

The scattering power of a certain atom  $j$  with the coordinates  $(x,y,z)$  depends on the atom type and is described with the atomic scattering factor  $f_j$ . Additionally the thermal disorder is taken into account by the B- or Debye-Waller factor (last term of the equation).

Because X-ray waves are diffracted by the electrons of the atoms the structure factor is related to the repetitive electron density pattern of the crystal by a Fourier transformation (FT). Applying a Fourier synthesis one can calculate the structure factors out of the electron density. In the diffraction experiment one wants to determine the electron density of a protein from the diffraction pattern to model its structure. By measuring the intensities of the reflection one obtains the amplitudes of the structure factors  $|F_{hkl}|$ . Before an inverse FT (equation 3) can be applied to the structure factors their complex value has to be completed by the phases ( $\alpha$ ) which are lost during recording the reflections (equation 4).

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

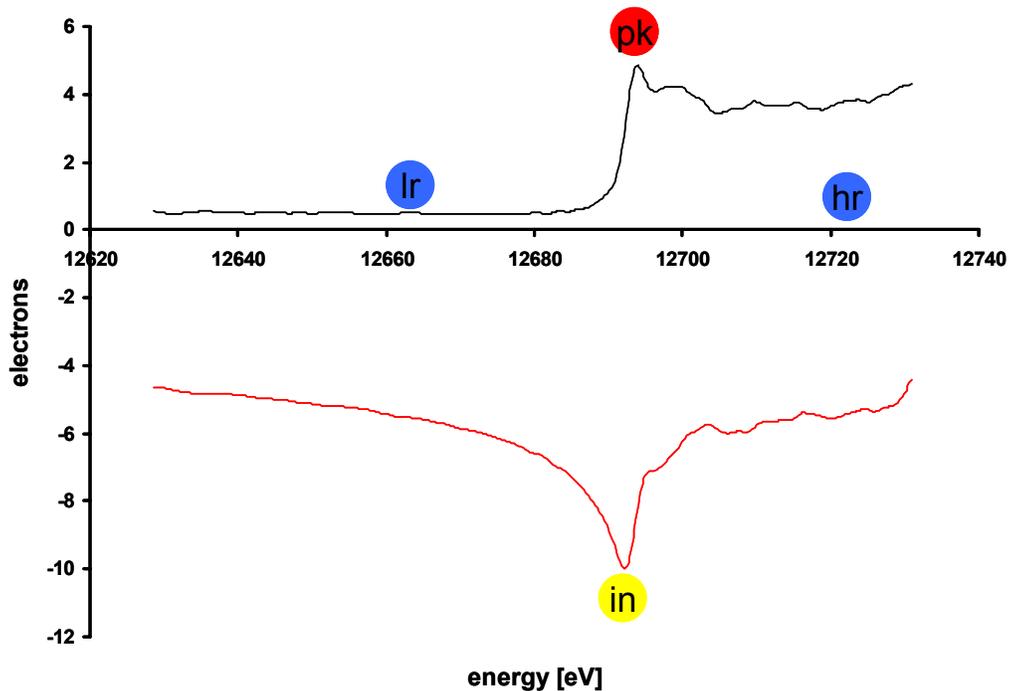
$$F_{hkl} = |F_{hkl}| \cdot \exp[i \cdot \alpha_{hkl}] \quad (4)$$

Obtaining this phase information is referred to as the “phase problem” in crystallography and can be achieved in different ways. If a structure of a similar molecule is known that model can be used to determine the phases by molecular replacement. *De novo* phasing procedures use the altered diffraction properties of heavy

metals for isomorphous replacement (SIR/MIR) or anomalous dispersion (SAD/MAD). The phasing methods used in this work are briefly described in the following chapters.

### 2.3.2 Single- and multi-wavelength anomalous diffraction

For normal diffraction experiments it is assumed that the emitted waves have the same phase and energy as the exciting beam (elastic scattering). The assumption that the scattering of the X-ray beam is elastic is only true to a certain extent because the electrons are not free oscillators but damped in an electric field and have an eigenfrequency. If the energy of the incident photons is close to this eigenfrequency of the electrons they can be absorbed and re-emitted at lower energy (fluorescence) or with altered phase (anomalous dispersion). If the absorption of a certain element is plotted against the energy of the X-ray beam the curve shows characteristic peaks at the so called absorption edges (Fig. 10).

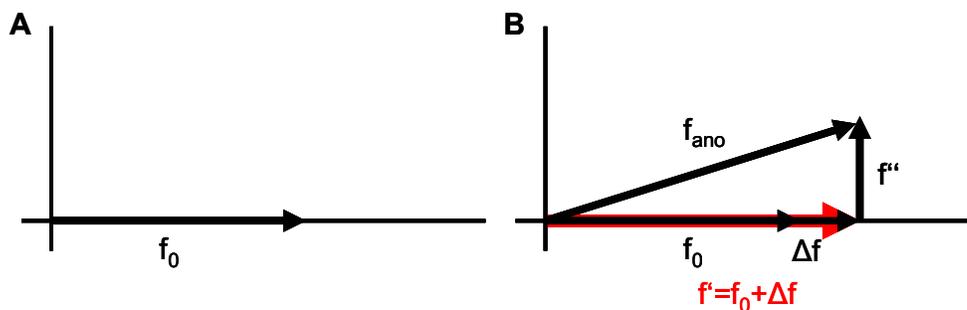


**Figure 10: Scattering factors as function of X-ray energy.**

The scattering factors  $f''$  (red) and  $f'$  (black) of a typical selenium crystal are shown as functions of the X-ray energy. The energy / wavelength for diffraction experiments are indicated as pk: peak wavelength (maximal  $f''$ -SAD+MAD); in: inflection point (minimal  $f'$ -MAD); lr and hr: low and high remote (anomalous contribution low-MAD).

At these edges the anomalous scattering is maximal whereas at lower and higher wavelength this contribution is minimal. Elements normally occurring in biological molecules such as carbon, nitrogen, oxygen do not have their absorption edges in the range of X-ray wavelengths used in diffraction experiments. For this reason heavy atoms (e.g. selenium, mercury, platinum) that have their transitions in the appropriate energy range are introduced into proteins to perform anomalous dispersion experiments. Because the actual energy where the transition occurs depends on the environment of the atoms the required wavelengths have to be determined for every experiment by a fluorescence scan. From this scan the values of  $f''$  are derived directly and are related to  $f'$  by the Kramers-Kronig equation.

$$f_{ano} = f_0 + \Delta f + if'' = f' + f'' \quad (5)$$



**Figure 11: The atomic scattering factor.**

A) Atomic scattering factor  $f_0$  for a free electron. B) Anomalous atomic scattering factor  $f_{ano}$  of a bound electron with the anomalous contribution consisting of a real part  $\Delta f$  and an imaginary part  $f''$ .

The anomalous diffraction results in a change of the atomic scattering factor that gains an anomalous contribution consisting of a real part  $\Delta f$  and an imaginary part  $if''$  that is dependent on the wavelength (Fig. 11). This anomalous contribution to the scattering factors leads to the break down of Friedel's law. In normal diffraction experiments the Friedel mates  $F_P(hkl)$  and  $F_P(-h-k-l)$  of a protein P have the same amplitudes but opposite phases. The imaginary  $f''$  contribution from a heavy atom H in an anomalous scattering experiment leads to a difference between  $F_H(hkl)$  and  $F_H(-h-k-l)$  and therefore in the whole  $F_{PH}$ .

This difference  $\Delta|F_{PH}|_{ano}$  (equation 6) between the now called Bijvoet pairs can be used to solve the heavy atom substructure by direct methods or using Patterson space.

$$\Delta|F_{PH}|_{ano} = \left( |F_{PH}(hkl)| - |F_{PH}(-h-k-l)| \right) \cdot \frac{f'}{2f''} \quad (6)$$

A Patterson function calculated with the squared Bijvoet difference amplitudes should contain only peaks corresponding to interatomic vectors between pairs of anomalous scattering atoms. The interatomic vectors of symmetry related atoms give the highest peaks and lie on the Harker sections whose coordinates (uvw) are depending on the space group. Using these peaks the anomalous scatterer can be located in the unit cell.

With the heavy atom substructure the phases  $\alpha_H$  and the whole structure factor of the heavy atoms are known. Together with the measured amplitudes for  $|F_{PH}|$  and  $|F_P|$  the phase contribution of the protein can be derived according to equation 7 and an electron density calculated by a FT.

$$F_{PH} = |F_P| \cdot \alpha_P + |F_H| \cdot \alpha_H \quad (7)$$

### 2.3.3 Molecular replacement

If a three dimensional model is known that is similar (approximately 1 Å rmsd) to the structure this has to be solved one can use that model for obtaining initial phases. To calculate the phases the model has to be oriented in the unit cell corresponding to the molecule whose structure has to be solved. The search for this orientation is a six dimensional problem because three rotational and three translational parameters have to be determined. Using the Patterson function (equation 8) the rotation and the translation of the model can be decoupled and performed one after each other.

$$P(uvw) = \frac{1}{V} \sum_{hkl} |F_{hkl}|^2 \exp[-2\pi \cdot i(hu + kv + lw)] \quad (8)$$

The calculation of the Patterson map for the rotational search is performed in an arbitrary large unit cell and only for peaks within a radius around the origin in the size of the model to eliminate intermolecular cross vectors. The remaining intramolecular self vectors are oriented on each other by the rotation function. The cross correlation

function of the calculated Patterson densities has a maximum if the model and the unknown structure are equally oriented in space. After the right rotation operators are found the translation function considering the cross vectors which have been excluded before is used to finally position the molecule in the unit cell. With the help of the correct positioned model the phases  $\alpha_{\text{calc}}$  can be calculated. This phase information together with the measured amplitudes  $|F_{\text{obs}}|$  determines the structure factor  $F$  of the protein (equation 9).

$$F = |F_{\text{obs}}| \cdot \exp[i\alpha_{\text{calc}}] \quad (9)$$

The phase information in this term is calculated with the help of a model that is different from the actual molecule and therefore bias the calculated electron density. To reduce this model error a weighted electron density map, e.g. a  $2F_o - F_c$  map, is used for model building. Additional and missing parts can be assigned in the difference electron  $F_o - F_c$  density map.

### 3 Objectives

Proteins of the structural maintenance of chromosomes (SMC) family play an important role in genome maintenance. They are present in all organisms and form the core of multiprotein complexes facilitating DNA repair and equal distribution of the genetic information onto the daughter cells. The current model presumes that the SMC proteins form large proteinous rings that trap the DNA to fulfill their function. For the cohesin complex it is known how the DNA is released from the ring during cell cycle but the question remains how the DNA enters the ring. Although it was shown that ATP binding and hydrolysis is essential for the function of SMC proteins its distinct role remained unclear.

To gain insights into the mechanism of SMC proteins on molecular level the aim of this work was to clone, express, purify and crystallize the nucleotide binding domain (NBD) of an archaeal SMC protein. The role of ATP binding and hydrolysis should be clarified by crystallizing the NBD in presence of ATP and transition state analogues.

Furthermore, after examination of the protein structure the role of important residues for biological function should be confirmed by *in vitro* assays of wild type and mutant protein. Possible non-SMC subunits of the archaea should be identified, cloned and expressed to study their influence on the biochemical activity of the NBD. Furthermore, the question whether DNA-binding modulates the activity of the SMC proteins should be addressed.

The protein should also be used to gain structural insights into the puzzling adenylate kinase activity described for the CFTR protein. This protein belongs to the family of the ABC ATPase as the SMC proteins do and therefore possess the same functional motifs that might be involved in this reaction.

## 4 Materials

### 4.1 Chemicals

Unless otherwise stated all chemicals were purchased from Merck AG (Darmstadt), Carl Roth GmbH (Karlsruhe) or Sigma-Aldrich (Deisenhofen) in the highest available purity. Crystallization screens and reagents were ordered from Hampton Research (Aliso Viejo, USA), Nextal Biotechnologies (now Qiagen, Hilden) and Jena Bioscience (Jena). Radioactive  $\gamma^{32}\text{P}$ -ATP was purchased from GE Healthcare.

### 4.2 Enzymes, standards, Kits and chromatographic material

BenchMark Protein Ladder	Invitrogen, Karlsruhe
Calf Intestine Alkaline Phosphatase	Fermentas, St. Leon-Rot
GeneRuler 1kb DNA Ladder	Fermentas, St. Leon-Rot
HiTrap Q HP	GE Healthcare, Uppsala Sweden
Ni-NTA-Agarose	Qiagen, Hilden
NucleoSpin Extract II	Macherey-Nagel, Düren
NucleoSpin Plasmid Quick Pure	Macherey-Nagel, Düren
Platinum <i>Pfx</i> DNA Polymerase	Invitrogen, Karlsruhe
Protein Molecular Weight Marker	Fermentas, St. Leon-Rot
Restriction Endonucleases	Fermentas, St. Leon-Rot
Superdex S200	GE Healthcare, Uppsala Sweden
Sephacryl S300	GE Healthcare, Uppsala Sweden
Source 15 S	GE Healthcare, Uppsala Sweden
T4 DNA Ligase	Fermentas, St. Leon-Rot

Taq DNA Polymerase

Fermentas, St. Leon-Rot

Unless otherwise stated all enzymes and kits were used according to the manufactures instructions.

### 4.3 Oligonucleotides

SMC for NdeI	5'-GGA AAA AAC <u>ATA TGC</u> CCT ACA TAG AGA AGC TTG AAC-3'
SMC rev NotI	5'-GAA AAA <u>AGC GGC CGC</u> TTA TCA ATT TCC ATG TTC CCA CCC-3'
SMC E182 rev EcoRI	5'-GAA <u>GAA TTC</u> TTA TCA TTC CTC TAA AGC TTT CTC TTT <u>CTT AG</u> -3'
SMC E1006 for NcoI	5'-GAA AAC <u>CAT GGA</u> AAA AGA AAA GAA AAA TGT C-3'
SMC K39A for	5'-GCT AAT GGA TCT GGA <u>GCG</u> TCT AAC ATT GGA GAT GCT ATT CTC-3'
SMC K39A rev	5'-GAG AAT AGC ATC TCC AAT GTT AGA <u>CGC</u> TCC AGA TCC ATT AGC-3'
SMC R59A for	5'-GTT ATC AGC AAA GGC CAT <u>GGC AGC</u> AAG TAG AAT CAG TGA TCT CAT ATT TG-3'
SMC R59A rev	5'-CAA ATA TGA GAT CAC TGA TTC TAC TTG <u>CTG</u> <u>CCA TGG CCT TTG CTG</u> ATA AC-3'
SMC Q145A for	5'-GGT TAT AAT ATA GTT CTC <u>GCG</u> GGA GAT ATA ACG AAG TTC-3'
SMC Q145A rev	5'-GAA CTT CGT TAT ATC TCC <u>CGC</u> GAG AAC TAT ATT ATA ACC-3'
SMC S1070R for	5'-GAG AAT TGA AGC TAT <u>GAG AGG</u> TGG GGA AAA AGC-3'
SMC S1070R rev	5'-GCT TTT TCC CCA <u>CCT CTC</u> ATA GCT TCA ATT CTC-3'
SMC E1098Q for	5'-CAT TTT ATC TCT TCG ATC <u>AGA</u> TAG ATG CTC ATT TAG ATG ATG CAA ACG-3'
SMC E1098Q	5'-GAG CAT CTA <u>TCT GAT</u> CGA AGA GAT AAA ATG GAG CAG G-3'
ScpA for NdeI	5'-AAA AAA <u>ACA TAT GGA</u> AAT TGA GGT TAC TCC TGT AG-3'
ScpA rev NotI	5'-AAA AGC GGC CGC TCA TTA CAT TGG AAC TAC TAA GAT CTC G-3'
ScpB for NdeI	5'-GGA AAA AAC <u>ATA TGG</u> GAC TAA TAG AGG ATA AAG CC-3'
ScpB rev NotI	5'-CCT TTT TTG <u>CGG CCG</u> CTT ATC ATT CCT TTT TCT TTT CAC TTT CC-3'
T7 promotor	5'-TAA TAC GAC TCA CTA TAG GG-3'

T7 terminator 5' -TAT GCT AGT TAT TGC TCA G-3'  
 holl for 5' -GGC GAC GTG ATC ACC AGA TGA TGC TAG ATG  
 CTT TCC GAA GAG AGA GC-3'  
 holl rev 5' -GGC GAC GTG ATC ACC AGA TGA TGC TAG ATG  
 CTT TCC GAA GAG AGA GC-3'

All oligonucleotides were purchased RP-HPLC-purified and lyophilized from Thermo Electron Corporation (Ulm). Restriction sites and mutated Codons are underlined.

#### 4.4 *E. coli* strains and Plasmids

<u><i>E. coli</i> strain</u>	<u>genotype</u>	<u>Source</u>
XL1 blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacIqZΔM15 Tn10 (Tetr)]	Stratagene, Heidelberg
Rosetta (DE3)	F <sup>-</sup> <i>ompT hsdSB</i> (rB <sup>-</sup> mB <sup>-</sup> ) <i>gal dcm</i> (DE3) pRARE2 (CamR)	Novagen, Madison USA
B834(DE3)	F <sup>-</sup> <i>ompT hsdSB</i> (rB <sup>-</sup> mB <sup>-</sup> ) <i>met gal dcm</i> (DE3)	Novagen, Madison USA

<u>Plasmid</u>	<u>source</u>
pET21b(+)	Novagen, Madison USA
pET28b(+)	Novagen, Madison USA
pETSH35	(Hoepfner, 2005)

## 5 Methods

### 5.1 Cloning

Genes encoding for full-length *Pyrococcus furiosus* SMC, ScpA and ScpB proteins were amplified by PCR from genomic DNA using SMC for NdeI – SMC rev NotI, ScpA for NdeI – ScpA rev NotI and ScpB for NdeI – ScpB rev NotI as primers (see 4.3). A typical PCR-reaction contained the following components:

water	35.5 $\mu$ l
10x <i>Pfx</i> Amplification buffer	5.0 $\mu$ l
10x PCR Enhancer solution	5.0 $\mu$ l
50 mM MgSO <sub>4</sub>	1.0 $\mu$ l
10 mM dNTP's	1.5 $\mu$ l
Template	0.5 $\mu$ l
Primer forward (50 pmol/ $\mu$ l)	0.5 $\mu$ l
Primer reverse (50 pmol/ $\mu$ l)	0.5 $\mu$ l
<i>Pfx</i> DNA Polymerase (2.5 U/ $\mu$ l)	0.5 $\mu$ l

For amplifying the DNA fragments of interest the following program was used:

Initial denaturation	94°C	180 s
Denaturation (30 times)	94°C	30 s
Annealing (30 times)	50°C	30 s
Extension (30 times)	68°C	60 s / 1000 kb
Final extension	68°C	300 s

The N- and C-terminal fragments of the SMC protein were amplified using SMC for NdeI – SMC E182 rev EcoRI and SMC E1006 for NcoI – SMC rev NotI as primers and the cloned full-length SMC gene as a template.

Mutations were introduced via a 2-step PCR reaction using the wild type gene as a template. In the first step the DNA was amplified using SMC for NdeI and the reverse mutagenesis Primer and SMC rev NotI and the forward mutagenesis primer resulting in two PCR fragments. The two PCR products of the first step were purified by agarose gel electrophoreses (see 4.1.2) and 0.5  $\mu$ l of each fragment were used as template in the second PCR reaction. During this overlap PCR in the first three cycles full-length template is synthesized without flanking Primers. After adding the amplification primers (SMC for NdeI - SMC rev NotI) the DNA fragment is amplified in additional 30 PCR cycles.

To analyze and purify the PCR reactions they were mixed with loading dye and DNA fragments were separated with an 1% agarose gel at 10 V/cm with 1x TAE as running buffer. To visualize DNA with UV light (254 nm) 0.01% (v/v) of a 10 mg/ml ethidium bromide solution was added to the gel. Bands of interest were cut out and the DNA was extracted with the NucleoSpin Extract II Kit.

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

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

The extracted PCR products were digested with the appropriate restriction endonucleases and purified with the NucleoSpin Extract II Kit. The target vectors were digested with the corresponding restriction enzymes, dephosphorylated, purified by agarose gel electrophoresis and extracted with the NucleoSpin Extract II Kit. For a typical ligation the following components were mixed and incubated over night at 16°C:

Water	10 $\mu$ l
10x ligation buffer	2 $\mu$ l

---

Plasmid	2 $\mu$ l
Insert	5 $\mu$ l
T4 DNA Ligase (5 U/ $\mu$ l)	1 $\mu$ l

Chemical competent *E.coli* cells were prepared according to Hanahan (Hanahan, 1983). 200 ml LB media was inoculated with 5 ml over night culture and grown to an OD600 = 0.35. The cells were cooled down on ice and pelleted by centrifugation (1900 g, 4°C, 10 min). After resuspension in 30 ml of ice cold TBI buffer and incubation on ice for 10 min the cells were pelleted again. The pellet was resuspended in 4 ml ice cold TBII, aliquoted in 100  $\mu$ l, flash frozen in liquid nitrogen and stored at -80°C.

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

TBI: 100 mM KCl, 50 mM MnCl<sub>2</sub>, 30 mM potassium acetate, 10 mM CaCl<sub>2</sub>, 15% (v/v) glycerol, adjust pH to 5.8 with 0.2 M Acetate

TBII: 10 mM MOPS, 10 mM KCl, 75 mM CaCl<sub>2</sub>, 15% (v/v) glycerol, adjust pH to 7.0 with NaOH

For transformation the cells were thawed and either 20  $\mu$ l of a ligation reaction were mixed with XL1blue competent cells or 0.5  $\mu$ l of a plasmid preparation was mixed with Rosetta (DE3) competent cells and incubated on ice for 30 min. After heat shock at 42°C for 1 min cells were cooled down on ice, 800  $\mu$ l LB media was added to the cells and then incubated for 1 h at 37°C. For selection the cells were spread on LB agar plates containing the appropriate antibiotic (1:1000) and incubated over night at 37°C.

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

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

Kanamycin: 50 mg/ml in water

Chloramphenicol: 50 mg/ml in 100% ethanol

Tetracycline: 12.5 mg/ml in 100% ethanol

For plasmid isolation 5 ml LB media containing the appropriate antibiotics was inoculated with single colonies and incubated over night at 37°C. After pelleting the cells plasmids were isolated with the NucleoSpin Plasmid Quick Pure Kit. The plasmids were tested for successful insertion of the gene of interest by test PCR with *Taq* DNA polymerase:

Water	12.5 µl
10x <i>Taq</i> Buffer with (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	2.0 µl
25 mM MgCl <sub>2</sub>	2.0 µl
10 mM dNTPs	1.0 µl
Plasmid	0.5 µl
Primer forward (50 pmol/µl)	0.5 µl
Primer reverse (50 pmol/µl)	0.5 µl
<i>Taq</i> DNA Polymerase (1 U/µl)	1.0 µl

The cycling was analogous to the *Pfx* DNA polymerase PCR except that the extension temperature was 72°C and the extension time was calculated for 2000 bp/min. The PCR products were analyzed by agarose gel electrophoreses and positive clones were sequenced by Medigenomix (Martinsried). Correct plasmids were used to transform Rosetta (DE3) cells and after positive expression test clones were stored as glycerol cultures (400 µl over night culture + 100 µl 40% (v/v) glycerol).

## 5.2 Expression and purification

All proteins were expressed in Rosetta (DE3) cells growing in LB media. For expression 3 l LB media were inoculated with 50 ml of an over night culture and protein production was induced by addition of 0.05 mM IPTG at an OD<sub>600</sub>=0.6. Cells were harvested after 3 h of induction, flash frozen in liquid nitrogen and stored at -80°C. Selenomethionine substituted protein was expressed analogous using methionine auxotrophic *E.coli* B824(DE3) cells growing in LeMaster media (LeMaster and Richards, 1985).

Cells having expressed SMCcd were resuspended in 20 mM Tris (pH 7.5), 100 mM NaCl, 5 mM EDTA and disrupted by sonication. Insoluble proteins and cell debris were removed by centrifugation for 20 min at 40000 g. The crude extract was incubated at 70°C for 10 min to denature endogenous *E.coli* proteins and subsequently cleared by centrifugation (20 min, 40000 g). To remove nucleic acids, proteins were precipitated by addition of ammonium sulfate to the supernatant to 80% of saturation. After 1 h incubation at 4°C precipitated proteins were pelleted by centrifugation (10 min, 20000 g) and dissolved in buffer A. The protein solution was separated on a Sephacryl S300 26/60 size exclusion column with buffer B. SMCcd containing fractions were pooled and dialyzed over night against 50 volumes 20 mM Tris (pH 8.0), 1 mM EDTA. The dialyzed protein was loaded on a HiTrap Q HP ion exchange column and eluted with a linear gradient over 10 column volumes (cv) from 100% buffer A to 100% buffer B. For further purification the flow-through and SMCcd containing fractions were pooled, loaded on a Source 15 S ion exchange column and eluted with a linear gradient from 100% buffer A to 100% buffer B over 20 cv. Protein containing fractions were pooled, concentrated and finally purified by Superdex S200 16/60 size exclusion chromatography in buffer C. Selenomethione substituted protein was purified analogous with addition of 1 mM DTT to all buffers.

Buffer A: 20 mM Tris (pH 8.0), 100 mM NaCl, 1 mM EDTA

Buffer B: 20 mM Tris (pH 8.0), 1 M NaCl, 1 mM EDTA

Buffer C: 5 mM Tris (pH 8.0), 100 mM NaCl, 0.1 mM EDTA

ScpA containing cells were resuspended in Ni-NTA-A and disrupted by sonication. After centrifugation (20 min, 40000g) the supernatant was loaded on a Ni-NTA gravity flow IMAC column. The column was washed with 10 cv Ni-NTA-W, 5 cv Ni-NTA-A and the protein was eluted with 100% Ni-NTA-B. ScpA containing fractions were pooled and dialyzed over night against 50 volumes 20 mM MES (pH 6.0), 50 mM NaCl, 1 mM EDTA. The protein was loaded on a HiTrap Q HP IEX column and eluted with a linear gradient from 50 mM NaCl to 1 M NaCl in 20 mM MES (pH 6.0), 1 mM EDTA. As a final purification step the protein was concentrated and loaded on a Superdex S200 16/60 size exclusion column with buffer C as running buffer. For complex formation the N-terminal His-tag was cleaved with Thrombin, uncleaved

protein separated by Ni-NTA IMAC and ScpA without His-tag mixed 1:1 with SMCcd protein. After incubation for 30 min on ice formed complexes were separated from free protein via size exclusion chromatography with buffer C.

Ni-NTA-A: 50 mM NaH<sub>2</sub>PO<sub>4</sub> (pH 8), 300 mM NaCl, 10 mM Imidazol

Ni-NTA-B: 50 mM NaH<sub>2</sub>PO<sub>4</sub> (pH 8), 300 mM NaCl, 250 mM Imidazol

Ni-NTA-W: 50 mM NaH<sub>2</sub>PO<sub>4</sub> (pH 8), 1 M NaCl, 10 mM Imidazol

For purification of ScpB cells were resuspended in Ni-NTA-A and disrupted by sonication. The crude extract was cleared by centrifugation (20 min, 40000g) and purified by Ni-NTA IMAC according to ScpA. ScpB containing fractions were pooled, concentrated and diluted with water to 50 mM NaCl before being loaded on a HiTrap Q HP IEX column. After elution the ScpB containing fraction were pooled, concentrated and the protein further purified by size exclusion chromatography (Superdex S200) with 5 mM Tris (pH 8.5), 300 mM NaCl, 0.1 mM EDTA as buffer.

Protein samples were analyzed by discontinuous SDS-PAGE (Laemmli, 1970). After mixing with Laemmli buffer the samples were denatured for 5 min at 95°C, separated in a 15% gel (4% stacking gel) with 1x TGS as running buffer at 30 V/cm and stained with Coomassie.

4x lower buffer: 3 M Tris (pH 8.5), 0.4% (w/v) SDS

4x upper buffer: 0.5 M Tris (pH 6.8), 0.4% (w/v) SDS

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

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

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

Proteins were concentrated by centrifugation (850 g, 4°C) using Amicon Ultra-15 Centrifugal Filter Units (Millipore, Billerica, USA). Concentrations of the protein solutions were determined by UV absorption at 280 nm and the corresponding extinction coefficient obtained by primary sequence analysis at [www.expasy.ch](http://www.expasy.ch).

### 5.3 Analytical size exclusion chromatography

ATP-dependent dimerization of SMCcd wt and E1098Q was tested by analytical size exclusion chromatography with 20 mM Tris (pH 8.0), 100 mM NaCl, 5 mM MgCl<sub>2</sub>, ± 0.5 mM ATP as running buffer. For interaction studies the retention volume of the proteins was determined alone and after mixing with ATP or ScpA and incubation in buffer C. Each time 50 µl of the protein solution was loaded on a Superdex S200 10/300 column at a flow rate of 0.5 ml/min.

### 5.4 ATPase activity assay

To test the ATPase activity 10 µM protein was incubated with and without 25 µM dsDNA (annealed holl for and holl rev) in 50 mM Tris (pH 7.5), 100 mM NaCl, 10 mM MgCl<sub>2</sub> and 1 mM ATP (containing 500 fold diluted  $\gamma^{32}\text{P}$ -ATP) at 55°C for 10 min. Aliquots of 1 µl were spotted on a PEI Cellulose F thin-layer chromatography plate (Merck, Darmstadt) and developed in 0.5 M LiCl, 1 M formic acid. After drying radioactive signals were recorded with a storage phosphor screen (GE Healthcare, Uppsala) and read out with a Storm imaging system (GE Healthcare, Uppsala). Data were processed with ImageQuant Software (GE Healthcare, Uppsala). The ability of ADP, ADP-AlF<sub>4</sub><sup>-</sup> and ADP-BeF<sub>3</sub><sup>-</sup> to inhibit these reactions was tested by addition of 5 mM of the inhibitors to the reaction described above (for preparation of the inhibitors see 5.7).

### 5.5 ATP binding assay

The ATP binding capability of the proteins was assayed by nitrocellulose filter binding assay. 2 µM protein was incubated in 50 mM Tris (pH 7.5), 100 mM NaCl, 10 mM MgCl<sub>2</sub> and 25 µM ATP (containing 500 fold diluted  $\gamma^{32}\text{P}$ -ATP) for 5 min on ice. The mixtures were applied on a nitrocellulose membrane with a dot blot device (Whatman, Brentford, UK). A storage phosphor screen was exposed to the dried membrane and read out with a Storm imaging system. Data were processed with ImageQuant Software.

## 5.6 Reverse adenylate kinase activity assay

1  $\mu$ M protein was incubated with 1 mM ADP in 50 mM Tris (pH 7.5), 100 mM NaCl and 10 mM MgCl<sub>2</sub> at 55°C for 2 hours. Wild type protein was assayed with and without addition of 2.5 mM AP5A (Sigma-Aldrich, Taufkirchen, Germany). To determine the amount of produced ATP, 100  $\mu$ l of the reaction mixture were mixed with 100  $\mu$ l of the “ATP determination kit” reagent from Biaffin (Kassel). The ATP generated in the reverse adenylate kinase reaction was used by firefly luciferase to oxidize D-luciferin. The resulting chemiluminescence was recorded and integrated over 10 s with a Lumat LB 9507 luminometer (Berthold Technologies, Bad Wildbad) immediately after mixing.

## 5.7 Crystallization

Initial screening for crystallization conditions was performed in CrystalEX 96 well sitting drop plates (Corning, New York, USA) using commercially available sparse matrix screens from Hampton, Jena Bioscience and Nextal. Setups were pipeted either by hand mixing of 1  $\mu$ l protein solution with 1  $\mu$ l crystallization buffer or by a Hydra II semi-automatic protein crystallization robot (0.2  $\mu$ l + 0.2  $\mu$ l) (Matrix Technologies Corporation, Hudson, USA). To improve crystal quality the conditions of initial hits were refined by varying the protein concentration, concentrations of the components of the crystallization buffer, pH, temperature, drop size, mixing ratio of the drop and ratio between drop and reservoir volume. Refinements were set up in 24 well format either in sitting drop Cryschem plates (Hampton Research, Aliso Viejo, USA) or hanging drop EasyXtal Tool plates (Nextal Biotechnologies - now Qiagen, Hilden).

SMCcd was crystallized by hanging-drop vapor diffusion method by mixing 2  $\mu$ l protein solution (8 mg/ml in buffer C) with 2  $\mu$ l precipitant solution (16% (w/v) PEG800, 15% (v/v) glycerol and 0.5 M KCl). Crystals appeared at 25°C within a few days and were directly flash frozen in liquid nitrogen.

Crystals of SMCcdE1098Q-ATP were obtained by mixing 2  $\mu$ l protein solution (5 mg/ml in buffer C supplemented with 5 mM ATP and 20 mM MgCl<sub>2</sub>) with 2  $\mu$ l precipitant solution (0.1 M Na-acetate [pH 4.6], 6% (w/v) PEG400 and 12% (v/v)

ethanol) at 25°C using the hanging-drop vapor diffusion method. As a cryoprotectant 30% (v/v) (±)-2-methyl-2,4-pentandiol was added to the precipitant solution. After a few seconds of incubation in this solution the crystals were flash frozen in liquid nitrogen.

SMCcd-ADP- $\text{AlF}_4^-$  and SMCcd-ADP- $\text{BeF}_3^-$  were crystallized as described for SMCcdE1098Q-ATP. The buffers were identical except that ATP was replaced by either 5 mM ADP, 5 mM  $\text{AlCl}_3$ , 20 mM NaF or 5 mM ADP, 5 mM  $\text{BeCl}_2$ , 20 mM NaF.

SMCcd-AP5A (9 mg/ml in buffer C with 2.5 mM AP5A, 25 mM  $\text{MgCl}_2$ ) was crystallized by sitting-drop vapor diffusion at 25°C after mixing 2  $\mu\text{l}$  protein solution with 2  $\mu\text{l}$  precipitant solution (BisTris [pH 6.6], 16% (w/v) PEG3350). Crystals were flash frozen in liquid nitrogen after the addition of 30% (w/v) PEG400.

## 5.8 Data collection and processing

Diffraction data were recorded at ESRF (European Synchrotron Radiation Facility – Grenoble, France) and SLS (Swiss Light Source - Villigen, Switzerland) synchrotron radiation sources. Optimal starting angle, total oscillation range and rotation angle per frame were determined using the “Strategy” option in MOSFLM. MAD and SAD datasets were recorded from selenomethionine substituted protein crystals starting with the peak wavelength. Best wavelengths for the peak, inflection and high remote data sets were determined by fluorescence scan on each crystal (for details see results).

Data of SMCcd, SMCcdE1089Q-ATP, SMCcd-ADP- $\text{AlF}$  and SMCcd-ADP- $\text{BeF}$  were indexed and integrated using DENZO (Otwinowski, 1997). Scaling was performed with SCALEPACK. The SMCcd-AP5A Data were processed using XDS (Kabsch, 1993).

## 5.9 Structure solution and refinement

The structure of the SMCcd protein was solved by MAD phasing. 10 of 11 possible anomalous scatterers were located with Shake'n'Bake (Weeks and Miller, 1999). With

this information SHARP (Bricogne et al., 2003) calculated the phases which were improved by SOLOMON (CCP4, 1994). An initial model was autobuilt in the resulting electron density with repeated cycles of RESOLVE and REFMAC. After bulk solvent and overall anisotropic B-factor correction the model was completed and refined by repeated cycles of manual model building with MAIN (Turck, 1992) and positional minimization and B-factor refinement with CNS (Brunger et al., 1998). Waters were automatically built with ARP/WARP (Morris et al., 2003).

The phases of the SMCcdE1098Q-ATP structure were obtained according to the apo protein using only the peak wavelength. The model of the SMCcd protein was manually positioned in the experimental density and refined as described above.

The structures of SMCcd-ADP- $\text{AlF}_4^-$  and SMCcd-ADP- $\text{BeF}_3^-$  were solved by molecular replacement using MOLREP (CNS) and SMCcd-ATP as search model. Refinement of the structures was performed as described above. Solvent molecules were positioned by WATERPICK a part of CNS.

The phases for the SMCcd-AP5A complex were obtained by molecular replacement with PHASER (CCP4, 1994) using SMCcdE1098Q-ATP lacking the residues 1155-1169 as a search model. Model refinement was performed by repeated cycles of manual model building in COOT (Emsley and Cowtan, 2004) and positional minimization, occupancy (only waters and COOH-groups) and B-factor refinement with CNS. Water molecules were built by WATERPICK (CNS).

CNS refinement was performed with the maximum likelihood method using amplitudes as refinement target and taking the Engh & Huber parameters into account (Engh, 1991). The quality of the refinement was monitored by calculating the  $R_{\text{work}}$  and  $R_{\text{free}}$  of the model. The  $R_{\text{free}}$  was calculated for 5% of the reflection omitted from the refinement. The geometry of the refined model was checked with the program PROCHECK (CCP4, 1994). (for details see results)

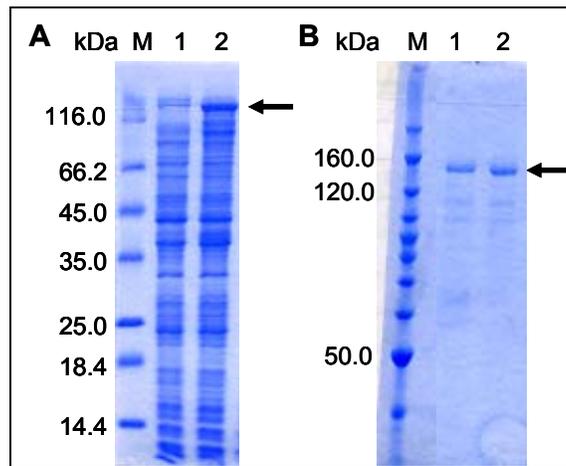
All pdb, topology and parameter files for the heterocompounds were downloaded from the HicUp database (Kleywegt and Jones, 1998).

## 6 Results

### 6.1 Cloning and expression

#### 6.1.1 *Pyrococcus furiosus* SMC

Database search for a homologue of SMC proteins in *Pyrococcus furiosus* yielded one single hit (www.expasy.org). The protein with the locus name pf1843 and the primary accession number Q8TZY2 has high sequence homology to SMC proteins from other archaea, bacteria yeast and human especially in the N- and C-terminal region containing the ABC-ATPase motifs (see appendix Fig. 42). Interestingly the deposited sequence shows an N-terminal extension not found in other SMC proteins except yeast SMC4. After cloning this gene into a pET21 vector no expression could be observed. Cloning a truncated version of this gene lacking the above mentioned extension a protein of the appropriate size (~135 kDa) was expressed (Fig.12 A).

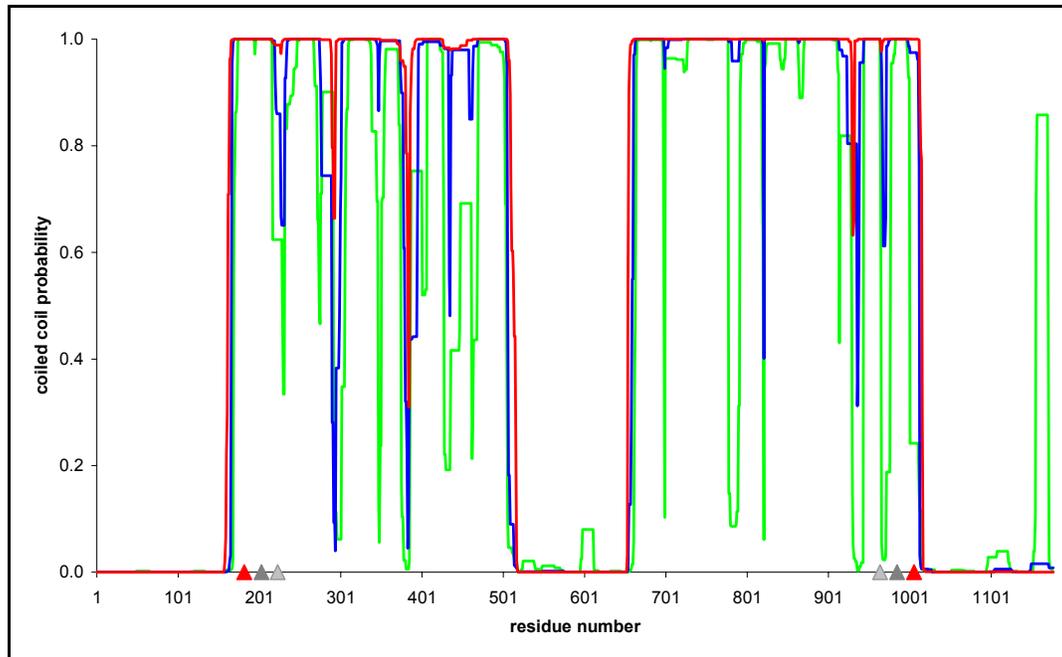


**Figure 12: Expression and purification of full-length SMC.**

A) M: MW-marker; 1: total cell lysate before induction; 2: total cell lysate 3 h after induction. B) M: BenchMark protein ladder; 1+2: elution fraction from final size exclusion column

The cells were grown at 37°C and soluble SMC protein was expressed after 3 h of induction. After disrupting the cells by sonication the protein could be purified with a heat step, ammonium sulfate precipitation, high salt size exclusion and anion exchange

chromatography according to SMCcd. In contrast to SMCcd the full-length protein does not bind to the cation exchange resin Source S and therefore after the final size exclusion chromatography contaminating proteins remained in the protein sample (Fig. 12 B). The protein was not further purified, because its long and flexible coiled-coil domain makes it not suitable for crystallization. To obtain a more globular protein for crystallization, constructs consisting only of the terminal head domain and a short part of the coiled-coil of different length were cloned. The lengths of the proteins were designed after analyzing the sequence of *Pfu* SMC for predicted coiled-coil regions with the COILS program ([http://www.ch.embnet.org/software/COILS\\_form.html](http://www.ch.embnet.org/software/COILS_form.html)).

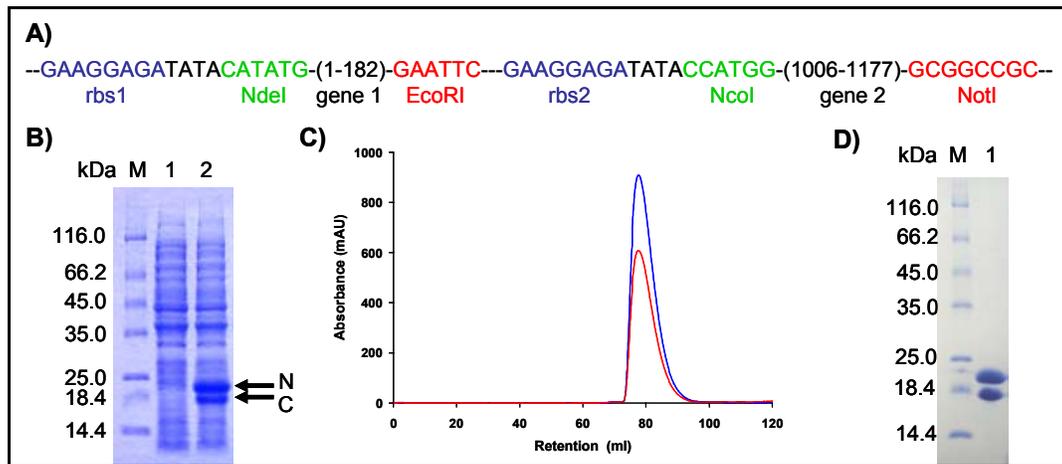


**Figure 13: Coiled-coil prediction for the full-length *Pfu* SMC protein**

Coiled-coil probability for the full-length *Pfu* SMC protein calculated with COILS. In green calculated with a window of 14 residues (green), in blue 21 and in red 28 residues. The red triangles mark the ends of the 2 peptide chains of SMCcd. Triangles in dark and light grey highlight 2 more constructs that were not used in this thesis.

The plot shows three regions with low coiled-coil probability: the N- and C-terminal end of the protein forming the ABC-ATPase domain and in the middle the hinge domain. With this theoretical information three different constructs were designed that comprise the N- and C-terminal part up to the beginning of the coiled-coil plus additional 20, 40 and 60 residues. The DNA fragments were cloned into the pETSH35 vector. This vector is derived from pET21 and allows a bicistronic expression of two

proteins. In the case of SMCcd the N- terminal residues 1 – 182 and C-terminal residues 1006 – 1177 were cloned with NdeI/EcoRI and NcoI/NotI into the vector to express the ABC-ATPase domain of *Pfu* SMC without additional amino acids from a tag or linker (Fig. 14 A).



**Figure 14: Expression and purification of SMCcd**

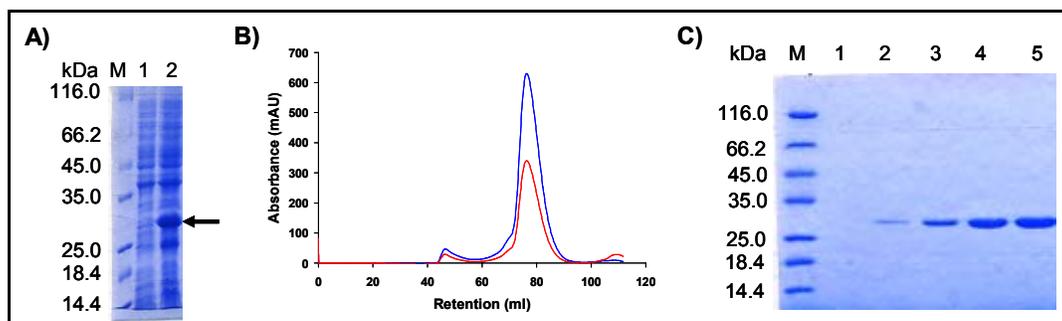
A) Cloning strategy for the bicistronic expression of the two peptide chains (1-182 / 1006-1177) of SMCcd. In blue the two ribosome binding sites, in green the 2' 5' and in red the 2' 3' restriction sites. B) Expression test of SMCcd analyzed in a 15% SDS gel. Arrows N and C mark the N- and C-terminal fragment. M: MW-Marker; 1: total cell lysate before induction; 2: total cell lysate 3 h after induction. C) Typical elution profile for SMCcd on a S200 16/60 column. Blue line: absorbance at 280 nm; red line: absorbance at 260 nm. D) Purified wt SMCcd. M: MW-Marker; 1: purified SMCcd

*E. coli* Rosetta (DE3) cells transformed with the resulting plasmid strongly expressed the two peptide chains in an approximately 1:1 stoichiometry after induction with IPTG (Fig. 14 B). The two fragments are soluble and heat stable indicating that they fold to their native conformation. This feature was used for the first purification step, after cell disruption by sonication *E. coli* proteins were denatured at 70°C and separated by centrifugation. In a next step proteins were precipitated with ammoniumsulfat, resuspended and separated by size exclusion chromatography with a high salt running buffer to remove of remaining nucleic acids and nucleotides. After reducing the salt concentration by dialysis the protein was further purified using an anion exchange chromatography. At pH 8 the vast majority of SMCcd does not bind the material indicating again that the two chains bind together and have an actual pI close to the theoretical 8.4 of a single chain protein with identical sequence. The C-terminal part alone has a theoretical pI of 6.77 and therefore should bind to the column, only the N-

terminal part with a pI of 8.67 should be in the flow through (calculated with ProtParam / [www.expasy.ch](http://www.expasy.ch)). For further purification the protein was bound to a cation exchanger, eluted and concentrated. The final purification step was performed by size exclusion chromatography. SMCcd eluted as a single peak at a retention volume corresponding to a single protein consisting of both peptide chains (Fig. 14 C). The elution profile and the analysis via SDS-Page (Fig. 14 D) show clearly that the bicistronic expression of the N- and C-terminal part of *Pfu* SMC bind in a 1:1 stoichiometry to each other, forming a protein that could be purified to homogeneity.

### 6.1.2 *Pyrococcus furiosus* ScpA

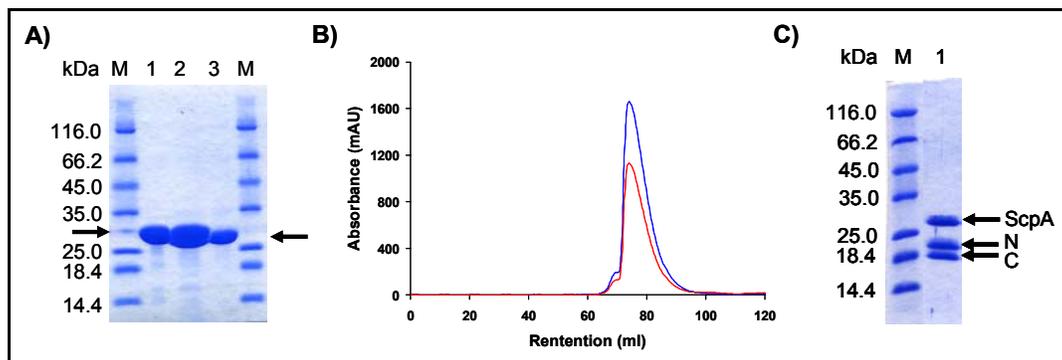
Database search for a *Pyrococcus furiosus* homologue of the bacterial ScpA protein resulted in one single hit. In the *Pfu* genome the gene for the protein with the ordered locus name pf1842 and the primary accession number Q8TZY3 overlaps with the gene for the SMC protein indicating a close relation between the two proteins (<http://www.ncbi.nlm.nih.gov>). The gene encoding *Pfu* ScpA was cloned into a pET28 vector using NdeI/NotI restriction sites to fuse an N-terminal 6xHis-tag to the protein. The fusion protein could be expressed in Rosetta (DE3) cells at 37°C for 3 h (Fig. 15 A) and was in the soluble fraction. Unlike SMCcd it is not heat stable and therefore the crude extract was loaded directly onto a Ni-NTA column.



**Figure 15: Expression and purification of His-ScpA**

A) Expression test of His-ScpA. M: MW-Marker; 1: total cell lysate before induction; 2: total cell lysate 3 h after induction. B) Typical elution profile for His-ScpA on a S200 16/60 column. Blue line: absorbance at 280 nm; red line: absorbance at 260 nm. C) Elution fractions from S200 size exclusion column containing purified His-ScpA. M: MW-Marker; 1-5: elution fractions.

The eluted protein was dialyzed against low salt buffer, further purified by an ion exchange chromatography and finally polished on a S200 size exclusion column. His-ScpA eluted as a single peak (Fig. 15 B) and showed a high degree of purity on a Coomassie stained acrylamid gel (Fig. 15 C). For crystallization and complex formation with SMCcd the His-tag was cleaved from the protein using a thrombin cleavage site between the tag and the protein. As shown in figure 16 B thrombin cleaves of the tag from His-ScpA (lane 1) resulting in a smaller protein without tag (lane 3). To obtain a homogenous protein species uncleaved protein was removed by a second Ni-NTA column. After the incubation of cleaved ScpA with SMCcd a complex containing one molecule ScpA and one molecule SMCcd could be eluted from a S200 size exclusion column (Fig. 16 B and C)

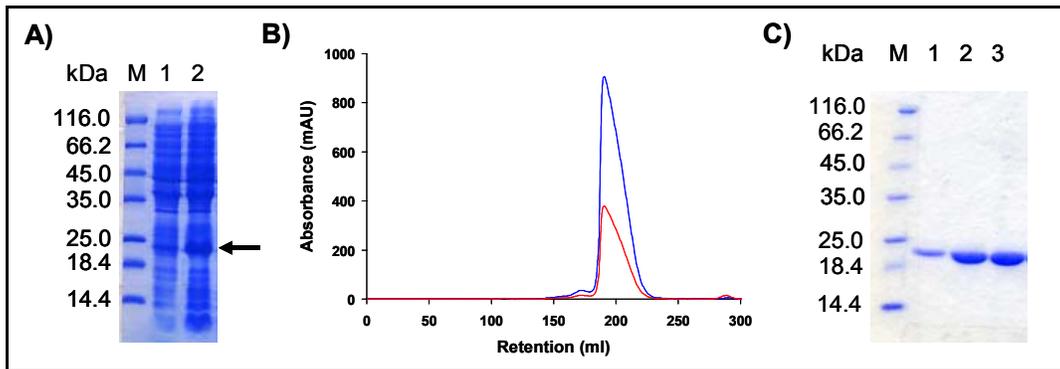


**Figure 16: Thrombin cleavage of His-ScpA and complex formation with SMCcd**

A) Analyses of the proteolytic removal of the His-tag from ScpA. M: MW-Marker; 1: His-ScpA before cleavage; 2: mixture of uncleaved and cleaved protein; 3: cleaved ScpA without His-tag. B) Elution profile of SMCcd-ScpA complex on a S200 size exclusion column. Blue line: absorbance at 280 nm; red line: absorbance at 260 nm. C) SMCcd-ScpA complex eluted from S200 column. Arrows indicating the 2 proteins (N and C indicating the two fragments of SMCcd) M: MW-Marker; 1: SMCcd-ScpA complex.

### 6.1.3 *Pyrococcus furiosus* ScpB

Database search for a possible ScpB homologue in *Pyrococcus furiosus* resulted in a protein with the ordered locus name pf2021 and the primary accession number Q8TZH3. The gene was cloned into pET28 to fuse a C-terminal 6x His-tag for purification. Expression in Rosetta (DE3) cells resulted in soluble overexpressed protein with the expected size. His-ScpB could be purified by IMAC, IEX and size exclusion chromatography.



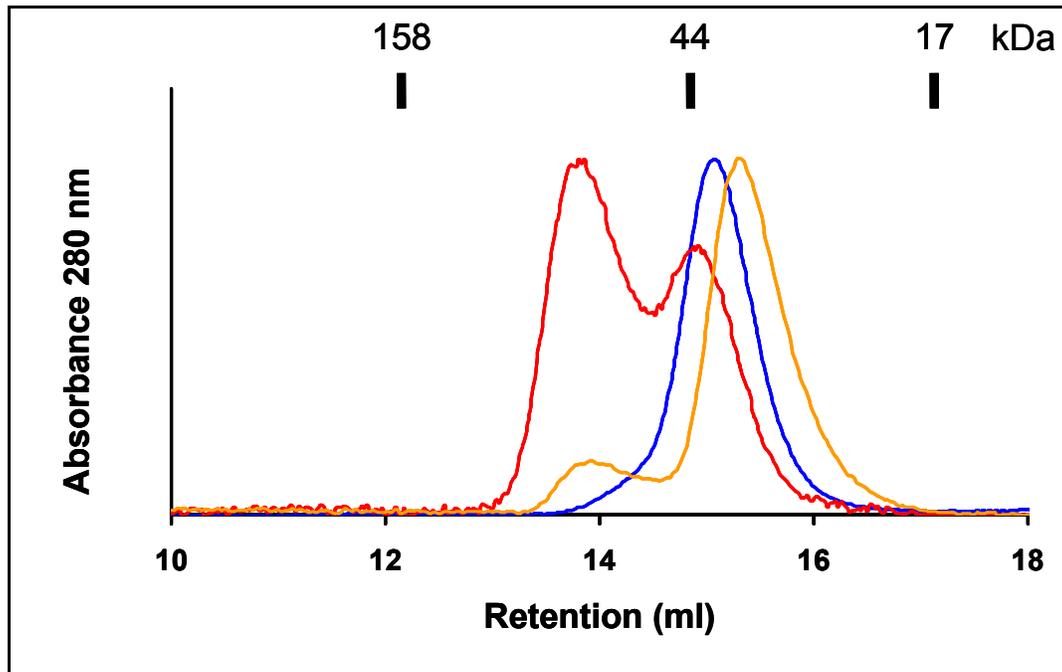
**Figure 17: Expression and purification of His-ScpB**

A) Expression test of His-ScpB. M: MW-Marker; 1: total cell lysate before induction; 2: total cell lysate 3 h after induction. B) Typical elution profile for His-ScpB on a S200 26/60 column. Blue line: absorbance at 280 nm; red line: absorbance at 260 nm. C) Elution fractions from S200 size exclusion column containing purified His-ScpB. M: MW-Marker; 1-3: elution fractions.

## 6.2 Biochemical studies

### 6.2.1 Analytical size exclusion chromatography

SMCcd although consisting of two peptide chains and being only a part of the *Pfu* SMC protein could be purified as a soluble protein. The Stoke's radius of SMCcd determined by analytical size exclusion chromatography corresponds to a protein with a molecular weight of roughly 40 kDa which coincides with a globular and monomeric protein (Fig. 18). This indicates that the designed SMCcd resembles the globular ABC ATPase domain and could be used to study the biochemical and structural features of the catalytic domain of the full-length protein.



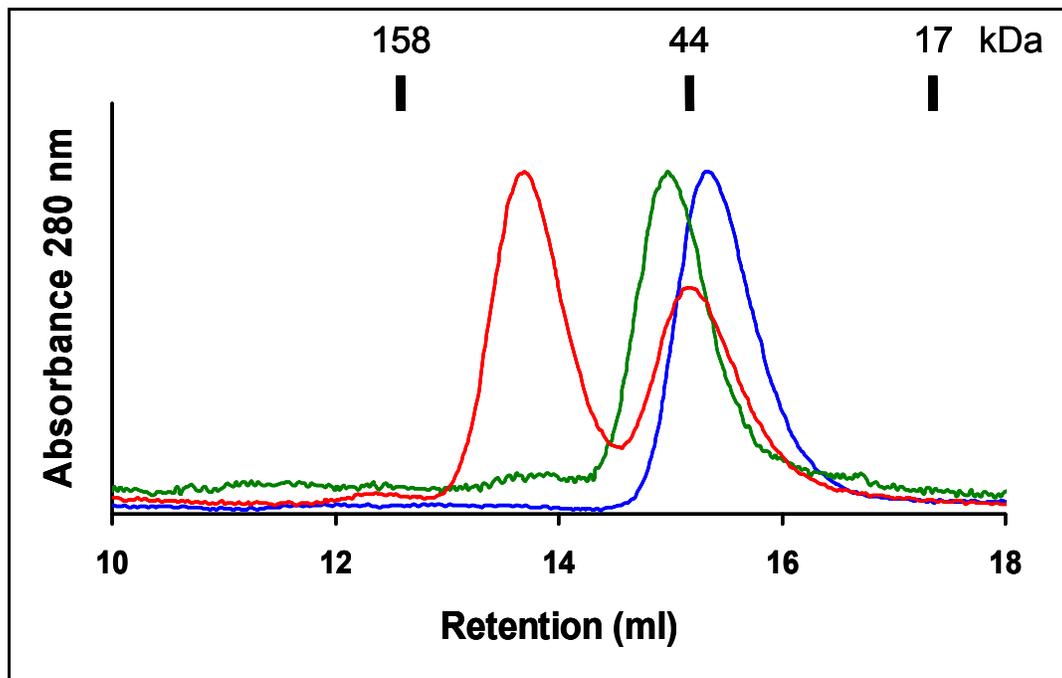
**Figure 18: Size exclusion profiles of SMCcd**

SMCcd analyzed on a S200 10/300 size exclusion column with indicated molecular weight standards. SMCcd elutes in a monomeric peak (blue line). Addition of slowly hydrolysable ATP analogue ATP $\gamma$ S leads to formation of dimers (orange line). The fraction of dimeric protein could be increased using ATP in the buffer and the active site mutant SMCcd-E1098Q (red line).

Presence of ATP leads to engagement of two ATP ABC-ATPase domains. To test if this is also true for SMCcd, wild type protein was analyzed on a size exclusion column in the presence of 0.5 mM ATP in the running buffer. Probably due to ATP hydrolysis no dimerization was observed in this experiment. To overcome the problem with ATP hydrolysis one can use slowly hydrolyzable ATP analogues such as ATP $\gamma$ S to study the effect of ATP. Preincubation of SMCcd with ATP $\gamma$ S lead to formation of a protein species with a Stoke's radius corresponding to a dimer (Fig. 18 orange line). The fraction of dimeric protein could be increased using a Walker B motif mutant of SMCcd (Fig. 18 red line). Mutating the active site glutamic acid 1098 to glutamine reduces the ATPase activity (Fig. 20) and traps the protein in the ATP bound state thereby stabilizing the dimer.

Monitoring the retention volume can also be used to verify interaction between proteins. If two or more proteins stably interact they stay together during a size exclusion chromatography and elute according to the size of the complex. To test for

the physical interaction SMCcd, ScpA and ScpB were analyzed alone, two of them and all three together on a S200 10/300 size exclusion column. All three proteins alone elute as monomers from the column. ScpB shows in this experiment interactions with neither SMCcd nor ScpA. SMCcd and ScpA form a stable complex that could be separated from residual monomeric protein. The complex consists of one SMCcd and one ScpA protein as determined by the retention volume (Fig. 19).



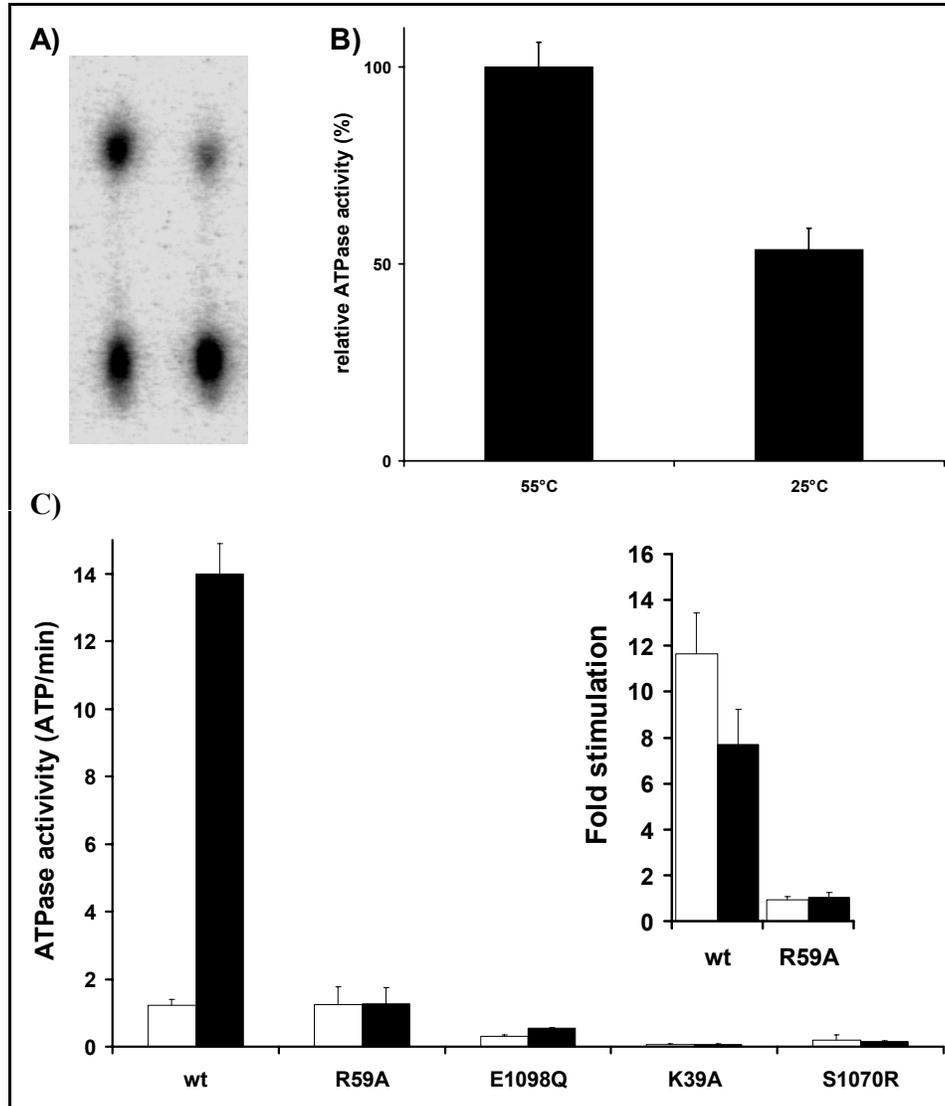
**Figure 19: Size exclusion profiles of SMCcd and ScpA**

Stoke's radius determined on a S200 10/300 size exclusion column with indicated molecular weight standards. In blue the monomeric peak of SMCcd alone. The green line shows the elution profile of ScpA. After preincubation of the two proteins they form a stable 1:1 complex (red line).

### 6.2.2 ATPase activity assay

SMCcd comprises the ABC-ATPase domain of *Pyrococcus furiosus* SMC protein. To test if this domain on its own is an active enzyme the ATPase activity of SMCcd was assayed using radioactive  $\gamma^{32}\text{P}$ -ATP as substrate. The reaction products were separated with thin layer chromatography (Fig. 20 A). Signal from the radioactive phosphate was detected and quantified. All reactions were incubated at 55°C because the source organism *Pyrococcus furiosus* is thermophilic and grows at temperatures around

100°C. Decreasing the temperature to 25°C reduces the reaction rate to 50% but does not completely abolish it (Fig. 20 B).



**Figure 20: ATPase activity assay of wt and mutant SMCcd**

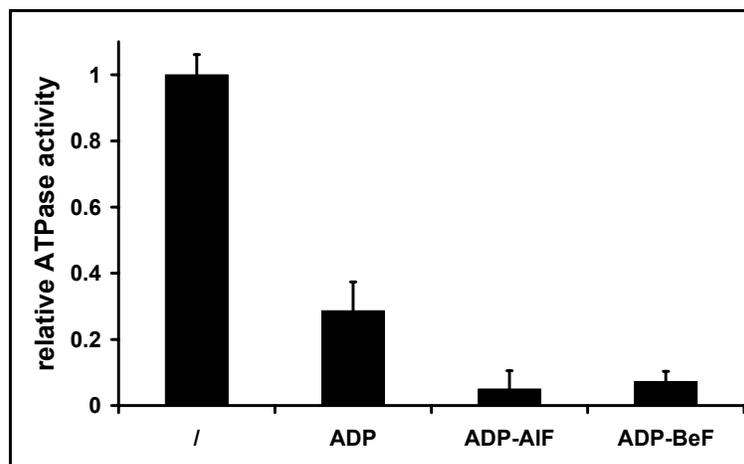
A) Exemplary result from a thin layer chromatography analysis of an ATPase reaction. The upper spots correspond to the free radioactive  $\gamma$  phosphate. ATP migrates very slowly and remains close to the spotted point at the bottom. B) Relative ATPase activity of SMCcd assayed at 25°C and 55°C. C) ATPase activity in absence (white) and presence (black) of DNA. Insert: relative stimulation of the ATPase activity of SMCcd (white) and SMCcd:ScpA (black). Error bars represent the standard deviation of three experiments.

To clarify the mechanism of ATP hydrolysis by SMCcd wild type and mutant protein was assayed for ATPase activity. The basal turn over rate of wild type SMCcd was determined to  $1 \pm 0.18$  ATP/min. Mutating arginine 59 to an alanine does not change

the basal activity of SMCcd (Fig. 20 C). All other three mutations strongly reduce the basal ATPase activity of SMCcd. Replacing the water activating glutamic acid in the Walker B motif by a glutamine reduces the activity of SMCcd. The K39A mutation in the Walker A motif interferes with ATP binding (see 6.2.3) and strongly decreases the activity almost to zero. Also the signature motif mutant S1070R nearly completely abolishes the hydrolysis of ATP by SMCcd.

To test whether DNA can activate SMCcd, the ATPase activity was determined in the presence of DNA. Addition of a 42mer dsDNA to the reaction of the wild type protein strongly increases the activity to  $14 \pm 0.9$  ATP/min. In the Walker B mutant the addition of DNA slightly increases the activity but it is still lower than the wild type basal activity. Mutation of arginine 59 to alanine completely abolishes the DNA stimulation of the ATPase activity. The presence of DNA does not lead to an increased activity of the K39A and S1070R mutant.

ScpA and ScpB do not have any intrinsic ATPase activity. Addition of ScpA which interacts with SMCcd (see 6.2.1) to the ATPase reaction has no influence on the basal activity and only leads to a weak reduction of the DNA stimulated activity (Fig. 20 C insert).



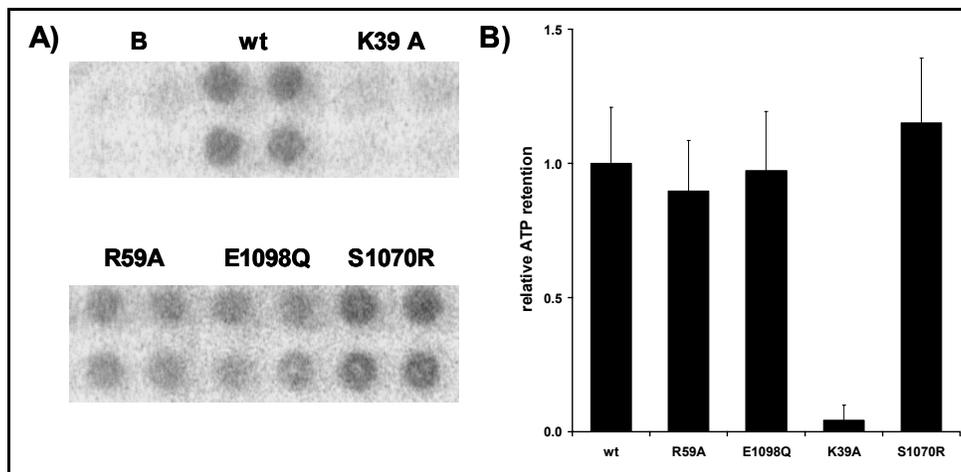
**Figure 21: ATPase activity of SMCcd in presence of inhibitors.**

The ATPase activity of 10  $\mu$ M wt SMCcd was assayed in the presence of 1 mM ATP and 5 mM ADP, ADP-AlF<sub>4</sub><sup>-</sup> or ADP-BeF<sub>3</sub><sup>-</sup>. Relative activity compared to the uninhibited reaction is shown in columns.

ADP as product of the ATPase reaction and ADP-AlF<sub>4</sub><sup>-</sup> and ADP-BeF<sub>3</sub><sup>-</sup> as transition state analogues of the ATP hydrolysis by ATPases are potent inhibitors of the reaction of ABC-ATPases. To test the ability of all three to inhibit the ATPase reaction of wt SMCcd the activity was assayed in the presence of 5 mM ADP, ADP-AlF<sub>4</sub><sup>-</sup> and ADP-BeF<sub>3</sub><sup>-</sup>. ADP as a product reduces under this reaction conditions the amount of hydrolyzed ATP to 1/3 of the uninhibited reaction (Fig. 21). The addition of either ADP-AlF<sub>4</sub><sup>-</sup> or ADP-BeF<sub>3</sub><sup>-</sup> to the reaction mixture and thereby formation of a complex analogues to the transition state of the hydrolysis reaction nearly completely abolishes the ATPase activity of wt SMCcd (Fig. 21).

### 6.2.3 ATP binding assay

To analyze if the differences in the ATPase activity are associated with defects in ATP binding rather than ATP hydrolysis the ATP binding properties of wild type and mutant SMCcd were determined. The proteins were incubated with radioactive  $\gamma^{32}\text{P}$ -ATP on ice to allow binding but reduce ATP hydrolysis.



**Figure 22: ATP binding assays of wild type and mutant SMCcd.**

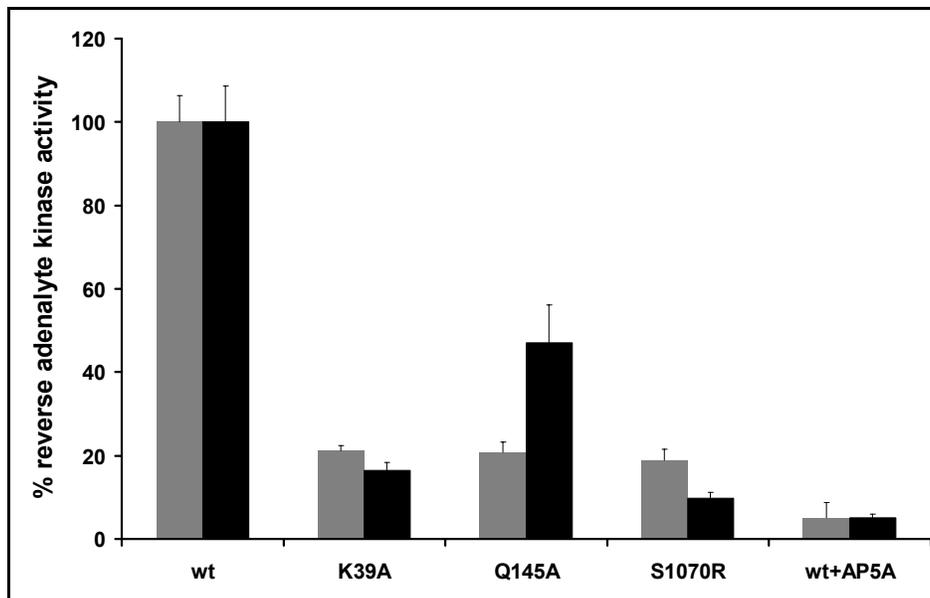
A) Typical result of a nitrocellulose filter binding assay to determine ATP binding capability of SMCcd. B: background; wt: wild type; K39A, R59A, E1098Q, S1070R: mutant SMCcd. B) ATP binding capacity of wt and mutant SMCcd assayed in triplicate.

While sucking the reaction mixture through a nitrocellulose membrane proteins bind to the membrane and retain bound ATP that could be detected, unbound ATP does not bind to the membrane and is in the flow through. Comparing the retention of wild type

and mutant SMCcd proteins shows that only the Walker A mutant K39A is deficient in ATP binding and all other proteins show virtual identical ATP binding (Fig. 22).

#### 6.2.4 Reverse adenylate kinase activity

The ability of SMCcd to catalyze the reverse adenylate kinase reaction was assayed by incubating SMCcd with ADP. The developed ATP was used by firefly luciferase to oxidize D-luciferin and the resulting luminescence was used to quantify the amount of ATP.



**Figure 23: Reverse adenylate kinase activity of SMCcd.**

The amount of produced ATP was quantified by integrating the generated luminescence signal from luciferase. In grey the activity with 200 μM ADP and in black with 1 mM ADP in the reaction mixture are depicted. wt: wild type protein; K39A, Q145A, S1070R mutant SMCcd; wt+AP5A: wild type protein with 2.5 mM AP5A inhibitor added to the reaction.

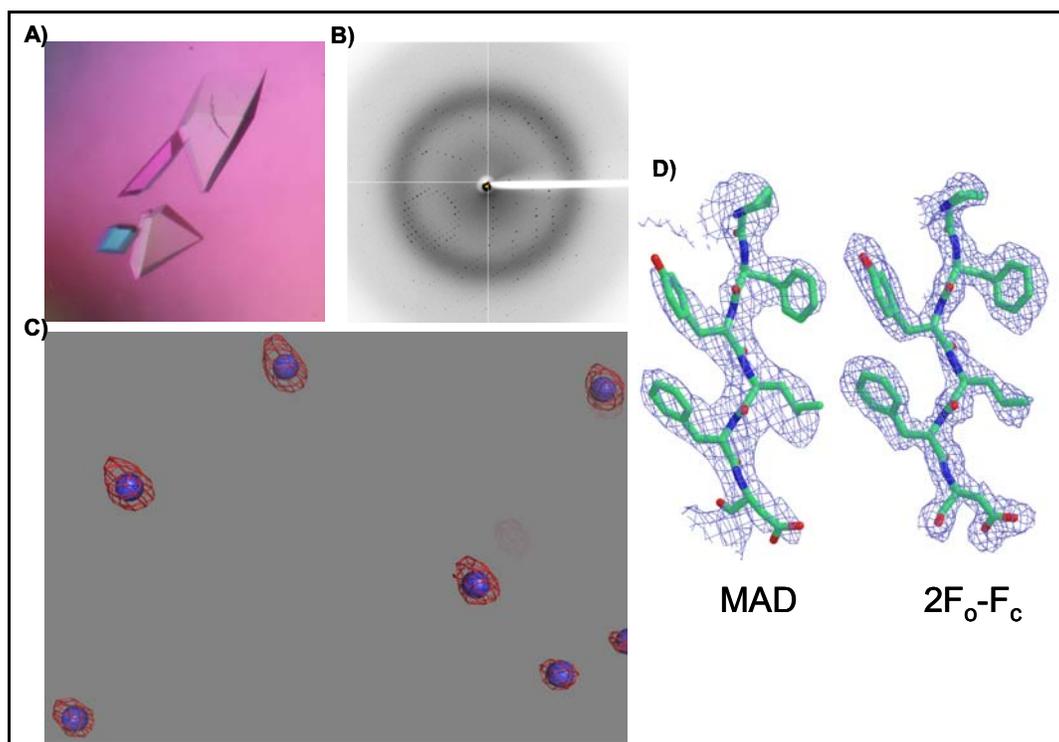
At a concentration of 200 μM the activity of all 3 mutant proteins is decreased 10 – 20% to that of wild type activity (Fig. 23). This background activity probably comes from residual impurities by *E.coli* adenylate kinases. Adding the inhibitor AP5A to the wild type reaction abolishes the activity nearly completely. Providing 1 mM ADP in the reaction mixture increases the amount of produced ATP from the Q145A mutant to 50% compared to the wild type reaction. Both other mutants have almost the same relative activity compared to reactions with lower ADP concentration. The addition of

inhibitor to the wild type reaction reduces the relative activity at 1 mM ADP concentration to the same level as in the 200 mM reaction.

## 6.3 Crystallization, structure solution and refinement

### 6.3.1 Apo wild type SMCcd

The prerequisite to determine the structure of a protein to atomic resolution using X-ray crystallography is growing crystals of the protein. For crystallization trials the purified wild type SMCcd was concentrated to 8 mg/ml in 5 mM Tris (pH 8.0), 100 mM NaCl and 0.1 mM EDTA. Screening was performed mixing 1  $\mu$ l protein solution with 1  $\mu$ l precipitant solution from commercially available sparse matrix screens at 25°C. After one to two days crystals were grown in several different precipitant solutions. Most promising with four single three dimensional crystals was the condition A2 of the Jena Bioscience screen 5 (12% (w/v) PEG8000, 10% (v/v) glycerol, 0.5 M KCl). This initial hit was refined to 16 % (w/v) PEG 8000, 15 % (v/v) glycerol, 0.5 M KCl. To avoid radiation damage during data collection at synchrotron X-ray sources the protein crystals are kept at 100 K, therefore and for storage the crystals were flash frozen without further cryo protocol. 15% (v/v) glycerol in the crystallization condition was enough to avoid formation of ice crystals. Crystals of SMCcd diffracted to a limiting resolution of 2.0 Å at beamline ID29 (ESRF). Data were recorded using an ADSC Q210 2D detector. Indexing with DENZO determined the unit cell to  $a = 102.3 \text{ \AA}$   $b = 56.7 \text{ \AA}$   $c = 78.9 \text{ \AA}$   $\alpha = 90^\circ$   $\beta = 123.0^\circ$   $\gamma = 90^\circ$  with a C2 symmetry. For further details of data collection see appendix table 2.



**Figure 24: Images of the structure solution steps of SMCcd apo form.**

A) Crystals of SMCcd. B) Diffraction pattern of SMCcd. C) 7 of the 8 Selenium sites found by Shake'n'Bake in blue spheres with anomalous difference electron density map contoured at  $4\sigma$  as red mesh. D) MAD and  $2F_o-F_c$  electron densities contoured at  $1\sigma$  around a portion of the refined model (color-coded sticks).

To be able to perform a MAD experiment for phasing the protein was additionally expressed in a methionine auxotroph *E.coli* strain in selenomethionine containing media. Selenomethionine substituted protein behaved as the wild type protein during expression, purification and crystallization. Anomalous data were recorded at 100 K with a marCCD detector at the PX beamline of the SLS. The exact wavelength of the peak, inflection and high remote dataset for the MAD experiment were determined by a fluorescence scan (for details see appendix table 1). All three datasets were recorded at one single crystal to a limiting resolution of 2.5 Å. Data were indexed, integrated and scaled with DENZO and SCALEPACK (for details see appendix table 2).

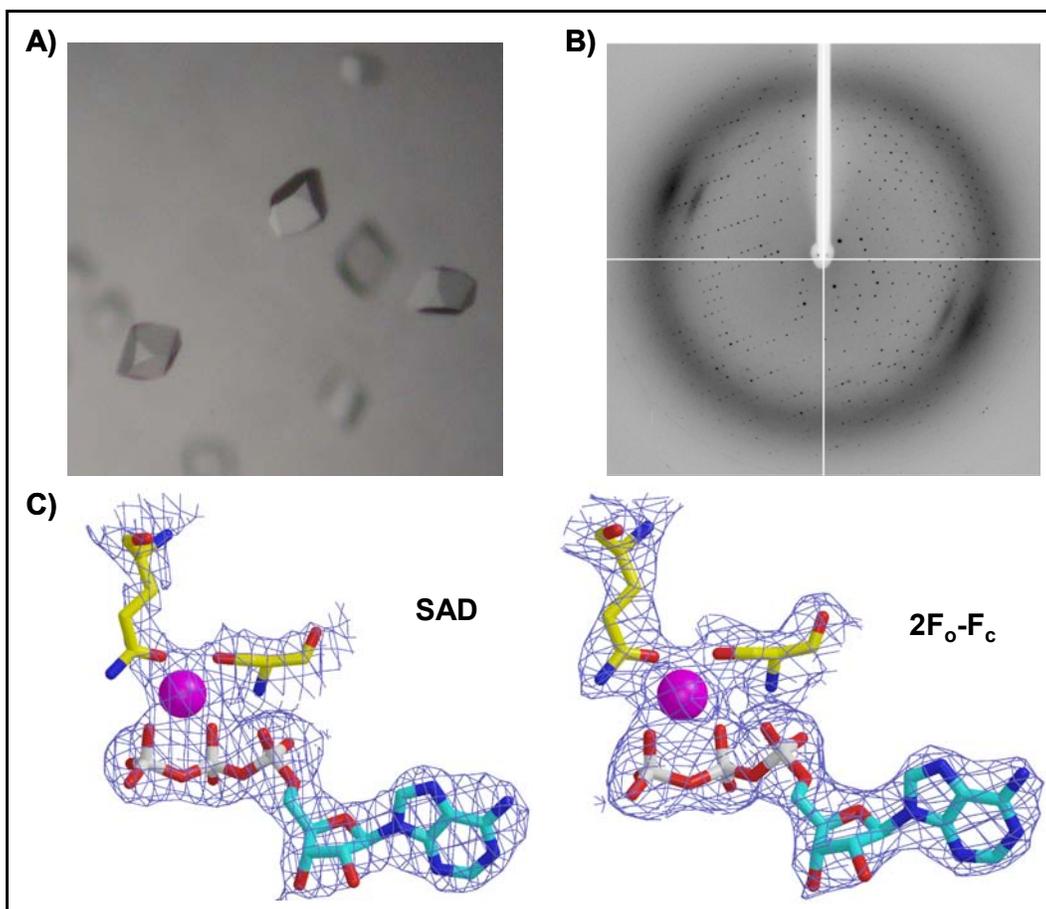
The number of molecules per asymmetric unit was determined by calculating the Matthews coefficient with the unit cell parameters, the symmetry and the molecular weight of SMCcd. The calculated Matthews coefficient is 2.4 for one molecule per asymmetric unit with a solvent content of roughly 50%. Taking this into account one

asymmetric unit should contain eleven methionines and therefore in the case of the selenomethionine substituted protein eleven anomalous scatterers. Eight out of the eleven anomalous scatterers were located with the help of Shake'n'Bake using the three datasets of the MAD experiment. With these sites the phases were calculated with SHARP and improved using density modification tools of SOLOMON. The initial model built into the resulting density by RESOLVE and REFMAC was completed with repeated cycles of manual model building in MAIN and refinement with CNS. The final model contains residues 1-56, 77-167, 1012-1145 and 1148-1173 and in total 2748 non hydrogen atoms and 205 solvent molecules and was refined to a final R factor of 20.7 ( $R_{\text{free}}$  26.1). For further details of the refined model see appendix table 2.

### 6.3.2 ATP bound E1098Q SMCcd

Wild type SMCcd could not be crystallized in complex with ATP, AMPPNP or ATP $\gamma$ S. As in the biochemical assays E1098Q SMCcd behaved identical to the wild type protein except the decreased ATPase activity this mutant was used to determine the structure of the ATP bound dimerized form of SMCcd. After concentrating the purified E1098Q SMCcd to 10 mg/ml (in 5 mM Tris (pH 8.0), 100 mM NaCl, 0.1 mM EDTA) it was mixed 1:1 with 10 mM ATP, 40 mM MgCl<sub>2</sub> in the same buffer and incubated on ice allowing binding and dimerization. Initial screening resulted in some hits that could be refined to optimal precipitation solution containing 0.1 M sodium acetate (pH 4.6), 6% (w/v) PEG400 and 12% ethanol. Crystals grew over night but dissolved after a couple of days perhaps due to residual ATP hydrolysis activity of the E1098Q mutant. Crystals were cryoprotected by adding 30% (v/v) MPD to the precipitation solution and flash frozen in liquid nitrogen. To obtain experimental phases selenomethionine substituted protein was used to collect a data set at the peak wavelength determined by a fluorescence scan at beamline PX (SLS). The data with a limiting resolution of 2.8 Å were indexed, integrated and scaled with the HKL suite. ATP bound E1098Q SMCcd crystallized in space group P3<sub>1</sub>21 with the unit cell parameters  $a = 102.6 \text{ \AA}$ ,  $b = 102.6 \text{ \AA}$ ,  $c = 87.2 \text{ \AA}$ ,  $\alpha = 90^\circ$ ,  $\beta = 90^\circ$ ,  $\gamma = 120^\circ$  and one molecule per asymmetric unit (Matthews coefficient 3.31 and 63% solvent). SnB was able to locate eight selenium sites and with these coordinates phases were obtained analogously to the apo form.

After positioning the nucleotide free model into the density clear  $F_o-F_c$  density for the bound ATP and a Magnesium ion appeared. ATP and the ion were placed in the density and refined together with the protein. The structure was refined as described above to a final R-factor of 21.0 ( $R_{\text{free}}$  26.9) and comprises residues 2-167, 1009-1172, one ATP molecule, one  $\text{Mg}^{2+}$  ion and 205 solvent molecules. For refinement statistics see appendix table 2.



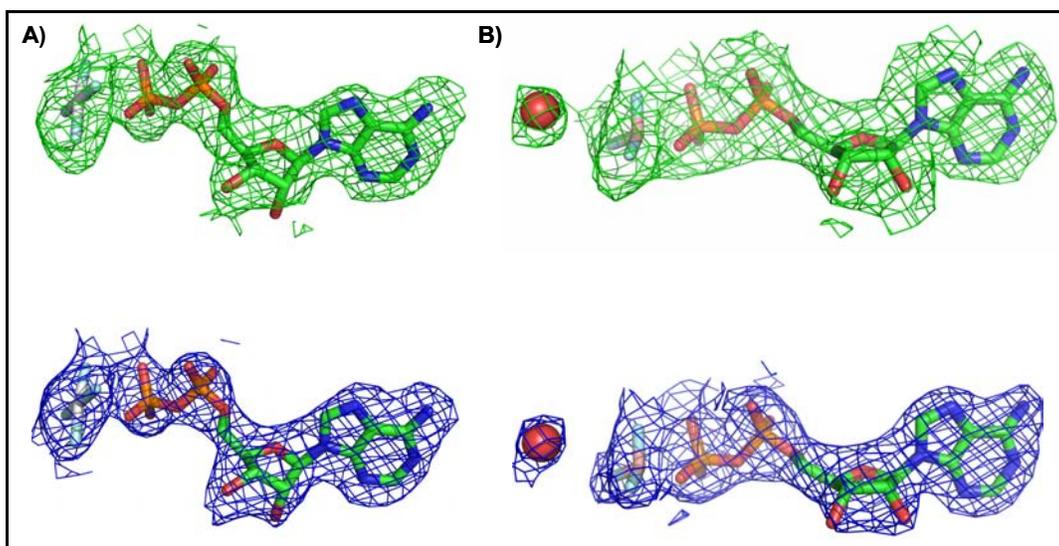
**Figure 25: Crystals, diffraction and density images of ATP bound E1098Q SMCcd.**

A) Crystals of ATP bound E1098Q SMCcd. B) Diffraction pattern of ATP bound E1098Q SMCcd. C) SAD and  $2F_o-F_c$  electron densities contoured at  $1\sigma$  around the ATP moiety superimposed with the refined model (color-coded sticks, magenta sphere:  $\text{Mg}^{2+}$ ).

### 6.3.3 ADP- $\text{AlF}_4^-$ and ADP- $\text{BeF}_3^-$ bound to SMCcd

To trap an ATPase in the transition state of the hydrolytic reaction one can use ADP together with orthovanadate ( $\text{VO}_4^{3-}$ ),  $\text{BeF}_3^-$  and  $\text{AlF}_4^-$  as transition state analogues of

the  $\gamma$  phosphate. SMCcd did not crystallize in complex with ADP and orthovanadate. SMCcd together ADP-BeF<sub>3</sub><sup>-</sup> or ADP- AlF<sub>4</sub><sup>-</sup> crystallized under the same conditions as dimeric E1098Q SMCcd. Both crystals belong to the same space group and had comparable unit cell dimensions as the ATP bound dimer. Molecular replacement using one monomer of the ATP bound SMCcd resulted in a clear positive F<sub>o</sub>-F<sub>c</sub> density for ADP-BeF<sub>3</sub><sup>-</sup> and ADP- AlF<sub>4</sub><sup>-</sup> bound at the active center (Fig. 26). In the case of the ADP-BeF<sub>3</sub><sup>-</sup> form the protein was better ordered and the structure could be refined to final R factors of 20.6 (24.1) whereas in the ADP- AlF<sub>4</sub><sup>-</sup> larger parts of the protein were disordered and the structure was refined to slightly higher R factors of 22.6 (27.1). Statistics of data collection, structure solutions and refinement are listed in table 3 of the appendix.



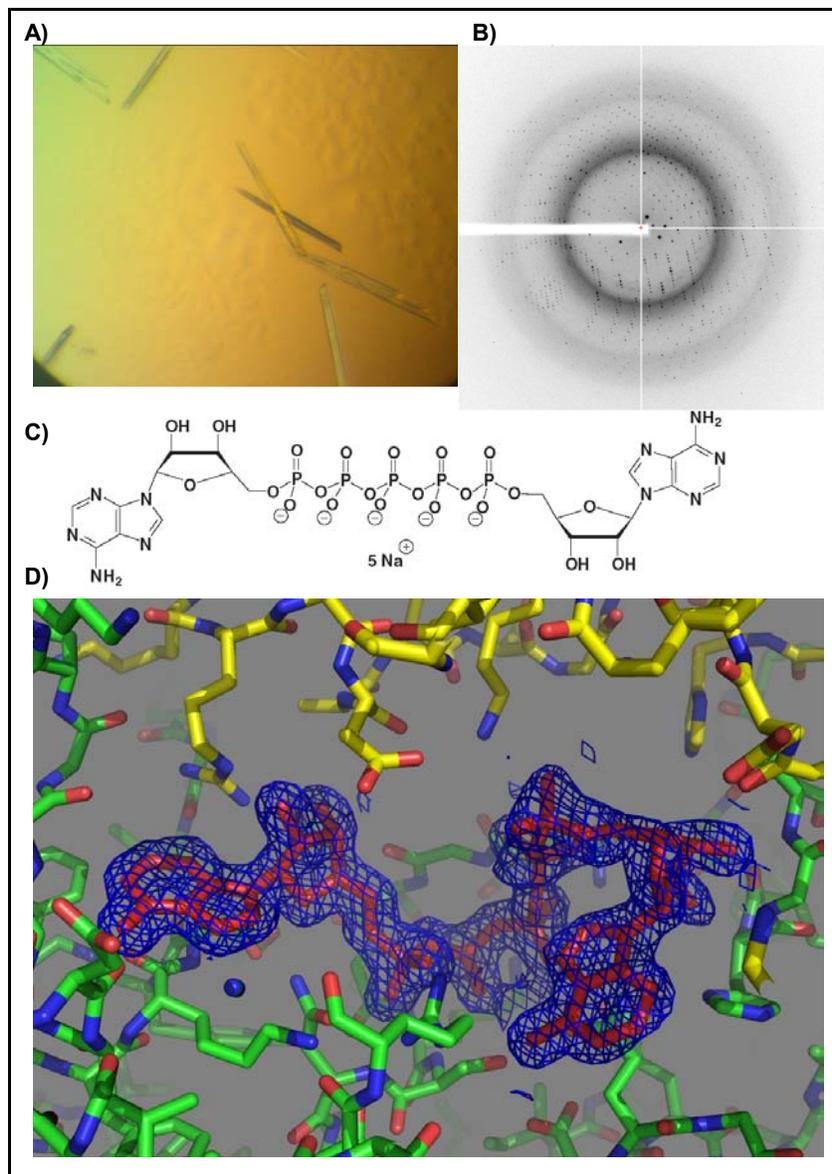
**Figure 26: Electron densities around ADP- AlF<sub>4</sub><sup>-</sup> and ADP-BeF<sub>3</sub><sup>-</sup>.**

A) In green the F<sub>o</sub>-F<sub>c</sub> density around ADP- AlF<sub>4</sub><sup>-</sup> after replacement, rigid body refinement and 1 cycle of positional refinement countered at 1 $\sigma$ . In blue the final 2F<sub>o</sub>-F<sub>c</sub> density at the same position countered at 1 $\sigma$ . B) Same figures as in A) for ADP-BeF<sub>3</sub><sup>-</sup> with an additional water molecule positioned at the same place as in the ATP bound structure.

#### 6.3.4 SMCcd in complex with AP5A

To investigate the molecular mechanism of the adenylate kinase activity of SMCcd the protein was co-crystallized with AP5A, a bisubstrate inhibitor of adenylate kinases. The screening for crystallization conditions resulted in long needles in almost all PEG

conditions. The conditions could be refined and crystals with the monoclinic space group  $P2_1$  diffracted to a limiting resolution of 1.6 Å.



**Figure 27: Crystals, diffraction and structure images of AP5A-SMCcd.**

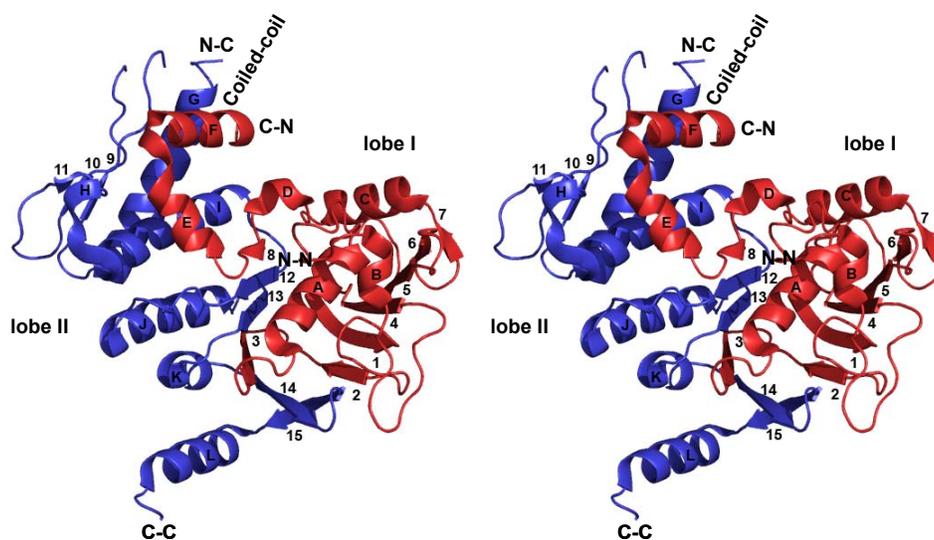
A) Crystals of SMCcd bound to AP5A, B) Diffraction image of AP5A-SMCcd recorded at ESRF ID14-2. C) Chemical structure of AP5A (Na<sup>+</sup> salt / image from Jena Bioscience). C) Well defined AP5A (red sticks) bound to one of the two molecules per asymmetric unit (color coded sticks with C atoms in green / C atoms in yellow: symmetry related molecule). The refined  $2F_o-F_c$  electron density contoured at  $1\sigma$  around the AP5A moiety is shown as blue mesh.

Molecular replacement with a monomer of ATP bound SMCcd resulted in clashes in the region of the C-terminal helix. Using a truncated version of this molecule lacking

residues 1155-1169 yielded a clear  $2F_o-F_c$  and  $F_o-F_c$  electron density for a whole AP5A moiety in one of the two molecules in the asymmetric unit. In the second molecule AP5A was partially disordered, presumably because the binding site of the second adenosine is located at a crystal contact. The data of structure solution and refinement are summarized in table 4 in the appendix.

## 6.4 Structural analysis

### 6.4.1 Overall structure of SMCcd



**Figure 28: Ribbon presentation of the SMCcd crystal structure.**

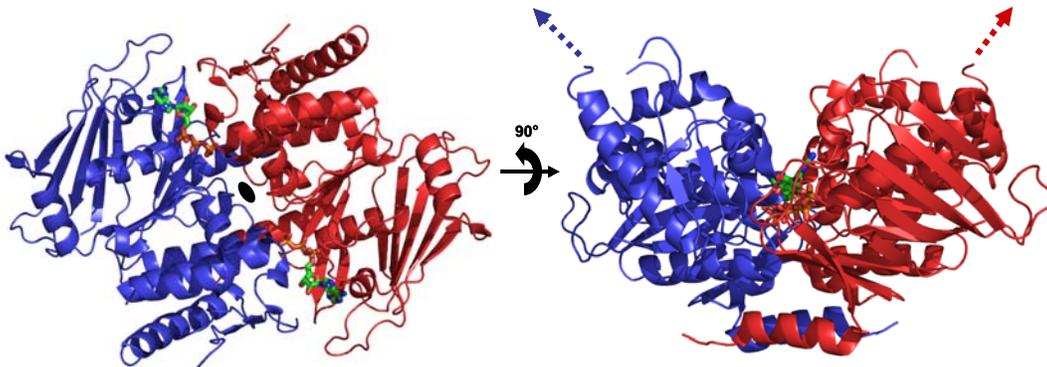
Stereo view of SMCcd with annotated secondary structure elements. In red color the N-terminal part from N-N to C-N, in blue the C-terminal part starting at N-C and ending at C-C. The two lobes and the direction of the coiled-coil are labeled.

The two peptide chains of SMCcd fold into one single globular domain consisting of two lobes with overall dimensions of  $60 \times 50 \times 40$  Å. Both lobes are joined by the central mixed parallel  $\beta$ -sheet formed by strand 3 and 8 of the N-terminal part and strand 12 and 13 of the C-terminal part. Lobe I is mainly formed by the N-terminal segment of the protein and comprises the flanking antiparallel  $\beta$ -sheet comprising strands 2, 1, 4, 5, 6 and 7. The central helix A is packed against this  $\beta$ -sheet and is connected via helix B

to strand 4 in this sheet. From the end of the sheet the peptide chain protrudes via helices C and D to the strand 8 of the central  $\beta$ -sheet and from there via E and F to the coiled-coil. The coiled-coil missing in this structure protrudes from lobe II and is inserted between helix F of the N-terminal fragment and helix G from the C-terminal fragment. Helix G already points in the direction of the coiled-coil whereas the peptide chain of helix F has to undergo an approximately  $120^\circ$  turn upwards to form the coiled-coil. Lobe II comprises mainly the C-terminal part of the protein. Helices G, I and J are sandwiched between the central  $\beta$ -sheet and the outer sheet formed by strands 9, 10 and 11. From the central  $\beta$ -sheet the chain continues through helix K and strands 14 and 15 to a long C-terminal helix L.

#### 6.4.2 Structure of the ATP bound dimer

Upon ATP binding E1098Q SMCcd dimerizes as shown by size exclusion chromatography (Fig. 18) but in the crystal the asymmetric unit contains only one molecule. After analyzing the crystal packing a dimer formed by two molecules related by a 2-fold crystallographic axis could be found (Fig. 29).



**Figure 29: Two orthogonal views of dimeric SMCcd.**

Ribbon models of ATP bound SMCcd with one monomer colored in blue and the other in red. The two ATP molecules sandwiched in the interface are shown in color coded sticks. On the left side a top view of the dimer with the 2-fold symmetry axis indicated in the middle of the molecule. On the right hand side after a  $90^\circ$  rotation upwards a front view along the interface is shown with arrows indicating the direction of the coiled-coils.

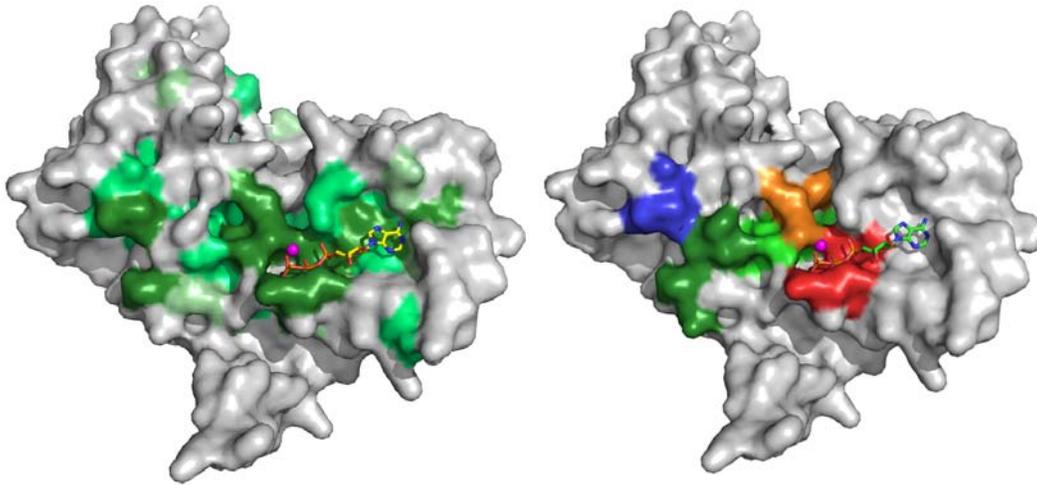
In the dimer two SMCcd molecules are oriented “head-to-tail” in which lobe I of one monomer binds to lobe II of the other monomer sandwiching two ATP molecules in their interface. This interface is situated along the long axis of the overall elliptic dimer.

The two coiled-coils protrude with an angle of approximately  $45^\circ$  and roughly  $27 \text{ \AA}$  space at their root from the same face of the dimer. This concave surface is therefore the inner side of a protein ring formed by two SMC molecules and is defined as “top” of the molecule. The opposite convex side is the outer side of the ring and is formed on the bottom of the molecules by the two C-terminal helices extending to the other monomer. Dimerization is achieved by the binding of ATP to the Walker A motif of one protein and the signature motif of the opposing SMCcd. The buried surface of ATP-protein interactions contributes with  $1800 \text{ \AA}^2$  in a similar amount to the dimer contact as the protein-protein contacts with  $2600 \text{ \AA}^2$ . In addition to the ATP-protein contacts there are some protein-protein interactions through direct hydrogen bonds (N35 to H1102/D1104; G36 to K1061/S1070; R1168 to D1105/D1131; H1102 to E/Q1098) and water mediated hydrogen bond networks.

### 6.4.3 The active site

The SMCcd dimer contains two active sites. These active sites are located at the dimer interface and are composed of residues from each SMCcd molecule. ATP is bound in a groove with the adenine pointing to the outside and the three phosphates pointing towards the middle of the complex. Residues surrounding this ATP binding site are highly conserved among SMC proteins and share high homology to the motifs characterizing an ABC ATPase (Fig. 30).

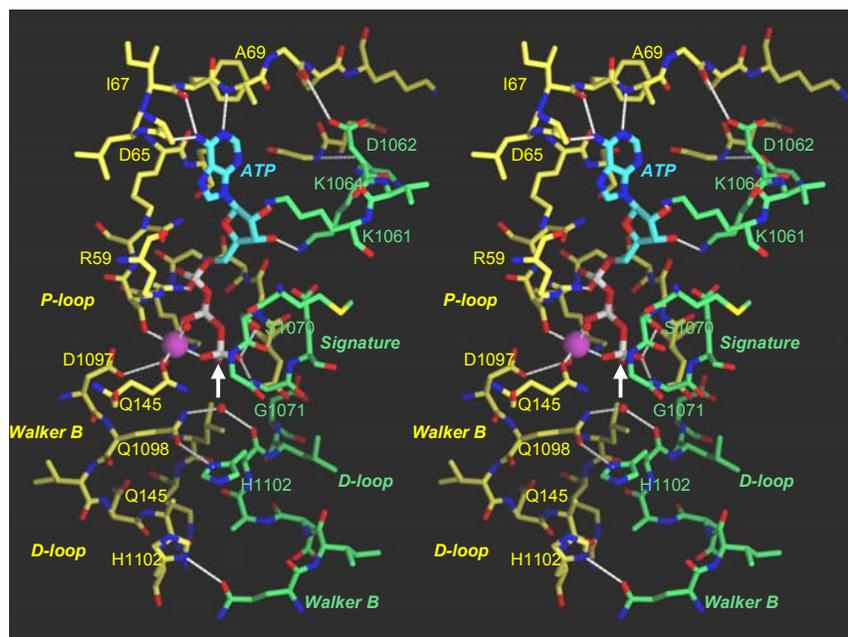
The adenine moiety is specifically recognized by three main-chain hydrogen bonds, namely between N1 of the adenine and the main-chain N of residue A69, between N6 of the adenine and the carbonyl O of residues I70 and D65. No interactions of the ribose with the protein molecule occur. The  $\alpha$  and  $\beta$  phosphates are bound by the Walker A / P-loop. A  $\text{Mg}^{2+}$  ion binds to the  $\beta$  and  $\gamma$  phosphates. To saturate the octahedral coordination sphere of the  $\text{Mg}^{2+}$  ion it is additionally coordinated by S40 and Q145 in one plane with the two phosphate oxygens and two water molecules above and below this plane. Residue D1097 stabilizes the coordination of the ion by binding to the lower water molecule and residue S40.



**Figure 30: Conservation and ABC ATPase motifs mapped on the surface of SMCcd.**

Surface representation of the dimer interface of one molecule with bound ATP and  $Mg^{2+}$  (color coded sticks and magenta sphere). On the left side the degree of conservation is mapped on the surface with dark green for identical amino acids to white for non conserved residues (Alignment of SMC from *Pyrococcus furiosus*, *Thermotoga maritima*, *Bacillus subtilis*, human [SMC 1], *Saccharomyces cerevisiae* [SMC 1+2+3+4] with ClustalW). On the right hand side same view with ABC ATPase motifs mapped on the surface with red: Walker A; green: Walker B; orange: Q-loop; blue: Signature; dark green: D-loop. The coiled-coil protrudes upwards on the left side of the molecule.

From the opposing molecule the adenine is not recognized, but there are two specific hydrogen bonds between residue K1064 and the 2'-OH and K1061 and the 3'-OH of the ribose. In addition the  $\gamma$  phosphate is bound by two hydrogen bonds from the signature motif S1070-O $\gamma$  and the main-chain N of G1072.



**Figure 31: Stereo view of ATP bound to the active center.**

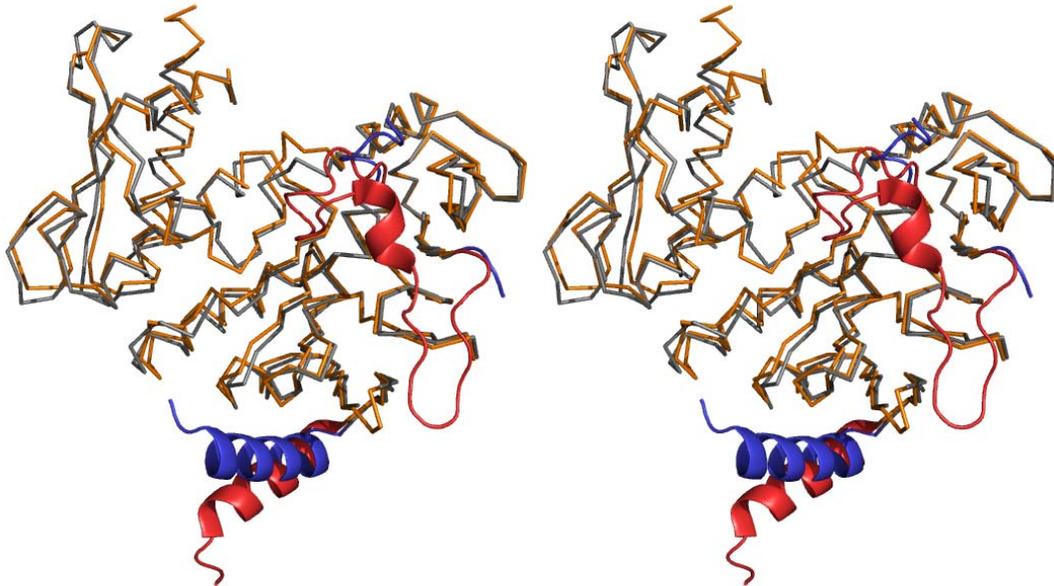
Color coded stick model of the active center of SMCcd with bound ATP, Mg<sup>2+</sup> is shown as magenta sphere and the attacking water molecule as red sphere. Only one out of two symmetrically related composite active sites in the dimer interface is shown. One subunit is depicted with yellow, the other with green C atoms. Key side chains are labeled (see text for details) and notable hydrogen bonds are shown as dashed lines.

For the hydrolysis of the ATP a water molecule carries out a nucleophilic attack on the  $\gamma$  phosphate. In the crystal structure a water molecule is observed that is poised for co-linear attack on the scissile P $\gamma$ -O bond. This likely nucleophile is optimally positioned and activated by hydrogen bonds to Walker B E1098 (mutated to Q to facilitate crystallization) and the main-chain carbonyl O of H1102 from the H-loop / H-switch from the opposing subunit.

#### 6.4.4 Structural comparison between monomeric and dimeric SMCcd

The nucleotide free and the ATP bound SMCcd can be superimposed with Isqman (Kleywegt, 1996) to a rmsd of only 0.99 Å for the C $\alpha$  atoms, indicating clearly that there are only minor conformational changes upon ATP binding and dimerization (Fig. 32). Only two regions show larger differences in the two SMCcd structures. This is on the one hand the C terminal helix that rotates away from the protein in the dimeric form (red helix in Fig. 32). Thereby steric clashes with the other protein are avoided and

after dimerization the helix spans towards the other subunit and takes part in protein-protein interaction.



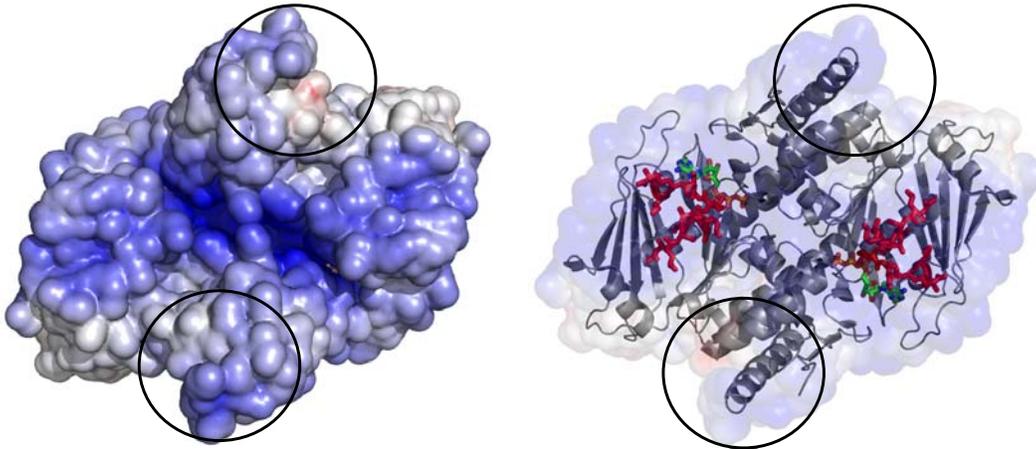
**Figure 32: Overlay of nucleotide free and ATP bound SMCcd.**

Stereo view of the backbone traces of nucleotide free SMCcd (gray) and ATP bound SMCcd (orange) superimposed using lsqman. The striking different parts, namely the R-loop and the C-terminal helix are colored blue for the apo protein and red in the ATP bound structure.

The largest rearrangement occurs at a surface loop in the vicinity of the ATPase site. In the nucleotide free crystal structure this loop spanning residues 50-80 is disordered and not visible in the electron density map (Fig. 32). After ATP binding this loop is ordered and forms helix B (Fig 32 and Fig. 28).

#### 6.4.5 The conserved arginine finger

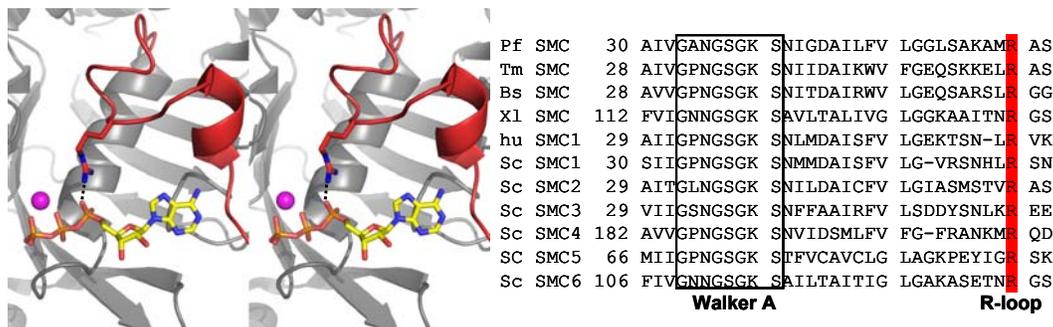
The R-loop that is ordered upon ATP binding and dimerization is located in the center at the concave side of the SMCcd dimer. As shown in figure 33 the R-loop contributes mainly to the positively charged surface region spanning the whole dimer in the inner side of a SMC ring.



**Figure 33: Surface of dimeric SMCcd with mapped electrostatic potential and R-loop.**

On the left side top view of the ATP bound dimer (see Fig. 29). The electrostatic potential (-10 kT/e [blue] to +10 kT/e [red]) is mapped on the solvent accessible surface. Bound ATP is shown in color coded sticks. On the right hand side same view with transparent surface and ribbon representation of the protein. The R-loop is shown as red sticks. The areas of the protruding coiled-coils are encircled.

The loop is stabilized by R59 that is fixed in the dimer after ATP binding. This arginine finger directly hydrogen bonds to the  $\alpha$  phosphate of the ATP. Sequence alignment of different SMC proteins from various species shows that this arginine finger is conserved among all SMC proteins (Fig.34).

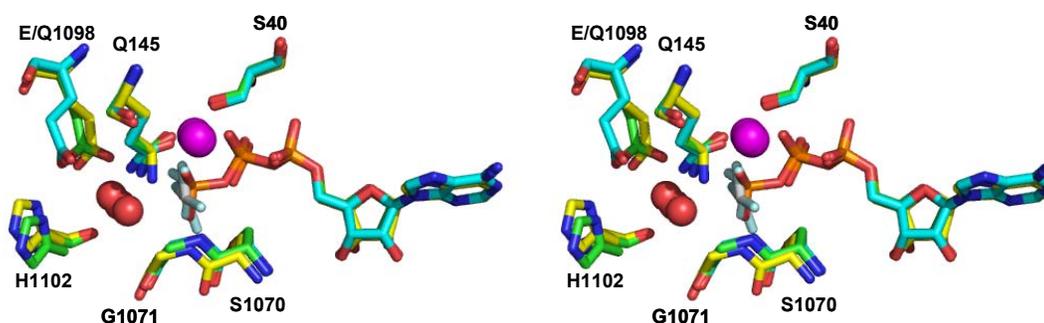


**Figure 34: Detail view of the R-loop and alignment of the arginine finger.**

Detailed stereo view of a ribbon model of SMCcd with the R-loop colored in red. The arginine finger R59 and ATP are shown as color coded sticks with the hydrogen bond as dashed line ( $Mg^{2+}$ : magenta sphere). On the right side sequence alignment (ClustalW (Thompson et al., 1994)) showing that the arginine finger (highlighted in red) in the R-loop is conserved among SMC proteins. Pf: *Pyrococcus furiosus*; Tm: *Thermatoga maritima*; Bs: *Bacillus subtilis*; hu: human; and Sc: *Saccharomyces cerevisiae*.

### 6.4.6 Structures with bound transition state analogues

When the SMCcd structures with bound transition state analogues are superimposed with ATP bound structure the overall rmsd for the  $C\alpha$  atoms varies between 0.36 (ATP to ADP-BeF<sub>3</sub><sup>-</sup>) and 0.53 (ADP-AlF<sub>4</sub><sup>-</sup> to ADP-BeF<sub>3</sub><sup>-</sup>) indicating highly similar structures with only minor reorganizations. In addition SMCcd with bound ADP- BeF<sub>3</sub><sup>-</sup> and ADP-AlF<sub>4</sub><sup>-</sup> is packed in the crystal in the same way as the ATP bound SMCcd. Applying the 2-fold symmetry operation used in the ATP form results in a dimer similar to the ATP induced dimer.



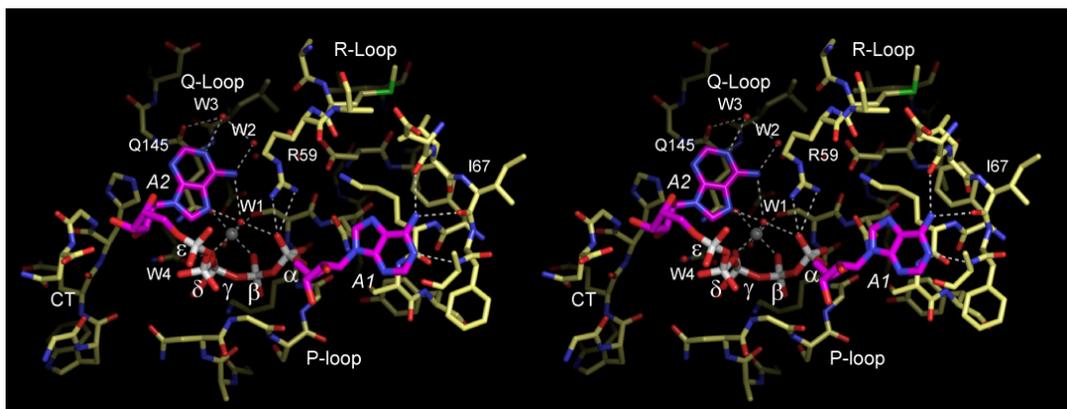
**Figure 35: Superimposed active site of SMCcd bound to ATP, ADP-AlF<sub>4</sub><sup>-</sup> and ADP-BeF<sub>3</sub><sup>-</sup>.** Stereo view of the active site with bound ATP, ADP-AlF<sub>4</sub><sup>-</sup> and ADP-BeF<sub>3</sub><sup>-</sup> with key residues shown as color coded sticks green: ATP form; yellow: ADP-AlF<sub>4</sub><sup>-</sup> form; blue: ADP-BeF<sub>3</sub><sup>-</sup> form. The Mg<sup>2+</sup> ions are depicted as magenta spheres and the attacking water molecules in red.

In the ADP-AlF<sub>4</sub><sup>-</sup> bound structure helices E and F are not ordered. Inspection of the active center reveals no indication for a conformational change which might lead to rearrangement of this region. The key residues at the active site responsible for the ATP binding and hydrolysis adopted equal conformation in all three structures and hold ATP / ADP-AlF<sub>4</sub><sup>-</sup> / ADP-BeF<sub>3</sub><sup>-</sup> and the Mg<sup>2+</sup> ion in quasi identical positions. Only the activating E1098 and the attacking water molecule vary slightly in their position in response to the additional charge and the altered shape of the ligands.

### 6.4.7 Structure of SMCcd in complex with AP5A

AP5A as a bi-substrate inhibitor of adenylate kinases blocks two sites at the enzyme: one adenosine binds the ATP site (from now adenosine 1), while the other adenosine binds the AMP site (from now adenosine 2) in the situation of the forward reaction. In

the crystal structure of SMCcd in complex with AP5A the inhibitor adopts a double bend “S” like conformation (Fig. 36). Adenosine 1 plus  $\alpha$ ,  $\beta$  and  $\gamma$  phosphates and one  $Mg^{2+}$  ion bind to the canonical ATP binding site of lobe I of SMCcd with an identical arrangement as in the ATP bound structure. The adenine 1 is specifically recognized and bound by hydrogen bonds to the main chain at residues 66-68. The  $\alpha$  and  $\beta$  phosphates are bound by the Walker A motif / P-loop and the conserved R59 of the arginine finger.



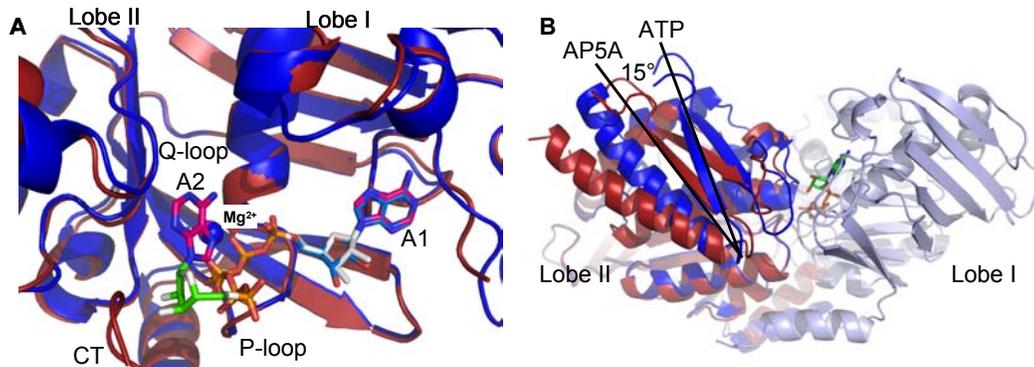
**Figure 36: Binding of AP5A to SMCcd.**

Stereo view of the bi-substrate inhibitor AP5A bound to SMCcd both represented as color coded stick models. Selected motifs, amino acids, water molecules (red spheres), hydrogen bonds (dashed lines) the  $Mg^{2+}$  ion (grey sphere) are annotated.

The coordination of the  $Mg^{2+}$  ion is similar to that of the ATP bound structure. Two coordination sites of the  $Mg^{2+}$  are occupied by oxygens of the  $\beta$  and  $\gamma$  phosphates. The coordination sphere of the  $Mg^{2+}$  ion is completed by side chain oxygen of Q145, S40 (Walker A) and two water molecules. The  $\delta$  and  $\epsilon$  phosphates are not recognized by the protein. Ribose 2 is bound by the C-terminal helix of the protein. This tail is not conserved among ABC proteins and thus this interaction might be due to crystal contacts. However, the C-terminal helix occupies the position of the D-loop from the opposing NBD in the dimeric form and thus the D-loop interacts with the AMP site.

Adenine 2 is bound to the interface of lobe I and II. The AMP site is formed on the one hand by the side chain of Q145 (Q-loop). Adenine 2 stacks onto the carboxamide moiety of Q145 in a similar manner as the interaction of adenine 1 with the K13-S14 peptide bond. In addition the N6 amine and N7 nitrogen of adenine 2 form two

hydrogen bonds to a  $Mg^{2+}$  ion water ligand (W1 in Fig. 36). Finally the N4 nitrogen binds the Q145 main chain carbonyl via a water molecule (W3 in Fig. 36).

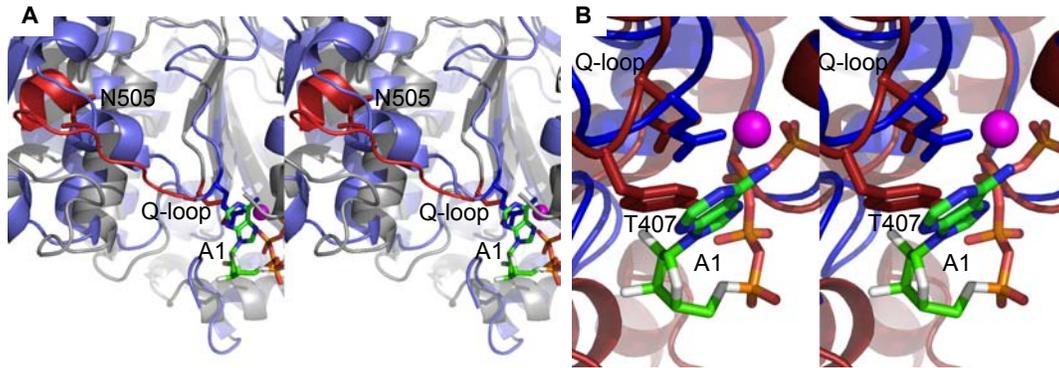


**Figure 37: Comparison of ATP and AP5A bound structures of SMCcd.**

Ribbon representation of the ATP bound (blue) and AP5A bound (red) structure of SMCcd with the ligand in color coded stick model. A) Active site view of the overlaid structures. The two Adenosins are denoted with A1 and A2, the two  $Mg^{2+}$  ions are shown as spheres and the three important binding sites P-loop, Q-loop and C tail (CT) are labeled. B) Domain view of the superposition with the color code of A). The opposing NBD in the ATP bound dimer is shown in light blue. Lobe II rotates in the AP5A bound structure  $\sim 15^\circ$  relative to lobe I in comparison to the ATP bound structure.

Superposition of the AP5A bound structure to the ATP bound structure shows a virtual identical arrangement of lobe I, the ATP and the adenosine I plus the  $\alpha$ ,  $\beta$  and  $\gamma$  phosphates respectively in both structures. However, in the AP5A structure lobe II is rotated  $\sim 15^\circ$  relative to lobe I in respect to the ATP structure due to the fact that the conformation of the Q-loop is different in both structures.

The relevance of these findings for the adenylate kinase activity of other ABC ATPases was further supported by comparison of the structure with known structures of two other ABC ATPases. In the CFTR protein the mutation N1303K in the NBD2 involved in cystic fibrosis disease abolishes the adenylate kinase activity (Randak and Welsh, 2003). Comparing the structure of AP5A bound SMCcd with the known structure of the NBD1 of mouse CFTR reveals that this in both NBD conserved glutamine (N505 in NBD1) binds the main chain thereby stabilizing the  $\beta$ -strand that follows the Q-loop (Fig. 38 A). Mutations of this residue in the NBD2 might lead to a destabilization of the whole region and perturb the AMP binding site by affecting the Q-loop.



**Figure 38: Comparison of AP5A bound SMCcd to CFTR NBD1 and ABCE1.**

A) Stereo view of the overlay of the AP5A bound SMCcd structure (light blue) to the structure of CFTR NBD1 (grey - PDB ID 1ROX). AP5A, Q-loop glutamine and N505 are depicted as color coded sticks and Mg<sup>2+</sup> as magenta sphere. B) Stereo view of the superposition of AP5A bound SMCcd and ABCE1 (red - PDB ID 1YQT) as ribbon presentation. AP5A, the Q-loop glutamine and Tyr407 are shown as color coded sticks and Mg<sup>2+</sup> as magenta sphere.

The ABCE1 ABC ATPase has a conserved tyrosine next to the Q-loop glutamine. Superposition of ABCE1 with the AP5A bound SMCcd shows that this residue points towards the putative AMP binding site (Fig. 38 B). The tyrosine thereby specifically blocks the adenine binding surface of the Q-loop glutamine and impedes adenylate kinase activity of ABCE1.

## 7 Discussion

Accurate propagation of genetic information during cell division requires proper DNA replication and three other important processes: DNA repair by recombination, sister chromatid cohesion and chromosome condensation. Proteins of the structural maintenance of chromosomes (SMC) protein family play a key role in each of these three processes by forming the core of the involved multiprotein complexes. SMC proteins are highly conserved and present in all kingdoms of life. The outstanding role of this protein family arises from their unique architecture that allows them to span long distances and form large proteinous rings. Sequence analysis revealed a ~500 Å long coiled-coil joining an ABC ATPase domain on one end and a hinge domain on the other end. SMC proteins require ATP binding and hydrolysis to promote cohesion and condensation, which is suggested to proceed via formation of SMC ring assemblies (Arumugam et al., 2003; Bazett-Jones et al., 2002; Kimura and Hirano, 1997; Weitzer et al., 2003).

To gain insights into the function of SMC proteins the molecular structure of the protein should be solved. However, the expression and purification of eukaryotic proteins is challenging due to complexity and different codon usage. A good way to overcome this problem in protein crystallography is the use of archaeal homologous. Archaeal proteins show high sequence similarity to eukaryotic proteins but have better properties such as thermo stability and lack of flexible regions. Furthermore, eukaryotic genomes contain several genes for SMC proteins that form heterodimers. In archaeal genomes only a single SMC protein is encoded and therefore they function as homodimers. This fact facilitates structural and biochemical analysis. *Pyrococcus furiosus* (*Pfu*) lives at elevated temperatures that request highly stable proteins and was therefore chosen as source organism in this work. For crystallization another problem arises as the SMC proteins share a characteristic three domain architecture with a central coiled-coil domain that has a high flexibility in regard to its arrangement as seen in electron microscopy (Fig. 5)(Haering et al., 2002; Hopfner et al., 2000; Melby et al., 1998). This conformational heterogeneity would impede crystallization where great homogeneity is a prerequisite for proper three dimensional arrangement of the proteins.

To overcome this problem only the catalytically active nucleotide binding domain (NBD) was used for crystallization. The residues building up the NBD were determined computer based using a web tool for coiled-coil prediction (Lupas et al., 1991). The coiled-coil and hinge domain forming the central part of the polypeptide chain were omitted from the protein. Amino acids forming the globular NBD can be found at the very N- and C-terminal part of the protein and were expressed as two fragments without a linker in a modified vector for bicistronic expression (Hoepfner, 2005). The two protein chains (aa 1-182 and 1006-1177) could be expressed in *E.coli* and form a single globular domain as indicated by heat stability, association throughout the whole purification procedure and the apparent molecular weight on the size exclusion column.

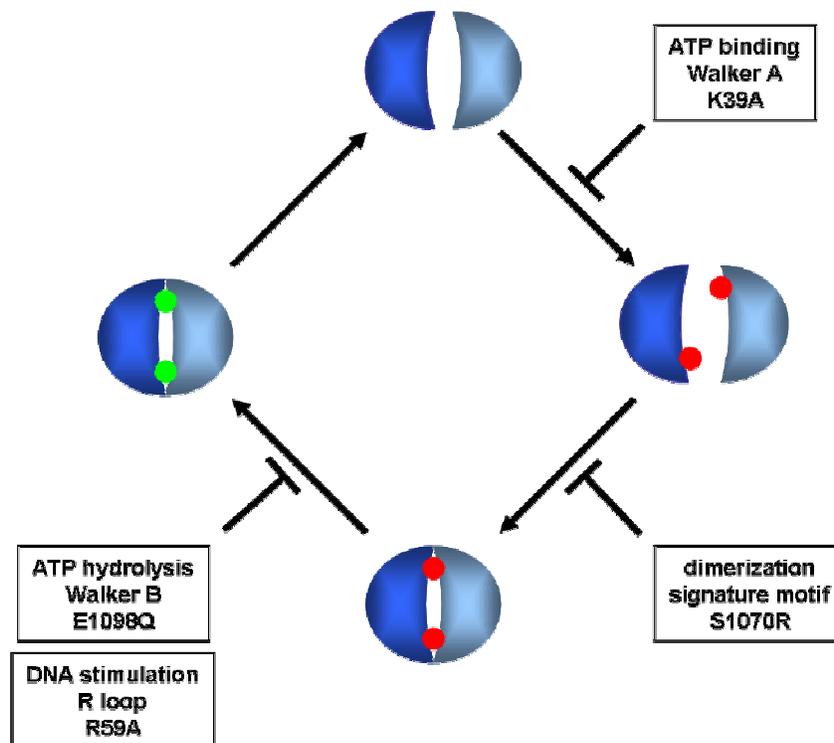
The resulting protein was crystallized and the structure was solved by *de novo* phasing to obtain best possible phases to a limiting resolution of 2.0 Å. The nucleotide binding domain of *Pfu* SMC adopts a two lobe conformation. Lobe I is formed by the N-terminal part of the protein and contains a central helix wrapped by a large  $\beta$  sheet. The C-terminal residues form the mainly  $\alpha$  helical lobe II. The two lobes are connected via a central four stranded  $\beta$  sheet composed of two  $\beta$  strands of both of the protein chains. Together with helices E and F, which extend from lobe I across lobe II, these extensive interactions explain why the N- and C-terminal part of the *Pfu* SMC proteins form a stable nucleotide binding domain upon co-expression. This stable interaction of the two parts of the protein is essential in nature to ensure functionality of the protein at elevated temperatures and thereby allow growth of *Pyrococcus furiosus*. The coiled-coil missing in the construct protrudes from lobe II in extension of helix G. Helix F is not yet oriented in direction of the coiled-coil and therefore the protein chain has to turn upwards to be able to interact with the extension of helix G. Comparison with other ABC ATPase structures reveals that this two helices are conserved among them and the coiled-coil can be therefore regarded as an insertion into the loop between these two helices. The structural similarity of the *Pfu* SMC NBD to other ABC ATPase structures is striking. Especially it has not only a high similarity to the structures of the NBDs of *Pfu* Rad50 and *Thermotoga maritima* SMC, both archaeal proteins, but also to the eukaryotic *Saccharomyces cerevisiae* SMC1 indicating that results obtained on the *Pfu* protein can serve as a model for all SMC proteins (Haering et al., 2004; Hopfner et al., 2000; Lowe et al., 2001).

To understand the role of ATP the protein was crystallized in the presence of ATP. No Crystals were obtained with none or slowly hydrolysable ATP analogues such as ATP $\gamma$ S and AMP-PNP, which might be due to their slightly different sterical properties. For crystallization of the ATP bound complex, it was necessary to introduce the Walker B motif mutation E1098Q, which traps the ATP complex of proteins. This mutation strongly reduces the ATPase activity of the NBD as shown in the ATPase assays. The reduced ATPase activity leads to formation of a dimeric ATP-protein complex as judged by the increased Stoke's radius on a size exclusion column. A complex of the same size but in a lower fraction compared to monomeric SMCcd could be obtained by incubating wild type protein with ATP $\gamma$ S. This lower occurrence of dimeric protein in the presence of ATP $\gamma$ S in comparison to ATP might explain the inability of crystal formation.

Inspection of the crystal packing revealed that a SMCcd-SMCcd homodimer with two sandwiched Mg<sup>2+</sup>-ATP moieties assembled along the crystallographic 2-fold axis. The two SMCcd domains are arranged "head-to-tail" creating two active sites in the dimer interface. This "head-to-tail" dimerization where lobe I of one monomer interacts with lobe II of the other was first shown for *Pfu* Rad50cd and MutS to be the functional arrangement of ABC ATPases (Hopfner et al., 2000; Lamers et al., 2000; Obmolova et al., 2000). Only in this orientation Walker A and Walker B motif from one SMCcd subunit and the signature motif from the opposing SMCcd come together and form a composite active site. The interaction of the Walker A motif of one subunit with the adenine ring and hydrogen bonding of the signature motif of the opposing molecule to the  $\gamma$  phosphate is the main driving force for the dimerization. Overall the buried surface between ATP and protein contributes with 1800 Å<sup>2</sup> in a similar amount to the dimer interface as the 2600 Å<sup>2</sup> buried surface between the proteins. Thus, assuming the ring model for the arrangement of the SMC dimer in the cell, ATP binding is the major driving force of ring closing and ATP hydrolysis the reason for ring opening at the nucleotide binding domains of the proteins.

ATP hydrolysis by ABC ATPases is achieved by the nucleophilic attack of a water molecule at the  $\gamma$  phosphate, subsequent disengagement of the dimer and release of ADP and P<sub>i</sub>. In the SMCcd complex structure a water molecule is poised for collinear

attack on the scissile bond between the  $\beta$  and  $\gamma$  phosphate. Positioning and activation of the water is achieved by cooperation of the two subunits. One SMCcd provides the Walker B glutamic acid (E1098) and the opposing subunit contributes a backbone carbonyl oxygen from a histidine (H1102). The contributions to positioning and activation of the nucleophile from both subunits suggest that ATP hydrolysis only takes place in the engaged SMCcd dimer. These findings are supported by mutagenesis data showing that disengagement of the dimer interaction by introducing a bulky arginine instead of a serine (S1070R) at the dimer interface abolishes ATP hydrolysis. The important role of the Walker B residue E1098 could be strikingly supported by mutating this residue to a glutamine as in the protein used for crystallization of the ATP complex. Changing the carboxyl group into an amide group strongly reduces ATPase activity of SMCcd.



**Figure 39: ATPase cycle of SMC proteins with functional mutations.**

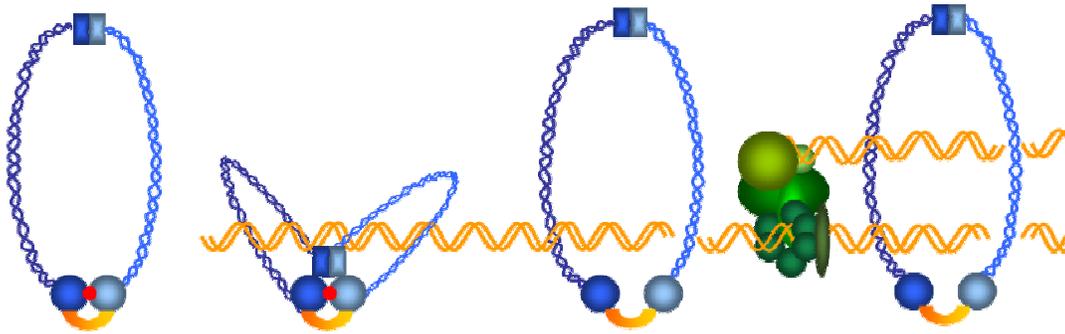
Each SMC head domain binds one ATP molecule (red; green: ADP). Mutating the Walker A lysine impairs nucleotide binding. ATP binding promotes dimerization which is prevented by introducing a bulky arginine at the position of the signature motif serine. The following ATP hydrolysis reaction can be disturbed by mutating the Walker B glutamic acid to a glutamine. DNA stimulation of the ATPase activity is abolished when the arginine finger is mutated. (adopted from (Losada and Hirano, 2005))

The cycle of ATP binding, ATP hydrolysis and nucleotide release (Fig.39) induces a conformational change within the ABC domain in many ABC enzymes, especially in ABC transporter and Rad50. This conformational switch was denoted as “power stroke” and is transmitted to the substrate specific domains of the ABC ATPase to drive the enzymatic cycle (Hopfner and Tainer, 2003). Comparison of the nucleotide free form, the ATP bound form and SMCcd bound to transition state analogues revealed very similar overall conformations in all states of the catalytic cycle. This indicates that SMC proteins do not possess a “power stroke” that lead to conformational changes throughout the molecule. In fact ATP binding and hydrolysis by SMC proteins seems to be limited to the closing and opening of the SMC protein rings or other higher order assemblies. However, there are two regions that change their conformation upon ATP binding and dimerization. The C-terminal helix L rotates away from the ABC domain and binds to the opposing subunit. This rearrangement points towards a role of this helix in the control of SMCcd association and ATPase activity by additional non-SMC subunits (Hirano and Hirano, 2004). A second conformational change is observed in a surface loop. Upon ATP binding this loop is ordered by the binding of an arginine to the  $\alpha$  phosphate. In the nucleotide free form of SMCcd this loop is partially disordered. The R-loop is located at the center of the positively charged surface at the concave side of the SMCcd dimer on lobe I and is strictly conserved among all SMC proteins. The ATP driven conformational variability and the location in the center of a positively charged surface suggests that the R-loop is a good candidate for the coupling of ATPase activity to interaction of the protein with DNA. An equivalent positively charged region has been implicated in DNA binding by Rad50 (Hopfner et al., 2001). To test the effect of DNA on SMCcd the ATPase activity was assayed in the absence and presence of dsDNA on wild type and mutant protein. The basal ATPase activity of wt *Pfu* SMCcd in the absence of DNA is weak but strongly stimulated in the presence of dsDNA. A similar DNA dependent stimulation of ATPase activity was observed for both bacterial and eukaryotic multisubunit SMC protein complexes, suggesting that this DNA stimulation directly acts on the ABC domains in SMC complexes (Hirano et al., 2001; Kimura and Hirano, 1997). Mutating the conserved arginine finger (R59) to an alanine has no effect on the basal ATPase activity but completely abolishes the stimulating effect of dsDNA approving its role in coupling DNA interaction to ATP

hydrolysis. The R-loop might be rearranged upon DNA interaction and R59 directed towards the ATP where it most likely counterbalances the developing negative charge during ATP hydrolysis. The structural and mechanistic role of the arginine finger shows remarkable analogy to the activation of small GTPases by GTPase activating proteins (GAPs). GAPs insert into the active site of G proteins an arginine finger that both binds GTP and stimulates GTP hydrolysis (Ahmadian et al., 1997; Scheffzek et al., 1997).

*In vivo* it was shown that this arginine finger plays indeed an important role in the biological function of yeast cohesin (Lengronne et al., 2006). Strains lacking the arginine finger in either SMC1 or SMC3 are moderately threatened by chromosome loss. Cells with the arginine mutated to an alanine in both SMC1 and SMC3 are viable but show dramatically increased chromosome instability. The fact that the cells with altered DNA stimulation are still viable provided a great tool for investigating the establishment of sister chromatid cohesion in the living cell. Measuring the kinetics of wild type and mutant cohesin loading onto chromosomes showed dramatically delayed association of the SMC complexes lacking the arginine in comparison to the wild type complex during G1 phase. However, mutating the arginine finger has no effect on the establishment of sister chromatid cohesion in S phase. These findings point towards a role of the R-loop in the initial loading of the SMC complexes onto DNA. Recent data suggest that this loading might be achieved by stepwise transport of the DNA through the hinge (Gruber et al., 2006). To allow DNA entry the hinge domain presumably has to interact with the head domains that are dimerized upon ATP binding. After passage of a DNA strand ATP is hydrolyzed and the hinge to head interaction is reversed (Fig. 40). In the case of cohesin this loading might be restricted to only one DNA strand per SMC ring whereas in the case of condensin several cycles of ATP binding – DNA loading – ATP hydrolysis are conceivable that trap loops of DNA to condensate the chromosome. Head domains missing the arginine finger have presumably an altered and less defined surface. They provide an insufficient interaction area to the hinge and thereby hinder an efficient loading of SMC rings onto DNA. After ATP hydrolysis the closing of the ring might be achieved by non-SMC proteins, e.g. Scc1 in the cohesin complex. For the establishment of sister chromatid cohesion two models are possible. The whole or parts of the replication machinery slide through the ring or on the other

hand the ring is transiently opened and closed again after the replication fork. So far no evidence for the latter has been found but it must be different from the first loading because the arginine finger mutations have no influence at this stage (Lengronne et al., 2006).



**Figure 40: Model for the establishment of sister chromatid cohesion.**

ATP (red) binding to the head domains leads to their dimerization. The hinge domain interacts with the inner surface of the dimerized head domains. Upon this interaction DNA can be passed sequentially through the hinge in the inner side of the SMC ring. The DNA now interacts with the arginine finger and allosterically activates the hydrolysis of ATP. This reaction impairs the head to head and head to hinge interactions. The ring is now closed by Scc1 proteins (orange). After the replication machinery (green) duplicated the chromosomes both sister chromatids are held together till the cleavage of Scc1 at the onset of anaphase. In the case of condensin the cycle of ATP binding – DNA gating – ATP hydrolysis might be run through several times to loop the DNA.

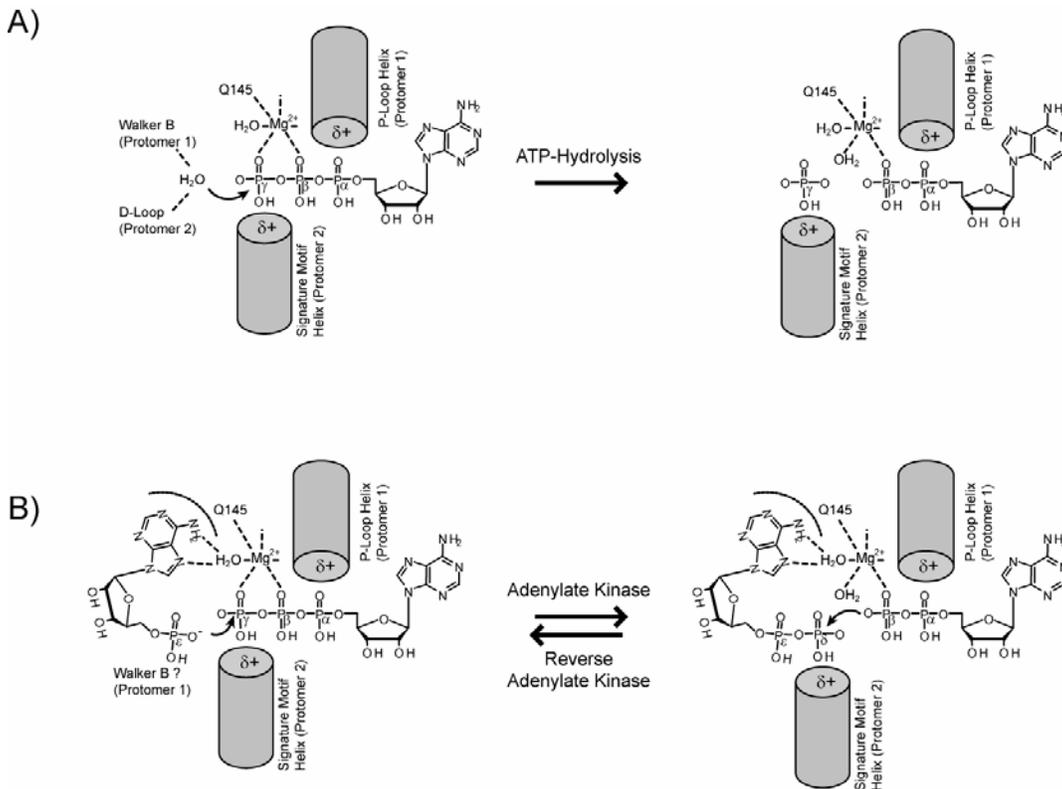
However, other modes of association of SMC complexes with DNA are possible. So far it is not clear if the huge structural rearrangement of the coiled-coil needed for the hinge to head interaction really occurs, although it was observed in some atomic force microscopy studies (Sakai et al., 2003; Yoshimura et al., 2002). Furthermore, this model does not explain why ATP hydrolysis is essential for cohesin loading. For the *Bacillus subtilis* SMC a sitting model is proposed where the DNA interacts with the hinge and is subsequently loaded through the head domains (Shintomi and Hirano, 2007). Taking the high sequence and structure similarity between the different SMC proteins from different organisms into account it is unlikely that their function is based on such different mechanisms. In addition the role of essential loading factors, in yeast for example Eco1, Ctf4 and Ctf18, is completely unknown so far (Lengronne et al., 2006; Milutinovich et al., 2007; Toth et al., 1999). The DNA might be directly gated through the head domains across the non-SMC subunits assisted by loading factors. Besides for Scc1 not much is known at structural and functional level about the non-

SMC proteins involved in the SMC complexes, too (Haering et al., 2004). Scc1 serves in the cohesin complex as a bridge between the head domains of SMC1 and SMC3 and is cleaved by a protease named separase on anaphase onset to allow sister chromatid separation (Uhlmann, 2001; Uhlmann et al., 2000). In bacteria ScpA and ScpB are the known subunits of the SMC complexes that share only weak sequence homology to eukaryotic non-SMC subunits. In *Pyrococcus furiosus* the gene for ScpA overlaps with the gene for SMC protein. This protein forms a tight complex with SMCcd but does not lead to higher order complexes *in vitro* than ScpA SMCcd dimers indicating that it has not a bridging function like Scc1. The interaction of SMCcd with ScpA had no significant influence on its ATPase activity although in *Bacillus subtilis* a moderately reduced activity was observed (Hirano and Hirano, 2004). In any case the SMCcd-ScpA complex also shows pronounced DNA stimulated ATPase activity, which is specifically abolished by the arginine finger mutation. Thus the DNA stimulated ATPase activity of SMCcd appears to be independent of its association with ScpA. A gene for a presumable homologue of ScpB was found apart from the two other genes in the genome. The protein encoded by this gene does neither form a stable complex with SMCcd nor with ScpA. Assuming that it is really the ScpB homologue it might interact with the missing coiled-coil of the SMC protein to form a complex or additional proteins mediate the complex formation.

One important question remains when analyzing the mechanism of SMC proteins. Why does the process of DNA trapping require the energy of ATP hydrolysis although no power stroke could be observed at the ABC ATPase domain? A similar question arose by investigations on the CFTR ion channel. This transmembrane protein regulates the passive flow of anions down an electrochemical gradient (Vergani et al., 2005a). CFTR belongs to the ABC ATPase family of proteins and its two NBD can hydrolyse ATP *in vitro* (Vergani et al., 2005b). At physiological nucleotide concentrations, however, the channel is efficiently regulated by the reversible adenylate kinase reaction (Gross et al., 2006; Randak and Welsh, 2003; Randak and Welsh, 2005). The adenylate kinase reaction releases virtually no free energy under cellular nucleotide concentrations consistent with the lacking necessity for energy consumption during a passive ion gating function. The presence of the distinct ATP hydrolysis and adenylate kinase activities within the ABC protein family is puzzling because CFTR possesses all the

typical ATP hydrolysis motifs. Furthermore the molecular mechanism of adenylate kinase activity remains unclear. Thus, to provide a mechanistic framework for understanding the evolution of adenylate kinase activity in ABC enzymes we biochemically and structurally analyzed the adenylate kinase properties on *Pfu* SMCcd. While the forward adenylate kinase reaction ( $\text{ATP} + \text{AMP} \rightarrow 2 \text{ADP}$ ) was masked by the much more robust ATPase activity, the formation of ATP out of ADP could be shown using a luciferase assay. The much higher rate of ATP hydrolysis argues that SMC proteins *in vivo* hydrolyze ATP rather than being regulated by an adenylate kinase activity. Nevertheless, the finding that SMCcd is able to catalyze the reverse adenylate kinase reaction raises the possibility that in principle all ABC ATPases have this activity. For structural analysis SMCcd was crystallized in complex with AP5A a bisubstrate inhibitor of adenylate kinases that also inhibited the adenylate kinase reaction of SMCcd. Inspection of the structure clearly showed that residues responsible for the binding of AP5A and therefore involved in the catalytic reaction belong to the conserved ABC ATPase motifs. One adenosine plus three phosphates bind to the canonical ATP binding site of lobe I at the Walker A and B motifs. Adenosine 2 stacks onto the side chain of the conserved Q-loop glutamine and is discriminated by hydrogen bonding to the  $\text{Mg}^{2+}$  and the glutamine. Together the later interaction could explain the inability of GMP to substitute AMP in CFTR (Gross et al., 2006; Randak and Welsh, 2003). Mutant analysis revealed that not only Walker A and the Q-loop but in addition the signature motif is required for activity suggesting that the reaction occurs in the engaged dimer as it was shown for the ATPase reaction. In the structure the attacking  $\text{P}_\epsilon$  is not correctly positioned. This might be due to the additional phosphate in AP5A in contrast to the substrates in this reaction containing only four phosphates. Comparing the AP5A bound structure with the ATP bound form revealed a second possible way of regulation upon adenylate kinase reaction besides the dimerization. The presence of AP5A leads to a  $15^\circ$  rotation of lobe II in respect to lobe I accompanied by a reorientation of the Q-loop. Thus, AMP binding to CFTR could also help to orient lobes I and II, a process that may follow or facilitate engagement of the NBDs in response to ATP binding. Since the TMDs of ABC transporters bind near the lobe interface, both NBD engagement and rotation of the lobes relative to each other could contribute to channel opening. Comparison of the structure of SMCcd

bound to AP5A with the structure of the NBD1 of CFTR explains the effect of the N1303K mutation in NBD2 that abolishes adenylate kinase activity of CFTR *in vitro* and lead to cystic fibrosis. The corresponding residue in NBD1 (N505) binds the main chain and stabilizes a strand that follows the Q-loop supposing that a mutation could perturb the proposed AMP binding site. In ABCE1 the possible AMP binding site is blocked and thereby the adenylate kinase reaction is disabled. Taken together these findings suggest that all ABC enzymes can catalyze both the ATPase and the adenylate kinase reaction but differently utilize them in nature. During evolution the existing motifs were altered in a way that one or the other reaction is favored.



**Figure 41: Unifying model for the ATPase and adenylate kinase activity.**

Selected protein motifs, nucleotides and the active site magnesium are schematically shown. The proposed nucleophilic attacks of the water during hydrolysis (A) or phosphate oxygens for the adenylate kinase reaction (B) are indicated. In both cases the ATP is bound by the Walker A / P-loop of one protomer and by the signature motif of the opposing NBD via the  $\gamma$  phosphate. For ATP hydrolysis a water molecule is positioned and activated for nucleophilic attack by the Walker B motif of one protomer and the D-loop of the second protein. In the proposed adenylate kinase reaction the adenine of AMP/ADP binds the Q-loop and coordinates the magnesium facilitating the nucleophilic attack of the phosphate oxygen. In both cases the helices of the P-loop and signature motif compensate for the increased negative charge during reaction.

## 8 References

- Ahmadian, M. R., Stege, P., Scheffzek, K., and Wittinghofer, A. (1997). Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. *Nat Struct Biol* *4*, 686-689.
- Arumugam, P., Gruber, S., Tanaka, K., Haering, C. H., Mechtler, K., and Nasmyth, K. (2003). ATP hydrolysis is required for cohesin's association with chromosomes. *Curr Biol* *13*, 1941-1953.
- Bazett-Jones, D. P., Kimura, K., and Hirano, T. (2002). Efficient supercoiling of DNA by a single condensin complex as revealed by electron spectroscopic imaging. *Mol Cell* *9*, 1183-1190.
- Bisbal, C., Martinand, C., Silhol, M., Lebleu, B., and Salehzada, T. (1995). Cloning and characterization of a RNase L inhibitor. A new component of the interferon-regulated 2-5A pathway. *J Biol Chem* *270*, 13308-13317.
- Blow, D. (2002). *Outline of Crystallography for Biologists* (New York: Oxford University Press).
- Bradley, G., Juranka, P. F., and Ling, V. (1988). Mechanism of multidrug resistance. *Biochim Biophys Acta* *948*, 87-128.
- Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr* *59*, 2023-2030.
- Britton, R. A., Lin, D. C., and Grossman, A. D. (1998). Characterization of a prokaryotic SMC protein involved in chromosome partitioning. *Genes Dev* *12*, 1254-1259.
- Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* *54*, 905-921.
- Carson, D. R., and Christman, M. F. (2001). Evidence that replication fork components catalyze establishment of cohesion between sister chromatids. *Proc Natl Acad Sci U S A* *98*, 8270-8275.
- CCP4, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* *50*, 760-763.
- Chuang, P. T., Albertson, D. G., and Meyer, B. J. (1994). DPY-27: a chromosome condensation protein homolog that regulates *C. elegans* dosage compensation through association with the X chromosome. *Cell* *79*, 459-474.
- Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., and Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Mol Cell* *5*, 243-254.
- Dervyn, E., Noirot-Gros, M. F., Mervelet, P., McGovern, S., Ehrlich, S. D., Polard, P., and Noirot, P. (2004). The bacterial condensin/cohesin-like protein complex acts in DNA repair and regulation of gene expression. *Mol Microbiol* *51*, 1629-1640.
- Drenth, J. (1999). *Principles of Protein X-ray Crystallography* (Heidelberg: Springer Verlag).
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* *60*, 2126-2132.

- Engh, R. H., R. (1991). Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallogr D Biol Crystallogr* *47*, 392-400.
- Fujioka, Y., Kimata, Y., Nomaguchi, K., Watanabe, K., and Kohno, K. (2002). Identification of a novel non-structural maintenance of chromosomes (SMC) component of the SMC5-SMC6 complex involved in DNA repair. *J Biol Chem* *277*, 21585-21591.
- Gerlich, D., Hirota, T., Koch, B., Peters, J. M., and Ellenberg, J. (2006). Condensin I stabilizes chromosomes mechanically through a dynamic interaction in live cells. *Curr Biol* *16*, 333-344.
- Gillespie, P. J., and Hirano, T. (2004). Scc2 couples replication licensing to sister chromatid cohesion in *Xenopus* egg extracts. *Curr Biol* *14*, 1598-1603.
- Graumann, P. L. (2000). *Bacillus subtilis* SMC is required for proper arrangement of the chromosome and for efficient segregation of replication termini but not for bipolar movement of newly duplicated origin regions. *J Bacteriol* *182*, 6463-6471.
- Graumann, P. L. (2001). SMC proteins in bacteria: condensation motors for chromosome segregation? *Biochimie* *83*, 53-59.
- Gross, C. H., Abdul-Manan, N., Fulghum, J., Lippke, J., Liu, X., Prabhakar, P., Brennan, D., Willis, M. S., Faerman, C., Connelly, P., *et al.* (2006). Nucleotide-binding domains of cystic fibrosis transmembrane conductance regulator, an ABC transporter, catalyze adenylate kinase activity but not ATP hydrolysis. *J Biol Chem* *281*, 4058-4068.
- Gruber, S., Arumugam, P., Katou, Y., Kuglitsch, D., Helmhart, W., Shirahige, K., and Nasmyth, K. (2006). Evidence that Loading of Cohesin Onto Chromosomes Involves Opening of Its SMC Hinge. *Cell* *127*, 523-537.
- Haering, C. H., Lowe, J., Hochwagen, A., and Nasmyth, K. (2002). Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol Cell* *9*, 773-788.
- Haering, C. H., Schoffnegger, D., Nishino, T., Helmhart, W., Nasmyth, K., and Lowe, J. (2004). Structure and stability of cohesin's Smc1-kleisin interaction. *Mol Cell* *15*, 951-964.
- Hagstrom, K. A., Holmes, V. F., Cozzarelli, N. R., and Meyer, B. J. (2002). *C. elegans* condensin promotes mitotic chromosome architecture, centromere organization, and sister chromatid segregation during mitosis and meiosis. *Genes Dev* *16*, 729-742.
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* *166*, 557-580.
- Higgins, C. F., Haag, P. D., Nikaido, K., Ardeshir, F., Garcia, G., and Ames, G. F. (1982). Complete nucleotide sequence and identification of membrane components of the histidine transport operon of *S. typhimurium*. *Nature* *298*, 723-727.
- Hirano, M., Anderson, D. E., Erickson, H. P., and Hirano, T. (2001). Bimodal activation of SMC ATPase by intra- and inter-molecular interactions. *Embo J* *20*, 3238-3250.
- Hirano, M., and Hirano, T. (1998). ATP-dependent aggregation of single-stranded DNA by a bacterial SMC homodimer. *Embo J* *17*, 7139-7148.
- Hirano, M., and Hirano, T. (2002). Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. *Embo J* *21*, 5733-5744.
- Hirano, M., and Hirano, T. (2004). Positive and negative regulation of SMC-DNA interactions by ATP and accessory proteins. *Embo J* *23*, 2664-2673.
- Hirano, M., and Hirano, T. (2006). Opening closed arms: long-distance activation of SMC ATPase by hinge-DNA interactions. *Mol Cell* *21*, 175-186.
- Hirano, T. (2006). At the heart of the chromosome: SMC proteins in action. *Nat Rev Mol Cell Biol* *7*, 311-322.

- Hirano, T., Kobayashi, R., and Hirano, M. (1997). Condensins, chromosome condensation protein complexes containing XCAP-C, XCAP-E and a *Xenopus* homolog of the *Drosophila* Barren protein. *Cell* *89*, 511-521.
- Hirano, T., and Mitchison, T. J. (1994). A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro. *Cell* *79*, 449-458.
- Hirota, T., Gerlich, D., Koch, B., Ellenberg, J., and Peters, J. M. (2004). Distinct functions of condensin I and II in mitotic chromosome assembly. *J Cell Sci* *117*, 6435-6445.
- Hoeppner, S. (2005) Structure of the Mediator subunit Cyclin C and subunit interaction studies within the Mediator head module, PhD Thesis, LMU, Munich.
- Hopfner, K. P., Karcher, A., Craig, L., Woo, T. T., Carney, J. P., and Tainer, J. A. (2001). Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. *Cell* *105*, 473-485.
- Hopfner, K. P., Karcher, A., Shin, D. S., Craig, L., Arthur, L. M., Carney, J. P., and Tainer, J. A. (2000). Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* *101*, 789-800.
- Hopfner, K. P., and Tainer, J. A. (2000). DNA mismatch repair: the hands of a genome guardian. *Structure* *8*, R237-241.
- Hopfner, K. P., and Tainer, J. A. (2003). Rad50/SMC proteins and ABC transporters: unifying concepts from high-resolution structures. *Curr Opin Struct Biol* *13*, 249-255.
- Husain, I., Van Houten, B., Thomas, D. C., and Sancar, A. (1986). Sequences of *Escherichia coli* uvrA gene and protein reveal two potential ATP binding sites. *J Biol Chem* *261*, 4895-4901.
- Jun, S. H., Kim, T. G., and Ban, C. (2006). DNA mismatch repair system. Classical and fresh roles. *Febs J* *273*, 1609-1619.
- Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* *21*, 916-924.
- Karcher, A., Buttner, K., Martens, B., Jansen, R. P., and Hopfner, K. P. (2005). X-ray structure of RLI, an essential twin cassette ABC ATPase involved in ribosome biogenesis and HIV capsid assembly. *Structure* *13*, 649-659.
- Kim, S. T., Xu, B., and Kastan, M. B. (2002). Involvement of the cohesin protein, Smc1, in Atm-dependent and independent responses to DNA damage. *Genes Dev* *16*, 560-570.
- Kimura, K., and Hirano, T. (1997). ATP-dependent positive supercoiling of DNA by 13S condensin: a biochemical implication for chromosome condensation. *Cell* *90*, 625-634.
- Kimura, K., and Hirano, T. (2000). Dual roles of the 11S regulatory subcomplex in condensin functions. *Proc Natl Acad Sci U S A* *97*, 11972-11977.
- Kireeva, N., Lakonishok, M., Kireev, I., Hirano, T., and Belmont, A. S. (2004). Visualization of early chromosome condensation: a hierarchical folding, axial glue model of chromosome structure. *J Cell Biol* *166*, 775-785.
- Kitagawa, R., Bakkenist, C. J., McKinnon, P. J., and Kastan, M. B. (2004). Phosphorylation of SMC1 is a critical downstream event in the ATM-NBS1-BRCA1 pathway. *Genes Dev* *18*, 1423-1438.
- Klein, F., Mahr, P., Galova, M., Buonomo, S. B., Michaelis, C., Nairz, K., and Nasmyth, K. (1999). A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. *Cell* *98*, 91-103.
- Kleywegt, G. J. (1996). Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr D Biol Crystallogr* *52*, 842-857.
- Kleywegt, G. J., and Jones, T. A. (1998). Databases in protein crystallography. *Acta Crystallogr D Biol Crystallogr* *54*, 1119-1131.

- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.
- Lamers, M. H., Perrakis, A., Enzlin, J. H., Winterwerp, H. H., de Wind, N., and Sixma, T. K. (2000). The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature* 407, 711-717.
- Larionov, V. L., Karpova, T. S., Kouprina, N. Y., and Jouravleva, G. A. (1985). A mutant of *Saccharomyces cerevisiae* with impaired maintenance of centromeric plasmids. *Curr Genet* 10, 15-20.
- LeMaster, D. M., and Richards, F. M. (1985). 1H-15N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. *Biochemistry* 24, 7263-7268.
- Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G. P., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* 430, 573-578.
- Lengronne, A., McIntyre, J., Katou, Y., Kanoh, Y., Hopfner, K. P., Shirahige, K., and Uhlmann, F. (2006). Establishment of sister chromatid cohesion at the *S. cerevisiae* replication fork. *Mol Cell* 23, 787-799.
- Lieb, J. D., Albrecht, M. R., Chuang, P. T., and Meyer, B. J. (1998). MIX-1: an essential component of the *C. elegans* mitotic machinery executes X chromosome dosage compensation. *Cell* 92, 265-277.
- Lieb, J. D., Capowski, E. E., Meneely, P., and Meyer, B. J. (1996). DPY-26, a link between dosage compensation and meiotic chromosome segregation in the nematode. *Science* 274, 1732-1736.
- Locher, K. P., Lee, A. T., and Rees, D. C. (2002). The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* 296, 1091-1098.
- Losada, A., and Hirano, T. (2005). Dynamic molecular linkers of the genome: the first decade of SMC proteins. *Genes Dev* 19, 1269-1287.
- Lowe, J., Cordell, S. C., and van den Ent, F. (2001). Crystal structure of the SMC head domain: an ABC ATPase with 900 residues antiparallel coiled-coil inserted. *J Mol Biol* 306, 25-35.
- Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* 252, 1162-1164.
- Martinez-Balbas, M. A., Dey, A., Rabindran, S. K., Ozato, K., and Wu, C. (1995). Displacement of sequence-specific transcription factors from mitotic chromatin. *Cell* 83, 29-38.
- Mascarenhas, J., Soppa, J., Strunnikov, A. V., and Graumann, P. L. (2002). Cell cycle-dependent localization of two novel prokaryotic chromosome segregation and condensation proteins in *Bacillus subtilis* that interact with SMC protein. *Embo J* 21, 3108-3118.
- McDonald, W. H., Pavlova, Y., Yates, J. R., 3rd, and Boddy, M. N. (2003). Novel essential DNA repair proteins Nse1 and Nse2 are subunits of the fission yeast Smc5-Smc6 complex. *J Biol Chem* 278, 45460-45467.
- McPherson, A. (2001). *Crystallization of Biological Macromolecules* (New York: Cold Spring Harbor Laboratory Press ).
- Melby, T. E., Ciampaglio, C. N., Briscoe, G., and Erickson, H. P. (1998). The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge. *J Cell Biol* 142, 1595-1604.

- Michaelis, C., Ciosk, R., and Nasmyth, K. (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* *91*, 35-45.
- Milutinovich, M., Unal, E., Ward, C., Skibbens, R. V., and Koshland, D. (2007). A Multi-Step Pathway for the Establishment of Sister Chromatid Cohesion. *PLoS Genet* *3*, e12.
- Morris, R. J., Perrakis, A., and Lamzin, V. S. (2003). ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol* *374*, 229-244.
- Nasim, A., and Smith, B. P. (1975). Genetic control of radiation sensitivity in *Schizosaccharomyces pombe*. *Genetics* *79*, 573-582.
- Nasmyth, K., and Haering, C. H. (2005). The structure and function of SMC and kleisin complexes. *Annu Rev Biochem* *74*, 595-648.
- Niki, H., Imamura, R., Kitaoka, M., Yamanaka, K., Ogura, T., and 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*, 5101-5109.
- Niki, H., Jaffe, A., Imamura, R., Ogura, T., and Hiraga, S. (1991). The new gene mukB codes for a 177 kd protein with coiled-coil domains involved in chromosome partitioning of *E. coli*. *Embo J* *10*, 183-193.
- Obmolova, G., Ban, C., Hsieh, P., and Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* *407*, 703-710.
- Oh, E. Y., Claassen, L., Thiagalingam, S., Mazur, S., and Grossman, L. (1989). ATPase activity of the UvrA and UvrAB protein complexes of the *Escherichia coli* UvrABC endonuclease. *Nucleic Acids Res* *17*, 4145-4159.
- Ono, T., Losada, A., Hirano, M., Myers, M. P., Neuwald, A. F., and Hirano, T. (2003). Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells. *Cell* *115*, 109-121.
- Oswald, C., Holland, I. B., and Schmitt, L. (2006). The motor domains of ABC-transporters. What can structures tell us? *Naunyn Schmiedebergs Arch Pharmacol* *372*, 385-399.
- Otwinowski, Z. a. M., W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode *Methods in Enzymology* *276*, 307-326.
- Phipps, J., Nasim, A., and Miller, D. R. (1985). Recovery, repair, and mutagenesis in *Schizosaccharomyces pombe*. *Adv Genet* *23*, 1-72.
- Randak, C., and Welsh, M. J. (2003). An intrinsic adenylate kinase activity regulates gating of the ABC transporter CFTR. *Cell* *115*, 837-850.
- Randak, C. O., and Welsh, M. J. (2005). Adenylate kinase activity in ABC transporters. *J Biol Chem* *280*, 34385-34388.
- Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J. L., and et al. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* *245*, 1066-1073.
- Sakai, A., Hizume, K., Sutani, T., Takeyasu, K., and Yanagida, M. (2003). Condensin but not cohesin SMC heterodimer induces DNA reannealing through protein-protein assembly. *Embo J* *22*, 2764-2775.
- Scheffzek, K., Ahmadian, M. R., Kabsch, W., Wiesmuller, L., Lautwein, A., Schmitz, F., and Wittinghofer, A. (1997). The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* *277*, 333-338.
- Shintomi, K., and Hirano, T. (2007). How are cohesin rings opened and closed? *Trends Biochem Sci*.
- Sixma, T. K. (2001). DNA mismatch repair: MutS structures bound to mismatches. *Curr Opin Struct Biol* *11*, 47-52.

- Skibbens, R. V., Corson, L. B., Koshland, D., and Hieter, P. (1999). Ctf7p is essential for sister chromatid cohesion and links mitotic chromosome structure to the DNA replication machinery. *Genes Dev* *13*, 307-319.
- Smith, P. C., Karpowich, N., Millen, L., Moody, J. E., Rosen, J., Thomas, P. J., and Hunt, J. F. (2002). ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol Cell* *10*, 139-149.
- Steffensen, S., Coelho, P. A., Cobbe, N., Vass, S., Costa, M., Hassan, B., Prokopenko, S. N., Bellen, H., Heck, M. M., and Sunkel, C. E. (2001). A role for *Drosophila* SMC4 in the resolution of sister chromatids in mitosis. *Curr Biol* *11*, 295-307.
- Sunako, Y., Onogi, T., and Hiraga, S. (2001). Sister chromosome cohesion of *Escherichia coli*. *Mol Microbiol* *42*, 1233-1241.
- Takahashi, T. S., Yiu, P., Chou, M. F., Gygi, S., and Walter, J. C. (2004). Recruitment of *Xenopus* Scc2 and cohesin to chromatin requires the pre-replication complex. *Nat Cell Biol* *6*, 991-996.
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* *22*, 4673-4680.
- Torres-Rosell, J., Machin, F., Farmer, S., Jarmuz, A., Eydmann, T., Dalgaard, J. Z., and Aragon, L. (2005). SMC5 and SMC6 genes are required for the segregation of repetitive chromosome regions. *Nat Cell Biol* *7*, 412-419.
- Toth, A., Ciosk, R., Uhlmann, F., Galova, M., Schleiffer, A., and Nasmyth, K. (1999). Yeast cohesin complex requires a conserved protein, Eco1p(Ctf7), to establish cohesion between sister chromatids during DNA replication. *Genes Dev* *13*, 320-333.
- Turck, D. (1992) Weiterentwicklung eines Programms fuer Molekuelgraphik und Elektronendichte-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklaerungen., PhD Thesis, Munich, Germany.
- Uhlmann, F. (2001). Secured cutting: controlling separase at the metaphase to anaphase transition. *EMBO Rep* *2*, 487-492.
- Uhlmann, F., Lottspeich, F., and Nasmyth, K. (1999). Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. *Nature* *400*, 37-42.
- Uhlmann, F., Wernic, D., Poupard, M. A., Koonin, E. V., and Nasmyth, K. (2000). Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. *Cell* *103*, 375-386.
- Van Houten, B., Croteau, D. L., DellaVecchia, M. J., Wang, H., and Kisker, C. (2005). 'Close-fitting sleeves': DNA damage recognition by the UvrABC nuclease system. *Mutat Res* *577*, 92-117.
- Vergani, P., Basso, C., Mense, M., Nairn, A. C., and Gadsby, D. C. (2005a). Control of the CFTR channel's gates. *Biochem Soc Trans* *33*, 1003-1007.
- Vergani, P., Lockless, S. W., Nairn, A. C., and Gadsby, D. C. (2005b). CFTR channel opening by ATP-driven tight dimerization of its nucleotide-binding domains. *Nature* *433*, 876-880.
- Wakeman, T. P., Kim, W. J., Callens, S., Chiu, A., Brown, K. D., and Xu, B. (2004). The ATM-SMC1 pathway is essential for activation of the chromium[VI]-induced S-phase checkpoint. *Mutat Res* *554*, 241-251.
- Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (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*, 945-951.

- Watanabe, Y., and Nurse, P. (1999). Cohesin Rec8 is required for reductional chromosome segregation at meiosis. *Nature* *400*, 461-464.
- Weeks, C. M., and Miller, R. (1999). Optimizing Shake-and-Bake for proteins. *Acta Crystallogr D Biol Crystallogr* *55*, 492-500.
- Weitzer, S., Lehane, C., and Uhlmann, F. (2003). A model for ATP hydrolysis-dependent binding of cohesin to DNA. *Curr Biol* *13*, 1930-1940.
- Yamanaka, K., Ogura, T., Niki, H., and Hiraga, S. (1996). Identification of two new genes, mukE and mukF, involved in chromosome partitioning in *Escherichia coli*. *Mol Gen Genet* *250*, 241-251.
- Yamazoe, M., Onogi, T., Sunako, Y., Niki, H., Yamanaka, K., Ichimura, T., and Hiraga, S. (1999). Complex formation of MukB, MukE and MukF proteins involved in chromosome partitioning in *Escherichia coli*. *Embo J* *18*, 5873-5884.
- Yazdi, P. T., Wang, Y., Zhao, S., Patel, N., Lee, E. Y., and Qin, J. (2002). SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. *Genes Dev* *16*, 571-582.
- Yoshimura, S. H., Hizume, K., Murakami, A., Sutani, T., Takeyasu, K., and Yanagida, M. (2002). Condensin architecture and interaction with DNA: regulatory non-SMC subunits bind to the head of SMC heterodimer. *Curr Biol* *12*, 508-513.
- Young, L., Leonhard, K., Tatsuta, T., Trowsdale, J., and Langer, T. (2001). Role of the ABC transporter Mdl1 in peptide export from mitochondria. *Science* *291*, 2135-2138.
- Zhao, X., and Blobel, G. (2005). A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization. *Proc Natl Acad Sci U S A* *102*, 4777-4782.
- Zimmerman, C., Klein, K. C., Kiser, P. K., Singh, A. R., Firestein, B. L., Riba, S. C., and Lingappa, J. R. (2002). Identification of a host protein essential for assembly of immature HIV-1 capsids. *Nature* *415*, 88-92.

## 9 Abbreviations

Å	Angstrom (1 Å = 10 <sup>-10</sup> m)
ABC ATPase	ATP-binding cassette ATP hydrolase
ADP	adenosine-5p-diphosphate
AMP	adenosine-5p-monophosphate
AP5A	P1,P5-di(adenosine-5')pentaphosphate
ATP	adenosine-5p-triphosphate
<i>B.subtilis</i>	<i>Bacillus subtilis</i>
bp	base pair(s)
<i>C.elegans</i>	<i>Caenorhabditis elegans</i>
cv	column volume
dsDNA	double stranded deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DTT	dithiothreitol
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
FT	Fourier transformation
g	gramm
g	multiples of the earth's gravitational field (centrifugation)
h	hour
IEX	ion exchange chromatography
IMAC	immobilized metal affinity chromatography
kDa	kilo Dalton
keV	kilo electron Volt
kT/e	Boltzman's constant * 300 Kelvin / 1 electron charge
LB	Luria-Bertani
M	molar (mol / l)
MAD	multiple anomalous diffraction
MES	2-(N-morpholino)ethanesulfonic acid
MR	moleculare replacement
min	minute
NBD	nucleotide binding domain
OD	optical density
P <sub>i</sub>	inorganic phosphate
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
<i>Pfu</i>	<i>Pyrococcus furiosus</i>
pI	isoelectric point
s	second
SAD	single anomalous diffraction
ScpA	Segregation and condensation protein A
ScpB	Segregation and condensation protein B
SDS	sodium dodecyl sulfate

---

SMC	structural maintenance of chromosomes
SMCcd	SMC catalytic domain
SnB	Shake'n'Bake
<i>T.maritima</i>	<i>Thermotoga maritima</i>
TMD	trans membrane domain
Tris	Tris(hydroxymethyl)aminomethane
U	unit
v	volume
v/v	volume per volume
w	weight
w/v	weight per volume
wt	wild type

## Appendix

**Table 1: Components of eukaryotic SMC complexes**  
(from (Losada and Hirano, 2005))

	<i>S. cerevisiae</i>	<i>S. pombe</i>	<i>C. elegans</i>	<i>D. melanogaster</i>	<i>A. thaliana</i>	<i>X. laevis</i>	<i>H. sapiens</i>
<b>Cohesin</b>							
SMC1	Smc1	Psm1	SMC-1	DmSMC1	AtSMC1	XSMC1	hSMC1 $\alpha$
SMC3	Smc3	Psm3	SMC-3	DmSMC3	AtSMC3	XSMC3	hSMC3
kleisin	Sec1/Mcd1	Rad21	SCC-1/COH-2 <sup>a</sup>	DmRAD21	SYN2-4 <sup>b</sup>	XRAD21	hSec1/hRAD21
?	Sec3	Psc3	SCC-3	DmSA,CG13916	CAB45374	XSA1, XSA2	hSA1, hSA2
SMC1 $\beta$ (meiosis)	-	-	-	-	-	-	hSMC1 $\beta$
kleisin (meiosis)	Rec8	Rec8	REC-8	C(2)M <sup>c</sup>	SYN1/DJF1	AAH87346	hiRec8
?(meiosis)	-	Rec11	-	-	-	?	hSA3/STAG3
<b>Condensin</b>							
SMC2 (I&II)	Smc2	Cut14	MIX-1	DmSMC2	AtCAP-E1, E2	XCAP-E	hCAP-E/hSMC2
SMC4 (I&II)	Smc4	Cut3	SMC-4	DmSMC4/gluon	AtCAP-C	XCAP-C	hCAP-C/hSMC4
HEAT (IA)	Yes4	Cnd1	-	CG1911	CAB72176	XCAP-D2	hCAP-D2/CNAPI
HEAT (IB)	Yes5/Ycg1	Cnd3	-	CG17054	BAB08309	XCAP-G	hCAP-G
kleisin (IC)	Bm1	Cnd2	-	Barren	AAC25941	XCAP-H	hCAP-H
HEAT (IIA)	-	-	HCP-6	CG31989	At4g15890	XCAP-D3	hCAP-D3
HEAT (IIB)	-	-	F55C5.4	-	At1g64960	XCAP-G2	hCAP-G2
kleisin (IIC)	-	-	C29E4.2	CG14685	At3g16730	XCAP-H2	hCAP-H2
<b>SMC5-6 complex</b>							
SMC5	Smc5	Spt18/Smc5	C27A2.1	CG32438	CAC01791	BAC56936	hSMC5
SMC6	Rhc18	Rad18/Smc6	C23H4.6, F54D5.14	CG5524	MIM	BAC56937	hSMC6
Ubiquitin ligase?	Nse1	Nse1	T23F6.3?	CG11329	At5g21140	MGC68739	hNse1
SUMO ligase	Mms21	Nse2	2G118	CG13732	At3g15150	MGC53049	hNse2
(MAGE) <sup>e</sup>	YDR288W	Nse3	-?	(CG10059) <sup>e</sup>	(At1g34770) <sup>e</sup>	(many) <sup>e</sup>	(>55 paralogs) <sup>e</sup>
?	Qri2	Nse4/Rad62	H21P03.2	CG13142	At1g51130, At3g20760	CA983359	hQri2, hNse4
?	YML023Cf	-	?	?	?	?	?
?	Kre29f	-	?	?	?	?	?

<sup>a</sup> The genome of *C. elegans* encodes two additional SCC-1/REC-8-like proteins, COH-1 and COH-3. COH-1 may have a non-mitotic role in development while depletion of COH-3 shows no apparent phenotype (Mito et al. 2003).

<sup>b</sup> Among the four SCC1/REC8 paralogs present in the genome of *A. thaliana*, only SYN1 has been shown to have a role in cohesion during meiosis. The putative cohesion function of SYN2, SYN3 and SYN4 remains to be determined (Cai et al. 2003).

<sup>c</sup> There is no obvious Rec8 ortholog in *D. melanogaster*. Instead, C(2)M encodes a distant kleisin- $\alpha$  family member that associates with DmSMC3 and has a role in SC formation (Heidmann et al. 2004).

<sup>d</sup> The dosage compensation complex (DCC) is unique to *C. elegans* (Hagstrom and Meyer 2003).

<sup>e</sup> MAGE (melanoma antigen-encoding) is a family of proteins widely expressed in human cancer cells (Barker and Salehi 2002).

<sup>f</sup> No obvious orthologs of YML023C or Kre29 are found in species other than those closely related to *S. cerevisiae*.

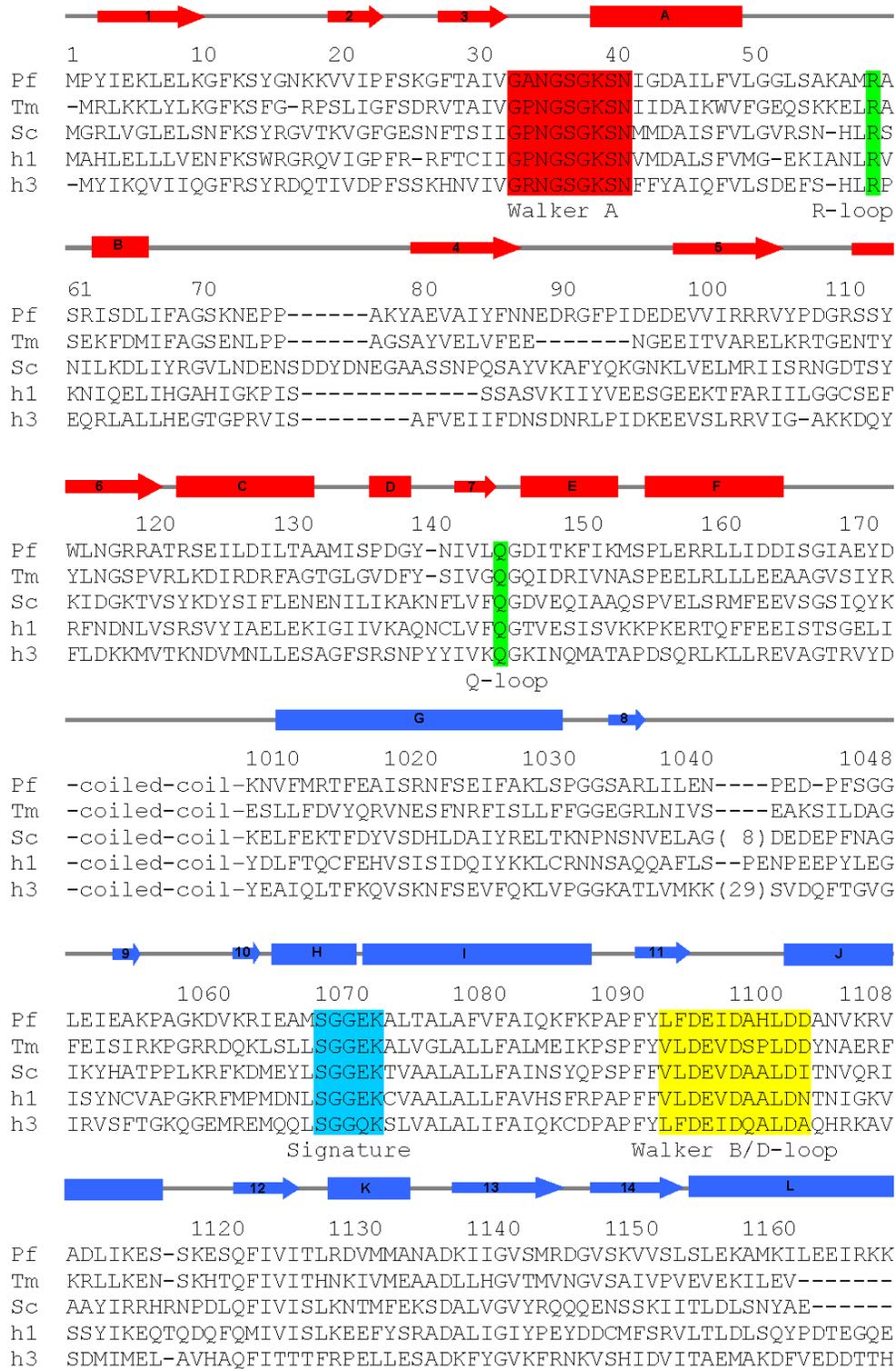


Figure 42: Structure based sequence alignment of SMC proteins. Aligned SMC proteins from *Pyrococcus furiosus* (Pf), *Thermatoga maritima* (Tm), *Saccharomyces cerevisiae* SMC1 (Sc) and human SMC1 and 2 (h1 / h2) with secondary structure elements (denoted as in Fig. 27) and conserved motifs.

**Table 2: Crystallographic table of SMCcd and ATP-SMCcd**

<b>Crystallographic data collection and analysis</b>				
ATP free crystal form: unit cell (C2) (Å): a=102.3 b=56.7 c=78.9 $\alpha=90$ $\beta=123.0$ $\gamma=90$				
Dataset	peak	inflection	high	native
X-ray source	PX (SLS)	PX (SLS)	PX (SLS)	ID29 (ESRF)
Wavelength (Å)	0.97913	0.97931	0.97121	0.91677
Data range (Å)	40.0-2.5	40.0-2.5	40.0-2.5	40.0-2.0
Observations (unique)	57254 (11997)	60450 (11782)	53270 (10847)	73963 (22600)
Completeness (%) (last shell)	89.3 (51.9) <sup>a</sup>	87.0 (43.8) <sup>a</sup>	87.6 (17.1) <sup>a</sup>	88.3 (55.5)
R <sub>sym</sub> <sup>b</sup>	0.061 (0.105)	0.059 (0.142)	0.064 (0.188)	0.070 (0.182)
ATP bound crystal form: unit cell (P3 <sub>1</sub> 21) (Å): a=102.6 b=102.6 c=87.2 $\alpha=90$ $\beta=90$ $\gamma=120$				
Dataset	peak	native		
X-ray source	ID14-4 (ESRF)	PX (SLS)		
Wavelength (Å)	0.97931	0.97547		
Data range (Å)	40-2.8	40.0-2.5		
Observations (unique)	178626 (16351)	98474 (18600)		
Completeness (%)	98.7 (94.8) <sup>a</sup>	99.5 (95.2)		
R <sub>sym</sub> <sup>b</sup>	0.080 (0.429)	0.054 (0.366)		
<b>Refinement</b>	ATP bound		ATP free	
Data set	native		native	
Data range (Å)	20.0-2.5		20.0-2.0	
Reflections F>0 (cross validation)	18583 (907)		22291 (1080)	
non-hydrogen atoms (solvent molecules)	2788 (210)		2748 (205)	
R <sub>cryst</sub> <sup>c</sup> (R <sub>free</sub> <sup>d</sup> )	21.0 (26.9)		20.7 (26.1)	
Rms bond length (Å)/bond angles	0.009 (1.3)		0.019 (2.3)	
% core (disallowed) in Ramachandran plot	90.2 (0.0)		94.7 (0.0)	
<sup>a</sup> Anomalous completeness				
<sup>b</sup> R <sub>sym</sub> is the unweighted R value on I between symmetry mates				
<sup>c</sup> $R_{\text{cryst}} = \sum_{\text{hkl}} \   F_{\text{obs}}(\text{hkl})  -  F_{\text{calc}}(\text{hkl})  \  / \sum_{\text{hkl}}  F_{\text{obs}}(\text{hkl}) $				
<sup>d</sup> R <sub>free</sub> = the cross validation R factor for 5% of reflections against which the model was not refined				

**Table 3: Crystallographic table of ADP-AlF<sub>4</sub><sup>-</sup> and ADP-BeF<sub>3</sub><sup>-</sup> bound SMCcd**

<b>Crystallographic data collection and analysis</b>		
Space group: P3(1)21		
unit cell AlF <sub>4</sub> <sup>-</sup> : a=120.12 Å b=120.12 Å c=86.88 Å α=90° β=90° γ=120°		
unit cell BeF <sub>3</sub> <sup>-</sup> : a=120.552 Å b=120.552 Å c=87.516 Å α=90° β=90° γ=120°		
	AlF <sub>4</sub> <sup>-</sup>	BeF <sub>3</sub> <sup>-</sup>
X-ray source	PX (SLS)	PX (SLS)
Wavelength (Å)	0.9999	0.9999
Data range (Å)	50.0-2.3	50-2.3
Observations (unique)	126551 (21501)	107533 (23477)
Completeness (%) (last shell)	91.1 (74.5)	97.9 (89.3)
R <sub>sym</sub> <sup>a</sup> (last shell)	0.046 (0.313)	0.050 (0.324)
<b>Refinement</b>		
	AlF <sub>4</sub> <sup>-</sup>	BeF <sub>3</sub> <sup>-</sup>
Data range (Å)	20.0-2.3	20.0-2.3
Reflections F>0 (cross validation)	21376 (1090)	22318 (1134)
non-hydrogen atoms (solvent molecules)	2451 (194)	2612 (201)
R <sub>work</sub> <sup>b</sup> (R <sub>free</sub> <sup>c</sup> )	23.0 (27.4)	20.6 (24.1)
Rms bond length (Å)/bond angles	0.0122 / 1.6378	0.0136 / 1.5799
% core (disallowed) in Ramachandran plot	90.7 (0.0)	92.0 (0.0)
<sup>a</sup> R <sub>sym</sub> is the unweighted R value on I between symmetry mates		
<sup>b</sup> R <sub>work</sub> = $\sum_{hkl} \ F_{obs}(hkl) - F_{calc}(hkl)\  / \sum_{hkl} F_{obs}(hkl)$		
<sup>c</sup> R <sub>free</sub> = the cross validation R factor for 5% of reflections against which the model was not refined		

**Table 4: Crystallographic table of AP5A-SMCcd**

<b>Crystallographic data collection and analysis</b>	
Space group: P2(1); unit cell: a=65.0Å b=86.1Å c=70.1Å α=90° β=116.6° γ=90°	
X-ray source	ID14-2 (ESRF)
Wavelength (Å)	0.933
Data range (Å)	50.0-1.6
Observations (unique)	306506 (82978)
I / σ (last shell)	16.3 (3.8)
Completeness (%) (last shell)	95.7 (83.0)
R <sub>sym</sub> <sup>a</sup> (last shell)	0.047 (0.255)
<b>Refinement</b>	
Data range (Å)	20.0-1.6
Reflections F>0 (cross validation)	79746 (4269)
non-hydrogen atoms (solvent molecules)	5324 (727)
R <sub>work</sub> <sup>b</sup> (R <sub>free</sub> <sup>c</sup> )	18.3 (21.4)
Rms bond length (Å)/bond angles	0.198 (1.8)
% core (disallowed) in Ramachandran plot	94.5 (0.0)
<sup>a</sup> R <sub>sym</sub> is the unweighted R value on I between symmetry mates	
<sup>b</sup> R <sub>work</sub> = $\sum_{hkl} \ F_{obs}(hkl) - F_{calc}(hkl)\  / \sum_{hkl} F_{obs}(hkl)$	
<sup>c</sup> R <sub>free</sub> = the cross validation R factor for 5% of reflections against which the model was not refined	

## Acknowledgements

I would like to thank all the people who have contributed to this work!

First of all I am very grateful to Prof. Dr. Karl-Peter Hopfner for giving me the opportunity to work in his lab, for his scientific advice and friendly support.

Many thanks to all current and former members of the “Happy Hopfner” lab for scientific support, pleasant atmosphere and unforgettable social events.

In particular I would like to thank Angelo Manzan for passing on this nice project to me, Alexandra Schele for her great help in this project, Claudio Shah my bachelor student for his enthusiasm during his work and Dr. Harald Dürr and Dr. Katja Lammens for the help in theoretical and practical crystallography.

Furthermore, I thank Annette Karcher, Axel Kirchhofer, Christian Schiller, Derk Bemeleit, Diana Pippig, Dr. Gregor Witte, Katharina Büttner, Katja Schulz, Nora Assenmacher, Dr. Sheng Cui, Song Gao, Sophia Hartung and all rotation, bachelor, master and diploma students.

Thanks to all people at the Gene Center for their help and the nice working atmosphere. Especially I would like to thank Toni for productive synchrotron trips, Karim for stimulating the social life and all the other members of the Cramer group for excellent collaboration in solving structure and other problems. I would like to thank Aaron Alt from the Carell group and Robert Lewis from the Michaelis group for promising collaborations.

I want to thank my wife Katja for going along with me on the hard way to solve structures during both working time and spare time. You are an incredible support for my work and my whole life. I also want to say Hello to my little son Jakob for enriching my life since 7<sup>th</sup> December 2006.

Zum Schluss möchte ich meiner Familie danken für die Hilfe und Unterstützung die sie mir auf meinem Lebensweg entgegen gebracht hat.

My apologies to all others who I have not mentioned by name – thank you for your support!

## Curriculum vitae

### Personal Details

Name	Alfred Armin Lammens
Date of Birth	30 <sup>th</sup> august 1976
Place of Birth	Würzburg
Nationality	German
Marital Status	married
Child	Jakob

### Education

1982-1986	Hammerhütter Grundschule, Siegen
1986-1990	Gymnasium am Rosterberg, Siegen
1990-1996	Johann-Schöner-Gymnasium, Karlstadt
1996	Abitur

### Civil Service

1996-1997	Sozialstation St. Nikolaus, Arnstein
-----------	--------------------------------------

### Academic Training

1997-2002	Studies in Biochemistry Martin-Luther-University Halle-Wittenberg, Halle/Saale
2001-2002	Diploma Thesis “Klonierung, Expression und Reinigung der Hormonbindungsdomäne des humanen Glp-1-Rezeptors” Institute of Biotechnology Martin-Luther-University Halle-Wittenberg, Halle/Saale Supervisors: Prof. Dr. Rainer Rudolph, Dr. Antje Kuntzsch
June 2002	Diplom / Master of Biochemistry
since 10.2002	PhD Thesis in structural biology Gene Center and Institute of Biochemistry Ludwig-Maximilians-University, München Supervisor: Prof. Dr. Karl-Peter Hopfner