Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

X-Ray Structures

of the

Sulfolobus solfataricus SWI2/SNF2 ATPase Core

and its Complex with DNA

Harald Dürr

aus Donauwörth

2005

Erklärung:

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

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig und ohne unerlaubte Hilfsmittel angefertigt.

München, den

_____

Dürr Harald

Dissertation eingereicht am :   08.07.2005

Erstgutachter: Prof. Dr. Karl-Peter Hopfner
Zweitgutachter: Prof. Dr. Patrick Cramer

Mündliche Prüfung am: 20.09.2005

The presented thesis was prepared in the time by November 2001 to June 2005 in the laboratory of Professor Dr. Karl-Peter Hopfner at the gene center of the Ludwig Maximilians University of Munich.

The most beautiful thing we can experience is the mysterious. It is the source of all true art and all science. He to whom this emotion is a stranger, who can no longer pause to wonder and stand rapt in awe, is as good as dead: his eyes are closed.

ALBERT EINSTEIN

Table of Content

Summary

1. Introduction

2. Material and Methods

2.1. Material

2.2 Methods

2.3 X-ray crystallography

3. Results

3.1 Biochemistry of SsoRad54cd

3.2 Structural analysis

Summary

Dynamic remodeling of chromatin or other persistent protein:DNA complexes is essential for genome expression and maintenance. Proteins of the SWI2/SNF2 family catalyze rearrangements of diverse protein:DNA complexes. Although SWI2/SNF2 enzymes exhibit a diverse domain organisation, they share a conserved catalytic ATPase domain that is related to superfamily II helicases through the presence of seven conserved sequence motifs. In contrast to helicases, SWI2/SNF2 enzymes lack helicase activity, but use ATP hydrolysis to translocate on DNA and to generate superhelical torsion into DNA. How these features implicate remodeling function or how ATP hydrolysis is coupled to these rearrangements is poorly understood and suffers from the lack of structural information regarding the catalytic domain of SWI2/SNF2 ATPase

In this PhD thesis I characterized the catalytic domain of *Sulfolobus solfataricus* Rad54 homolog (SsoRad54cd). Like the eukaryotic SWI2/SNF2 ATPases, SsoRad54cd exhibits dsDNA stimulated ATPase activity, lacks helicase activity and has dsDNA translocation and distortion activity. These activities are thereby features of the conserved catalytic ATPase domain itself. Furthermore, the crystal structures of SsoRad54cd in absence and in complex with its dsDNA substrate were determined. The *Sulfolobus solfataricus* Rad54 homolog catalytic domain consists of two RecA-like domains with two helical SWI2/SNF2 specific subdomains, one inserted in each domain. A deep cleft separates the two domains. Fully base paired duplex DNA binds along the domain 1: domain 2 interface in a position, where rearrangements of the two RecA-like domains can directly be translated in DNA

manipulation. The binding mode of DNA to SsoRad54cd is consistent with an enzyme that translocate along the minor groove of DNA.

The structure revealed a remarkable similarity to superfamily II helicases. The related composite ATPase active site as well as the mode of DNA recognition suggests that ATP-driven transport of dsDNA in the active site of SWI2/SNF2 enzymes is mechanistically related to ATP-driven ssDNA in the active site of helicases. Based on structure-function analysis a specific model for SWI2/SNF2 function is suggested that links ATP hydrolysis to dsDNA translocation and DNA distortion.

The represented results have structural implications for the core mechanism of remodeling factors. If SWI2/SNF2 ATPases are anchored to the substrate protein:DNA complex by additional substrate interacting domains or subunits, ATP-driven cycles of translocation could transport DNA towards or away from the substrate or generate torsional stress at the substrate:DNA interface.

Finally, I provide a molecular framework for understanding mutations in Cockayne and X-linked mental retardation syndromes. Mapping of the mutations on the structure of SsoRad54cd reveal that the mutations colocalize in two surface clusters: Cluster I is located adjacent to a hydrophobic surface patch that may provide a macromolecular interaction site. Cluster II is situated in the domain 1 : domain 2 interface near the proposed pivot region and may interfere with ATP driven conformational changes between domain 1 and domain 2.

Introduction

Dynamic remodeling of chromatin or other persistent protein:DNA complexes

Molecular machines that carry out fundamental processes of DNA replication, recombination, repair and transcription need to recognize and bind to their DNA substrates. In eukaryotic cells, however, DNA is packed into nucleosomes and higher order chromatin. A nucleosome, the building block of chromatin, is composed of about ~146 base pairs of DNA that is wrapped in almost two superhelical turns around a core histone octamer composed of two copies of each of the four histones, H2A, H2B, H3 and H4 (Dutnall and Ramakrishnan, 1997; Luger and Richmond, 1998). Neighbouring nucleosomes are spaced by short segments of linker DNA resulting in a nucleosomal array. The N termini of the histones protrude from the compact particle to contact DNA, other histones and nonhistone proteins contributing to the folding of the nucleosomal fiber into complex higher-order structures, collectively called chromatin (Bustin et al., 2005; Chakravarthy et al., 2005; Hayes and Hansen, 2001; Zlatanova et al., 1999). The chromatin architecture has the advantage to reach a high level of compaction and it forms a first fundament for epigenetic control, gene regulation and DNA maintenance (Kimmins and Sassone-Corsi, 2005; Soejima and Wagstaff, 2005). On the other side, the assembly of DNA in nucleosomes affects the accessibility of DNA for regulatory complexes and creates a potent obstacle for other protein:DNA interactions in transcription, replication, DNA repair or recombination. Therefore, a transient perturbation of the repressive chromatin organisation and a more dynamic chromatin structure must be created to facilitate genomic accessibility. Likewise, stalled transcription machineries or other

persistent protein:DNA complexes can interfere with gene expression and genome maintenance (Ehrenhofer-Murray, 2004; Ljungman and Zhang, 1996; van den Boom et al., 2002).

Chromatin Remodeling by covalent histone modification

In recent years, two broad classes of chromatin remodeling activities have received much attention. The two processes are able to generate a local chromatin topology with increased accessibility to DNA. The first class includes proteins or protein complexes that posttranslationally modify chromatin by acetylation, phosphorylation, methylation or other modifications and thereby modulate the contacts between histones and DNA (Cheung et al., 2000; Haushalter and Kadonaga, 2003; Lusser and Kadonaga, 2003; Strahl and Allis, 2000). The functional consequences of these modifications are currently explored with high intensity and seems to have profound effects on higher order chromatin topology (Imhof and Becker, 2001). For instance, acetylation of specific lysine residues in the N-terminal tails of histones is correlated with a more open chromatin structure by changing the nucleosome-nucleosome interactions involved in folding of the nucleosomal fiber (Turner, 2000; Yang and Seto, 2003). In contrast, repressive chromatin structure is commonly characterized by DNA methylation (Paulsen and Ferguson-Smith, 2005). Alternatively, histone modifications provide signals that regulate the activities of other factors that mediate chromatin dynamics (i.e. "histone code"). Since these covalent modifications are reversible, they can act as chromatin-based "on/off" switches that regulate a multitude of DNA-related processes.

Chromatin Remodeling by ATP dependent activities

The second class of chromatin remodeling activities comprises large multiprotein complexes that use energy derived from ATP hydrolysis to modulate histone:DNA interactions in chromatin in a more subtle and transient way (Becker and Horz, 2002; Vignali et al., 2000). Many ATP-dependent chromatin remodeling complexes have been described and were attributed to be involved in a wide range of processes

including transcription, replication, repair, recombination and nucleosome assembly (Haushalter and Kadonaga, 2003; Havas et al., 2001; Varga-Weisz and Becker, 1998). The ATP dependent chromatin remodeling complexes use the energy from ATP hydrolysis to weaken the tight binding of DNA around the histones. The enfeebled histone:DNA interactions in turn facilitate nucleosome repositioning or even displacement of histone octamers. This relocation leads to an accumulation of patches of accessible DNA establishing a "fluid" state of chromatin, but maintains the overall chromatin organisation (Becker and Horz, 2002; Kingston and Narlikar, 1999). This subtle and transient mode of chromatin remodeling demands for disruption of a number of DNA:histone interactions and as a result requires energy. The diverse set of complexes that qualify as molecular machines for ATP dependent chromatin remodeling activities have a common denominator. Although they are compositionally and functionally diverse, they all share a conserved motor subunit that belongs to the SWI2/SNF2 family of ATPases (Eisen et al., 1995). The SWI2/SNF2 family of ATPases can be further classified into several subfamilies according to the presence of domains outside the conserved ATPase region. For example the SNF2 subfamily contains an additional bromodomain and includes e.g. the yeast Swi2/Snf2 subunit of the SWI/SNF complex (yeast mating-type switching/sucrose non-fermenting), the yeast Sth1 subunit of the RSC complex (remodels the structure of chromatin) and the *Drosophila* Brahma or human BRG-1/hBRM subunits of the SWI/SNF-related complexes in *Drosophila* and humans (Cairns et al., 1996; Elfring et al., 1994; Wolffe, 1994). The ISWI (imitation switch) subfamily is represented by for instance the ATPase subunit (ISWI) of the NURF (nucleosome-remodeling factor), CHRAC (chromatin-accessibility complex) or ACF (ATP-utilizing chromatin assembly and remodeling factor) complex or by ISW1 and ISW2 catalytic subunit of chromatin remodeling complexes in yeast (Ito et al., 1997; Poot et al., 2000; Tsukiyama et al., 1995; Varga-Weisz et al., 1997). The ATPase subunits of the ISWI subfamily are characterised by a conserved motif called SANT (Swi3, Ada2, N-CoR, TFIIIB) domain and the lack of the bromodomain (Grune et al., 2003). CHD1, a further SNF2 subfamily, includes for instance CHD1, Mi-2α/CHD3 or Mi-2β/CHD4 as subunit of the NuRD or Mi-2 complexes and has additionally a chromodomain and a DNA-binding motif (CHD for chromodomain, helicase-like motifs and DNA-binding domain) (Delmas et al., 1993; Woodage et al., 1997). Finally, the yeast INO80, the catalytic ATPase subunit of the INO80 complex, is the

most recently described member of the SWI2/SNF2-family of ATPases (Shen et al., 2000). INO80 has very little similarity with other SWI2/SNF2 proteins outside the conserved ATPase domain and has been grouped to an individual subfamily. Whether these diverse domain compositions reflect specific intra-complex interactions, mechanistic divergence or preferences for modified substrates remains unclear. However, all these SWI2/SNF2 ATPases have been found to participate in ATP-dependent chromatin remodeling processes.

In addition, SWI2/SNF2 ATPases are not only involved in the alteration of chromatin structure. In fact, they are implicated in many different „remodeling" like processes in transcription, recombination or DNA repair. There, they are suggested to promote clearing of persistent protein:DNA complexes that may interfere with normal cell progression or act in regulatory pathways that require transient rearrangements of protein:DNA complexes.

Remodeling processes in transcription, recombination and DNA repair

Several proteins have been described that are clearly related to SWI2/SNF2 proteins, although they can not be assigned to one of the three major subfamilies nor are involved in classical chromatin remodeling processes. For example, the yeast transcription regulator Mot1 modulates transcription by catalysing the ATP-dependent dissociation of TATA box binding protein from promotors (Auble et al., 1997). The human CSB protein or yeast Rad26 displace stalled RNA polymerase II from DNA lesions in transcription-coupled repair (TCR) to make the DNA lesion accessible to the DNA repair machinery (Svejstrup, 2003). The nucleotide excision repair protein Rad16 is specifically involved in the removal of lesions from transcriptionally silent regions of the genome. Finally, Rad54 is a key component of the homologous recombination machinery. Rad54 interacts with the Rad51 nucleoprotein filament and stimulates homologous DNA pairing within the context of chromatin, maybe by protein clearing at the homology search step of genetic recombination (Alexeev et al., 2003; Alexiadis and Kadonaga, 2002; Jaskelioff et al., 2003)

Furthermore, SWI2/SNF2 enzymes are not restricted to eukaryotes, but are also found in archaea and bacteria. For instance the prokaryotic transcription activator HepA, which remodels the posttranscriptional complex and stimulates RNA-polymerase recycling (Sukhodolets et al., 2001) or the transcription repair coupling factor (TCRF), the prokaryotic counterpart of CSB/Rad26 (Park et al., 2002). Several other prokaryotic and archaeal proteins have been linked to the SWI2/SNF2 family based on the sequence homology of their ATPase domain including *Sulfolobus solfataricus* Rad54 homolog.
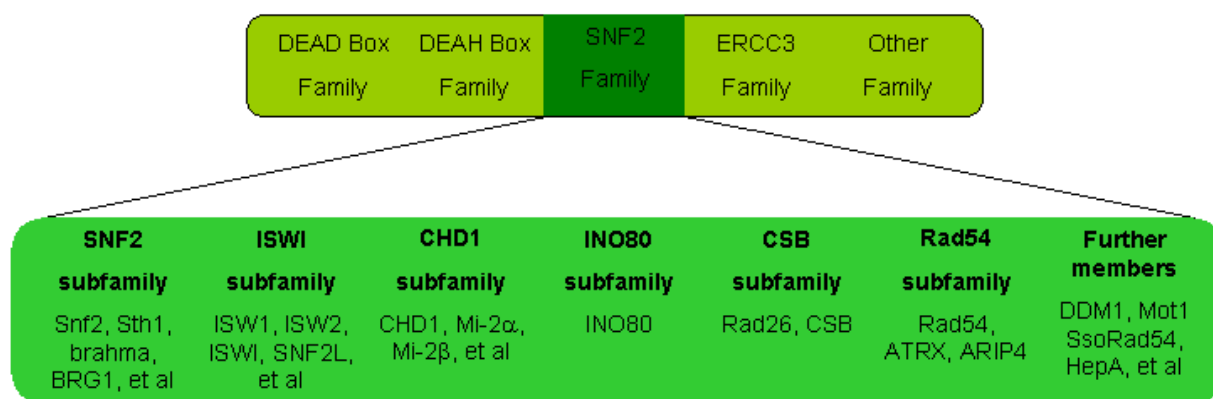
Figure 1: The SWI2/SNF2 family of ATPases can be further divided into distinct subfamilies. The diagram includes the classical proteins involved in chromatin remodeling (SNF2, ISWI and CHD1 subfamilies) and additional proteins that belongs to the SWI2/SNF2 family based on their conservation of the ATPase domain (adapted from (Lusser and Kadonaga, 2003).

SWI2/SNF2 ATPases are involved in all major protein:DNA remodeling reactions, like transcription control, hence affecting cell cycle or cell differentiation (Muller and Leutz, 2001), replication and DNA repair (Ehrenhofer-Murray, 2004; Svejstrup, 2003). Correspondingly, mutations in SWI2/SNF2 enzymes have serious consequences to the organism. For instance, mutations of a SWI2/SNF2 ATPase homolog in *Caenorhabditis elegans* abolishes asymmetric division of certain T cells (Sawa et al., 2000), ISWI mutations in *Drosophila* affect both cell viability and gene expression during development (Deuring et al., 2000) and mutations of BRG1 is

associated with cell transformations (Muchardt and Yaniv, 2001). Significantly, human diseases were identified that are associated to mutations of SWI2/SNF2 ATPases. For instance, mutations in the CSB protein result in the human disease Cockayne Syndrome B, which is characterized by increased UV sensitivity and neurodevelopmental abnormalities (Christiansen et al., 2003; Osterod et al., 2002). The alpha thalassemia / mental retardation syndrome (ATR-X syndrome), which is characterized by severe mental retardation, microcephaly, distinct facial dysmorphism, and genital abnormalities, as well as a wide spectrum of other pathological features, is caused by mutations of the ATRX gene (Borgione et al., 2003; Villard and Fontes, 2002). Finally, Schimke immuno-osseous dysplasia is caused by mutations in SMARCAL1 (SWI2/SNF2 related matrix associated, actin dependent regulator of chromatin, subfamily a-like 1) and displays phenotypically growth retardation, defective cellular immunity and progressive nephropathy (Boerkoel et al., 2002).

Mechanisms for protein:DNA remodeling

Although all SWI2/SNF2 enzymes display compositionally and functionally their individual characteristics, they all are molecular motors that are suggested to function by catalysing rearrangements in diverse protein:DNA contacts in an ATP-dependent manner. The ability of ATP dependent remodeling factors to modulate or disrupt protein:DNA interactions has been intensively studied using a variety of biochemical assays. Different types of remodeling reactions, including generation of nuclease accessibility to nucleosomal DNA, histone octamer relocation or alteration of DNA supercoiling have been analysed (Hamiche et al., 1999; Langst et al., 1999). The probably best studied examples are the classical chromatin remodeling factors, but increasing data demonstrate that also other SWI2/SNF2 enzymes, like CSB or Rad54, function in these classical "nucleosome mobility assays" (Alexeev et al., 2003; Beerens et al., 2005). Although some differences were noted, a common outcome of these experiments is that all tested remodeling factors transiently catalyse the disruption of histone:DNA interactions resulting in the repositioning of histone octamers on DNA in cis ("sliding") (Brehm et al., 2000; Jaskelioff et al., 2000; Langst et al., 1999) or even in displacement and transfer of the histone octamer

(Lorch et al., 1999; Narlikar et al., 2001). These findings support the idea that SWI2/SNF2 enzymes act on protein:DNA interfaces in a common fashion. In theory, nucleosome repositioning could provide a means by which appropriate DNA is exposed to interaction with gene regulatory elements. For chromatin remodeling factors the target protein:DNA complex is likely to be nucleosomes, but in the case of CSB/Rad26, HepA or TCRF, the primary target could be some other protein:DNA complexes, for instance stalled polymerase. The action of these proteins may allow the polymerase:DNA interface to be remodeled, thereby providing access for the DNA repair machinery. Similarly, Mot1 is directed towards the modulation of TBP binding to the TATA box and plays a role in regulation of gene expression.

The mechanism by which remodeling factors rearrange or disrupt protein:DNA interactions, remains controversial. A mechanism that is associated with a transient "unpeeling" of DNA from the surface of the histone octamer or from other persistent protein:DNA interfaces is proposed for energetical reasons instead of a simultaneous breakage of all protein:DNA contacts. Two major mechanistic models were postulated for the chromatin remodeling activities, but may also be applicable for other SWI2/SNF2 enzymes: "bulge diffusion" and "twist defect diffusion" (for reviews see (Becker and Horz, 2002; Owen-Hughes, 2003). The first model involves the formation of a loop or bulge on the surface of a nucleosome that may be transmitted around the surface of the nucleosome. Transient dissociation of DNA, followed by reformation of histone:DNA interactions at a different site would result in an initial formation of a loop or bulge that can then propagate around the nucleosome. The second model (twist defect diffusion) assumes that DNA is distorted through localized alterations to twist. According to this model, histone:DNA interactions may be disrupted by (un)twisting of DNA in a small domain, presumed to be topologically constrained by the remodeling enzyme itself and the nucleosome substrate. The rotation of DNA around its long axis may enable DNA to propagate incrementally over the surface of nucleosomes in a "cork-screw" like action, leading to a replacement of "canonical" histone DNA interactions by equivalent ones involving neighbouring base pairs. Both models are reasonable for the relocation of nucleosomes or other persistent protein:DNA complexes.

SWI2/SNF2 ATPases on the job

Although the two proposed mechanisms give adequate concepts on the macroscopic level, the question of how SWI2/SNF2 ATPases actually catalyze the remodeling of protein:DNA interfaces is poorly understood. Although phenomenological differences have been interpreted to indicate fundamentally different mechanisms, they alternatively result from variations of one basic remodeling reaction (Langst and Becker, 2004).

The SWI2/SNF2 ATPases like ISWI, Sth1, CSB/Rad26 or Rad54 have remodeling activity themselves, so the ATPase subunits are suggested to be the catalytic core of the remodeling complexes (Alexeev et al., 2003; Beerens et al., 2005; Corona et al., 1999). Since the homology of SWI2/SNF2 like proteins is mainly restricted to the catalytic ATPase domain, this conservation indicates that a principal ATP-driven event is conserved among the SWI2/SNF2 family that may account for the remodeling function. Biochemical characterisation of SWI2/SNF2 enzymes indeed revealed a common set of shared activities, supporting the concept that SWI2/SNF2 enzymes act in a common fashion on the biochemical level. This underlying fundamental principle might be qualitatively or quantitatively regulated or targeted to specialized functions by the non-ATPase domains or subunits. Interestingly, the position of the conserved ATPase domain within the SWI2/SNF2 enzymes is not conserved as well and ranges from the N-terminus, the central part to the C-terminus (Fig. 2A). This indicates that the SWI2/SNF2 ATPase domain resembles a motor module and the different location of the ATPase domain may be important for the translations of the SWI2/SNF2 ATPase generated force, which could explain the variations of the remodeling activities.

Initial sequence comparisons suggested that SWI2/SNF2 ATPases fall within the large superfamily II of helicase-related proteins (Eisen et al., 1995; Gorbalenya et al., 1988). In particular, SWI2/SNF2 enzymes share seven sequence motifs that are closely related to those of superfamily II helicases and cluster in a region averaging ≈ 400 amino acids in length (Fig. 2). These "helicase motifs" are implicated in ATP binding, ATP-hydrolysis and DNA interaction.
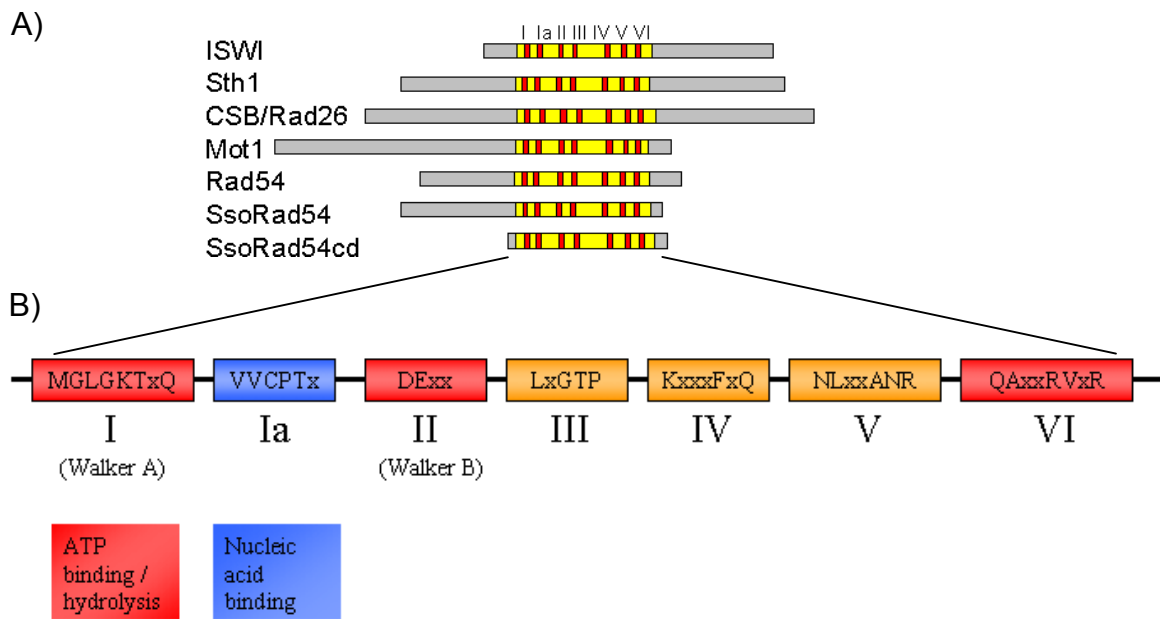
Figure 2:   A) Schematic drawing of a selection of SWI2/SNF2 enzymes. The homology between SWI2/SNF2 enzymes is mainly restricted to the ATPase domain (yellow) that contains the conserved helicase motifs (red boxes). The location of the ATPase domain within the enzymes is not conserved. B) Diagram of the seven classical conserved helicase motifs defined by Gorbalenya and Koonin (annotated I to VI). The consensus sequence is adapted to SWI2/SNF2 related proteins.

As mutations within these motifs have been shown to ablate the function of the remodeling factors, they are likely to play an important role in their action (Richmond and Peterson, 1996). Although the ATPase activity is essential for their function both in vitro as well as in vivo, all SWI2/SNF2 ATPases exhibit a low or undetectable intrinsic ATPase activity. However, the ATPase activity is stimulated by DNA. The optimal DNA substrate for ATPase stimulation is variable and depends on the enzyme. For instance, the yeast SWI/SNF complex and the most related yeast and human complexes are activated by single-stranded DNA, double-stranded DNA as well as nuleosomal DNA (Cairns et al., 1996; Cote et al., 1994). In contrast, ISWI is maximally stimulated by an assembly of DNA with histones, indicating that ISWI recognizes a specific structural feature of nucleosomes (Corona et al., 1999). The yeast Rad54 is only stimulated by double-stranded DNA or chromatin, but not by single-stranded DNA (Swagemakers et al., 1998).

Since helicases are molecular "motor" enzymes that use the energy of ATP hydrolysis to unwind transiently stable duplex DNA, it was initially postulated that disruption of chromatin structure involves unwinding of DNA. But surprisingly, despite of their apparent relationship to helicases, all SWI2/SNF2 enzymes studied so far lack detectable helicase activity using classical strand displacement assays for DNA helicase activity (Cote et al., 1994). Since SWI2/SNF2 ATPases do not appear to function as helicases, an alternative model suggests that these proteins are ATP-dependent DNA-tracking enzymes. X-ray analyses of helicases revealed that helicases are organized in a modular structure consisting of an ATP hydrolysing domain, which encompass the canonical helicase sequence motifs, and additional dsDNA destabilisation or ssDNA binding domains (Kim et al., 1998; Subramanya et al., 1996; Velankar et al., 1999). Mechanistically it has been proposed that helicases translocate on the downstream ssDNA product and destabilize the upstream DNA duplex through their duplex nucleic acid destabilization domains (Soultanas et al., 2000). Moreover, the crystal structures of helicases indicate that DNA unwinding activity can be uncoupled from the translocase activity. Therefore, it has been proposed that translocation may be a common feature of all members of the superfamily II helicase-like proteins (Soultanas et al., 2000; Tuteja and Tuteja, 2004; Whitehouse et al., 2003). Two approaches gave hints for the ability of SWI2/SNF2 enzymes to translocate along DNA. Firstly, the ATPase activity of Sth1 (the catalytic subunit of RSC complex), ISWI, Swi2/Snf2 and Rad54 depends on the length of the DNA fragment used for stimulation (Jaskelioff et al., 2003; Saha et al., 2002; Whitehouse et al., 2003). These observations were interpreted that the ATP hydrolysis rate is increased while they can progress along DNA. Secondly, it has been reported that Sth1, ISWI, Mot1 and Rad54 have the ability to displace a triplex forming oligonucleotide, which forms a triple-helical structure via Hoogsteen base-pairing with its target sequence (Auble and Steggerda, 1999; Jaskelioff et al., 2003; Ristic et al., 2001; Saha et al., 2002). This displacement was previously used to monitor the motion of the restriction enzyme EcoR124I (Firman and Szczelkun, 2000). These studies provide accumulating evidence that ATP-dependent translocation along DNA is a basic mechanistic feature of SWI2/SNF2 enzymes. Yet, this translocation seems not to be highly processive, since SWI2/SNF2 enzymes can readily be removed from template DNA by an excess of competitor DNA (Whitehouse et al., 2003). However, translocation along the DNA backbone is also

associated with rotation of DNA. Presumed that the rotation induced torsion is topologically constrained by the remodeling enzyme itself and the remodeled substrate, rotation of DNA by the translocating enzyme would affect the topology of DNA. For instance if translocation occurs within a constrained loop, torsion would be expected to accumulate. Electron microscopy studies with the SWI/SNF complex revealed a tendency for loop formation with DNA (Bazett-Jones et al., 1999; Kassabov et al., 2003). Further support comes from cruciform extrusion experiments, an approach that can be used to monitor distortion of DNA (Havas et al., 2000). These experiments show that SWI2/SNF2 enzymes are capable of distorting DNA by introducing negative superhelical torsion. DNA translocation and associated DNA rotation could provide means by which SWI2/SNF2 enzymes could alter both twist and writhe of DNA at protein:DNA interface and by this way may disrupt protein:DNA contacts. Furthermore, these activities may be the driving force for the "twist defect diffusion" or "bulge diffusion" mechanisms.

Sulfolobus solfataricus *Rad54 homolog*

SWI2/SNF2 like proteins have been found in all kingdoms of life. In prokaryotes, however, the homology is restricted essentially to the helicase motifs comprising ATPase domain. Sequence analysis of *Sulfolobus solfataricus* revealed the presence of a gene that shows homology to eukaryotic SWI2/SNF2 ATPases. This protein from *Sulfolobus solfataricus* was assigned to be homologous to the yeast Rad54 protein, since the ATPase domain is located at the C-terminus, similarly to Rad54. It is not known whether this protein is involved in recombination like Rad54 protein, in transcription like Snf2 or Mot1 or in other remodeling processes. *Sulfolobus solfataricus* belongs to the subdomain of Crenarchaeota that in contrast to Euryachaeota does not possess true eukaryal-like histones. Instead, *Sulfolobus solfataricus* exhibits small architectural DNA binding proteins, named Sul7d and Alba. Thus, *Sulfolobus solfataricus* Rad54 homolog could also be responsible for remodeling like activity of Sul7d/Alba:DNA complexes.

Structure determination by X-ray crystallography

Over the last years, substantial progresses in the understanding of remodeling processes have been made. But despite the apparently widespread utilization of the conserved SWI2/SNF2 ATPase domain, there is currently little mechanistic understanding of how ATP hydrolysis is coupled to DNA translocation and distortion by SWI2/SNF2 enzymes. This circumstance suffers form the lack of structural information regarding the SWI2/SNF2 ATPase motor subunit.

Since proteins fold into specific three-dimensional conformations that are intrinsically related to their functions, the knowledge of the atomic structure of macromolecules is essential for a fundamental understanding of their function and a prerequisite for structure based functional studies. The ability to determine the precise molecular structure of proteins or nucleic acids has revolutionized the study of their functions in many fields of biology. X-ray crystallography and NMR spectroscopy are currently the only techniques capable of determining the structures of biological macromolecules at atomic resolution. The structure determination by NMR, an indirect spectroscopic method, is limited to small molecules (< 30 kDa), but provide in addition to structural insights information about intramolecular dynamics or reaction kinetics. In contrast, X-ray crystallography has no size limitation for molecules or complexes, but displays only a "snapshot" of the macromolecules and the high accuracy of crystallographic structures has to be paid with high quality crystals. X-ray crystallography exploits the fact that X-rays are diffracted by the periodic repetition of the electrons in the crystal that can be described by a Fourier transformation. In return, Fourier transformation can be used to calculate the electron density from measured diffraction data.

However, growth of high quality crystals forms the basic prerequisite for X-ray crystallography. This requires large amounts of homogeneous and pure protein (> 95%), which is concentrated to a supersaturated state to obtain crystals. SWI2/SNF2 enzymes are often large multidomain proteins that can interact with one or more partner proteins. These features often result in an inherently flexible molecule when detached from their binding partners that hampers crystallisation. To increase the probability to get crystals, a commonly used approach is to remove

flexible regions that might hamper crystal packing. Archaeal proteins share many sequence and functional features with eukaryotic proteins, but are often smaller and more robust, and thus often serve as excellent model systems for complex processes and structural investigations. Therefore, we decided to use the conserved catalytic domain of the *Sulfolobus solfataricus* Rad54 homolog for structural analysis of the conserved catalytic ATPase domain of SWI2/SNF2 enzymes.

Objectives

The aim of this PhD thesis was to reveal structural insights into the ATPase subunit of SWI2/SNF2 enzymes by X-ray crystallography. Furthermore, for a full understanding of how SWI2/SNF2 enzymes modulate their protein:DNA substrates, it is important to examine the structural interplay between the enzyme and its DNA substrate. Therefore, the PhD thesis was also addressed to structural analysis of the interaction between the SWI2/SNF2 ATPase domain with its DNA substrate. Finally, structure-function analysis may reveal important residues for SWI2/SNF2 function and reveal insights into the mechanisms for DNA dependent ATP hydrolysis, DNA translocation and distortion.

Sulfolobus solfataricus Rad54 hereby served as an example of a SWI2/SNF2 ATPase. I characterized the biochemical properties of the catalytic domain of *Sulfolobus solfataricus* Rad54 (SsoRad54cd) and showed that it displays like eukaryotic remodeling factors ATP dependent DNA translocation and distortion activity. Using X-ray crystallography we determined the crystal structures of SsoRad54cd in the absence of DNA and in complex with its double-stranded DNA substrate. Our findings in conjunction with structure-function analysis provide first structural insights into the core mechanism and DNA interaction of the conserved SWI2/SNF2 ATPase and suggest a model for SWI2/SNF2 ATPase function. Furthermore, these data provide a first structural framework for any sequence and structure unspecific dsDNA translocase in complex with its dsDNA substrate.

Material and Methods

Materials

Chemicals, enzymes, radioactive material and chromatographic material

Unless otherwise stated, all common chemicals were ordered by Merck (Darmstadt), Roth (Karlsruhe) or Sigma (Deisenhofen). The enzymes were obtained from Invitrogen (Karlsruhe), MBI Fermentas (St. Leon-Rot) or New England Biolabs (Frankfurt). All chromatographic material and columns as well as radioactive material (γ-^{32}P-ATP) were purchased from Amersham Bioscience (Freiburg). Crystallisation screens, crystallisation grade reagents and crystallisation tools were obtained from Hampton Research (Laguna Niguel, USA), Nextal (Montreal, Canada) or Jena Bioscience (Jena). The oligonucleotides for cloning were ordered from Thermo Electron Corporation (Ulm) and the oligonucleotides for crystallisation were synthesized by the Genecenter in-house facility (Dr. Arnold). All oligonucleotides used were HPLC purified.

Methods

Standard molecular biology procedures like polymerase chain reaction (PCR), restriction enzyme digestion, ligation of DNA fragments, preparation of competent *Escherichia coli* cells, transformation, amplification of plasmid DNA and analysis by

agarose gel were performed according to standard protocols (Ausubel, 1999; Sambrook and Rusell, 2001). In addition, plasmid DNA was prepared with plasmid purification kit (Qiagen, Hilden) and isolation and purification of DNA fragments were performed using the gel extraction kit (Qiagen, Hilden) according to the manufacture´s instructions. Further employed standard methods for working with and analysis of proteins like SDS-PAGE, determination of protein concentration, protein concentrating, limited proteolysis and so on were used as described in Sambrook and Russell 2001.

Cloning and expression of *Sulfolobus solfataricus* Rad54 and variants

The gene encoding the N-terminal (ORF SSO1653) and C-terminal (ORF SSO1655) fragment of the *Sulfolobus solfataricus* Rad54 homologous protein were amplified by PCR from genomic DNA using N-term_for and N-term_rev or C-term_for and C-term_rev, respectively, as primers (table 1). The two fragments were combined by overlapping PCR using the N- and C-terminal PCR products as DNA templates and the flanking oligonucleotides as primers (N-term_for and C-term_rev). The region encoding amino acid 790-809 (aa 790–802 of the N-terminal fragment and aa 1-7 of the C-terminal fragment) were omitted, since these amino acids are not found in homologous proteins and the two fragments are likely interrupted by the ORF SSO1654 encoded transposase. Using the "full-length" gene as DNA template, N-terminal truncated constructs ((Δ1-284) and (Δ1-429)) were generated by PCR (table 1) and cloned into pET28b expression vector (Novagen, Darmstadt) at NheI and NotI as restriction sites.

For site-directed single amino acid mutagenesis, specially designed complementary oligonucleotides encoding the desired mutation were used to generate two DNA fragments with overlapping ends. The corresponding primers are shown in table 1. In the first amplification reactions, the N- and C-terminal overlapping PCR products are generated, for instance by using the primer combination aa430_for and "site-directed mutagenesis primer"_rev for the N-terminal part and the "site-directed mutagenesis primer"_for with the C-term_rev for the C-terminal part, respectively. These two PCR fragments have an overlapping region with a melting temperature of 60-75°C. The

subsequent overlap PCR (using the flanking primer aa430_for and C-term_rev) resulted in the gene with an incorporated nucleotide exchange. The overlap PCR was started with an annealing temperature of 48°C for the first 5 cycles to allow effective annealing of the two template fragments ($\approx$ 100 ng each) and subsequently increased to 55°C for the next 25 cycles to enhance specificity. The correct insertion of the mutations was confirmed by DNA sequencing.

Some notes on primer design: The oligonucleotides for cloning consist of a complementary part with a melting temperature (T_M) between 68-72°C, an attached endonuclease restriction site and four A overhangs. The mutagenic oligonucleotide primers comprise a complementary part with T_M of about 30°C before and about 60°C after the desired mutation triplet codon (5´-3´direction). The melting temperatures were calculated following $T_m = 22 + 1{,}44 \cdot (\sum (A+T) + 2 \cdot \sum (G+C))$. Additionally, the primers were made as short as possible to minimize the likelihood of secondary structure formation and should have a minimum GC content of 40 % with G or C as terminal base.

Oligonucleotide	Sequence (5´-3´direction)
Cloning	
N-term_for	AAAAGCTAGCATGTGGATAGTCCATGGCGTATG
N-term_rev	GAAATTATATCATCTCTCTCTTTTTTAGATAGTTCTCC
C-term_for	AAAAAAGAGAGAGATGATATAATTTCAAAGTTTCAGAAC
C-term_rev	AAAAGCGGCCGCTTAATATCCACCTACAGAGAGTTCT
aa285_for	AAAAGCTAGCAAACTGAATGGATTGTTTCTTGGTGAAG
aa430_for	AAAAGCTAGCAAATCCTTTCAATTATTGGAGCCC
aa663_for	AAAAAAACATATGGATCTGCCGGATAAAATAGAGACAAATG
aa657_rev	AAAGCGGCCGCTTATAAATCGTTGATTATTGCCTTAT

Site directed mutagenesis

E563Q_for	ATATATTGTAATAGATCAGGCACAGAATATTAAAAATCCC
E563Q_rev	TATTCTGTGCCTGATCTATTACAATATATTTCCATTCAAC
N569I_for	GCACAGAATATTAAAATCCCCCAAACTAAAATTTTTAAAGCTGTC
N569I_rev	GACAGCTTTAAAAATTTTAGTTTGGGGGATTTTAATATTCTGTGC

Oligonucleotide	Sequence (5´-3´direction)
(continue)	
R586W_for	GTCTAAATACTGGATAGCTCTTACGGGGACG
R586W_rev	CGTAAGAGCTATCCAGTATTTAGACTTTAACTCTTTG
K700E_for	GTTACAGGAATAAAGAGAGAGGGTATGATTTTATCTACTTTGCTC
K700E_rev	GATAAAATCATACCCTCTCTCTTTATTCCTGTAACAGAGTC
K711E_for	CTTTGCTCAAATTAGAACAAATCGTGGATCATCCTG
K711E_rev	GATCCACGATTTGTTCTAATTTGAGCAAAGTAGATAAAATC
H716A_for	CAAATCGTGGATGCGCCTGCACTACTGAAAGGAG
H716_rev	GTAGTGCAGGCGCATCCACGATTTGTTTTAATTTG
G722Q_for	GCACTACTGAAACAGGGCGAACAGTCTGTACGAAG
G722Q_rev	GACTGTTCGCCCTGTTTCAGTAGTGCAGGATGATCC
Q755A_for	GCTATTTTCACAGCATTTGTAGATATGGGGAAAATAATAAG
Q755A_rev	CCATATCTACAAATGCTGTGAAAATAGCTATTTTATCTCC
S783_for	TATATGGAGAACTAGTGAAAAAAGAGAGAGATGATATAATTTC
S783_rev	CTCTCTCTTTTTTTCACTAGTTCTCCATATAAGAAAGGAAC
R788E_for	CTATCTAAAAAGAGGAAGATGATATAATTTCAAAGTTTCAG
R788E_rev	GAAATTATATCATCTTCCTCTTTTTTAGATAGTTCTCCATAT
K808E_for	GTTCTCTCCGTTGAGGCAGGAGGTTTCGGAATAAATC
K808E_rev	CCTCCTGCCTCAACGGAGAGAACTATAAATTTCAC
R840E_for	CAAGCTACAGATGAAGTTTACAGAATAGGGCAAACTAG
R840E_rev	CTATTCTGTAAACTTCATCTGTAGCTTGATCCTCGAC
R843E_for	GATAGGGTTTACGAAATAGGGCAAACTAGGAACGTC
R843E_rev	GTTTGCCCTATTTCGTAAACCCTATCTGTAGCTTGATC
K872E-for	CAACTATTAGCATTTGAAAGATCGTTATTTAAGGATATAATATCTT
K872_rev	AAGATATTATATCCTTAAATAACGATCTTTCAAATGCTAATAGTTG
V850G_for	GGCAAACTAGGAACGGCATTGTCCATAAACTGATAAGCGTAG
V850C_rev	CTACGCTTATCAGTTTATGGACAATGCCGTTCCTAGTTTGCC

Table 1: List of oligonucleotides used for cloning and site-directed mutagenesis. Underscored sequences correspond to the attached restriction sites, blue coloured triplet codons highlight the desired mutation.

The proteins were synthesized in *E.coli* BL21 Rosetta (DE3) cells (Novagen). For this purpose the cells were grown in LB media to OD_{600} = 0.4-0.6 and cooled down to 18°C. The expression was induced by addition of 0.05 mM IPTG and carried out overnight. The cells were harvested by centrifugation, resuspended in 50 mM K_2HPO_4/KH_2PO_4 [pH=7.2], 300 mM KCl, 1 mM DTT and disrupted by sonication. Cell debris and insoluble proteins were removed by centrifugation. The proteins were purified by an initial heat step (10 min 65°C) to denature endogenous *E.coli* proteins and cleared by centrifugation (5 min, 20000 g). The supernatant was further purified by immobilized Ni affinity utilizing a N-terminal 6x His-tag. After extensive washing with equilibration buffer containing 50 mM K_2HPO_4/KH_2PO_4 [pH=7.2], 300 mM KCl, 10 mM imidazole and 1 mM β-mercaptoethanol, the protein was eluted in one step with buffer supplemented with 250 mM imidazole. The elution fraction was dialysed against low salt buffer (20 mM Tris/HCl [pH=7.3], 80 mM NaCl, 1 mM EDTA) and loaded on an equilibrated Resource S ion exchange column (Amersham Bioscience, Freiburg). Elution from the cation exchange chromatography column was performed by a 20 column volume gradient against 20 mM Tris/HCl [pH=7.3], 1 M NaCl and 1 mM EDTA using the Äkta system (Amersham Bioscience, Freiburg). The final purification step was done by Sephadex S200 size exclusion chromatography in 20 mM Tris/HCl [pH=7.5], 150 mM NaCl, 1 mM EDTA and 1 mM DTT.

Expression of Selenomethionine substituted protein

The selenomethionine labeling of proteins were performed using a methionine auxotrophic *E. coli* strain. The cells were grown to an OD_{600} = 0.5 – 0.6 in LeMaster medium (see appendix) supplemented with selenomethionine, cooled down to 18°C and induced with 0.05 mM IPTG. The expression was carried out overnight, cells harvested by centrifugation and treated and purified as described above, but special caution was exercised at all steps to keep a reductive environment by addition of 1 mM DTT to all buffers.

Functional assays

ATPase assay

The DNA-stimulated ATPase assays were performed in reaction mixtures (20 µl) containing 20 mM K_2HPO_4/KH_2PO_4 [pH=5.5], 50 mM NaCl, 5 mM $MgCl_2$, 100µg/ml bovine serum albumin (BSA), 1 mM DTT, 0.1 mM unlabeled ATP, 20 nM γ-^{32}P-ATP (3000 Ci/mmol, Amersham Bioscience) and 1 µM ds/ssDNA. The reaction was kept on ice, started by addition of typically 0.1 µM protein and incubated at different temperatures (30°C to 70°C) for 15 minutes. The reactions were stopped by addition of 1 µl of 0.5 M EDTA to remove magnesium ions that are required for ATP hydrolysis. Aliquots of 1 µl of the reaction mixture were spotted onto a thin layer chromatography polyethyleneimine-cellulose plate (Merck, Darmstadt), air-dried and developed in 0.5 M LiCl / 1.0 M formic acid. The liberated radioactive γ-phosphate almost runs with the front of the developing solution. After the front of the developing solution has reached ¾ of the plate height, the TLC plate was removed form the container and air dried. The products were visualized by Phosphorimager (Amersham Bioscience) and analysed using Image Quant software. The upper spot corresponds to liberated γ-^{32}P and the lower visible spot to non-hydrolysed ATP. The percentage of hydrolysed ATP was calculated according to following equation:

$$\% \text{ hydrolysed ATP} = \frac{\text{amount of liberated phosphate}}{\text{amount of unhydrolysed ATP} + \text{liberated phosphate}}$$

DNA binding – DOT BLOT double filter binding assay

The DNA binding affinity was analysed using a double-filter nitrocellulose/nylon binding assay (Manzan et al., 2004), a powerful technique commonly used to study protein-nucleic acid interactions. The standard reaction buffer for the DNA binding studies contained 20 mM MES [pH=5.5], 50 mM NaCl, 5 mM $MgCl_2$, 100µg/ml bovine serum albumin (BSA), 1 mM DTT, 25 nM unlabeled DNA and $\approx$ 2 nM of ^{32}P labeled DNA. A 50 mer oligonucleotide was 5′ ^{32}P labeled by T4 Polynucleotide kinase (MBI Fermentas) according to manufacture´s instructions and used to generate double-

stranded, partial duplex or holiday junction DNA substrates (table 2). Typically a concentration series ranging from 5 µM to 0 µM protein were incubated with the DNA substrate for 20 minutes at 25°C and then filtered through a double layer of one nitrocellulose filter (PROTRAN BA85, Schleicher&Schuell) and one nylon filter (Roti®-Nylon plus, Roth; nitrocellulose on top of nylon membrane) in a 96 well dot blot apparatus (Schleicher&Schuell). Proteins and protein:DNA complexes bind to the nitrocellulose filter. The remaining free DNA binds to the nylon filter allowing complete recovery of labeled DNA. For quantification the radioactivity was recorded using STORM Phosphorimager (Amersham) and the ratio of protein-bound to unbound DNA was determined using Image Quant software.

DNA	oligonucleotides
dsDNA	5´-GCTCTCTCTTCGGAAAGCATCTAGCATCCTGTCAGCTGCATGGAACG-3´
	5´-CGTTCCATGCAGCTGACAGGATGCTAGATGCTTTCCGAAGAGAGAGC-3´
ssDNA	5´-GCTCTCTCTTCGGAAAGCATCTAGCATCCTGTCAGCTGCATGGAACG-3´
holiday	5´-GGCGACGTGATCACCAGATGATGCTAGATGCTTTCCGAAGACAGACC-3´
	5´-GCTCTCTCTTCGGAAAGCATCTAGCATCCTGTCAGCTGCATGGAACG-3´
	5´-CGTTCCATGCAGCTGACAGGATGCTAGTCAAGGCGAACTGCTAACGG-3´
	5´-CCGTTAGCAGTTCGCCTTGACTAGCATCATCTGGTGATCACGTCGCC-3´
Y-substrate	5´-GCTCTCTCTTCGGAAAGCATCTAGCATCCTGTCAGCTGCATGGAACG-3´
	5´-CGTTCCATGCAGCTGACAGGATGCTAGTCAAGGCGAACTGCTAACGG-3´

Table 2:   List of oligonucleotides used for DNA binding assays. The corresponding oligonucleotides were annealed by heating up to 80°C for 10 minutes and slow cooling to 4°C in 20 mM Tris/HCl [pH=8.0].

Prior to filtration, the nitrocellulose filter was activated by soaking in 0.4 M KOH for 10 minutes followed by extensive rinsing with water. Both membranes were subsequently equilibrated in assay buffer (20 mM MES [pH=5.5], 50 mM NaCl, 5 mM MgCl$_2$, 100µg/ml bovine serum albumin (BSA), 1 mM DTT) for 10 min.
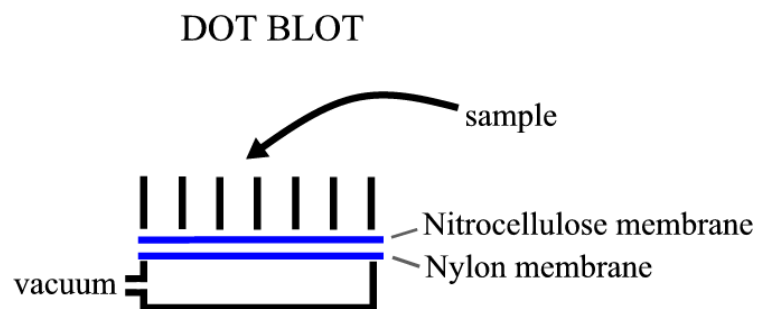
Figure 3:   Schematic assembly of a typical dot blot double filter binding assay. The sample is filtered through a double layer of one nitrocellulose filter (top) and one nylon filter (below) in a 96 well dot blot apparatus.
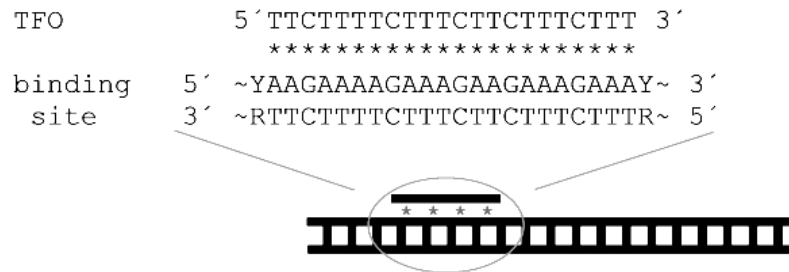
Helicase assay

The unwinding activity of the catalytic domain of *Sulfolobus solfataricus* Rad54 (SsoRad54cd) was tested with ^{32}P-labeled oligonucleotides annealed to M13mp19 (+) strand as substrate. Annealing was performed by mixing 10 pmol 5´-^{32}P labelled oligonucleotides, complementary to nucleotides 6311-6327 of M13mp19 (+) strand DNA, with 2 pmol of M13mp19 (+) strand DNA in 50 mM Tris/HCl [pH=7.5], 50 mM NaCl, 10 mM MgCl$_2$ and 1 mM DTT, heating up to 80°C for 10 minutes and slowly cooling to room temperature. The substrate was subsequently purified by gel filtration through a 2 ml Sephadex-G50 column (see triplex displacement assay). The helicase assay was performed at 37°C for 30 minutes in 10 µl reaction mixtures containing 20 mM MES [pH 5.5], 50 mM NaCl, 5 mM MgCl$_2$, 3 mM ATP, 100 µg/ml bovine serum albumin (BSA), 1 mM DTT, 2,5 nM ^{32}P-labeled helicase substrate and 50 – 250 nM SsoRad54cd. The reaction was terminated by the addition of 0.3% SDS, 10 mM EDTA, 5% glycerol and 0.1% bromphenol blue and analysed by electrophoresis on a 12% nondenaturing polyacrylamide gel. The extent of DNA unwinding was quantified by electronic autoradiography (STORM Phosphorimager, Amersham) and Image Quant software.

Triplex displacement assay

To evaluate protein translocation on duplex DNA, the displacement of a triplex forming oligonucleotide (TFO) from double-stranded DNA (triplex) by a translocating enzyme was assayed. Triplex substrates are generated by annealing a TFO to its target duplex DNA sequence. The TFO interact with its target sequence through the formation of a triple helical structure, whereby the pyrimidine rich (CT) TFO occupies the major groove parallel to a purine-rich (GA) strand and utilizes Hoogsten hydrogen bonds to form T_AT and C+_GC triplets (Gowers et al., 1999). The triplex DNA formation was performed as described in Firman and Szczelkun (Firman and Szczelkun, 2000). Briefly, the TFO was ^{32}P labeled by T4 polynucleotide kinase according to manufacture´s instructions and purified with commercially available oligonucleotide purification kits (Qiagen). To generate the triplex substrate, equimolar concentration (100nM) of NotI linearized pMJ5 plasmid and ^{32}P-labeled TFO were mixed in 25 mM MES [pH=5.5] and 10 mM MgCl$_2$, incubated at 57°C for 15 minutes and left to cool to 4°C overnight. Additionally to the protocol described, the resulting triplex was purified using a 2 ml Superdex G50 column (Amersham). Therefore, I established a quick and inexpensive procedure: as a small disposable "gel filtration" column I used a Pasteur pipette and sealed it at the bottom with a glass bead. The column is gravity packed with 2-3 ml degased Superdex G50 fine media suspension (Amersham). The column is equilibrated with about 5 ml of 25 mM MES [pH=5.5] and 10 mM MgCl$_2$ before use. The triplex was loaded on the column and allowed to enter. Subsequently equilibration buffer was carefully applied and 0.1 to 0.2 ml fractions were collected. The desired triplex elutes rapidly due to its large size and can be separated from free labeled TFO that migrates slower. The elution profile can be monitored using a szintillation counter. The efficiency of triplex formation and purity of the fraction was subsequently checked by agarose gel electrophoresis (see below). For the triplex displacement reaction, 2.5 nM of triplex substrates were incubated at 40°C for 30 minutes with the indicated amount of protein in buffer containing 25 mM MES [pH=5.5], 50 mM NaCl, 10 mM MgCl$_2$, 100μg/ml BSA, 1 mM DTT and 4 mM ATP. The reactions were quenched with GSMB buffer (15 % (w/v) glucose, 3% SDS, 250 mM MES [pH=5.5], 0.1 mg/ml proteinase K and 0,4 mg/ml bromphenol blue) and analysed with a 1.5% agarose gel (40 mM Tris-acetate [pH=5.5], 5 mM Na-acetate, 2.5 mM Mg-acetate) at 10 V/cm for 1.5 hour. The

proportion of bound and free TFO was determined using STORM Phosphorimager (Amersham) and Image Quant software.

```
TFO                5´ TTCTTTTCTTTCTTCTTTCTTT 3´
                      *********************
binding      5´ ~YAAGAAAAGAAAGAAGAAAGAAAY~ 3´
   site      3´ ~RTTCTTTTCTTTCTTCTTTCTTTR~ 5´
```


Figure 4:   Schematic drawing of a triplex substrate. The sequences of the CT rich triplet forming oligonucleotide (TFO) and the corresponding target sequence are depicted. The TFO binding is mediated by Hoogsten hydrogen bonds (*).

Cruciform extrusion assay

The cruciform extrusion assay was used to detect the generation of altered superhelical torsion in DNA by *Sulfolobus solfataricus* Rad54 protein. In this approach, cruciform formation from a linear DNA that contains $[AT]_{34}$ inverted repeat as a cruciform forming element, is used as a means to measure superhelical tension in DNA. The idea behind the cruciform extrusion assay is that remodeling enzymes are incubated with linearized pXG540 plasmid DNA that contains the $[AT]_{34}$ inverted repeat. In the presence of ATP, the remodeling enzymes generate superhelical torsion that results in the formation of the cruciform. The cruciform structure can be specifically cleaved by the structure specific T4 endonuclease VII (ENDO VII) (Lilley and Kemper, 1984). ENDO VII introduces bilateral cleavages and generates a double-stranded break in the DNA close to the junction. The resulting two DNA fragments can be analysed by gel electrophoresis. The amount of ENDO VII cleavage products provides a direct measure of the DNA distortion activity of the remodeler. The assay was performed according to Havas et al (Havas et al., 2000). 30 ng (1.5 nM) AvaI linearized pXG540 plasmid (containing a $[AT]_{34}$ inverted repeat) were incubated in 20 mM Hepes [pH=6.5], 50 mM NaCl, 3 mM $MgCl_2$, 5 % glycerol and 0.1 mM DTT with the indicated quantities of protein in the presence of 1 mM

ATP and 0.15 µg/ml endonuclease VII (Giraud-Panis and Lilley, 1996). The reactions were incubated at 40°C for 45 minutes and stopped by addition of 0.6 % SDS, 20 mM EDTA and 0,5 mg/ml proteinase K followed by incubation at 50°C for 15 minutes. The reactions were analysed by separation on a 1.2% agarose gel, stained with SYBR gold (Molecular Probes) and visualized with Typhoon™ 9400 Imager (Amersham).
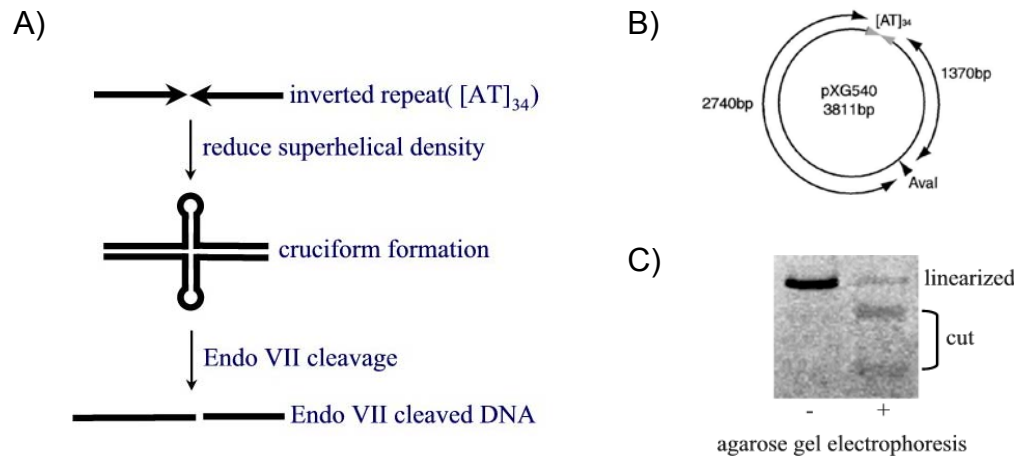
Figure 5:   A) Schematic illustration of the principle of the Cruciform extrusion assay. Generation of unconstrained superhelical torsion by SWI2/SNF2 enzymes leads to cruciform extrusion that is recognised by the structure specific endonuclease VII (ENDO VII). B) Schematic drawing of pXG540. pXG540 plasmid contains [AT]$_{34}$ inverted repeats that serves as cruciform reporter. The unique AvaI restriction site is used for linearization. Figures adapted from (Havas et al., 2000) C) The cleavage products (cut) are analysed by agarose gel electrophoresis. The gel shows an example of an negative (-) and positive (+) cruciform extrusion assay obtained with *Sulfolobus solfataricus* Rad54 catalytic domain in the absence (-) and presence (+) of ATP.

X-ray crystallography

Crystallisation

Growth of high quality crystals form the basic prerequisite for X-ray crystallography and obtaining diffraction quality crystals are still a major obstacle in protein crystallographic research. Crystals are three dimensional ordered structures that can be described by a repetition of identical unit cells. The crystallisation process is a thermodynamically driven reaction and analogous to conventional chemical reactions, proteins have to overcome an energy barrier in order to crystallize. This higher energy intermediate corresponds to the critical nucleus, the starting point of crystallisation. Crystal growth basically involves three steps: nucleation, growth and cessation of growth. Especially crystallisation of macromolecules is a complex process based on finding individual conditions and parameters that lead to formation of crystals. To find conditions that favour protein nucleation and crystallisation is the goal of the first screening step. Typically these initial crystals are not suited for diffraction experiments. A second step comprises the optimisation of the crystallisation conditions to obtain high quality diffracting crystals.

Crystallisation requires homogenous and pure protein, which is concentrated to a supersaturated state to obtain crystals. Different techniques are employed for setting crystallisation trials, which include sitting drop vapor diffusion or hanging drop vapor diffusion techniques. Thereby, a drop of protein solution is suspended over a reservoir containing buffer and precipitant. The drop slowly equilibrates with the reservoir solution by diffusion, leaving the drop with optimal crystal growth conditions. The search for crystallisation conditions were performed with commercially available crystallisation sparse matrix screens from Hampton, Jena Bioscience or Nextal. Usually, drops of 1 µl crystallisation buffer + 1 µl of proteins solution were mixed and suspended over 100 µl reservoir solution using 96 well crystallisation plates (Corning®). For more extended screening 0.2 µl + 0.2 µl drops were set using the Hydra II semi-automatic protein crystallisation robot (Apogent Discoveries, USA). In the second step, the quality of the crystals were improved by extensive variation of

the initial crystallisation conditions by variation of the protein concentration, pH, drop size or by adding additives like glycerol, salts or sugars at different concentrations.

Crystallisation of SsoRad54cd

SsoRad54cd was crystallized by sitting drop vapor diffusion at 25°C after mixing 2 μl protein (30 mg/ml in 20 mM Tris/HCl [pH=8.5], 200 mM NaCl, 1 mM DTT, 1 mM EDTA) and 2 μl precipitant (2 M $(NH_4)H_2PO_4$, 100 mM Tris/HCl [pH=3.9], 50 mM sodium malonate, 5 % glycerol).

Crystallisation of SsoRad54cd C-terminal domain.

SsoRad54cd C-terminal domain (aa 663-906) was crystallized by mixing 2 μl protein (20 mg/ml in 20 mM Tris/HCl [pH=7.5], 200 mM NaCl, 1 mM DTT, 1 mM EDTA) and 2 μl precipitant (12 % PEG 4000, 100 mM Tris/HCl [pH=8.5], 5 % ethylene glycol) at 25°C using hanging drop vapor diffusion.

Crystallisation of SsoRad54cd:DNA complex

The oligonucleotides were synthesized in-house (Dr. Arnold), HPLC purified and dissolved in 10 mM Tris/HCl [pH=8.0] at a concentration of 3 nmol/μl. The double-stranded DNA was generated by mixing the two corresponding oligonucleotides in a molar ratio of exactly 1:1, heating up to 85°C for 5 minutes followed by slow cooling to 4°C. The binary SsoRad54cd:DNA complex was produced by incubating SsoRad54cd with the duplex DNA substrate at a protein concentration of 2 mg/ml in 20 mM Tris/HCl [pH=7.5], 150 mM NaCl, 1 mM EDTA and 1 mM DTT and a protein : DNA molar ratio of 1:1.2 for 1 hour on ice. Subsequently the protein:DNA complex was concentrated to 20 mg/ml using ultrafiltration devices (Nanosep 10K, PALL Corporation, Dreieich). The resulting SsoRad54cd:DNA complex was directly used for crystallisation without further purification of the complex. Initial screens were performed in sitting drop vapour diffusion with commercial sparse matrix screens

from Hampton using 96 well crystallisation plates (Corning®). The screening for appropriate crystallisation conditions for SsoRad54cd:DNA complex also included various DNA substrates that differed primarily in length (table 3). Finally, SsoRad54cd:DNA complex were crystallized by mixing 2 µl SsoRad54cd:DNA complex (20 mg/ml in 20 mM Tris/HCl [pH=7.5], 150 mM NaCl, 1 mM EDTA and 1 mM DTT) with 2 µl precipitant (15 % PEG 3350, 0.1 M magnesium formate and 5 % glycerol). The final double-stranded DNA substrate was composed of 5´-AAAAA AATTGCCGAAGACGAAAAAA–3´ and 5´-TTTTTTTCGTCTTCGGCAATTTTT–3´.

Crystals were cryo-protected by addition of 20 % glycerol to the mother liquor and flash-frozen in liquid nitrogen for storage and data collection.

Mercury derivatisation of SsoRad54cd

Heavy metal derivatisation of proteins is a common technique in X-ray crystallography used for phasing. SsoRad54cd was mercury derivatized by soaking SsoRad54cd crystals in mother liquid supplemented with 0.5 mM HgCl$_2$ for 2 hours. To minimize background diffraction, the crystals were subsequently backsoaked in mother liquid without HgCl$_2$ for 15 minutes, transferred into precipitant supplemented with 20 % glycerol for 10 minutes and flash frozen in liquid nitrogen.

Structure determination by X-ray diffraction

Theory of X-ray structure determination

X-ray crystallography is a powerful experimental technique that exploits the fact that X-rays are diffracted by the electrons of the periodically arranged atoms in a crystal. Based on the diffraction pattern and additionally measured phase information, the electron density can be reconstructed using Fourier synthesis. An atomic model is then progressively built into the experimental electron density and refined in an iterative process. The refined atomic coordinates provide a good model for the protein structure. The following part briefly describes the theoretical background for structure determination based on X-ray diffraction experiments. For more detailed information, see crystallography textbooks (see references)

X-rays are electromagnetic radiation that interact with the electrons of atoms. The electric field of X-rays excite the electrons into forced dipole oscillations that in turn act as an emitter of electromagnetic radiation (diffraction). Diffraction applications depend on the elastic scattering (Thompson Scattering), i.e. the incident beam is diffracted without energy or phase changes. The diffracted X-rays from different atoms in the crystal interfere with each other and result in distinct reflections. The diffraction pattern is directly related to the electron distribution in the crystal and can be described by Bragg´s law. The reflections are recorded as an array of spots on the detector, whereby each reflection represents the overall scattering from a particular set of Bragg planes. In a diffraction experiment the position of reflections and their intensities are measured. The positions of the reflections are related to the diffraction from a particular set of Bragg planes and are labeled with the index triple (h,k,l) (Miller indices, reciprocal space coordinates). Additionally, each reflection can be assigned with the appropriate intensity that is proportional to the square of the structure amplitude $\left| F_{hkl} \right|$. The X-ray scattering power of an atom depends on the number of electrons in the particular atom. Based on the Don Cromer and J. Mann equation the (normal) atomic scattering factor (f) of an atom can be calculated. In addition, the scattering power is further influenced by the so called temperature-, B-

or Debye-Waller factor that is taking into account the thermal "vibration" (isotropic/anisotropic) of the electrons.

The diffraction from the unit cell of a crystal results from the overall scattering contributions of all atoms in the unit cell and is described by structure factors. The structure factor F(hkl) of a reflection h,k,l is a complex number derived from the following equation:

$$F(h,k,l) = \sum_{j=1}^{atoms} f_{(j)} \cdot \exp[2\pi \cdot i(hx_{(j)} + ky_{(j)} + lz_{(j)}] \quad (1)$$

This is a simple summation which extends over all atoms j , with x, y and z as their fractional coordinates. The f(j) is the atomic scattering factor of atom j and depends on the atom type and the diffraction angle of the corresponding reflection (h,k,l). Using this formula, the structure factors from a known structure can easily be calculated. In crystallography, however, we are dealing with the inverse problem. Structure factors (precisely, only their amplitude from the measured intensities) are known and the structure has to be determined.

Crystals display a periodic assembly of the molecules, therefore the electron density resembles a periodic function, so that we can apply a Fourier transformation (FT) to obtain the electron density from the periodic complex exponential function. The corresponding FT to formula (1) has the structure factor F(hkl) as its Fourier coefficients:

$$\rho_{(x,y,z)} = \frac{1}{V} \sum_h \sum_k \sum_l F_{(h,k,l)} \cdot \exp[-2\pi \cdot i(hx + ky + lz)] \quad (2)$$

This formula represents a summation over all structure factors for each position x, y, z in the normalized unit cell, divided by the cell volume. In fact, the reciprocal space (h,k,l), which is represented by the diffraction pattern, is transformed into the real space (x,y,z), the electron density. But in order to perform the FT, the knowledge of the complex structure factors is needed, but only their amplitudes $\left|F_{(h,k,l)}\right|$, but not their phases $\alpha_{(h,k,l)}$ are known from normal diffraction experiments. This fact is the central problem in structure determination by X-ray crystallography and is referred as the "phase problem".

Equation (2) can be reformulated so that the phase term $\alpha_{(h,k,l)}$ becomes evident.

$$\rho_{(x,y,z)} = \frac{1}{V}\sum_h\sum_k\sum_l \left|F_{(h,k,l)}\right| \cdot \exp[-2\pi \cdot i(hx + ky + lz - \alpha_{(h,k,l)})] \quad (3)$$

This formula separated the actually measured $\left|F_{(h,k,l)}\right|$ and its corresponding unknown phase part $\alpha_{(h,k,l)}$. Over the years many methods have been developed to deduce the phases for reflections, including computationally based direct methods, isomorphous replacement, and multi-wavelength anomalous dispersion (MAD) methods. Following, the two techniques applied in this PhD thesis are briefly described.

Single-wavelength anomalous dispersion (SAD) and multi-wavelength anomalous dispersion (MAD)

These experimentally derivations of the phases exploit the fact that the scattering factor contains additional complex contributions from anomalous dispersion effects (essentially resonance absorption), which become substantial in the vicinity of the X-ray absorption edge of the scattering atom. Heavy atoms usually have an absorption edge in the range of the wavelengths that are useful for X-ray crystallography. Therefore, heavy atoms are introduced into proteins. At the absorption edge, some of the incident radiation is used to create transitions in the atoms that reduce the total scattering at this wavelength and also change the phases of the coherently scattered X-rays. The changes in scattering power and phases of the heavy metal atoms result in differences between the diffracted intensities measured at different wavelengths as well as differences between reflections of the type (hkl) and (-h-k-l) (anomalous or Bijvoet pairs) measured at the same wavelength (known as breakdown of the Friedel´s law). These anomalous contributions to the atomic scattering factor can be expressed through a real and an imaginary component. The atomic scattering factor at any wavelength can be described by the following equation:

$$f(\lambda) = f_0 + \delta f'(\lambda) + if''(\lambda) \quad (4)$$

f_0 is the conventional atomic scattering factor for wavelength far from an absorption edge, $\delta f'(\lambda)$ represent the amount by which the normal scattering is reduced at

wavelength λ, and $f''(\lambda)$ account for the amount of anomalous scattering, the out-of-phase component of the scattering at this wavelength.

The anomalous differences can be used to determine the heavy metal coordinates in the unit cell. A common used method implies the Patterson function by which the heavy atom substructure is determined by deconvolution of the auto-correlation function of the heavy atoms electron density. A second technique used for determination of the positions of anomalous scatterers, are direct methods that use phase triplet relations for phase calculation. Direct methods are only applicable to small molecules, but are superior methods for the determination of complex substructures, for instance in MAD or SAD. Once the heavy atom substructure is determined, their contribution to the structure factors can be calculated. Assuming isomorphism, the structure factors for the derivative crystal (F_{PH}) are equal to the sum of the protein structure factors (F_P) and the heavy atom structure factors (F_H).

$$F_{PH} = F_P + F_H \quad (5)$$

Since the structure factors of F_H as well as the amplitudes of the structure factors for F_{PH} and F_P are known, one can deduce possible values for the protein phase angles. The Fourier transform of the amplitudes IF_PI and phases ($\Delta\varphi+\varphi_H$) yield an electron density map corresponding to all atoms in the structure ($\Delta\varphi$ = difference in phase angle between normal and anomalous scattering components; φ_H = phase angle of F_H)

Molecular replacement

Molecular replacement is commonly used if a model exists for a reasonably large fraction of the structure in the crystal, from which estimates of the phases can be computed. For this purpose it is necessary to place the model structure in the correct orientation and position in the "new" unit cell that include a 6 dimensional search problem (actually two 3 dimensional searches): three rotation angles and three translational parameters. Traditional molecular replacement methods are based on the properties of the Patterson function (6). The Patterson function or Patterson map is a Fourier transform of the intensities (amplitudes squared, no phase information

necessary, $\alpha_{hkl}=0$) and gives a map of all cross vectors between each atom pair in the unit cell.

$$P(uvw) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}|^2 \exp[-2\pi(hu + kv + lw)] \quad (6)$$

The rotation and translation search are based on the comparison of the Patterson maps of the model structure and of the crystal diffraction data. The "fit" are calculated by the rotation function or translational function, respectively. Since intramolecular vectors depend only on the orientation of the molecule, they are used for the rotation function. The intramolecular vectors are "extracted" by using only the Patterson peaks within the unit cell "radius" that mostly exclude the intermolecular cross vectors. Once the correct orientation is determined, intermolecular cross vectors that depend on both, the orientation of the molecule and the position, can be exploited in the translation search. Subsequently, the phases for the reflections of the unknown structure can be estimated by calculating the phase angles implied by the model structure that has been appropriately oriented and positioned in the cell of the unknown structure. The calculated phase angles $(\alpha_{hkl})_{calc}$ are attached to the measured structure amplitudes $|F_{hkl}|_{obs}$ resulting in an estimate of the structure factor (or the electron density).

$$F(h) = |F_{obs}(h)| \cdot \exp[i\alpha_{calc}(h)] \quad (7)$$

The use of $(\alpha_{hkl})_{calc}$ for the electron density map calculation implicates a bias towards the model structure used to generate the phase angles. Therefore, a $2|F_O| - |F_C|$ electron density map or other weighted electron density maps (e.g. $3F_O$-$2F_C$, Sigma-A weighted,…) are typically used to minimize model bias.

$$F(h) = (2|F_O(h)| - |F_C(h)|) \cdot \exp[i\alpha_{calc}(h)] \quad (8)$$

Subsequently, an atomic model is built into the electron density followed by an iterative refinement process, by which the model is adjusted to find a closer agreement between the calculated and observed structure factors. An agreement index, called R-factor, is used to monitor the quality of the model. Different methods are employed for refinement like density modifications, using constrains, positional refinement, manual model building and so on.

Results

Multiprotein or multidomain SWI2/SNF2 enzymes remodel protein:DNA interfaces in a mechanism that is poorly understood. These complexes show a quite diverse overall subunit composition, but they all contain a principal motor subunit that belongs to the SWI2/SNF2 type of ATPases (Becker and Horz, 2002; Peterson, 2002). The ATPase subunits in turn share a conserved catalytic ATPase core indicating that a principal ATP-driven event on DNA forms the mechanistic basis of the dynamic remodeling processes. To gain insights into the action of the conserved catalytic ATPase domain, we determined the molecular structure of the catalytic ATPase core of *Sulfolobus solfataricus* Rad54 homolog in absence of DNA and in complex with its dsDNA substrate. The c̲atalytic d̲omain of *S̲ulfolobus s̲olfataricus* Rad54 homolog (SsoRad54cd) comprises the residues 430-906 and represents the complete conserved part of SWI2/SNF2 ATPases. Since the function of SsoRad54cd is unknown, I examined its biochemical properties and demonstrated that the activities of SsoRad54cd resemble several SWI2/SNF2 typical properties. Therefore the *Sulfolobus solfataricus* Rad54 homolog is an appropriate object to study the mechanism of SWI2/SNF2 enzymes. Furthermore, structure-modeling and structure-function analysis were used to develop a model that could explain observed activities of eukaryotic SWI2/SNF2 enzymes.

Comparative sequence analysis identified *Sulfolobus solfataricus* open reading frames SSO1653 (N-terminal fragment) and SSO1655 (C-terminal fragment) to encode a protein with homology to SWI2/SNF2 family ATPases (Fig. 6). Although the homology extends mainly to the seven helicase motifs comprising ATPase domain,

the protein was assigned to be homologous to the yeast Rad54 since the conserved ATPase domain is located at the C-terminus, similarly to Rad54. Therefore we named the protein *Sulfolobus solfataricus* Rad54 homolog (SsoRad54), but it is probably equally homologous to Rad54 proteins or other members of the SWI2/SNF2 family.
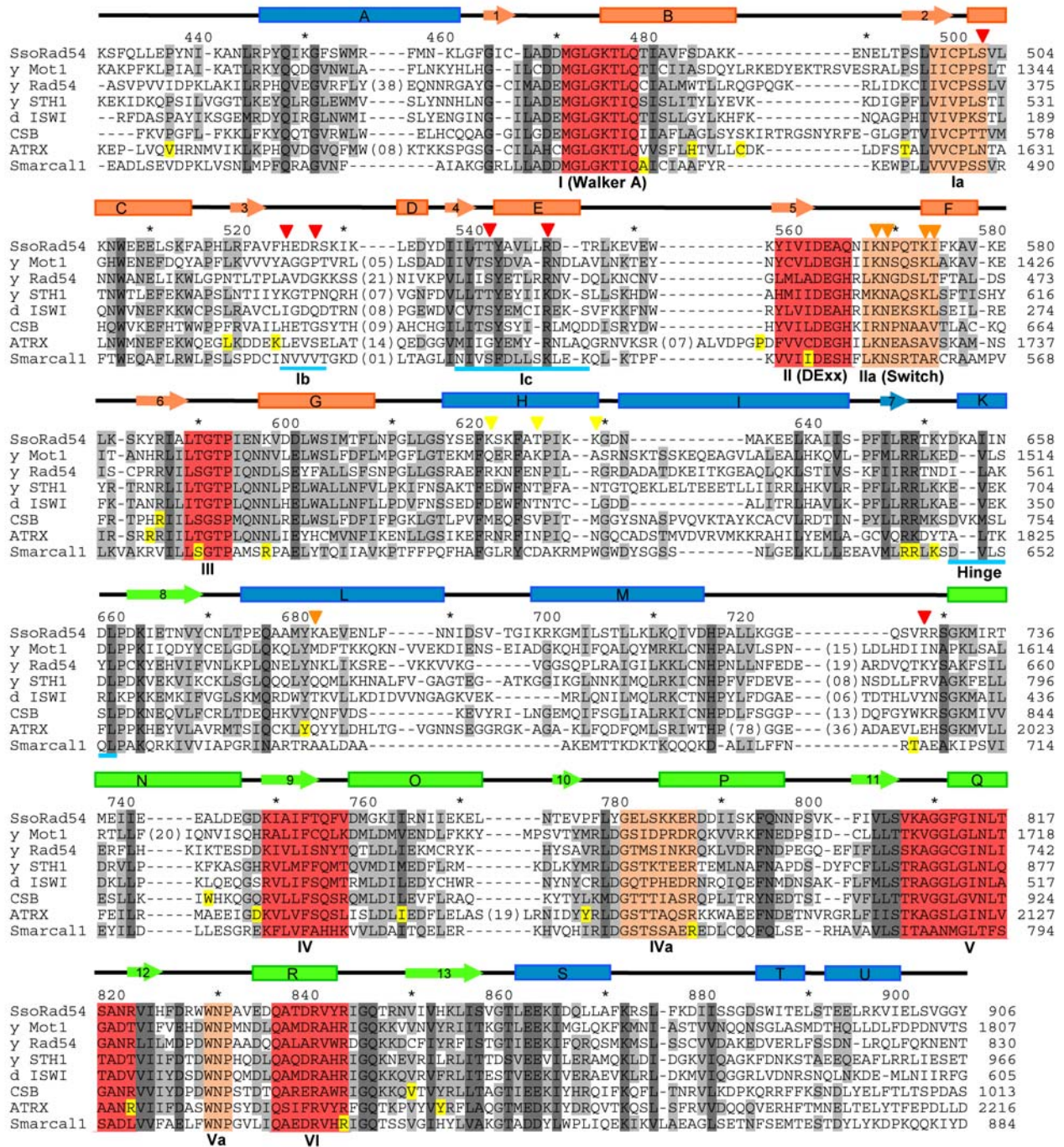

Figure 6: Structure based sequence alignment of the catalytic domain of SsoRad54 with several other SWI2/SNF2 enzymes. DExx box helicase sequence motifs (red) or structurally identified additional motifs of

SWI2/SNF2 enzymes (pale red) are highlighted and annotated. Additional conserved regions are shaded light (moderately conserved) or dark grey (strongly conserved). Human disease missense mutations are depicted yellow. The annotated secondary structure of SsoRad54cd is shown on top of the alignment (arrows: β-strands; boxes: α-helices). Residues implicated in DNA binding are highlighted with triangles: red: 3'-5' strand; orange 5'-3' strand; yellow: putative additional interactions. Abbreviations: SsoRad54: *Sulfolobus solfataricus* Rad54 homolog; y_Mot1: *Saccharomyces cerevisiae* Mot1; y_Rad54: *Saccharomyces cerevisiae* Rad54; y_STH1: *Saccharomyces cerevisiae* STH1; d_ISWI: *Drosophila melanogaster* imitation of switch; CSB: human Cockayne syndrome protein B; ATRX: human ATRX protein; Smarcal1: human SMARCAL1.

Cloning and expression of SsoRad54cd

Sequence comparison of SsoRad54 with various SWI2/SNF2 ATPases reveals that *Sulfolobus solfataricus* open reading frames SSO1653 (N-terminal fragment) and SSO1655 (C-terminal fragment) are separated by a transposon. This insertion comprises 20 amino acids (N-terminal aa 789-802 and C-terminal aa 1-7) that are missing in all other SWI2/SNF2 enzymes. Based on the sequence comparisons, the full-length open reading frame for SsoRad54 was cloned omitting the putative transposon associated recombination site. In a first step, the two fragments were amplified by PCR and subsequently combined by overlap PCR. The PCR product was cloned into pET28 expression vector (Novagen) and expressed in *E.coli* BL21 Rosetta strain (Novagen). Since expression of the full-length SsoRad54 yielded insoluble protein, N-terminal truncated SsoRad54 variants (∆1-284 and ∆1-429) were created using the SsoRad54cd full-length gene as template. The SsoRad54 variants were cloned into pET28 vector and expressed in *E.coli* BL21 Rosetta. Both constructs showed high-yield expression in *E.coli*. SsoRad54cd (aa 430-906) was purified to near homogeneity by an initial "heat-step", immobilized Ni affinity (utilizing a N-terminal 6x His tag), Resource S ion exchange and Superdex S200 gel filtration

chromatography by standard procedures (see Material and Methods). The same protocol was used for purification of SsoRad54_(285-906).
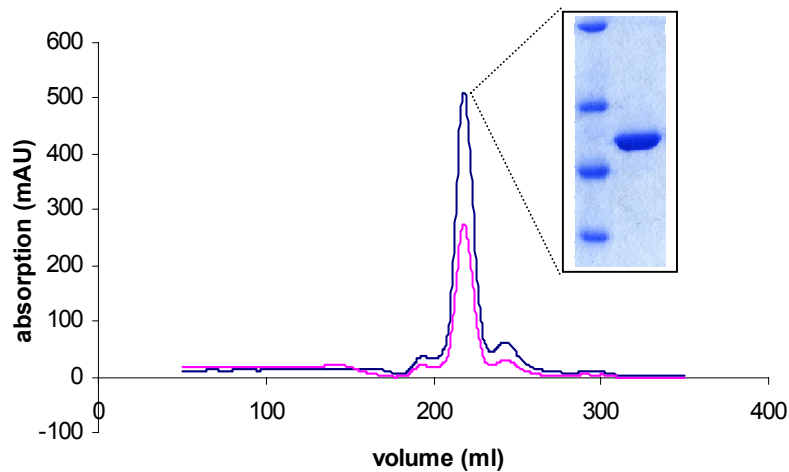
Figure 7:   Superdex S200 gel filtration elution profile of SsoRad54cd. Blue corresponds to the absorption at 280 nm, pink to absorption at 260 nm. A single peak elutes at a volume according to a size of 57 kDa indicating that SsoRad54cd exists as monomer in solution. Inlet: SDS-PAGE of the peak fraction.

Biochemical characterisation of the ATPase domain of *Sulfolobus solfataricus* Rad54 homolog (SsoRad54cd)

Biochemical studies on many SWI2/SNF2 complexes or individual SWI2/SNF2 related ATPase subunits revealed a set of shared biochemical activities. These activities include DNA stimulated ATPase activity, translocation on DNA and generation of negative superhelical torsion into DNA (Beerens et al., 2005; Havas et al., 2000; Ristic et al., 2001). Based on the recent biochemical results, DNA translocation and distortion activities have been suggested to be appropriate to remodeling function. Both activities rely on ATP-hydrolysis by the conserved helicase-like ATPase subunit, indicating that a common ATP driven event on DNA is conserved among SWI2/SNF2 enzymes. The sequence conservation, which is mainly restricted to the catalytic ATPase core, and the findings that the activities depend on ATP hydrolysis suggest that these activities may be features of the conserved ATPase domain itself. To test this idea, DNA stimulated ATPase activity,

DNA binding properties, ATP dependent translocase and DNA distortion activity of the catalytic domain of *Sulfolobus solfataricus* Rad54 homolog (SsoRad54cd) were analysed. Moreover, these experiments may confirm the homology of SsoRad54 to eukaryotic SWI2/SNF2 ATPases in addition to the sequence similarity.

dsDNA stimulated ATPase activity

All SWI2/SNF2 enzymes contain an ATPase domain that is essential for in vitro and in vivo function (Laurent et al., 1993). The intrinsic ATPase activity of these proteins is very low, but is stimulated in the presence of DNA. The optimal DNA substrate for stimulation, however, is variable and depends on the enzyme. For instance, the ATPase activity of the yeast SWI/SNF and RSC complex is stimulated equally by double-stranded DNA, single-stranded DNA and nucleosome fragments (Cairns et al., 1996; Cote et al., 1994). In contrast, *Drosophila* ISWI protein shows maximal stimulation only in the presence of nucleosome particles (Brehm et al., 2000; Tsukiyama and Wu, 1995). Yeast Rad54 is stimulated in the presence of double-stranded DNA or chromatin substrate, but not by single-stranded DNA (Petukhova et al., 1998).

We tested the stimulation of the ATPase activity of SsoRad54cd by single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA). Therefore, SsoRad54cd was incubated with γ-^{32}P-ATP in the absence or presence of ssDNA or dsDNA. ATP hydrolysis was then analysed by thin layer chromatography. No significant hydrolysis of ATP was observed without DNA. However, the addition of 50 mer double-stranded DNA and to a less amount single-stranded 50 mer oligonucleotides (random sequence) stimulated ATP hydrolysis. Single-stranded DNA substrates tend to form secondary structures that make a distinction of true ssDNA dependent activity quite difficult and can lead to false positive results. For instance, ϕ174 single-stranded DNA resulted in substantial stimulation of the ATPase activity of yeast Rad54, whereas Rad54 is not stimulated by poly(dA) or poly(dT) (Petukhova et al., 1998). To exclude that the activation of the ATPase activity is due to secondary structures, the ATPase stimulation of SsoRad54cd was tested with poly(dA):poly(dT) duplex as double-stranded DNA and poly(dT) as single-stranded DNA substrate. Poly(dT) is

not able to form secondary structures and represents a true single-stranded DNA substrate. Poly(dA):poly(dT) duplex DNA greatly stimulated the ATPase activity of SsoRad54cd, whereas poly(dT) was not effective (Fig. 8). These findings indicate that the previously observed stimulation by addition of the single-stranded 50 mer oligonucleotide resulted from secondary structure in the DNA. The results demonstrate that SsoRad54cd has dsDNA stimulated, but not ssDNA stimulated ATPase activity, which is consistent with yeast Rad54.

Typically, the ATPase assays are performed at pH = 7.0 and 30°C. At these conditions SsoRad54cd exhibits a low dsDNA stimulated ATP hydrolysis with a turn over rate of $\approx 0.2 \dfrac{ATP}{sec \cdot active\_site}$. Since *Sulfolobus solfataricus* lives in an environment with pH = 2-4 at about 80°C, the reaction conditions were optimized to fit closer the physiological conditions. Optimal conditions for the ATPase activity assay with SsoRad54cd were determined at pH = 5.5 and 70°C. Higher temperatures tended to be impractical due to high background ATP hydrolysis. Consistently, studies on DNA primase and hydroxymethylglutaryl-coenzyme A-reductase from *Sulfolobus solfataricus* demonstrated best activities at pH = 5.5 and 85°C.
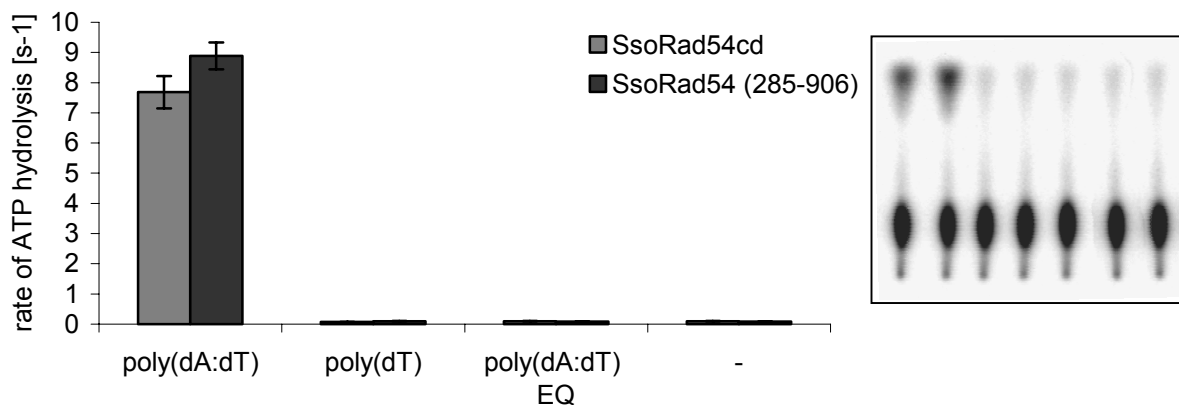
Figure 8:   ATPase assays of SsoRad54cd (grey bars), SsoRad54 (285-906) (black bars) and the respective E563Q mutations in the presence of double-stranded poly(dA:dT), single-stranded poly(dT) substrates or in absence of DNA (-) reveal a specific dsDNA stimulated ATPase activity, which is abolished by a Walker B mutation (E563Q). Values (means ± standard deviation of three independent experiments) denote the rate of ATP hydrolysis per second and active site. Right panel: exemplary

image from thin layer chromatography analysis of an ATP hydrolysis reaction. The liberated radioactive γ-phosphate (upper spot) almost runs with the front of the developing solution. The lower visible spot correspond to non-hydrolysed ATP (same sequence as the bar chart).

As shown in figure 8, SsoRad54cd exhibits dsDNA stimulated ATPase activity with a turn over rate of $\approx 7.6 \frac{ATP}{sec \cdot active\_site}$ under optimal conditions. Compared to the ATP hydrolysis rate of $\approx 7.5 \frac{ATP}{sec \cdot active\_site}$ for Sth1 (Du et al., 1998), $\approx 1.1 \frac{ATP}{sec \cdot active\_site}$ for the SWI/SNF complex (Laurent et al., 1993) and $\approx 6.7 \frac{ATP}{sec \cdot active\_site}$ for Rad54 (Jaskelioff et al., 2003), SsoRad54cd possesses an effective ATPase activity in the presence of dsDNA that corresponds quantitatively to eukaryotic SWI/SNF family members. To test the influence of the N-terminal 429 amino acids of SsoRad54, additionally a longer construct was used to quantify the ATPase activity. Since I was not able to express soluble full-length SsoRad54, a SsoRad54 variant lacking "only" the N-terminal 284 amino acids was analysed. SsoRad54_(285-906) displayed only a moderately higher turn over rate (see figure 8), indicating that the N-terminal part of SsoRad54 does not have a strong influence on the ATPase activity. As a negative control, SsoRad54cd with the single amino acid substitution E563Q in the Walker B motif was included in the experiments. It has been shown that E→Q mutations in the Walker B motif abolishes ATPase activity of many P-loop ATPases such as ABC-, AAA[+] or DExx box ATPases (Lammens et al., 2004). The absence of ATP hydrolysis with SsoRad54cd_EQ substantiate that the observed ATPase activity is not due to the presence of copurifying contaminants.

Biochemical studies on Sth1 revealed that the stimulation of its ATPase activity depends on the length of the DNA fragment used (Cairns et al., 1996). Similar results were found for ISWI, SWI/SNF and Rad54 (Jaskelioff et al., 2003; Whitehouse et al., 2003). Maximum stimulation is reached with DNA of about 100 bp, but is typically not further stimulated with longer DNA, at least when assessing the catalytic subunit of the remodelers by itself. The dependence of the DNA length for ATPase stimulation

of SsoRad54cd was addressed using saturating amounts of dsDNA with 21bp, 50 bp and ≈300bp in length.
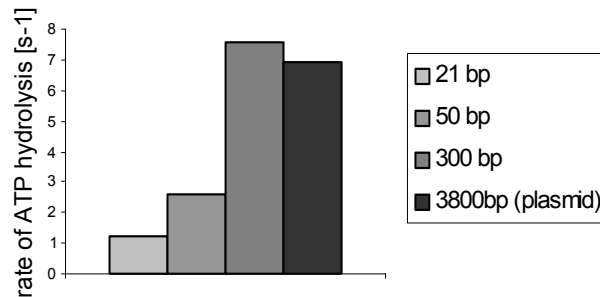
Figure 9: ATPase assays of SsoRad54cd in the presence of saturating amounts of dsDNA with different length indicate that SsoRad54cd need a minimum DNA length for optimal stimulation. Longer DNA than 300 bp does not further increase its ATPase activity, suggesting that the optimal DNA length for stimulation of SsoRad54cd ATPase activity range between 50 and 300 bp.

Figure 9 illustrates that SsoRad54cd ATPase activity significantly depends on DNA length. The stimulation increases proportional with the DNA length and could be enhanced about 7 fold changing from 21 bp to 300 bp substrate. Longer dsDNA (linearized plasmid, 3800 bp) does not further increase its ATPase activity. Hence, the optimal stimulation of SsoRad54cd ATPase activity requires dsDNA with lengths between 50 bp and 300 bp. No further evaluation of the minimal DNA length required for optimal ATPase stimulation was performed, but nevertheless, the range of dsDNA length required for optimal stimulation of SsoRad54cd corresponds well with values found for Sth1 (>80 bp), ISWI (>40 bp) and Rad54 (>70 bp) (Jaskelioff et al., 2003; Whitehouse et al., 2003). DNA-length dependent stimulation of ATP hydrolysis is a feature of processive DNA translocation and has been assessed as first hint that SWI2/SNF2 enzymes might have DNA translocating activity. In the literature it is argued, that ATP hydrolysis is coupled with progression along DNA. When the end of a DNA fragment is reached, a slow dissociation and recycling step may occur, during which the rate of ATP hydrolysis is reduced. If translocation along DNA is restricted due to short fragments, ATPase activity might be expected to be lower. However, no further stimulation of ATP hydrolysis occurred with DNA substrates longer than ≈100 nucleotides, indicating little processivity of SWI2/SNF2 enzymes. A second reason

for length dependent stimulation of ATP hydrolysis by DNA might be that DNA binding to an extended binding site or to multiple binding sites on large remodeling factors may be necessary for full stimulation.
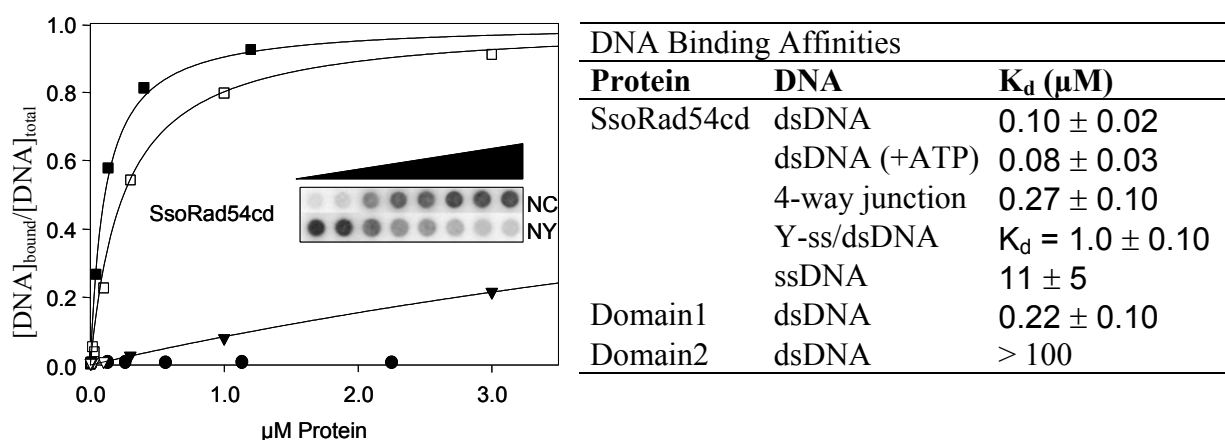
Lack of helicase activity

Initial sequence comparisons reveal that SWI2/SNF2 proteins have sequence motifs that are related to DExx box helicases. First, it was suggested that SWI2/SNF2 enzymes use the energy derived form ATP hydrolysis to unwind DNA. DNA unwinding could provide the means to remodel protein DNA interfaces. However, for all SWI2/SNF2 enzymes, no helicase activity could be detected using standard oligonucleotide displacement assays. Since the function of SsoRad54 is unknown, helicase activity for SsoRad54cd was tested. A ^{32}P-labeled 25 mer oligonucleotide was hybridized to circular single-stranded M13mp19 (+) DNA and used as substrate for the helicase assay. No displacement of the hybridized oligonucleotide was observed with SsoRad54cd in the presence of ATP, indicating that SsoRad54cd has no DNA unwinding activity. Consistent with the observation for other SWI/SNF2 proteins, this result suggests that ATP hydrolysis by SsoRad54cd is not involved in DNA unwinding.

DNA binding properties

Since the ATPase activity of SWI2/SNF2 enzymes is stimulated in the presence of DNA, it is not surprising that SWI2/SNF2 enzymes interact with DNA. Quinn et al showed, using gel retardation assays that the SWI/SNF complex has a high affinity for DNA ($K_d \approx 1-9$ nM) and demonstrated that the SWI/SNF complex has no intrinsic affinity to regulatory sequences or display other obvious sequence-specificity (Quinn et al., 1996). To evaluate the DNA binding properties of SsoRad54cd, I first examined the DNA affinity to dsDNA and ssDNA using double filter nitrocellulose/nylon binding assays and calculated its binding constants. The ATPase activity of SWI2/SNF2 enzymes depends on DNA length that might be due to an extended binding site. For ISWI it has been reported to require a minimal DNA length

of 41 bp for maximal DNA binding efficiency (Whitehouse et al., 2003).Therefore, in order to allow optimal DNA interaction, I used DNA with a length of 50 nucleotides to monitor DNA binding. For characterisation of the DNA binding properties of SsoRad54cd, 50 bp dsDNA and 50 mer ssDNA fragments (see table 2) were incubated with SsoRad54cd at 25°C for 20 minutes. The double filter binding assay revealed that SsoRad54cd strongly prefers dsDNA (K_d = 0.10 $\pm$ 0.02 µM) over ssDNA (K_d = 11 $\pm$ 5 µM), consistent with the dsDNA, but not ssDNA stimulated ATPase activity.

DNA Binding Affinities		
Protein	**DNA**	**K_d (µM)**
SsoRad54cd	dsDNA	0.10 $\pm$ 0.02
	dsDNA (+ATP)	0.08 $\pm$ 0.03
	4-way junction	0.27 $\pm$ 0.10
	Y-ss/dsDNA	K_d = 1.0 $\pm$ 0.10
	ssDNA	11 $\pm$ 5
Domain1	dsDNA	0.22 $\pm$ 0.10
Domain2	dsDNA	> 100

Figure 10: DNA binding properties of SsoRad54cd. Typical nitrocellulose (NC)/Nylon (NY) double filter binding assays of SsoRad54cd with a variety of DNA oligonucleotides: SsoRad54cd + dsDNA (■), SsoRad54cd + ssDNA (▼), SsoRad54cd_domain 1 (aa 430-657) + dsDNA (□) and SsoRad54cd_domain 2 (aa 659-906) + dsDNA (●). DNA:protein complexes bind to the NC, whereas remaining uncomplexed DNA binds to NY membrane. The data were fitted by non-linear least square fit according to $f=p^n/(K_d+p^n)$, where f is the fraction of DNA bound to SsoRad54cd relative to total DNA, K_d is the dissociation constant, p is the protein concentration and n is the Hill coefficient. Inset: Example of dot blot filter binding data. DNA concentration is held constant whereas the protein concentration increases from left to right (black triangle). The darkness of the dots is proportional to the amount of γ-^{32}P-labeled DNA bound. Right table: List of dissociation constants of SsoRad54cd with different DNA substrates.

The yeast SWI/SNF complex also binds to synthetic 4-way junction DNA with high affinity (1 nM) (Quinn et al., 1996). The recognition and binding to secondary structures have been suggested to be crucial for SWI/SNF function. Interestingly, it has been argued that 4-way junction DNA resembles the DNA at the entry / exit point of a nucleosome particle where the helices form a cross-over structure (Lilley, 1992). To examine if SsoRad54cd preferentially binds to secondary structures, the binding to a 4-way junction substrate and a Y-substrate were tested (Fig. 10). SsoRad54cd binds to 4-way junction substrate approximately with the same affinity than dsDNA ($K_d = 0.27 \pm 0.10$ µM), suggesting that SsoRad54cd does not prefer secondary structure containing DNA. The Y-substrate was bound with an affinity between dsDNA and ssDNA ($K_d = 1.0 \pm 0.10$ µM). This data can be explained by "dilution" of the duplex DNA part and is consistent with the lack of helicase activity, since helicases bind preferentially to ssDNA/dsDNA junctions.

The binding constants of SsoRad54cd are in agreement with the DNA binding constants reported for yeast Rad54 (approximately $K_d = 0.11$ µM for dsDNA) (Raschle et al., 2004). However, the dissociation constants for SsoRad54cd are higher than for other remodeling factors like the SWI/SNF complex or ISWI. These differences could be expected, since multidomain or multisubunit remodeling factors have additional DNA interacting domains. For instance, the SWI/SNF complex contacts DNA with at least 3 further subunits (Quinn et al., 1996) and ISWI has an additional nucleosome-binding site in the SANT domain that probably significantly adds to the DNA binding affinity. Thus, it is not surprising that multisubunit remodeling factors bind DNA more tightly than the catalytic domain of SsoRad54cd alone.

Based on the structure of SsoRad54cd:DNA complex (see below) we also tested DNA binding affinity of separately expressed and purified domain 1 and domain 2 of SsoRad54cd. The structure suggests that DNA is bound predominantly by domain 1. Concordantly, domain 1 of SsoRad54cd (430-658) showed only a moderately reduced DNA binding activity ($K_d = 0.22 \pm 0.10$ µM) compared to SsoRad54cd, whereas domain 2 (659-906) shows no significant DNA binding affinity for either dsDNA or ssDNA.

The ATPase activity of SWI2/SNF2 enzymes is stimulated by DNA. In return the presence of ATP might influence the affinity to DNA. Therefore the effect of the presence of the slow-hydrolysable ATPγS on the formation of SsoRad54cd:DNA complexes and the DNA binding of SsoRad54cd_E563Q in the presence of ATP were tested. The rational behind testing the Walker B mutation in the presence of ATP was that it has been reported that the interaction of many ATPases with their substrates is not stabilized by nucleotide analogs, but in several ATPase systems, the Walker B box mutation was successfully used to obtain a stable trapped ATP-bound state (Hirano and Hirano, 2004; Lammens et al., 2004). In both cases, SsoRad54cd exhibits only a moderate increased DNA binding affinity (K_d = 0.08 $\pm$ 0.03 µM) in the presence of ATPγS or ATP, indicating that DNA recognition is not substantially influenced by ATP. The finding that ATP does not significantly affect DNA binding of SsoRad54cd is consistent with observation from eukaryotic Rad54 (Petukhova et al., 1999) and similar results were obtained for the SWI/SNF complex, where the addition of ATP did not show any effect in gel retardation assays or in footprinting reactions (Quinn et al., 1996).

ATP-dependent DNA translocation

Since the ATPase activity of SWI2/SNF2 enzymes is not coupled to helicase activity, one proposed model of how SWI2/SNF2 enzymes catalyze the rearrangements in diverse protein:DNA complexes is that these ATPases are ATP-dependent DNA tracking enzymes (Pazin and Kadonaga, 1997). In this hypothesis, the SWI2/SNF2 enzymes may be targeted to a specific protein:DNA complex and translocate along DNA in an ATP-dependent manner. This model is attractive because enzymes that catalyze directed progression along DNA can generate both motion and force, i.e. they are true biological motors. This can potentially explain how SWI2/SNF2 enzymes may disrupt or remodel protein:DNA contacts. First hints for DNA translocation by SWI2/SNF2 enzymes were derived from the DNA-length dependent stimulation of their ATPase activity. This DNA length dependence was interpreted that the ATP hydrolysis is increased during progression along DNA. A second approach, the triplex displacement assay has successfully been used to monitor the

motion of SWI2/SNF2 enzymes along DNA in a more direct manner. For instance, the ability to disrupt triplex DNA in an ATP dependent manner has been described for the RSC complex, the SWI/SNF complex, Sth1, ISWI, Mot1 and Rad54 (Auble and Steggerda, 1999; Ristic et al., 2001; Saha et al., 2002; Whitehouse et al., 2003). Triplex substrates are generated by annealing a tiplex forming oligonucleotide (TFO) to its target DNA duplex sequence (Firman and Szczelkun, 2000). The TFO interacts with the major groove of its target duplex sequence through the formation of a triple-helical structure mediated by Hoogsten base-pairing. Translocation of a protein along the dsDNA disrupts the triple-helical structure that can be detected as dissociation of the ^{32}P-labeled TFO from the DNA triplex.

To evaluate the ability of SsoRad54cd to track along dsDNA, the triplex displacement assay was used. Different amounts of protein were incubated with triplex substrate in the presence or absence of ATP. The results, summarized in figure 11, demonstrate that SsoRad54cd can efficiently displace TFO from the triplex substrate in an ATP dependent manner. In the absence of ATP or in the presence of the slow-hydrolysable ATP analog ATPγS no significant TFO displacement occurs, indicating that DNA binding alone is not sufficient for TFO displacement. Consistently, the SsoRad54cd_E563Q mutant that is not able to hydrolyse ATP failed to displace TFO.
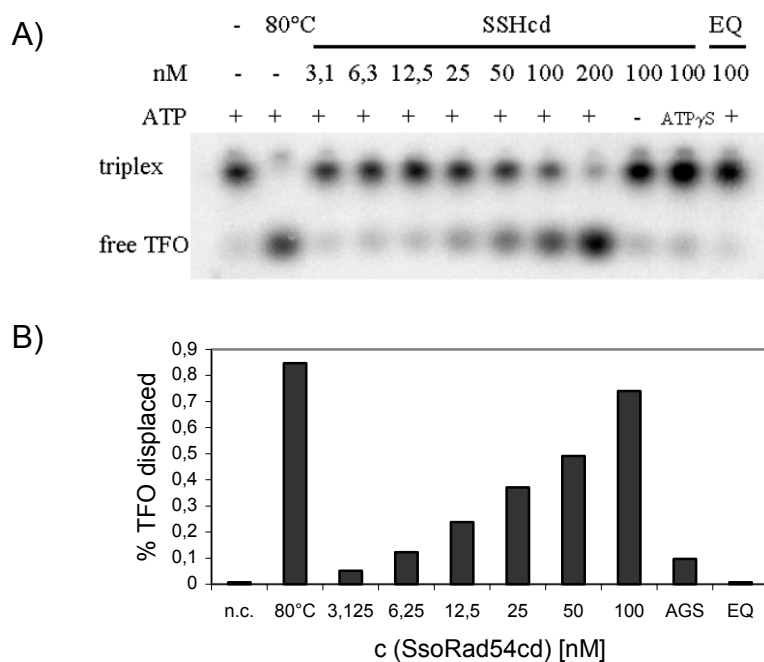
Figure 11: A.) Triplex displacement assay indicates an active translocation of SsoRad54cd on dsDNA. SsoRad54cd (indicated concentrations) can

displace triplex forming oligonucleotides (TFO) from plasmid DNA (triplex) in the presence of ATP, whereas in the absence of ATP, in the presence of the slow-hydrolysable ATP analog ATPγS or using SsoRad54cd_E563Q (EQ), no or weak displacement is detected. The upper band corresponds to the triplex structure (triplex), the lower band corresponds to the free TFO resulting form TFO displacement (free TFO). 80°C, control triplex disruption by heat. B.) Bar diagram presentation shows typical levels of triplex displacement (% TFO displacement) in the presence of ATP and indicated amounts of SsoRad54cd. Reactions contained 5 nM triplex DNA.

The activity of SsoRad54cd in the triplex displacement assay is only modestly reduced compared to yeast Rad54 (Jaskelioff et al., 2003) or other SWI2/SNF2 family enzymes. Some reduction may be due to the suboptimal temperature of 40°C for *Sulfolobus solfataricus* proteins (higher temperatures are incompatible with the assay). Noteworthy, these observations demonstrate that the ATPase domain itself is sufficient to efficiently translocate along DNA.

The ATP dependent TFO displacement by SWI2/SNF2 enzymes strongly argues for an active translocation along DNA. Alternatively, it is also possible that distortion of DNA by bending or twisting could account for triplex displacement without the requirement for DNA translocation. However, for instance in the case of Rad54, single-strand nicks containing substrates for the triplex displacement assay could not prevent triplex disruption suggesting that TFO displacement reflects translocation of Rad54 and is not due to generation of torsional stress. Similar findings were obtained with ISWI (Whitehouse et al., 2003). The displacement of triplex DNA by ISWI can be blocked by 5-10 bp gaps in the 3´-5´ DNA strand (but not in the 5´-3´ DNA strand) that is positioned between triplex and nucleosome, indicating that ISWI functions as a nucleosome-stimulated DNA translocase with specific polarity and translocates along DNA predominantly via contacts with the 3´-5´strand. These observations correspond well with the structural data obtained from the SsoRad54cd:DNA complex that suggest a putative 3´-5´ direction of translocation of SsoRad54cd on DNA (see below).

Although all these observations provide accumulating evidence for functional DNA translocation by SWI2/SNF2 enzymes, this is not universally accepted (Kassabov et al., 2003). In fact, both ATPase and triplex displacement assays do not support SWI2/SNF2 enzymes to be highly processive translocases and suggest a processivity limit of less than 100 bp.
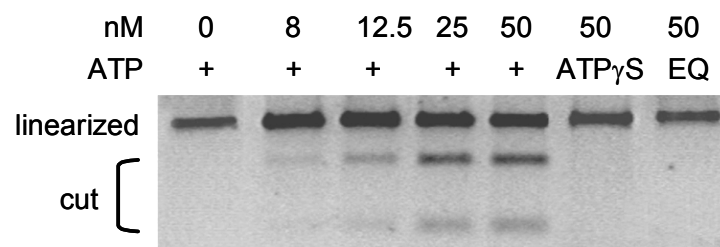
Generation of superhelical torsion

The ability to distort DNA in an ATP dependent manner is shared by many SWI2/SNF2 enzymes, including yeast SWI/SNF complex, *Xenopus* Mi-2 complex, ISWI, BRG1 or yeast Rad54 (Havas et al., 2000; Jaskelioff et al., 2003). This suggests that the generation of unconstrained negative superhelical torsion represents a primary biochemical activity shared by SWI2/SNF2-related ATPase motors. The generation of a superhelical torque provides a potent means by with SWI2/SNF2 proteins may manipulate protein:DNA interfaces.

Havas et al (Havas et al., 2000; Jaskelioff et al., 2003) developed a method that use cruciform extrusion from linear DNA as a means to measure superhelical tension in DNA. A cruciform structure is normally not stable in linear DNA fragments, since the formation of cruciform incurs a free energy cost. However, cruciform formation is also associated with a relaxation of negative superhelical density. This means that cruciform structures are stable in negatively supercoiled DNA, where the positive free energy of cruciform extrusion is compensated by the relaxation of superhelical density (Lilley, 1980). The pXG540 plasmid with an $[AT]_{34}$ inverted repeat has been used as cruciform forming reporter that has been shown to adopt cruciform structures at moderate levels of supercoiling and without a significant kinetic barrier (Greaves et al., 1985). Subsequently the cruciform structure can specifically detected by the structure specific T4 endonuclease VII (ENDOVII).

We tested the ability of SsoRad54cd to introduce superhelical torsion into linear DNA using this cruciform extrusion assay (Havas et al., 2000). Different amounts of SsoRad54cd were incubated with linearized pXG540 plasmid in the presence of ATP or ATPγS. Cruciform formation was detected by the structure-specific T4

endonuclease VII cleavage products (Fig. 12). SsoRad54cd is able to generate superhelical torsion into linear DNA, which leads to cruciform extrusion. Like the translocation activity, the extrusion of cruciforms is ATP dependent. In the absence of ATP, in the presence of the slow-hydrolysable ATP analog ATPγS or using SsoRad54cd_E563Q mutant no cruciform extrusion can be observed. These data indicate that SsoRad54cd can effectively alter DNA topology using the free energy from ATP hydrolysis.

Figure 12:   Cruciform extrusion assay indicates generation of superhelical torsion. In the presence of ATP, SsoRad54cd (indicated concentrations) can extrude cruciforms from linearized plasmid DNA, detected by structure-specific endonuclease cleavage products ("cut"). No activity is observed in the absence of ATP, in presence of ATPγS or using SsoRad54cd_E563Q (EQ).

The observation that SsoRad54cd is capable of altering superhelical twisting in linear DNA is somehow surprising. In linearized DNA, the superhelical torsion has to be restrained by the enzyme to maintain the superhelical tension and to support the stable existence of a cruciform. Havas et al demonstrated cruciform extrusion in the presence of large remodeling factors, which have multiple DNA binding sites. In these cases, the constrain may be achieved by loop formation as detected by electron microscopy of SWI/SNF complexes bound to DNA (Bazett-Jones et al., 1999). However, the data clearly show that the ATPase domain itself is sufficient to qualitatively generate superhelical torsion into DNA.

How can cruciform extrusion from linear DNA by SsoRad54cd be explained? To detect cruciform extrusion by SsoRad54cd a protein:pXG540 plasmid ratio of ≈ 10:1 is necessary. The activity of SsoRad54cd thereby is about 10 times lower

compared to SWI2/SNF2 complexes, where a protein:pXG540 plasmid ratio of $\approx$ 1:1 is sufficient (Havas et al., 2000; Jaskelioff et al., 2003). The tenfold difference in cruciform activity to large remodeling factors is probably due to their opportunity to restrain superhelical torsion by loop formation at multiple DNA binding sites. However, SsoRad54cd is comparable effective with Rad54. The generation of superhelical torsion by yeast Rad54 has been proposed to be a result of transient multimerization on DNA, paired with a dsDNA translocase activity (Ristic et al., 2001). Similarly, since it is unlikely that the relative small SsoRad54cd can bind large DNA loops at multiple DNA binding sites, the observed cruciform extrusion activity may be as well the result of transient multimerization in combination with an ATP-driven dsDNA translocation that could explain the higher required protein:DNA ratio.

The slightly larger amounts of SsoRad54cd required for cruciform extrusion compared to yeast Rad54 is probably due to the suboptimal temperature and pH. The optimal conditions for *Sulfolobus solfataricus* proteins are at 80°C and pH = 5-6. However, the cruciform extrusion activity cannot be performed at more than 40°C and pH = 6.5 due to ENDO VII inactivation and rapid background formation of cruciforms.

How SWI2/SNF2 enzymes, including SsoRad54cd introduces topological stress in DNA is unclear. It has been proposed that the ability to generate distortion of DNA involves rotation of the DNA that could be a result of ATP dependent translocation along the helical DNA backbone. The conjunction of DNA distortion and translocation is attractive, since even a relatively moderate translocation of DNA along the helical DNA backbone would include a substantial rotation of DNA along the helical axis. Alternatively, there is the possibility that SWI2/SNF2 enzymes stabilize cruciform structures rather than generate superhelical torsion by themselves and thereby contributing to cruciform formation. This idea raised from the fact that some SWI2/SNF2 enzymes preferably bind 4-way junction DNA. The DNA binding studies of SsoRad54cd does not support this idea, since SsoRad54cd does not show any preferences for 4-way junctions compared to dsDNA. Furthermore, in the absence of ATP, in the presence of ATPγS or using SsoRad54cd_E563Q mutant, no cruciform extrusion by SsoRad54cd was observed, indicating that DNA binding alone by SsoRad54cd is not sufficient. There is no lucid explanation, why cruciform

stabilisation by binding to 4-way junctions should be ATP dependent, particularly with regard to the fact that DNA binding is only moderately influenced in presence of ATP.

The findings described above demonstrates that SsoRad54cd recapitulates many of the conserved biochemical features of full remodeling factors including DNA stimulated ATPase activity, DNA translocation and DNA distortion activity. These activities have been suggested to account for the modulation of protein:DNA interfaces. The proved features of SsoRad54cd indicate that the conserved SWI2/SNF2 ATPase domain itself has the means to generate the torque in DNA that is used in remodeling processes. To gain insights into the mechanism of how SWI2/SNF2 ATPases use ATP hydrolysis for translocation and/or generation of superhelical torsion we determined the crystal structure of SsoRad54cd in absence and in presence of its dsDNA substrate.

Structural investigations of the catalytic domain of *Sulfolobus solfataricus* Rad54

In the last years significant progress has been made in analysing the biochemical properties of SWI2/SNF2 enzymes. Nevertheless, for a full understanding, the molecular mechanism of how these proteins remodel protein:DNA interfaces or how ATP hydrolysis is coupled with translocase and DNA distortion activity remains to be discovered. This circumstance suffers mainly from the lack of structural information. Recently, the crystal structure of the nucleosome recognition module of the remodeling factor ISWI has been solved (Grune et al., 2003) and provided the first detailed structural information about a SWI2/SNF2 enzyme. The structure comprises the C-terminal region outside the conserved ATPase domain including a SANT and SLIDE domain. These two domains are suggested two play a critical role in contacting and binding to the DNA/nucleosome substrate. However, so far no atomic structure of the catalytic ATPase domain of any SWI2/SNF2 enzyme is available. To reveal the structural basis for the ATPase core of SWI2/SNF2 enzymes, we determined the atomic structure of SsoRad54cd using X-ray crystallography.

Crystallisation and structure determination of SsoRad54cd

X-ray crystallography is currently the most powerful technique for determining structures of biological macromolecules at atomic resolution. For structure determination by X-ray diffraction experiments, high quality crystals are necessary that is still a major obstacle in protein crystallographic research.

Initial crystallisation experiments were performed by sitting drop vapor diffusion with purified SsoRad54cd at a concentration of 15 mg/ml in 20 mM Tris/HCl [pH = 7.5], 100 mM NaCl, 0.5 mM EDTA and 1 mM DTT using commercial available screens from Hampton Research and Jena Bioscience. Since most of the drops remained clear, the concentration was increased to 30 mg/ml. In only one crystallisation condition (2 M $(NH_4)H_2PO_4$, 100 mM Tris/HCl [pH=8.5]) 3 dimensional single crystals appeared after three days with orthorhombic morphology. The crystallisation condition was refined by intensive additive screening resulting in the final optimal

crystallisation condition of 2 M $(NH_4)H_2PO_4$, 100 mM Tris/HCl [pH=3.9], 100 mM NaCl, 50 mM sodium malonate and 5 % glycerol.
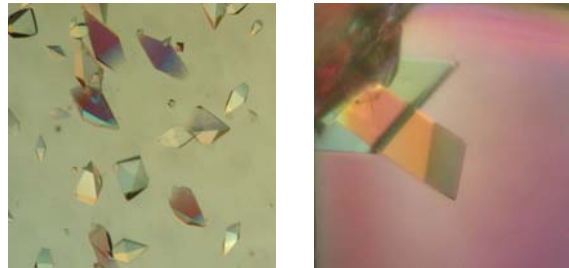
Figure 13:   Crystal images of SsoRad54cd and SsoRad54 C-terminal domain. Crystal of SsoRad54cd grew in 2 M $(NH_4)H_2PO_4$, 100 mM Tris/HCl [pH=3.9], 100 mM NaCl, 50 mM sodium malonate and 5 % glycerol. The crystal of SsoRad54 C-terminal domain appeared in 12% PEG 4000, 100 mM Tris/HCl [pH=7.5], 200 mM NaCl and 5 % ethylene glycol.

SsoRad54cd crystallized in space group $P4_12_12$ with one molecule per asymmetric unit and diffracted to 3.0 Å resolution. To improve the diffraction quality of the crystals, further SsoRad54cd constructs without a N-terminal 6x-His-Tag or variation of the N-terminal end were tested in crystallisation experiments, but no crystals were obtained. For X-ray diffraction experiments, the crystals were transferred into the refined crystallisation condition supplemented with 20 % glycerol and flash frozen. For phase determination by a two wavelength diffraction experiment, mercury derivatives of SsoRad54cd crystals were prepared by soaking the crystals in the appropriate 0.5 mM $HgCl_2$ containing crystallisation buffer for 2 hours, followed by backsoacking for 10 min and transferring into crystallisation buffer supplemented with 20 % glycerol as cryoprotectant.

The atomic structure of SsoRad54cd was determined by a two-wavelength anomalous dispersion experiment at Hg L_{III} edge at beamline ID14-4 (ESRF, Grenoble), using mercury as anomalous scatterer. Since the measurement of MAD datasets depends on accurate measurement of the datasets at accurate wavelengths, the optimal wavelengths were determined by fluorescence scan. Data were measured at two wavelengths from a single Hg derivatized crystal to minimize

systematic errors. The first dataset was collected at the peak wavelength (maximum f´´) at 1.00654 Å followed by the edge inflection point (minimum f´) at 1.0090 Å. The dataset for the peak wavelength was recorded redundantly to get a high overall completeness and by inverse beam geometry to make sure to obtain a most complete Bijvoet set for optimal phasing. We also measured at remote high energy wavelength, but because of crystal decay due to radiation damage the remote dataset was not included in data processing.
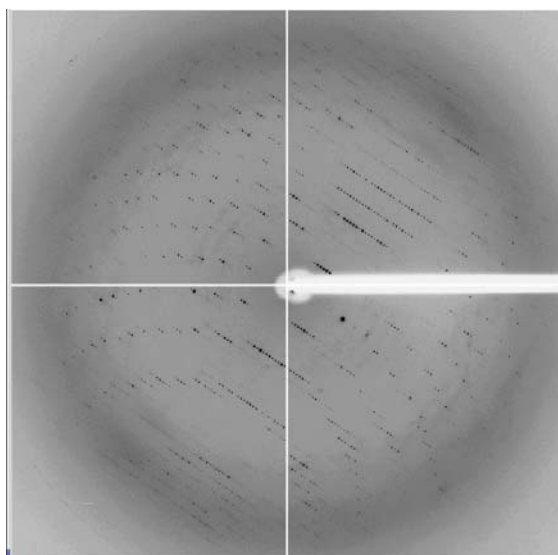
Figure 14:   Diffraction image of SsoRad54cd recorded at beamline ID14-4 (ESRF, Grenoble). Crystals of SsoRad54cd diffracted to 3 Å resolution.

The diffraction images showed defined spots and a moderately mosaicity of 0.6°. Datasets were collected to 3 Å resolution. The $HgCl_2$ soaked crystals even had a slightly better quality than the native crystals. Data collection revealed that the crystals were of space group $P4_12_12$, with unit-cell dimensions a = b = 84.1 Å, c = 227.2 Å and $\alpha = \beta = \gamma = 90°$. These unit-cell dimensions were consistent with the presence of one molecule per asymmetric unit. The native and heavy metal derivative crystal were isomorphous with variation of the cell dimensions smaller than 1%. The data were processed with XDS (Kabsch, 1993) and the three mercury sites could be located by direct methods using SnB (Smith et al., 1998). Phases were calculated with SHARP (Global Phasing) and improved with SOLOMON (Collaborative Computational Project, 1994), resulting in an interpretable electron density.
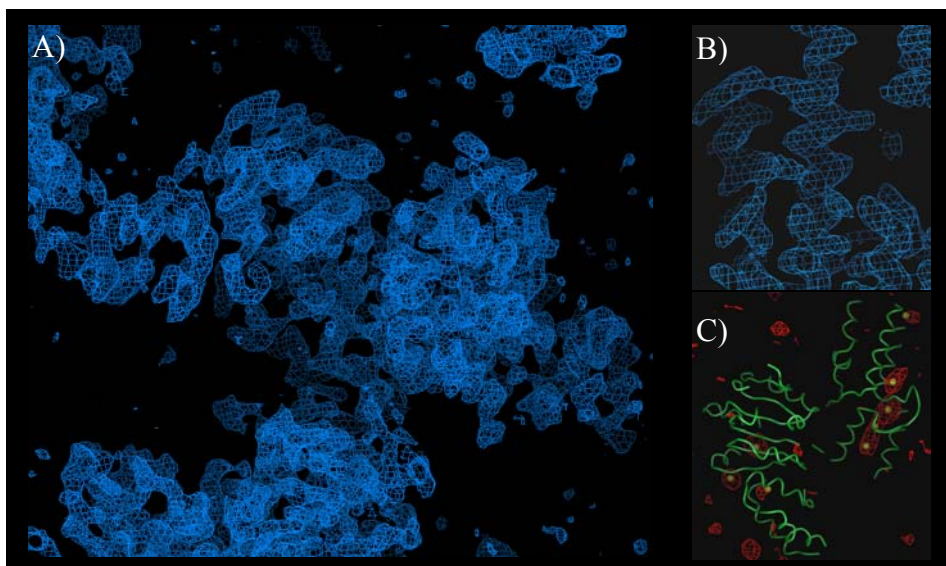
Figure 15:   A) 1σ contoured multiple anomalous dispersion (MAD) map of SsoRad54cd at 3 Å resolution. B) Zoomed section of the electron density map showing a α-helix. Note that the sidechains are difficult to identify. C) Initial model of SsoRad54cd with superimposed selenomethionines (yellow balls) and the corresponding difference 2 σ Fourier map (red) as obtained from a SAD experiment with selenomethionine substituted SsoRad54cd.

However, the phases could only be obtained to 3.5 Å resolution. The electron density map clearly defined the backbone and secondary structures, but it was difficult to impossible to identify the sidechains. Therefore we built an initial poly alanine model into the experimental electron density. The electron density was improved by phase combination using the partial model. Subsequently, the improved density was used to trace the peptide sequence starting from the mercury binding sites. At this stage we also used a dataset of a selenomethionine substituted SsoRad54cd crystal that was recorded at the Se L$_{III}$ edge. Although these crystals diffracted only to 6 Å resolution, we were able to locate the selenomethionine peaks by phase combination and anomalous difference Fourier analysis. The obtained selenomethionine sites were superimposed with our initial atomic model to verify the putative methionine positions (Fig. 15 C). Finally, the experimental electron density allowed tracing of most of the polypeptide chain except a large portion at the C-terminus. The preliminary model suggested that SsoRad54cd consists of two domains. Using the

structural information, the boundaries of the two domains could be exactly identified. In order to complete tracing the polypeptide chain for the entire SsoRad54cd, I determined the crystal structure of the C-terminal domain of SsoRad54cd (residues 663-906) separately.

Crystallisation and structure determination of the C-terminal domain of SsoRad54cd

For complete structure determination of SsoRad54cd we also crystallized separately expressed domain 2 of SsoRad54cd (residues 663-906) that was defined by the preliminary SsoRad54cd model. Unexpectedly, the purified domain 2 (25 mg/ml in 20 mM Tris/HCl [pH=7.5], 150 mM NaCl, 1 mM EDTA and 1 mM DTT) crystallized already in the reaction tube that was stored in the fridge (4°C) overnight, resulting in crystals of up to 3 mm in size. For diffraction experiments, however, domain 2 of SsoRad54cd was crystallized in sitting drop vapor diffusion by mixing 2 µl protein (20 mg/ml in 20 mM Tris/HCl [pH=7.5], 150 mM NaCl, 1 mM EDTA and 1 mM DTT) and 2 µl precipitant (12% PEG 4000, 100 mM Tris/HCl [pH=7.5], 200 mM NaCl, 5 % ethylene glycol). For X-ray diffraction experiments, the crystals were transferred into crystallisation condition supplemented with 20 % glycerol and flash frozen.

A dataset of domain 2 of SsoRad54cd were collected at PX-beamline (SLS, Villingen) with a resolution of 1.9 Å. Domain 2 of SsoRad54cd crystallized in space group C2 with two molecules in the asymmetric unit. The data were processed with Denzo and Scalepack (HKL Research, Inc) and phases were calculated by molecular replacement with AMoRe (Collaborative Computational Project, 1994) using the partial model of the C-terminal domain of SsoRad54cd as search molecule. The molecular replacement resulted in a single solution of the rotational and translational search. After bulk-solvent and overall anisotropic B-factor correction, we refined the model by repeated cycles of positional minimization and restrained B-factor refinement with CNS (Brunger et al., 1998) and manual model building with MAIN (Turck, 1992) (for refinement statistics, see crystallographic table, p. 61).

The 2.0 Å structure of the C-terminal domain was used to completely trace and refine the model of SsoRad54cd. After bulk solvent and overall anisotropic B-factor correction, we refined the model of SsoRad54cd by repeated cycles of positional minimization and restrained B-factor refinement with CNS (Brunger et al., 1998) and manual model building with MAIN (Turck, 1992). The refinement against 3.0 Å resolution resulted in an R-factor of 24.2 % (R_{free} = 29.6 %) with good stereochemistry (see crystallographic table p. 61).

Crystallisation of the SsoRad54cd:DNA complex

For a full understanding of how SWI2/SNF2 enzymes modulate their protein:DNA substrates, it is important to examine the structural interplay between the enzymes and its DNA substrate. In order to get structural information of how SsoRad54cd interact with DNA and how ATP hydrolysis is coupled with DNA translocation and distortion, I tried to crystallize SsoRad54cd in complex with its dsDNA substrate. SWI2/SNF2 enzymes bind duplex DNA in a predominantly sequence and structure independent manner. This property provides a formidable obstacle for crystallization of a specific protein:DNA complex. To overcome this problem, I used the strategy that employs DNA with a more flexible internal region combined with stiffer ends. The concept of sequence-directed structural softness or flexibility has been introduced some years ago and may provide a physical basis for protein:DNA recognition (Harrington and Winicov, 1994). The rational behind this was that a slightly twisting, bending or deformation of the DNA for optimal DNA binding by SsoRad54cd may be facilitated by a more flexible region that can serve as an anchor point. Using double-stranded DNA with a flexible central part and more inflexible ends may mimic a central "flexible wedge" that may promote specific DNA binding for crystallisation. Analysis of sequence-dependent dynamics in duplex DNA revealed, that AAAAAA tracts sequences have an increased stiffness compared to non-A tract sequences (Dlakic and Harrington, 1995). A number of specific sequences, in particular CA and TA dinucleotides, have been suggested to be more flexible. For crystallisation of a specific SsoRad54cd:DNA complex, I used dsDNA with a 12 bp internal sequence of predicted increased flexibility in combination with terminal poly(dA) runs of different length. DNA length is an important variable in crystallization of a protein:DNA

complex. Thus we used different sets of oligonucleotides ranging from 17 to 28 basepairs. For initial cocrystallisation experiments we used increments of 3-4 bp, followed by smaller steps for refinement. Another important variable for crystallisation of a protein:DNA complex is the design of the DNA ends. In many crystals of protein:DNA complexes, it has been observed that the DNA ends are involved in crystal packing by contacting another DNA end. In order to promote such crystal contacts, I used a 5´-dA overhang combined with a 3´-dT overhang.

oligonucleotides	length	sequence
DNA 1	17 nt	5´- CCTTTTAGTTTATTGGG -3´
DNA 2	17 nt	5´- CCCAATAAACTAAAAGG -3´
DNA 3	21 nt	5´- AAAAAACATTAGACGAAAAAA -3´
DNA 4	21 nt	5´- TTTTTTTCGTCTAATGTTTTT -3´
DNA 5	23 nt	5´- AAAAATTGCCGAAGACGAAAAAA -3´
DNA 6	23 nt	5´- TTTTTTTCGTCTTCGGCAATTTT -3´
DNA 7	25 nt	5´- AAAAAAATTGCCGAAGACGAAAAAA -3´
DNA 8	25 nt	5´- TTTTTTTCGTCTTCGGCAATTTTTT -3´
DNA 9	26 nt	5´- AAAAAAAATTGCCGAAGACGAAAAAA -3´
DNA 10	26 nt	5´- TTTTTTTCGTCTTCGGCAATTTTTTT -3´
DNA 11	26 nt	5´- GAAAAAAATTGCCGAAGACGAAAAAG -3´
DNA 12	26 nt	5´- CTTTTTCGTCTTCGGCAATTTTTTTC -3´
DNA 13	27 nt	5´- AAAAAAAATTGCCGAAGACGAAAAAAA -3´
DNA 14	27 nt	5´- TTTTTTTTCGTCTTCGGCAATTTTTTT -3´
DNA 15	28 nt	5´- AAAAAAATTTCGCCGAAGACGAAAAAA -3´
DNA 16	28 nt	5´- TTTTTTTCGTCTTCGGCGAAATTTTTT -3´

Table 4: List of the various DNA oligonucleotides used in the cocrystallisation experiments with SsoRad54cd. DNA1 and DNA2 are complementary and forming a pair, likewise DNA 3 and 4, DNA 5 and 6 and so on. The resulting duplex DNA has single nucleotide overhangs and a central region of predicted increased flexibility. Best quality crystals were obtained using the 25 mer duplex DNA (DNA 7/8).

The double-stranded DNA for cocrystallisation was generated by mixing the two corresponding oligonucleotide in a molar ratio of exactly 1:1, heating up to 90°C for 5 minutes followed by slow cooling to 4°C. The binary protein:DNA complex was produced by incubating SsoRad54cd (20 mg/ml in 20 mM Tris/HCl [pH = 7.5], 150 mM NaCl, 0.5 mM EDTA and 1 mM DTT) with the duplex DNA substrate at a molar ratio of 1:1.2 for one hour on ice. Before adding the DNA to the protein solution, the pH of the DNA solution was adjusted to 7.5 to prevent protein denaturation due to acidification. Without buffering of the DNA solution, protein precipitation was observed upon DNA addition.

Crystallization experiments for SsoRad54cd:DNA complex were performed by hanging or sitting drop vapor diffusion methods at 20°C and mixing 1 µl protein with 1 µl of crystallisation buffer. Using the initial collection of DNA, 2 dimensional plate-like crystals were obtained in the presence of the heteroduplex DNA 7/8 in 15% PEG 3350 and 0.1 M magnesium formate. The first crystals obtained were strongly intergrown and had a size of 100 x 100 x 5 µm. The crystals grew out of a smeary protein gel at the well bottom that probably served as nucleation points.
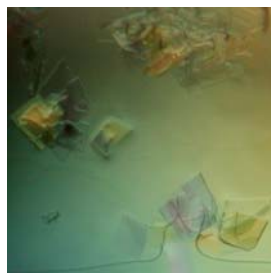
Figure 16:   Crystals of SsoRad54cd:DNA complex. The crystals grew in sitting drop vapor diffusion in 15% PEG 3350, 0.1 M magnesium formate and 20 % glycerol.

Extensive search for crystallisation conditions that favour the formation of single 3 dimensional crystals were performed. In addition to variation of additives and pH, refined DNA substrates were used. Although all attempts failed to produce 3 dimensional single crystals, yet the crystals could be improved. The addition of glycerol resulted in crystals that were still growing from the bottom sharing a

crystallisation starting point, but they were less intergrown and exhibit dimensions of 400 x 200 x 20 µm. From these crystals single crystal plates could be abandoned that diffracted to 3 Å resolution. The SsoRad54cd:DNA complex crystals belong to space group P2$_1$ with unit cell dimension of a = 86.93 Å, b = 83.74 Å, c= 106.39 Å, α = 90°, β = 109,78°, γ = 90°.

Structure determination of SsoRad54cd in complex with its dsDNA substrate

SsoRad54cd:DNA complex crystallized in space group P2$_1$ with two molecules in the asymmetric unit and diffracted to 3.0 Å resolution at the PX beamline (SLS, Villingen). Data were processed with XDS and the structure solved by molecular replacement with AMoRe, using domain 1 (SsoRad54cd 444-657) and domain 2 (SsoRad54cd 663-906) of the uncomplexed structure as search model. This strategy already proved to be successful in the case of the PcrA:DNA helicase structure (Velankar et al., 1999). A single molecular replacement solution was refined at 3.0 Å resolution, using the procedure described above. To exclude model bias, the replacement solution was verified using a 5.5 Å resolution dataset of selenomethionine containing crystals. Anomalous difference Fourier analysis with model derived phases verified the predicted position of the selenium atoms. Although the two molecules in the asymmetric unit have no significant differences we applied strict non-crystallographic symmetry constraints throughout refinement. The SsoRad54cd:DNA complex was refined at 3.0 Å resolution, resulting in a R-factor of 23.3% (R$_{free}$=28.1%) and good stereochemistry. The final model comprises the conserved SWI2/SNF2 type catalytic domain of SsoRad54 (residues 432-906) and 19 base pairs DNA duplex for each of the two molecules in the asymmetric unit. The electron density of the nucleotides corresponding to the DNA along the entire length of the protein plus one end of the DNA duplex that is stabilized by crystal lattice contacts were clearly evident even in the early stages of refinement. We could unambiguously trace the DNA sequence due to well ordered head-to-head crystal lattice contacts between two DNA molecules. The six terminal base pairs at the other duplex end, which is not stabilized by protein or crystal lattice contacts, are not visible in the electron density, presumably due to increases mobility.

Data Collection	SsoRad54cd[a]		SsoRad54cd:dsDNA[b]	Domain 2[c]
Data set	peak	inflection	native	native
X-ray source	ID14-4 (ESRF)	ID14-4 (ESRF)	PX (SLS)	PX (SLS)
Derivative	Hg	Hg	-	-
Wavelength (Å)	1.0065	1.0090	0.9780	1.0000
Data range (Å)	20-3.0	20-3.0	20-3.0	20-2.0
Observations (unique)	137929 (30251)	128376 (28997)	113578 (52560)	102837 (36943)
I/σ (last shell)	15.56 (3.18)	18.6 (3.05)	6.59 (1.93)	14.6 (2.7)
Completeness (%) (last shell)	97.7[d] (92.7[d])	97.7[d] (92.3[d])	93.1 (72.1)	95.6 (75.3)
R_{sym}[e] (last shell)	0.06 (0.39)	0.06 (0.46)	0.12 (0.44)	0.042 (0.277)

Refinement	SsoRad54cd	SsoRad54cd:dsDNA	Domain 2
Data range (Å)	20-3.0	20-3.0	20-2.0
Reflections F>0	16322	28481	32479
(crossvalidation)	(802)	(1401)	(1634)
Nonhydrogen atoms	3873	9198	3917
(nucleotides/solvent)	(-)	(78 nucleotides)	(260 water molecules)
R_{work}[f] (R_{free}[g])	0.242 (0.296)	0.233 (0.281)	0.219 (0.260)
RMS bond length (Å)	0.008	0.011	0.013
(bond angles)	(1.7)	(1.6)	(1.7)

[a]SsoRad54cd crystal cell constants (P4$_1$2$_1$2)(Å): a = 84.1, b = 84.1, c = 227.2; α = 90°, β = 90°, γ = 90°

[b]SsoRad54cd:DNA crystal cell constants (P2$_1$)(Å): a = 86.9, b = 83.7, c = 106.4; α = 90°, β = 109.8°, γ = 90° and two molecules per asymmetric unit.

[c]Domain 2 crystal cell constants (C2)(Å): a = 111.8, b = 61.1, c = 75.0; α = 90°, β=99.3°, γ = 90° and two molecules per asymmetric unit.

[d]Anomalous completeness

[e]R_{sym} is the unweigthed R value on I between symmetry mates

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

[g]R_{free} = the crossvalidation R factor for 5% of reflections against which the model was not refined.

**Table 4:** Crystallographic data collection and Refinement statistics

Structure of SsoRad54cd

SsoRad54cd consists of two structural domains (domain 1 and 2) that together form a χ-shaped structure of 70 Å x 70 Å x 40 Å dimension (Fig. 17 A). Domain 1 and domain 2 each contain a "RecA" type α/β subdomain (denoted 1A and 2A, respectively) with a central β-sheet, flanked by α-helices (Fig. 17 C). The seven conserved sequence motifs that are implicated in ATP-hydrolysis (I,II,III,VI) or DNA binding (Ia,IV,V) of helicases (Caruthers and McKay, 2002; Velankar et al., 1999) are located in loop regions of subdomains 1A (motifs I, Ia, II, III) and 2A (motifs IV, V, VI) (Fig. 18). The core architecture of SsoRad54cd is related to that found in helicase structures (Singleton et al., 2001; Subramanya et al., 1996), suggesting that SWI2/SNF2 enzymes and DExx box helicases share the basic ATP-hydrolysis mechanism.

Subdomains 1B and 2B (Fig. 17 A) are structurally not closely related to equivalent regions of helicases and could help to translate ATP-driven conformational changes between 1A and 2A into DNA distortion and translocation. Subdomain 1B extends the sheet of domain 1 by contributing a seventh β-strand and a helix-loop-helix cap. Subdomain 2B is a six membered helix bundle structure that is situated on the tip of sheet 2. It consists of a helical protrusion, originating from the central sheet of domain 2 that packs against a helical bundle, formed by the C-terminus of SsoRad54.

A deep cleft separates domain 1 and 2 (Fig. 17 A, 18). Motifs I (Walker A), II (DExx) and III on domain 1 of SsoRad54cd form the primary ATP-binding site in the active site cleft. Motifs IV-VI on domain 2, however, are not situated in the active site cleft, as observed for helicases (Subramanya et al., 1996), but positioned on the outside of SsoRad54cd. This location is due to a ~180° flip of domain 2, compared to the orientation of equivalent structural regions in crystal structures of ATP bound helicases. The unusual position of domain 2 suggests that SsoRad54cd probably undergoes a large conformational change in response to ATP-binding, which might be part of the power stroke of SWI2/SNF2 enzymes. The two domains are linked over αK, which packs in between the two domains. The αK helix, as well as the

β strand before and after (β7, β8) are conserved, indicating that SWI2/SNF2 enzymes have similar structural coupling of the two domains.
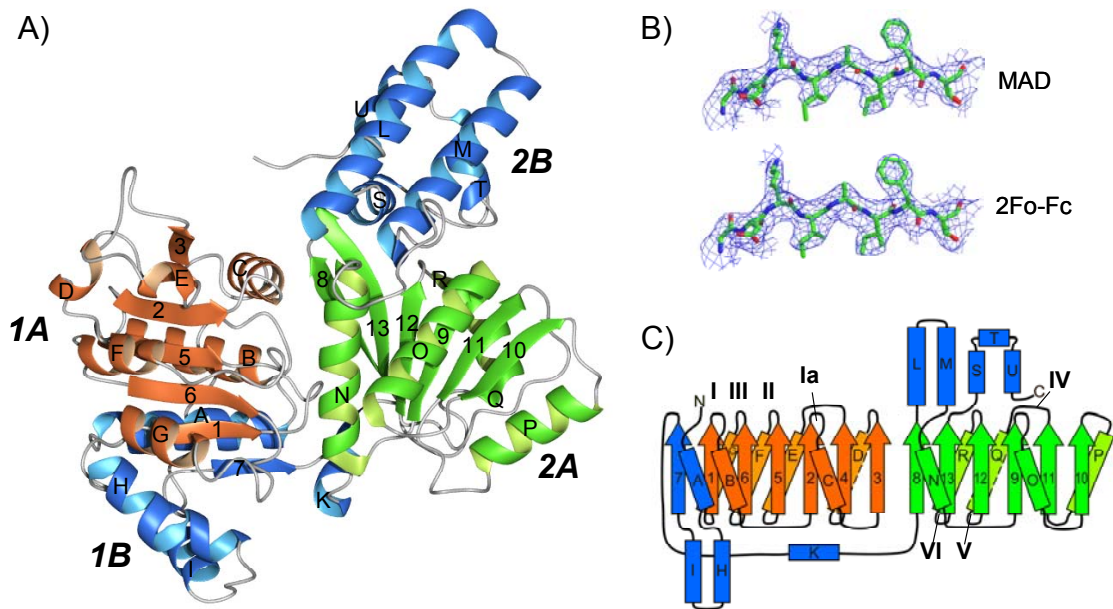
Figure 17: A) Ribbon presentation of the crystal structure of SsoRad54cd in the absence of DNA with annotated secondary structure. SsoRad54cd consists of two RecA-like subdomains (1A, orange, and 2A, green) that harbour the ATPase active site in the domain interface. Two additional SWI2/SNF2 specific subdomains (1B and 2B, blue) are attached to the core subdomains. B) Representative portion of 1σ contoured multiple anomalous dispersion (MAD) and $2F_0-F_c$ electron density maps of SsoRad54cd. C) Topology diagram of SsoRad54cd with secondary structure annotation and depicted conserved helicase motifs.

Structure of SsoRad54cd:DNA complex

The overall fold of the protein and the orientation of the two domains are very similar as that described for the DNA unbound SsoRad54cd structure (Fig. 17, 18). Only minor rearrangements are observed. The DNA duplex binds alongside the entrance of the active site cleft, in a position where ATP-driven changes between domain 1 and 2 could directly be linked to translocase or DNA distortion activity (Fig. 18). Both strands of the visible DNA duplex are fully base paired to each other, suggesting that SsoRad54cd and other SWI2/SNF2 enzymes translocate along duplex DNA without strand separation activity. The conformation of the structurally defined base pairs is most similar to B-DNA, indicating that DNA binding alone is not sufficient to significantly distort DNA. Thus, alteration of the DNA structure by SWI2/SNF2 enzymes might require ATP-driven conformational changes in the catalytic domain. For instance, the structures of PcrA helicase in two different DNA complexes revealed that binding and hydrolysis of ATP results in a large change in the relative orientation of the two RecA-like domains.
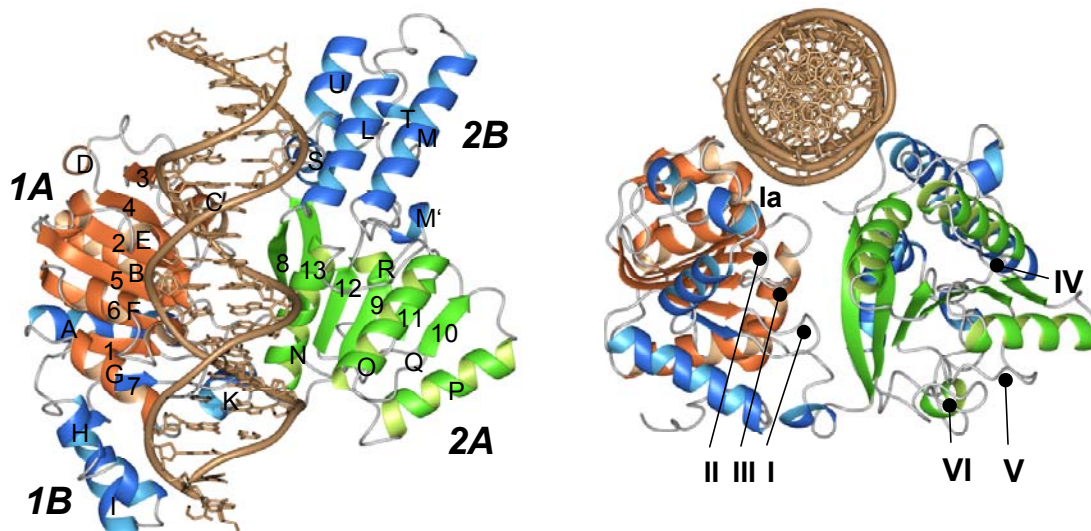

Figure 18: (Left panel) Top view of the ribbon representation of SsoRad54cd:DNA complex with annotated secondary structure. (Right panel) Side view along the active site cleft with annotated functional motifs. Sequence motifs I-VI are located in loop regions of subdomain 1A and 2A. DNA (brown ribbon/stick model) binds to domain 1 along the domain 1:2 interface. Colour code of figure 17.

A closer inspection of the crystal packing environment

SsoRad54cd binds DNA in a predominantly sequence and structure independent manner that makes the generation and crystallisation of a specific protein:DNA complex more difficult. The strategies I used to overcome that problem are discussed in chapter "Crystallisation of the SsoRad54cd:DNA complex" (p. 57). In fact, we observe a slight widening of the minor groove at the contact region and a slight bending of the DNA. Both distortions, although small, may have helped to place the protein onto the DNA in a more homogenous fashion

One further interesting point to note are unusual head-to-head crystal lattice contacts between two DNA molecules. In the SsoRad54cd:DNA crystals the nucleotide overhangs of one DNA molecule do not pair ordinary with the appropriate nucleotide overhang of the other DNA molecule, but form a triple helix interface by utilizing base stacking and extrahelical hydrogen bonds of the T-overhang to the symmetry related DNA molecule (Fig. 19). Remaining crystal lattice contacts are mediated by protein:protein interactions that are typical of those expected between molecules in a crystal lattice.
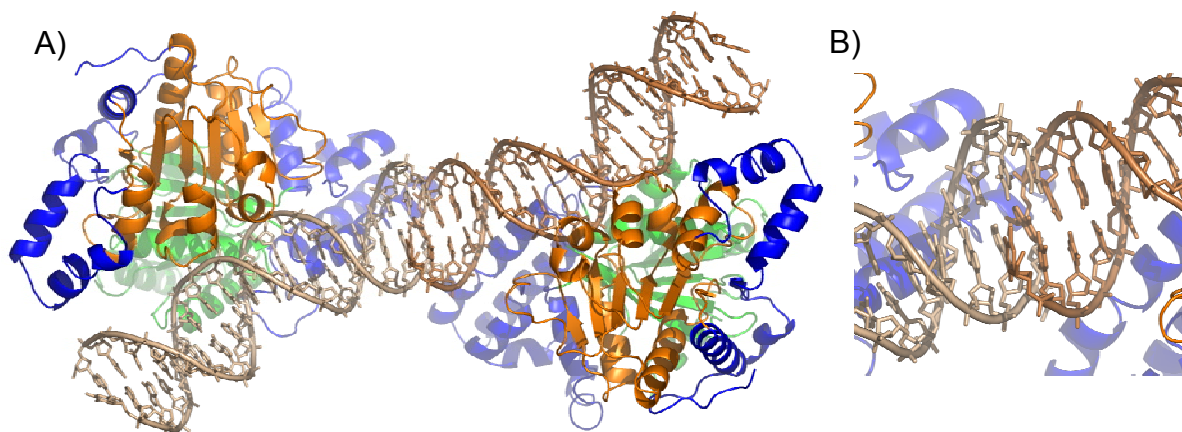

Figure 19:   A) Crystal lattice contacts in SsoRad54cd:DNA complex were to some extend mediated by unusual head-to-head crystal contacts between two DNA molecules which form a triple helix interface. B) Zoomed region of the triple helix interface. (Colour code of figure 17)

The unusual orientation of domains 2 in respect to domain 1 raised the concern that this might be influenced by crystallisation artefacts. As with any crystal structure, it formally cannot be excluded that the orientation of domain 2 is to some extend influenced by crystal packing. However, the crystallisation conditions and space groups of SsoRad54cd in the absence and in complex with DNA are completely different. Furthermore, we checked if the crystals of SsoRad54cd:DNA complex and SsoRad54cd share crystals contacts. Careful examination of the pattern of intermolecular crystal contacts revealed that both crystal forms are unrelated, apart from one shared crystal contact. This crystal contact amounts to only 16 % of the total surface area involved in crystal lattice contacts. In addition, this crystal contact is about 3 times smaller than the crystallographically observed interface between domain 1 and 2, arguing that this crystal contact is probably too small to energetically drive formation of an unphysiological domain 1 : domain 2 interface. In addition, mutational analysis demonstrates that a point mutation in the crystallographically observed interface (V850G) interferes with SsoRad54cd function (p. 84) and supports the functional relevance of the observed conformation. In combination, this suggests that the conformation of the SsoRad54cd in the DNA complex may display a functional role and might not be due to a crystallisation artefact.

DNA interaction with SsoRad54cd

39 out of 50 nucleotides of the DNA duplex are visible in the electron density (Fig. 20). DNA duplex is bound predominantly to subdomain 1A by recognition of the two phosphate chains along the minor groove (Fig.20), consistent with an enzyme that translocates along DNA in a sequence independent manner. The 3'-5' strand (named according to the putative direction of translocation of SsoRad54cd on DNA) is recognized by the loop between $\beta2$ and αC (motif Ia, Fig. 6, 20) and the loop between $\beta3$ and αD (denoted motif Ib). The 5'-3' strand is bound by a loop-helix motif consisting of the loop between motif II and αF, and the N-terminus of αF. Finally, αE (denoted motif Ic) binds into the minor groove via R547 (Fig. 20 A, C). R547 is highly conserved among SWI2/SNF2 enzymes. Its long side chain could serve as a flexible anchor for dsDNA yet allow DNA sliding during translocation.
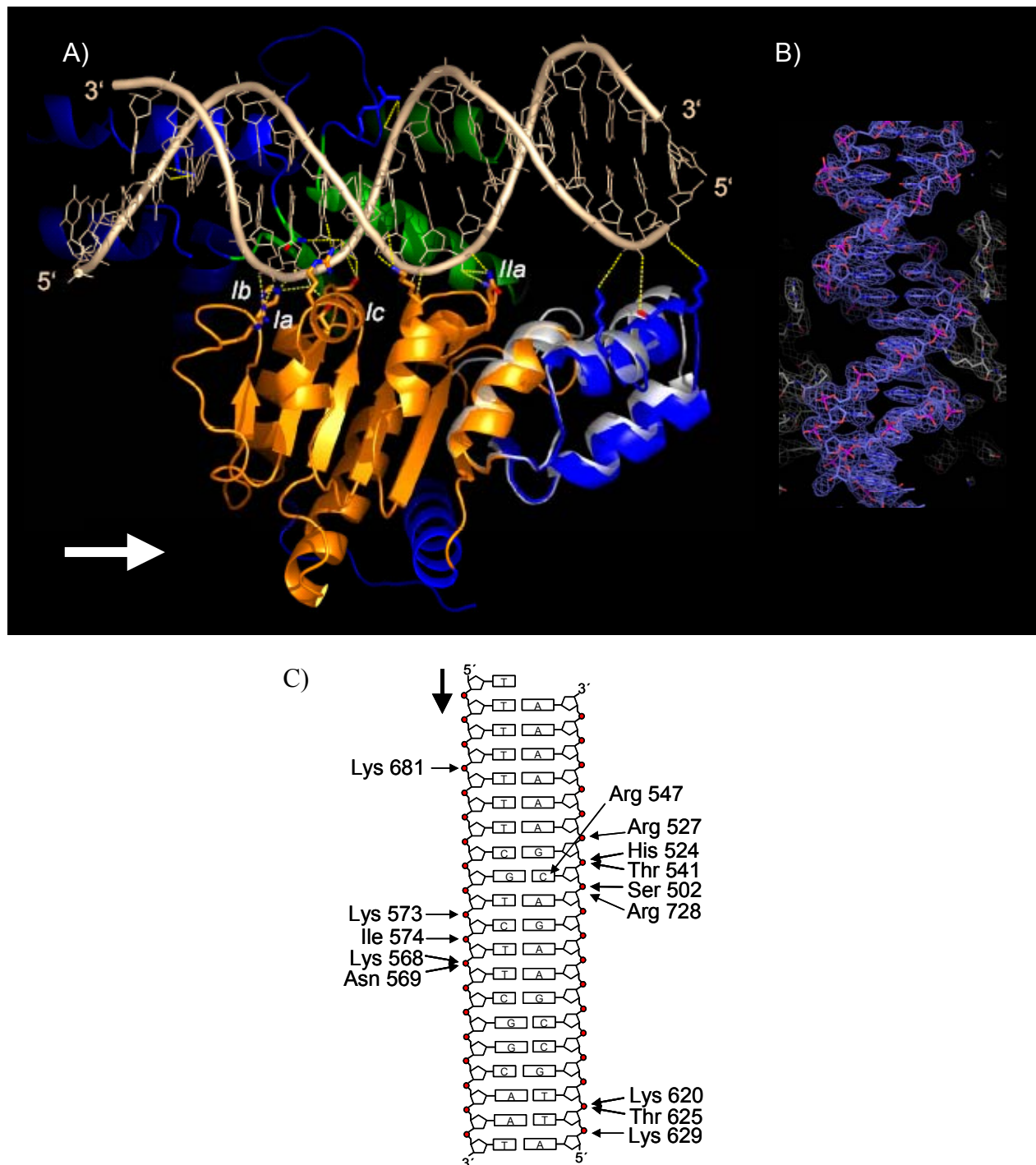
Figure 20: A) Ribbon presentation of the SsoRad54cd:DNA complex with highlighted protein:DNA contacts (colour code of Figure 17). Both DNA strands are recognized by several motifs on domain 1 (dashed yellow lines). Subdomain 1B is flexible (blue, DNA complex; grey, DNA free crystal structure), suggesting that it may also bind to DNA after slight rearrangements. The arrow indicates the putative direction of

translocation of the enzyme on DNA. B) Final 1σ contoured $2F_0$ - F_c electron density map at the protein:DNA interface. C) Sequence of the structurally defined DNA duplex and its recognition by protein residues. The arrow indicates the putative direction of translocation of the enzyme on DNA.

A potential third interaction site of SsoRad54cd with DNA is mediated by several lysine and threonine residues (K621, T626, K628, K629) in subdomain 1B. As judged by superposition of DNA-bound and DNA-free crystal structures, this region possesses enough conformational flexibility to participate in DNA binding. Domain 2 binds to DNA via R728 and K781. This interaction appears to be much weaker and less conserved, indicating that domain 2 does not have a high affinity DNA binding site. These structural observations are consistent with the DNA binding studies described above.

The binding mode of SsoRad54cd is consistent to DNA binding studies with the SWI/SNF complex. Some years ago, Quinn et al. already suggested that the yeast SWI/SNF complex interacts (at least in parts) with the minor groove of the DNA helix since the minor-groove binding drugs distamycin A and chromomycin A3 competed with the SWI/SNF complex for DNA binding (Quinn et al., 1996). The recognition of the phosphate backbone is also consistent with the finding that SWI2/SNF2 enzymes do not appear to recognize a consensus sequence. In addition, the structure of SsoRad54cd:DNA complex shows that SsoRad54cd spans a DNA length of 12-15 base pairs as estimated from the distance between the edge of SsoRad54cd and the full dsDNA binding site. This suggest, that remodeling factors require a minimal handle of at least 12-15 base pairs adjacent to the remodeled substrate in order to allow the ATPase domain to grab DNA. The estimated 12-15 bp correspond well to experimental values of a minimal 17 bp dsDNA handle required for Mot1 to remove TBP form TATA box (Darst et al., 2001) and the minimal 20 bp extranucleosomal DNA required by ISW2 (Zofall et al., 2004). Concordantly, ISWI or Sth1 require a minimal DNA length of 15-23 bp to trigger ATPase activity.

DNA stimulated ATPase activity

SsoRad54cd exhibits dsDNA activated ATPase activity, similar to Rad54 and other full-length SWI2/SNF2 enzymes (Laurent et al., 1993). Comparison of the structure of the SsoRad54cd:DNA complex with the structure of DNA free SsoRad54cd reveals a potential mechanism for the stimulation of the ATPase activity of SWI2/SNF2 enzymes by DNA, or at least may account for parts of the DNA dependent ATPase activation.
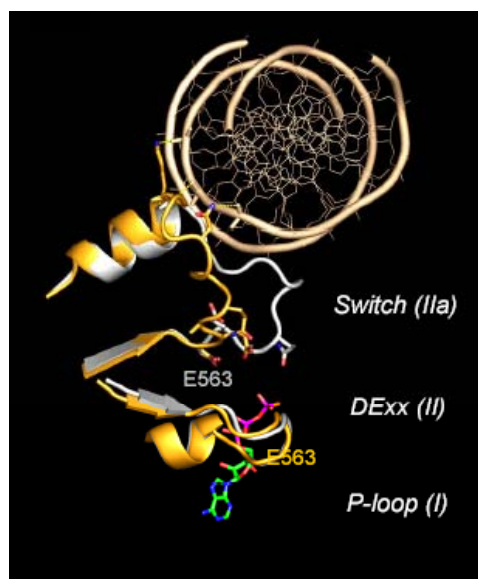
Figure 21:   Superposition of DNA bound (colour coded) and DNA-free (grey) crystal structures of SsoRad54cd shows that the loop (motif IIa) following the DExx motif switches in response to DNA binding. This rearrangement optimally positions E563 in the DExx motif for ATP hydrolysis. A putative ATP molecule (colour coded stick model) is modeled based on its location in related P loop enzymes.

In the absence of DNA, the ATP-hydrolyzing DExx motif (motif II) is in an unusual β-conformation, which is not compatible with a functional active site of Walker box ATPases (Fig. 21). In the presence of DNA, the DExx motif switches into the typical active α-conformation (Subramanya et al., 1996). Hereby, E563 flips 180° into a position where it can participate in polarizing the attacking water during ATP-hydrolysis. This structural switch is induced by direct interaction of a conserved loop (motif IIa, Fig. 20) behind the DExx motif with the phosphate backbone of the 5'-3'

strand. The structural switch could explain why SWI2/SNF2 ATPases are activated by DNA. The switch loop (motif IIa) is conserved in SWI2/SNF2 enzymes (Fig. 6), indicating that the switch loop might be relevant not only for SsoRad54cd function but also for other SWI2/SNF2 enzymes. In addition, a mutation in one of the conserved residues (N569I) reduces both DNA stimulated ATPase activity and DNA binding activity of SsoRad54cd (see below). Nevertheless, other structural changes may also contribute to DNA dependent ATPase stimulation.

Comparison with DExx box helicases

SsoRad54cd, like other SWI2/SNF2 ATPases, is related to DExx box helicases through the presence of seven conserved motifs. High resolution structures of a number of helicases have been determined by X-ray crystallography. These include for instance superfamily 1 helicases *Bacillus stearothermophilus* PcrA, *Thermus thermophilus* UvrB, and superfamily 2 helicases *Thermotoga maritima* RecG or hepatitis C virus NS3 (Kim et al., 1998; Velankar et al., 1999). The structures of these helicases revealed that SF1 and SF2 helicases contain a core of two RecA-like domains. The conserved helicase motifs are located on these two core domains and are closely associated in the tertiary structure of the enzyme. The seven helicase motifs together form the DNA and ATP binding sites.

While the ATPase core is conserved, the N- and C- terminal regions are characterized by a high degree of sequence and length variability, suggesting that these divergent regions are responsible for the individual functions of the helicases. The helicase structures suggest that helicases consist of duplex nucleic acid destabilization domains coupled with a translocase motor unit (Soultanas et al., 2000). For instance, RecG is composed of two principal structural parts. Firstly, a DNA translocase domain, which contains two RecA-like motifs "sandwiching" a nucleotide binding site. Mechanistically it is proposed that ATP binding and hydrolysis causes a structural transition between the RecA domains (rigid body motion) that allows the translocase motor to progress along duplex DNA. During a catalytic cycle, spatial arrangements between the domains by up to 120° lead to an repetitive alternation between open and close conformations (Singleton and Wigley,

2002). This translocation drives the movement of the second "wedge" domain along the arms of the forked DNA. The motion of the enzyme along DNA is thought to extrude DNA through the wedge domain, forcing the duplex apart. The two-step division suggests that strand separation activity can be uncoupled with DNA translocation activity.

In DExx box helicases, motifs in domain 2 participate in ATP (motif VI, Fig. 3) and DNA binding (motif IV, V) (Velankar et al., 1999). However, SsoRad54cd domain 2 is flipped by 180° and motifs IV-VI are remote from both ATP and DNA binding sites. In particular, motif VI of SWI2/SNF2 enzymes is highly conserved to superfamily II DExx box helicases and has been implicated in ATP binding and hydrolysis in a variety of helicase structures (Caruthers and McKay, 2002). This suggests that domain 2 of SsoRad54cd probably rearranges in response to ATP-binding. On the basis of the very similar fold topology of the core RecA-like subdomains of SsoRad54cd (1A and 2A) and helicases, the two subdomains of SsoRad54cd were superimposed independently with the equivalent structural regions of PcrA and NS3. The comparative superposition revealed a remarkable similarity of the tertiary structure, resulting in root mean square deviations of typically 1.0 – 1.5 Å at secondary structure elements. This similarity indicates that the composite active site of SWI2/SNF2 enzymes is highly related to DExx box helicases. The closely matched superpositions support the idea that ATP induces or stabilizes a ≈ 180° flip of domain 2 with respect to domain 1 (Fig. 22 A). Such a putative change in the relative orientation of the two subdomains favourably aligns motif VI in the active site cleft to participate in the for DExx box enzyme typical composite ATP binding site. Likewise, motifs IV and V are suitably positioned to participate in DNA binding. Therefore the crystallographically observed conformation of domain 2 in the absence of ATP might represent an "open" conformation to facilitate ATP or DNA binding.

Although the precise nature of ATP-driven conformational changes of SsoRad54cd requires determination of crystal structures in complex with ATP, the remarkable similarity of the two RecA-like subdomains of SsoRad54cd (1A and 2A) to equivalent domains of DExx box helicases legitimates the use of ATP-bound helicase structures to model the putative ATP-bound domain orientation of SsoRad54cd (Fig. 22 A). The substantial flip of domain 2, required to position motif VI near the ATP-binding site, is

both consistent with the small and relatively hydrophilic domain 1:2 interface and the location of the linker (αK). In the putative ATP-bound conformation, the SWI2/SNF2 specific subdomains 1B and 2B form a "tweezers" that are well positioned to directly interact with or manipulate upstream DNA (Fig. 22).
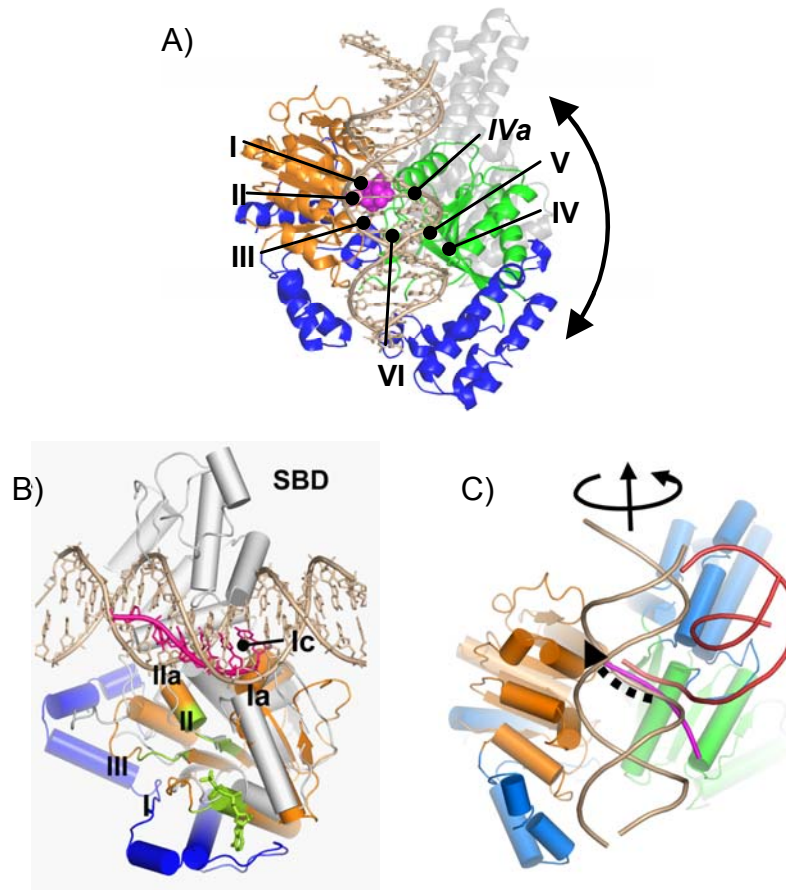
Figure 22: A) Putative ATP-driven conformational changes in SsoRad54cd. Domain 1 and 2 were independently superimposed with equivalent structural regions of the DExx box helicase PcrA. Domain 2 likely flips 180° (colour coded structure) from its conformation in the absence of ATP (grey). This putative conformational change would favourably position motif VI near the ATP binding site (magenta) and motifs IV, IVa and V near DNA. In the putative ATP-bound conformation, subdomains 1B and 2B (blue) form "tweezers" that are suited to relay ATP-driven domain rotations into alterations of DNA structure, while motifs IV, IVa and V could participate in dsDNA binding. B) Ribbon model of a superposition of SsoRad54cd domain 1 (colour code of Fig. 17) with the equivalent structural regions of the Hepatitis C virus helicase NS3

(grey). ATPase motifs (green) and the 3'-5' strands of SsoRad54cd dsDNA (brown) and NS3 ssDNA (magenta) overlay well, indicating that both enzymes translocate in the same direction on DNA. NS3 contains a specific ssDNA-binding domain (SBD). The absence of this domain in SsoRad54cd could explain why SWI2/SNF2 enzymes do not have helicase activity. C) Comparison of DNA bound to SsoRad54cd (colour coded ribbon model plus brown duplex DNA), NS3 (only oligo-dU is shown in magenta) and PcrA (only ssDNA/dsDNA is shown in red). The evident partial similarity of DNA recognition suggests that dsDNA is transported along SWI2/SNF2 enzymes in a similar manner than ssDNA by helicases (dashed arrow). Thus DNA may move along SWI2/SNF2 enzymes in a screw motion (arrows on top).

The closely matched ATP-binding site, formed by motifs I, II and III and, assuming the putative ATP bound SsoRad54cd model, complemented by motif VI, suggests a common mechanism for ATP-hydrolysis by SWI2/SNF2 ATPases and DExx box helicases (Fig. 22). In support of this idea, the 3'-5' strand of the dsDNA bound to SsoRad54cd overlays well with the 3'-5' oligo-dU bound to NS3, or the 3'-5' ssDNA tail bound to PcrA (Fig. 22 C). The similarity of DNA recognition by domain 1A raises the possibility that the powerstroke required for translocation of the dsDNA minor groove across the surface of SsoRad54cd is related to the powerstroke that drives ssDNA translocation by DExx box helicases. In helicases, this powerstroke is linked to the ATP dependent opening and closure of the cleft between the two RecA subdomains, in particular by interaction of motif VI with ATP.

Although SWI2/SNF2 enzymes share several motifs and structural features with DExx box helicases they nevertheless lack helicase activity. Comparison of SsoRad54cd with the PcrA and NS3 helicases also revealed important differences that could explain the lack of helicase activity. For instance, in NS3 and PcrA, additional domains firmly grab the bases of ssDNA and presumably help the enzyme to translocate along ssDNA. SsoRad54cd lacks this additional single-stranded DNA-binding domain (SBD, Fig. 22 B). Instead, SsoRad54cd has a more extended dsDNA binding site, for instance by interaction of the switch loop with the 5'-3' strand. The lack of the ssDNA-binding domain in SWI2/SNF2 enzymes indicates why

SsoRad54cd binds ssDNA and Y-DNA substrates with much less affinity than dsDNA (Fig. 9). Consistently, the SsoRad54cd ATPase activity is not stimulated by ssDNA, in contrast to DExx box helicases. Thus, SsoRad54cd cannot translocate on ssDNA in an ATP-hydrolysis dependent manner. Secondly, the structure of the SsoRad54cd:dsDNA complex does not provide any evidence for a helix-destabilizing region in SsoRad54cd and the crystallographically observed DNA is fully base paired along the entire side of SsoRad54cd. These findings argue that SWI2/SNF2 travel along dsDNA without strand separation activity. Although such domains (ssDNA binding site and helix destabilizing region) could in principle be part of the N-terminus of SsoRad54, which is lacking in this structure, or even provided by additional subunits, they are not part of the conserved core of SWI2/SNF2 enzymes. In addition, a longer SsoRad54 construct (SsoRad54cd_285-906) does not show substantially increased DNA binding or ATPase activity.

Comparison with eukaryotic Rad54 from zebrafish

Recently, a crystal structure of zebrafish Rad54 has been published (Thoma et al., 2005). Like SsoRad54cd, eukaryotic Rad54 core consist of two α/β-lobes, with two helical insertions in each of the two lobes. The two RecA-like domains are separated by a broad cleft and flanked by the helical insertions. In contrast to SsoRad54cd, the orientation of the RecA-like domains in the crystal structure of zebrafish Rad54 is similar to those seen in several structures of superfamily 2 helicases and agrees with the putative "closed" model of SsoRad54cd. Analogous to helicases and the suggested SsoRad54cd model, the seven helicase motifs of zebrafish Rad54 map to the domain 1:2 interface in a virtually identical rearrangement of the motifs compared to helicases. The high similarity supports a common mechanism of ATP-hydrolysis of SWI2/SNF2 ATPases and DExx box helicases. The overall structure of the domains of zebrafish Rad54 are very similar to the SsoRad54cd domains, as evident from the superposition of the individual domains of the two structures, including the SWI2/SNF2 specific accessory domains (1B and 2 B) (Fig. 23). The resulting domain orientation of SsoRad54cd hereby resembles the model derived from the superposition with superfamily 2 helicases (see above) and validates the approach described above.
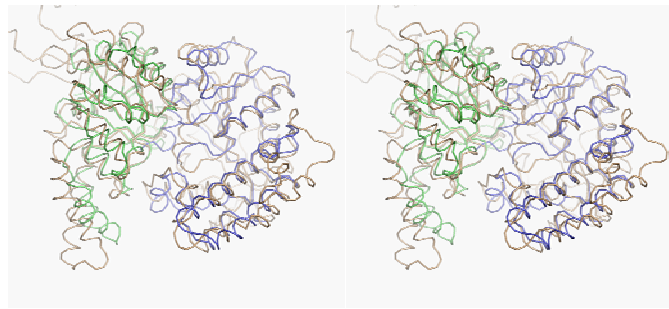
Figure 23:   Stereo view of the superposition of domain 1 and domain 2 of SsoRad54cd (green, blue) with equivalent regions of zebrafish Rad54 (brown). The remarkable structural similarity clearly demonstrates the homology of these two proteins and supports the proposed orientation of domain 2 of SsoRad54cd in the "closed" form.

The overlay demonstrates a remarkable structural similarity of SsoRad54cd with eukaryotic Rad54, as apparent from root mean square deviations of 0.9-1.3 Å at secondary structure elements. Moreover, the biochemical and structural analysis on SsoRad54cd definitely group SsoRad54 to SWI2/SNF2 enzymes. Thus, SsoRad54cd represents another example of for evolutionary conservation between archaea and eukaryotes.

Mutagenesis studies

Classical functional motifs

SWI2/SNF2 enzymes and helicases share a set of seven conserved motifs, known as helicase motifs. Expanding structural information on helicases provide information on the function of the conserved helicase motifs and their participation in the ATP hydrolysis mechanism and DNA binding.

Motif I and II
The contribution of motifs I and II (Walker A and B, respectively) to ATP binding and hydrolysis is well understood. The lysine of the Walker A motif, classically defined by GxxxxGK(T/S) consensus sequence, interacts with the α and β phosphates of

Mg^{2+}ATP/Mg^{2+}ADP and the threonine or serine coordinates a Mg^{2+} ion. The Walker B motif, characterized by the eponymous DExx consensus sequence, coordinates the Mg^{2+} via an aspartic acid through water mediated outer sphere interactions. The conserved glutamic acid polarizes the water molecule for the nucleophilic attack and catalyses ATP hydrolysis. The configuration of Walker A and B is not only shared by DExx box helicases, but also by other P-loop ATPases such as SWI2/SNF2 enzymes, AAA^{+} and ABC ATPases.

Motif VI

Next to Walker A and B motifs, motif VI of helicases, which is located on domain 2, plays an important role in the ATP hydrolysis cycle. Motif VI shows less conservation among DExx box helicases, but contains a highly conserved arginine in the middle of the motif (arginine finger). This arginine is suggested to interact with ATP by forming a salt bridge with the γ phosphate of the ATP. For instance, in the structure of PcrA, the arginine residue ligates the terminal phosphate of a nonhydrolysable ATP analog and mutational studies on several helicases support the model that the arginine in motif VI interacts with ATP. Therefore, motif VI seems to play an important role for interdomain communication. These observations suggest a mechanism, in which ATP binding and hydrolysis is coupled to conformational rearrangements in the core domains. Motif VI of SWI2/SNF2 enzymes is highly conserved between SWI2/SNF2 family members and contains likewise an invariant arginine. Analogous "arginine finger" residues have also been documented in other AAA+ and ABC ATPases, where they are proposed to have a role in transducing ATP hydrolysis and pyrophosphate release into specific interdomain conformational changes.
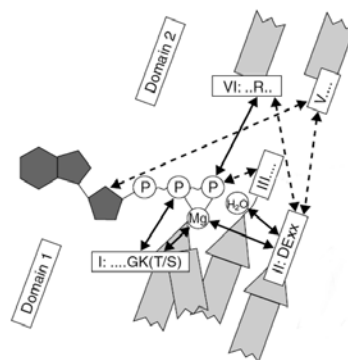
Figure 24: Schematic drawing of the implications of the helicase motifs in ATP binding and hydrolysis. The network of interactions (solid arrows) between motifs I, II and VI and Mg^{2+}ATP/Mg^{2+}ADP mainly account for

the ATP hydrolysis cycle and are shared by helicases and other ATPases. Further interactions occur, but are more specific to helicases (dotted arrows). (adapted from (Caruthers and McKay, 2002)

Motifs Ia, Ic and IV

The implications of the other motifs are much less understood. Helicase motifs Ia and Ic (TxGx) and IV are suggested to bind DNA, as observed in helicase:DNA crystal structures. Interestingly, helicase motif Ia and Ic in SsoRad54cd contact DNA via conserved residues (S502, T541 and R547) (Fig. 6, 20). Ic recognizes the 3´-5´ strand of the dsDNA, which overlays well with the 3´-5´strand of ssDNA bound to DExx box helicases (Fig. 22 C), suggesting a similar mode of DNA binding.

Motifs III and V

In helicases, motifs III and V show substantial divergence in the amino acid sequence. In some instances, these motifs participate in a complex network of interactions that include ligation of $Mg^{2+}ATP/Mg^{2+}ADP$, formation of specific salt bridges or hydrogen bonds between domain 1 and 2 and DNA binding. For instance, in helicase structures, motif III interacts with the γ phosphate of ATP, with DNA, or is involved in sensing of ATP hydrolysis and interdomain communication (Tuteja and Tuteja, 2004). Motif III, IV and V are highly conserved in SWI2/SNF2 enzymes and exhibit sequence and structural similarity to superfamily II helicases. In particular, motif IV and V on domain 2 may provide potential DNA interactions following the putative rearrangement of domain 2.

Structure-function analysis of SsoRad54cd

As described above, the conservation of the helicase motifs in SWI2/SNF2 enzymes and the related structural topology of the two domains of SsoRAd54cd with PcrA or NS3 suggest a similar ATP binding and hydrolysis mechanism. Implications of the conserved motifs in the function of SWI2/SNF2 enzymes have been already demonstrated for the yeast SWI2/SNF2 ATPase domain by mutagenesis studies (Richmond and Peterson, 1996). To illustrate the involvement of the conserved helicase motifs for SsoRad54cd, a set of exemplary single amino acid mutations

within the classical motifs were created and their functional consequences analysed. The selection of the mutations is based on the structure of SsoRad54cd:DNA complex, on conservation of the amino acids and on the implication of the amino acids in ATP-hydrolysis or DNA-binding. I analysed the behaviour of the SsoRad54cd mutants with respect of their dsDNA stimulated ATPase activity, their cruciform extrusion activity and their DNA binding activity.

Besides these "classical" helicase motifs, the structure of SsoRad54cd (Fig. 18) in combination with the sequence alignment (Fig. 6) also suggests other sequence elements that could be specifically involved in SWI2/SNF2 ATPase function. Mutations were constructed in these additional motifs to identify crucial residues that are relevant for the catalytic activity of the SWI2/SNF2 ATPase domain. The set of mutations that were performed are depicted in table 5 and their effects on SsoRad54cd activity are summarized in figure 25.

Classical helicase motifs

Mutation	motif	Mutation	motif
E563Q	II	N569I	IIa (switch)
Q755A	IV	G722Q	interface
K808E	V	R590W	cluster I
R840E	VI	K700E	domain 2B (αM)
R843E	VI	K711E	domain 2B (αM)
		S783V	IVa
		R788E	IVa
		K872E	
		V850G	cluster II

Table 5: Mutations in SsoRad54cd. A set of exemplary mutations within the classical motifs and additional motifs were created to confirm or identify crucial residues that are relevant for the catalytic activity of the SWI2/SNF2 ATPase domain.

The mutations were biochemically analysed by DNA stimulated ATPase and cruciform extrusion assay. The effects of the mutations correlate well in both assays, indicating that the ATP hydrolysis is mechanistically linked with cruciform extrusion.
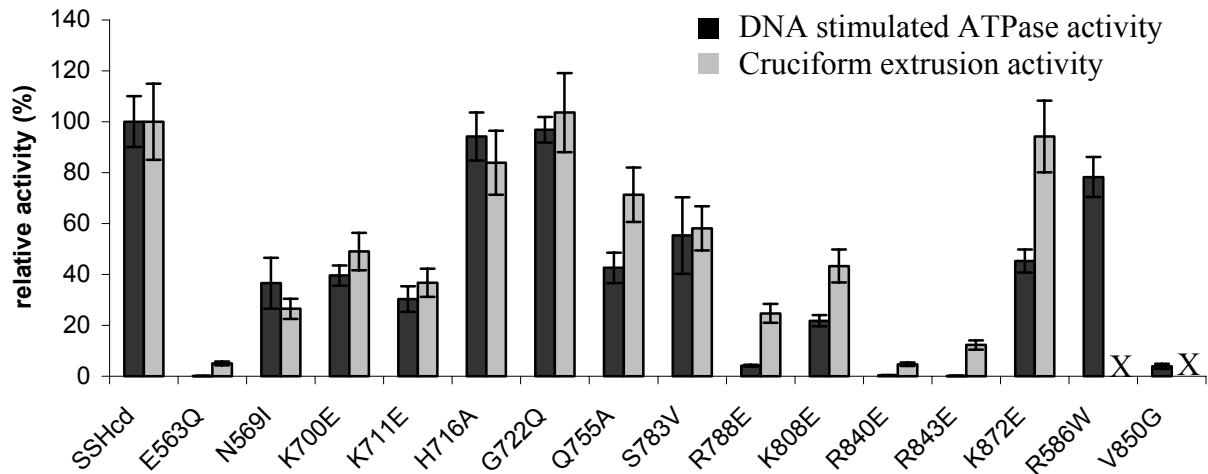
Figure 25: Effects of single amino acid mutations on DNA dependent ATPase (black bars) and cruciform extrusion assay (grey bars). The analyses confirm the involvement of the conserved motifs in SsoRad54cd function and also identify additional amino acids that are relevant of SsoRad54 activity. Shown is the relative activity compared to wtSsoRad54cd (in %) (means of three independent experiments ± standard deviation)

The mutations in motif II (E563Q) and motif VI (R840E and R843E) virtually abolish DNA stimulated ATPase activity and cruciform activity of SsoRad54cd. However, in the crystallographic observed SsoRad54cd conformation, motif VI is remote of the primary ATP binding site formed by motif I, II and III. The mutational analysis proves that motif VI is essential for SsoRad54cd activity. This effect can only be explained, if domain 2 reorients and positions motif VI in the active site cleft to allow participation in the ATP-hydrolysis cycle. Analogous, the mutation of Q755A (motif IV) or K808E (motif V) reduces the activity of SsoRad54cd and indicates the implication of these motifs in SsoRad54cd function, although the interpretation of the effects is more difficult. The residues are located outside the active cleft in the crystallized observed structure and again are positioned appropriately in the model to partake their function. In addition, these results indicate that the classical motifs are important for

SsoRad54cd function and support the idea that SWI2/SNF2 enzymes and helicases share a basic ATP hydrolysis mechanism.
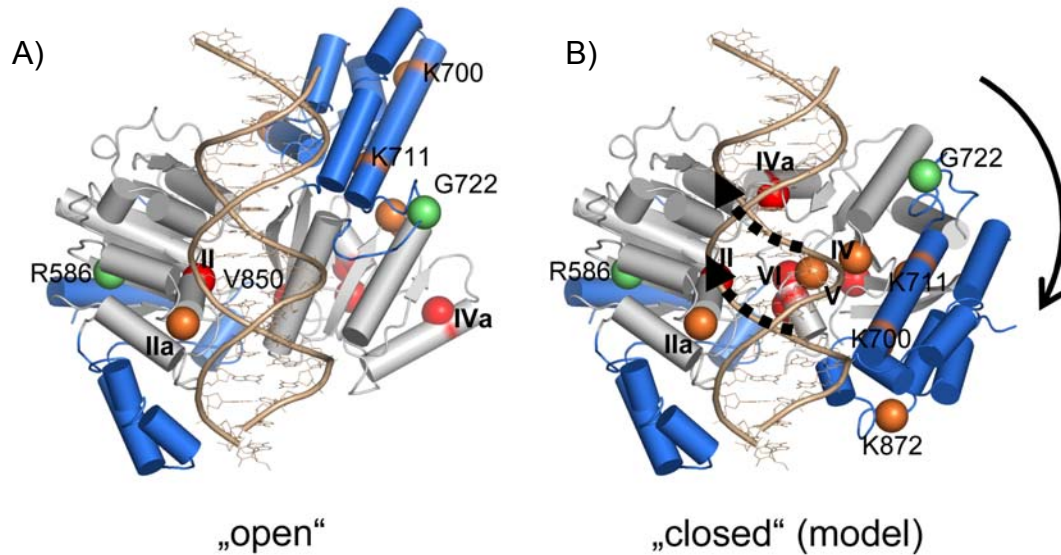
Figure 26: A) Crystallographically observed "open" conformation. B) Model for the putative "closed" conformation, based on superposition of domain 2 with the equivalent domain in the AMPPNP bound structure of PcrA (grey, 1A and 2A; blue 1B and 2B). The putative 180° flip of domain 2 (solid arrow) would favourably position motifs IV-VI along the active site cleft. The location and effect of mutations (colour coded spheres showing percent of residual ATPase activity; red, 0-10%; orange 10-50%; green >74%) are consistent with the modeled orientation of domain 2. In the closed form, subdomain 2B (via αM) may contact DNA. For instance, an ATP-binding induced push of 2B on the upstream minor groove could advance DNA by sliding across domain 1.

New functional motifs

Besides the classical helicase motifs, the structure and sequence alignment also identifies conserved, previously unnoticed SWI2/SNF2 specific motifs. For instance, the switch loop (motif IIa) appears to be a conserved feature of SWI2/SNF2 enzymes. A mutation in one of the conserved residues (N569I) reduces both DNA stimulated ATPase activity (Fig. 25) and DNA binding activity (data not shown). We

therefore denote the sequence motif that forms the switch loop as motif IIa (Fig. 6). An additional highly conserved motif among SWI2/SNF2 enzymes is formed by the domain 2 loop between $\beta10$ and αP (denoted hereby as motif IVa). Mutation of R788E, an invariant arginine in motif IVa also virtually abolishes DNA stimulated ATPase activity and a further mutations in motif IVa (S783V) also affect both activities. Interestingly, the equivalent arginine of the SWI2/SNF2 ATPase SMARCAL1 is mutated in Schimke immuno-osseous dysplasia, indicating that motif IVa is also relevant in other SWI2/SNF2 enzymes and important in vivo (Boerkoel et al., 2002).

The structural conservation of the two SWI2/SNF2 specific helical insertions (1B and 2B, respectively) indicates that they might have a functional role in DNA translocation or distortion. In particular, αM seem to be involved in contacting DNA. Three possible candidates of αM were mutated: the two basic amino acids (K700 and K711) and the highly conserved histidine (H716). Surprisingly, H716 do not show any effect on SsoRad54cd activity, whereas the K700E and K711E affect DNA stimulated ATPase and cruciform activity.

The mutational analyses demonstrate on the one hand the implications of the classical helicase motifs in SsoRad54cd function and on the other hand identified additional important residues and new functional SWI2/SNF2 specific motifs. The effects of the mutations in helicase motifs (E563Q, Q755A, R788E, K808E, R840E, R843E), as well as additional residues that are predicted to be near the DNA binding site (K700E, K711E) or remote from ATP and/or DNA binding sites (R586W, G722Q) on ATPase and cruciform extrusion activity of SsoRad54cd are consistent with the proposed model (Fig. 26). In this respect, domain 2 mutations might directly interfere with the ATPase activity or may prevent a proper orientation of domain 2 in the presence of ATP.

Structural basis for human diseases

In general, the location of human disease mutations could uncover important functional regions of SWI2/SNF2 enzymes. Several human disease syndromes are linked to mutations in SWI2/SNF2 enzymes. Cockayne syndrome is caused by mutations in the transcription/repair factor $\underline{C}$ockayne $\underline{s}$yndrome protein $\underline{B}$ (CSB) (Bradsher et al., 2002). The severe X-linked alpha thalassaemia/mental retardation (ATR-X) syndrome is caused by mutations in human ATRX gene (Picketts et al., 1996). Recently, Schimke immuno-osseous dysplasia was linked to mutations in SMARCAL1, a protein of poorly understood function that is related to the prokaryotic HepA protein.

SMARCAL1 mutations are mostly found in or directly adjacent to ATPase and DNA binding motifs, including the newly identified motif IVa, suggesting that Schimke immuno-osseous dysplasia is caused by a defective ATP-hydrolysis or DNA binding by SMARCAL1 (Fig. 6). Most of the ATRX and CSB missense mutations, however, are found remote from DNA- or ATP-binding motifs. Surprisingly, these mutations co-localize in two surface clusters on domain 1 and 2 (cluster I and II, Fig. 27 A) when mapped onto the structure of SsoRad54cd. Many of the mutations are conserved between the disease related protein and SsoRad54cd, allowing a precise structural localization.
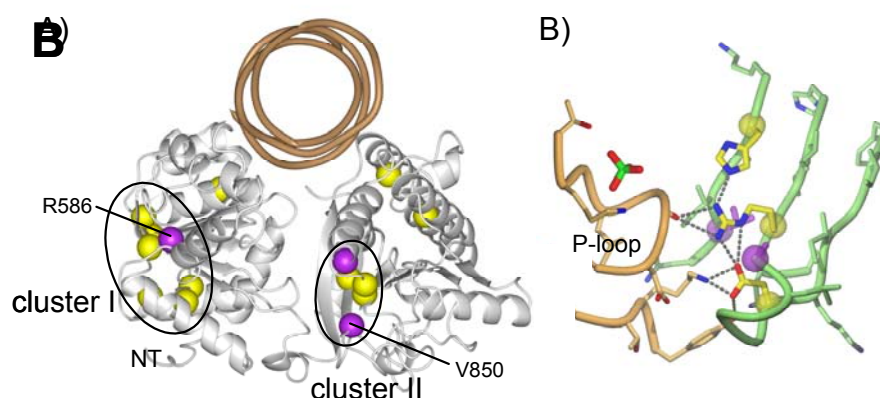
Figure 27: A) Mutations involved in ATRX syndrome (yellow spheres) and Cockayne syndrome (purple spheres) colocalize in two clusters (cluster I and II) on SsoRad54cd (grey ribbon model) B) Detailed view of cluster II reveals that several ATR-X (yellow) and Cockayne (purple)

syndrome mutations are in direct contact to each other, suggesting that they disrupt a conserved functional site on SWI2/SNF2 enzymes. The location of cluster II at the domain 1 (orange) : domain 2 (green) interface near the P loop (with crystallographically observed bound phosphate) indicates that these mutations could interfere with ATP-driven conformational changes.

Cluster I is remote from DNA and ATP-binding sites, but adjacent to a hydrophobic surface patch near the N-terminus of SsoRad54cd. Thus cluster I mutations could specifically perturb the interaction of the CSB or ATRX catalytic domain with additional N-terminal domains. Cluster II mutations are located in the crystallographically observed interface of domain 1 and 2 and are linked to each other by hydrogen bonds or hydrophobic contacts (Fig. 27 B). Since cluster II is located near the proposed pivot region of domain 1 and 2, mutations in this region could interfere with ATP-driven structural changes between domain 1 and 2.

The CSB mutation R670W (Cluster I) and V957G (Cluster II) were introduced into the equivalent residues of SsoRad54cd (R586 and V850) and the effect of the mutations on the ATPase activity were tested (Fig. 25). V850G abolished DNA stimulated ATPase activity of SsoRad54cd, suggesting that it interferes with conformational changes between domain 1 and 2 that are required for ATP-hydrolysis. In contrast, R586W barely reduces DNA stimulated ATPase activity of SsoRad54cd, consistent with the idea that cluster I mutations might disrupt a macromolecular interaction interface, such as additional substrate binding domains.

Thus, although ATR-X and Cockayne syndromes are genotypically and phenotypically unrelated, many of the underlying mutations structural cluster at surface or interface regions. Although the structure of SsoRad54cd is an approximation to some extent, this clustering indicates that these mutations do not disrupt a specialized functional aspect of CSB or ATRX, but interfere with conserved features of SWI2/SNF2 enzymes. The location of these mutations outside identified motifs and DNA interfaces argues that not only active site motifs but more global structural features of different SWI2/SNF2 enzymes are conserved, for example the communication of the ATPase domain with additional substrate binding domains.

Discussion:

Model for ATP-driven translocation and distortion by SWI2/SNF2 enzymes

Current models suggest that ATP dependent DNA translocation and distortion by SWI2/SNF2 enzymes provide the means to remodel protein:DNA interfaces (Langst and Becker, 2004). However, how ATP hydrolysis is coupled to translocation and distortion activity is still poorly understood. The structure of SsoRad54cd in complex with its dsDNA substrate provides first structural insights into the interplay between the catalytic core of SWI2/SNF2 enzymes with the DNA substrates. Our results reveal that the SWI2/SNF2 ATPase domain consists of two RecA-like domains, which are highly related to DExx box helicases. The two domains are separated by a deep cleft and flanked by two SWI2/SNF2 specific helical insertions. The DNA binds alongside the domain 1 : domain 2 interface, in a position where ATP-driven conformational changes between the two domains could directly be linked to DNA translocation and distortion. The sequence and structural homology of SWI2/SNF2 enzymes and DExx box helicases suggests that they share the basic ATP hydrolysis mechanism. In helicases, the translocase activity can be physically uncoupled from the strand separation activity (Soultanas et al., 2000). For instance, the crystal structures of PcrA suggest a rigid body movement, large conformational changes of the two RecA-like subdomains (1A and 2A) due to ATP binding and hydrolysis, that drive the enzyme along the ssDNA backbone. Consistently, the structure of RecG suggests that helicase motifs can function as part of a duplex DNA translocase domain. At present, we haven't been able to crystallize an ATP-bound complex of SsoRad54cd. However, the remarkable topologically and structurally similarity of the

two RecA like subdomains of SsoRad54cd (1A and 2A) to the equivalent domains of DExx box helicases including DNA recognition suggests that ATP-driven transport of dsDNA in the active site of SWI2/SNF2 enzymes is mechanistically related to ATP-driven ssDNA translocation in the active site of helicases (Fig. 22). Superpositions of the two RecA subdomains of SsoRad54cd with equivalent domains of the ATP-bound helicase structures in combination with our mutagenesis data help us to derive a mechanistical model for SsoRad54cd function. The putative model suggests that in the presence of ATP, domain 2 reorients ~180° from its crystallographically observed "open" conformation into a "closed" conformation (Fig. 22 A). The interdomain flexibility of SsoRad54cd would allow such substantial conformational changes. The putative "closed" conformation favourably aligns functional motifs along the composite ATP binding site in the active site cleft and is consistent with the mutational analysis (Fig. 22 and 26). In the model, domain 2 is situated near the upstream minor groove. Consistently, mutations along the proposed path of DNA reduce DNA stimulated ATPase activity in a similar way than mutations in motifs IV, IVa and V (Fig. 25, 26). Mutations in conserved residues that are remote from the putative DNA binding cleft (R586W and G722Q) in the ATP-bound conformation do not significantly interfere with DNA stimulated ATPase activity. In the putative ATP-bound conformation, SWI2/SNF2 specific subdomains 1B and 2B form a "tweezers" at the upstream DNA (Fig. 22). This tweezers is well positioned to bind the upstream minor groove. The structural conservation of the two SWI2/SNF2 specific helical insertions (1B and 2B, respectively) and mutations in the αM helix (K700E and K711E) support a functional role in DNA translocation.

Based on the structural and mutational results, we propose a specific and testable mechanism for the ATP-dependent translocation of SWI2/SNF2 enzymes on DNA. Prior to translocation, DNA is bound at domain 1 at the high affinity DNA binding site. Domain 2 could rearrange or is stabilized into a conformation that allows e.g. binding of motifs VI to ATP. In the putative "closed" ATP bound conformation, the SWI2/SNF2 specific accessory domains (1B and 2B, respectively) form the tweezers that contact DNA approximately one helical turn upstream of the dsDNA binding on domain 1. Thereby, domain 2 could for instance push on upstream DNA (solid arrow in Fig. 28) to facilitate sliding of domain 1 along the minor groove (dashed arrows in Fig. 28). If domain 2 simply pushes on upstream minor groove, domain 2 does not

need a high affinity DNA binding site by itself, a feature we observe experimentally (Fig. 10). Following advancement of DNA, ATP-hydrolysis might relax the structure to allow binding of domain 1 to DNA, ADP->ATP exchange and/or rebinding of domain 2 at a new "translocated" upstream DNA binding site.
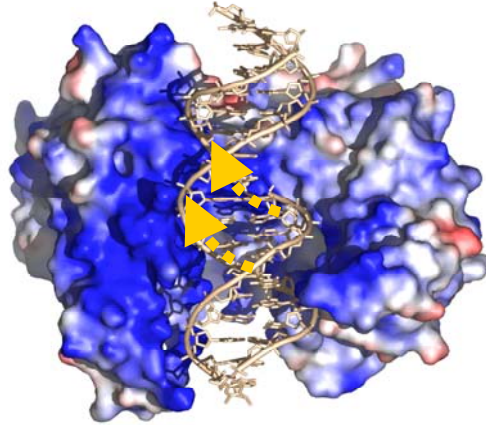

Figure 28:   Model for ATP-driven translocation on DNA. DNA binds to the principal binding site on domain 1. In the putative ATP-bound form of SsoRad54cd, domain 2 might push on upstream DNA. ATP-hydrolysis presumably weakens DNA interaction at domain 1, allowing relaxation of the structure by sliding of the DNA minor groove strands along domain 1 (dashed arrows).

This mechanism is both consistent with the observed step size and directionality of SWI2/SNF2 enzymes (Whitehouse et al., 2003). The proposed mechanism could explain how the characteristic biochemical activities of SWI2/SNF2 enzymes are generated by one mechanism that, in its core, is related to that of DExx box helicases. In addition, the model uses a specific feature of DNA, namely that rotation and translation are equivalent in helical structures. Thus, the translocation along the minor groove of dsDNA implicates rotational force that could explain DNA distortion activity.

Implication into remodeling of protein:DNA complexes

How can DNA translocation or distortion by SWI2/SNF2 enzymes mechanistically be linked to modulation of protein:DNA complexes? The best-studied examples of such

activities are nucleosome sliding by chromatin remodeling complexes. The mechanism by which remodeling factors induce nucleosome sliding, i.e. break and reform histone:DNA interactions, remains controversial and is likely to be different for the different types of remodeling factors. Since DNA interacts with the histone octamer through a large number of close contacts, a simultaneous breakage of all these contacts is highly unfavourable. A mechanism that is associated with transient disruption of DNA:histone contacts is more likely. The step-by-step redistribution implies the breaking and reforming of only a subset of histone:DNA contacts minimizing the free energy penalty to an acceptable level. Two major models of mechanisms are proposed and illustrated in figure 29: "twist defect diffusion" and "bulge diffusion" (for reviews (Becker and Horz, 2002; Owen-Hughes, 2003).
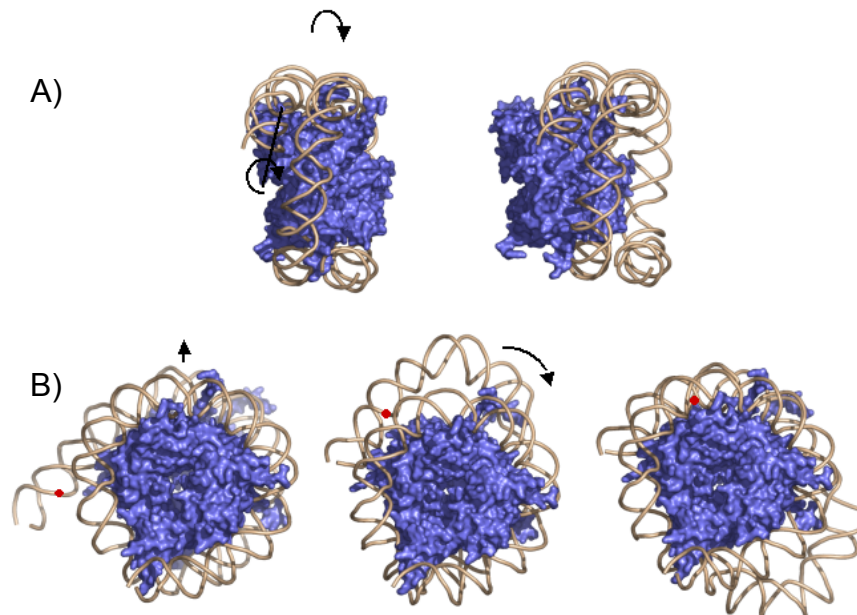
Figure 29: A) "twist defect diffusion" model. Histone:DNA interactions are disrupted by generation of superhelical torsion. The distortion propagates over the surface of the histone octamer resulting in a "screw-like" lateral cross-transfer of the nucleosomal DNA. B) "bulge diffusion" model. Local loop or bulge formation leads to a transient dissociation of nucleosomal DNA and subsequent rebinding of more distal sequences (accented by a red point). The bulge migrates around the histone octamer resulting in a relocation of the nucleosome.

The first model (twist defect diffusion) assumes that DNA is distorted through localized alterations to twist. According to this model, histone:DNA interactions may

be disrupted by the (un)twisting of DNA in a small domain, presumed to be topologically constrained by the remodeling enzyme itself and the nucleosome substrate. The rotation of DNA around its long axis may enable DNA to propagate incrementally over the surface of nucleosomes in a "cork-screw" like action. This twist defect diffusion would lead to a replacement of "canonical" histone:DNA interactions by equivalent ones involving neighbouring base pairs. Supporting this model, twist diffusion has been detected over long distances of nucleosomal DNA in high-resolution crystal structures (Suto et al., 2003). Further support comes from the findings that remodeling factors are capable of generating superhelical torsion in linear DNA. However, in order for nucleosome redistribution by this mechanism, a large number of these DNA distortions have to occur in a concerted orientation. In addition, there are experimental findings that argue against the "twist defect diffusion" mechanisms. For instance the observation that nucleosome sliding occurs in discrete step size is not consistent with the "twist defect diffusion" mechanism. In addition, insertion of nicks or gaps in nucleosomal DNA, that would block twist diffusion does not affect catalyzed nucleosome repositioning (Langst and Becker, 2001; Saha et al., 2002)

The second model involves the formation of a loop or bulge on the surface of a nucleosome that may be transmitted around the surface of the nucleosome. Since DNA at the edge of nucleosomes are less stably associated with nucleosomes than DNA at the center (Flaus and Owen-Hughes, 2003a), it is likely that remodeling factors initiate destabilisation of histone:DNA interaction at this Achilles heel. Transient dissociation of the DNA, starting at the entry/exit site of nucleosomes, followed by reformation of histone DNA interactions at a different site would result in an initial formation of a loop or bulge that can then propagate around the nucleosome. Several movements of the bulge around the octamer surface in the same direction would lead to complete translocation of the DNA relative to the octamer according to the loop size. The observations that the SWI/SNF complex tend to form loops and that nucleosomes slide in discrete steps are in accordance with the "bulge diffusion" mechanism (Bazett-Jones et al., 1999). A similar mechanism have also been described for the passage of RNA polymerase through nucleosomes (Studitsky et al., 1994).

There are many questions concerning the mechanism of how other remodeling factors like CSB/Rad26, Mot1 or Rad54 rearrange their protein:DNA substrates. However, similar remodeling mechanism as described for the nucleosome sliding may be also applicable for the disruption of other persistent protein:DNA interfaces. For instance, recently it has been shown that CSB actively wraps DNA by altering DNA double helix conformation (Beerens et al., 2005).

Our proposed mechanistic model can provide a framework for the two hotly debated features of chromatin remodeling factors. First, why do even diverse remodeling factors generate superhelical torsion in DNA and translocate on dsDNA (Havas et al., 2000; Jaskelioff et al., 2003; Ristic et al., 2001; Rouleau et al., 2002; Xue et al., 2003; Yu et al., 2004)? Second, how does a putatively common mechanism of the motor domain generate different activities (e.g. bulging, twist diffusion) for the mobilization of nucleosomes (Flaus and Owen-Hughes, 2003b; Langst and Becker, 2004)? The answer to the first problem is suggested by the proposed structural mechanism of the SWI2/SNF2 type ATPase itself. The biochemical activity and crystals structures of SsoRad54cd suggest that DNA distortion and translocation are two elements of the powerstroke of the conserved SWI2/SNF2 ATPase. In particular, even a relatively moderate translocation of DNA by sliding along the minor groove would include a substantial rotation of DNA along the helical axis. However, remodeling factors might use both translocation and rotation. The suggested model could also help to understand the answer to the second problem, namely the different mechanisms for the repositioning or mobilization of nucleosomes by chromatin remodelers. For instance, large remodeling factors may bind to substrate protein:DNA complexes both by the catalytic domain and by additional substrate binding domains, as shown for the ISWI complex (Grune et al., 2003). If the enzyme is anchored to the remodeled substrate, the screw motion of DNA at the catalytic domain could not only transport DNA toward (or away from) the substrate protein:DNA complex, but could also generate rotational torque that is well suited to disrupt or remodel the substrate protein:DNA interfaces (Fig. 30)
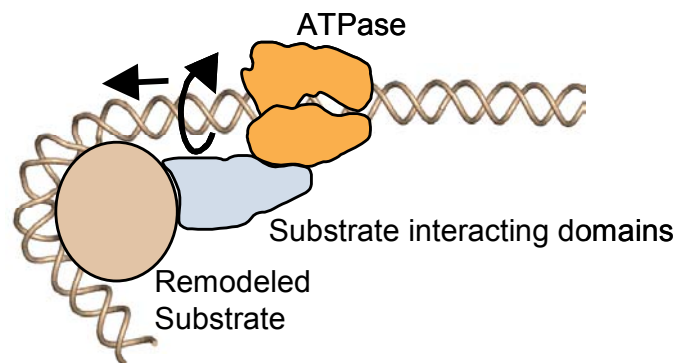
Figure 30:   Implications for remodeling of DNA:protein complexes by SWI2/SNF2
enzymes. Rotation and translocation (arrows) of DNA at the ATPase
domain (orange) may provide the torsional force that may be translated
in different protein:DNA complexes remodeling modes by additional
substrate interacting domains (light blue).

The apparent versatility in this mechanism may explain the different observed
activities of remodeling factors (Langst and Becker, 2004; Owen-Hughes, 2003). For
instance, transport of DNA towards the nucleosome by the motor domain might result
in bulges that can diffuse around the nucleosome ("bulge diffusion", Fig. 29 B). Such
a mechanism is suggested for ISWI containing remodelers (Langst and Becker,
2001). In contrast, rotation of DNA at the remodeler could peel DNA from
nucleosomes according to the "twist defect" mechanism (Fig. 29 A) as proposed for
SNF2 containing complexes (Flaus and Owen-Hughes, 2004) or wrap DNA as
suggested for the CSB protein (Beerens et al., 2005). Thus, it is possible that the
position of additional substrate interacting domains relative to the motor domain,
which strongly varies among different SWI2/SNF2 ATPase classes, translates a
conserved core activity of the motor domain into different remodeling mechanisms.
Clearly, future studies must clarify the interplay between motor domains and
substrate binding domains and the architecture of multisubunit remodeling factors.

Conclusions

Dynamic remodeling of nucleosomes or other protein:DNA complexes is implicated in many processes in genome expression and maintenance. ATP-driven remodeling activities are carried out by SWI2/SNF2 enzymes that form the engine of diverse and large remodeling factors. The X-ray structure of the catalytic domain of *Sulfolobus solfataricus* Rad54 homolog in the absence of DNA and in complex with dsDNA provides first structural insights for the core mechanism of remodeling factors. Together with structure-function analysis, the data suggest a specific mechanism for DNA translocation and distortion that could explain diverse activities observed with remodeling factors. In particular, the minor groove tracking presumably translocates DNA along the SWI2/SNF2 ATPase active site cleft in a screw motion, a process that is well suited to translocate and distort DNA at nearby protein:DNA interfaces. Our proposed model uses a specific feature of DNA, namely that rotation and translation are equivalent in helical structures. To our knowledge, the results report the first structural insight of an ATP-driven dsDNA translocase domain in complex with DNA. Thus, our results might provide a structural framework for the mechanistic understanding of ATP-driven dsDNA translocases in general.

Acknowledgments

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

I am most grateful to my supervisor Prof. Dr. Karl-Peter Hopfner for giving me the opportunity to work on a challenging and most exciting project in his laboratory, for his scientific guidance, inspiration and friendly advices.

My thanks to all present and former members of the Hopfner lab for the scientific support and the social and enjoyable lab atmosphere. Thanks to Annette Karcher, Angelo Manzan, Alfred Lammens, Nora Assenmacher, Katharina Büttner, Alexandra Schele, Derk Bemeleit, Sophia Hartung, Katja Wenig and Sheng Cui. In particular, I would like to thank "my" diploma students Christian Körner, Marisa Müller and Volker Hickmann for their great help in the project and their enthusiasm.

Special thanks to the Boehringer Ingelheim Fonds for the financial support, the uncomplicated bureaucracy, the nice atmosphere during the meetings and the personal support. I am grateful to BIF for giving me the opportunity to attend the international ASM conference on Bermuda, 2004. This was a fantastic experience.

Thanks to all my friends outside the "scientific world" for providing a non-scientific buffer that is essential to survive the rainy days of research.

In particular I would like to thank my wife Christina for the mental support and patience. I think I have to apologize for all the delays, when I forgot the time over the scientific work.

Zum Schluss möchte ich mich ganz besonders bei meinen Eltern für die stetige Unterstützung und ihr vollstes Vertrauen bedanken

My apologies to all others who I have not mentioned by name. I am indebted to them for the many ways they helped me.

References

Alexeev, A., Mazin, A., and Kowalczykowski, S. C. (2003). Rad54 protein possesses chromatin-remodeling activity stimulated by the Rad51-ssDNA nucleoprotein filament. Nat Struct Biol *10*, 182-186.

Alexiadis, V., and Kadonaga, J. T. (2002). Strand pairing by Rad54 and Rad51 is enhanced by chromatin. Genes Dev *16*, 2767-2771.

Auble, D. T., and Steggerda, S. M. (1999). Testing for DNA tracking by MOT1, a SNF2/SWI2 protein family member. Mol Cell Biol *19*, 412-423.

Auble, D. T., Wang, D., Post, K. W., and Hahn, S. (1997). Molecular analysis of the SNF2/SWI2 protein family member MOT1, an ATP-driven enzyme that dissociates TATA-binding protein from DNA. Mol Cell Biol *17*, 4842-4851.

Bazett-Jones, D. P., Cote, J., Landel, C. C., Peterson, C. L., and Workman, J. L. (1999). The SWI/SNF complex creates loop domains in DNA and polynucleosome arrays and can disrupt DNA-histone contacts within these domains. Mol Cell Biol *19*, 1470-1478.

Becker, P. B., and Horz, W. (2002). ATP-dependent nucleosome remodeling. Annu Rev Biochem *71*, 247-273.

Beerens, N., Hoeijmakers, J. H., Kanaar, R., Vermeulen, W., and Wyman, C. (2005). The CSB protein actively wraps DNA. J Biol Chem *280*, 4722-4729.

Boerkoel, C. F., Takashima, H., John, J., Yan, J., Stankiewicz, P., Rosenbarker, L., Andre, J. L., Bogdanovic, R., Burguet, A., Cockfield, S*., et al.* (2002). Mutant chromatin remodeling protein SMARCAL1 causes Schimke immuno-osseous dysplasia. Nat Genet *30*, 215-220.

Borgione, E., Sturnio, M., Spalletta, A., Angela Lo Giudice, M., Castiglia, L., Galesi, O., Ragusa, A., and Fichera, M. (2003). Mutational analysis of the ATRX gene by DGGE: a powerful diagnostic approach for the ATRX syndrome. Hum Mutat *21*, 529-534.

Bradsher, J., Auriol, J., Proietti de Santis, L., Iben, S., Vonesch, J. L., Grummt, I., and Egly, J. M. (2002). CSB is a component of RNA pol I transcription. Mol Cell *10*, 819-829.

Brehm, A., Langst, G., Kehle, J., Clapier, C. R., Imhof, A., Eberharter, A., Muller, J., and Becker, P. B. (2000). dMi-2 and ISWI chromatin remodelling factors have distinct nucleosome binding and mobilization properties. Embo J *19*, 4332-4341.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-

Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S*., et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54 (Pt 5)*, 905-921.

Bustin, M., Catez, F., and Lim, J. H. (2005). The dynamics of histone H1 function in chromatin. Mol Cell *17*, 617-620.

Cairns, B. R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Du, J., Laurent, B., and Kornberg, R. D. (1996). RSC, an essential, abundant chromatin-remodeling complex. Cell *87*, 1249-1260.

Caruthers, J. M., and McKay, D. B. (2002). Helicase structure and mechanism. Curr Opin Struct Biol *12*, 123-133.

Chakravarthy, S., Park, Y. J., Chodaparambil, J., Edayathumangalam, R. S., and Luger, K. (2005). Structure and dynamic properties of nucleosome core particles. FEBS Lett *579*, 895-898.

Cheung, P., Allis, C. D., and Sassone-Corsi, P. (2000). Signaling to chromatin through histone modifications. Cell *103*, 263-271.

Christiansen, M., Stevnsner, T., Modin, C., Martensen, P. M., Brosh, R. M., Jr., and Bohr, V. A. (2003). Functional consequences of mutations in the conserved SF2 motifs and post-translational phosphorylation of the CSB protein. Nucleic Acids Res *31*, 963-973.

Corona, D. F., Langst, G., Clapier, C. R., Bonte, E. J., Ferrari, S., Tamkun, J. W., and Becker, P. B. (1999). ISWI is an ATP-dependent nucleosome remodeling factor. Mol Cell *3*, 239-245.

Cote, J., Quinn, J., Workman, J. L., and Peterson, C. L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. Science *265*, 53-60.

Darst, R. P., Wang, D., and Auble, D. T. (2001). MOT1-catalyzed TBP-DNA disruption: uncoupling DNA conformational change and role of upstream DNA. Embo J *20*, 2028-2040.

Delmas, V., Stokes, D. G., and Perry, R. P. (1993). A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. Proc Natl Acad Sci U S A *90*, 2414-2418.

Deuring, R., Fanti, L., Armstrong, J. A., Sarte, M., Papoulas, O., Prestel, M., Daubresse, G., Verardo, M., Moseley, S. L., Berloco, M.*, et al.* (2000). The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. Mol Cell *5*, 355-365.

Dlakic, M., and Harrington, R. E. (1995). Bending and torsional flexibility of G/C-rich sequences as determined by cyclization assays. J Biol Chem *270*, 29945-29952.

Du, J., Nasir, I., Benton, B. K., Kladde, M. P., and Laurent, B. C. (1998). Sth1p, a Saccharomyces cerevisiae Snf2p/Swi2p homolog, is an essential ATPase in RSC and differs from Snf/Swi in its interactions with histones and chromatin-associated proteins. Genetics *150*, 987-1005.

Dutnall, R. N., and Ramakrishnan, V. (1997). Twists and turns of the nucleosome: tails without ends. Structure *5*, 1255-1259.

Ehrenhofer-Murray, A. E. (2004). Chromatin dynamics at DNA replication, transcription and repair. Eur J Biochem *271*, 2335-2349.

Eisen, J. A., Sweder, K. S., and Hanawalt, P. C. (1995). Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. Nucleic Acids Res *23*, 2715-2723.

Elfring, L. K., Deuring, R., McCallum, C. M., Peterson, C. L., and Tamkun, J. W. (1994). Identification and characterization of Drosophila relatives of the yeast transcriptional activator SNF2/SWI2. Mol Cell Biol *14*, 2225-2234.

Firman, K., and Szczelkun, M. D. (2000). Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement. Embo J *19*, 2094-2102.

Flaus, A., and Owen-Hughes, T. (2003a). Dynamic properties of nucleosomes during thermal and ATP-driven mobilization. Mol Cell Biol *23*, 7767-7779.

Flaus, A., and Owen-Hughes, T. (2003b). Mechanisms for nucleosome mobilization. Biopolymers *68*, 563-578.

Flaus, A., and Owen-Hughes, T. (2004). Mechanisms for ATP-dependent chromatin remodelling: farewell to the tuna-can octamer? Curr Opin Genet Dev *14*, 165-173.

Giraud-Panis, M. J., and Lilley, D. M. (1996). T4 endonuclease VII. Importance of a histidine-aspartate cluster within the zinc-binding domain. J Biol Chem *271*, 33148-33155.

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P., and Blinov, V. M. (1988). A conserved NTP-motif in putative helicases. Nature *333*, 22.

Gowers, D. M., Bijapur, J., Brown, T., and Fox, K. R. (1999). DNA triple helix formation at target sites containing several pyrimidine interruptions: stabilization by protonated cytosine or 5-(1-propargylamino)dU. Biochemistry *38*, 13747-13758.

Greaves, D. R., Patient, R. K., and Lilley, D. M. (1985). Facile cruciform formation by an (A-T)34 sequence from a Xenopus globin gene. J Mol Biol *185*, 461-478.

Grune, T., Brzeski, J., Eberharter, A., Clapier, C. R., Corona, D. F., Becker, P. B., and Muller, C. W. (2003). Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. Mol Cell *12*, 449-460.

Hamiche, A., Sandaltzopoulos, R., Gdula, D. A., and Wu, C. (1999). ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF. Cell *97*, 833-842.

Harrington, R. E., and Winicov, I. (1994). New concepts in protein-DNA recognition: sequence-directed DNA bending and flexibility. Prog Nucleic Acid Res Mol Biol *47*, 195-270.

Haushalter, K. A., and Kadonaga, J. T. (2003). Chromatin assembly by DNA-translocating motors. Nat Rev Mol Cell Biol *4*, 613-620.

Havas, K., Flaus, A., Phelan, M., Kingston, R., Wade, P. A., Lilley, D. M., and Owen-Hughes, T. (2000). Generation of superhelical torsion by ATP-dependent chromatin remodeling activities. Cell *103*, 1133-1142.

Havas, K., Whitehouse, I., and Owen-Hughes, T. (2001). ATP-dependent chromatin remodeling activities. Cell Mol Life Sci *58*, 673-682.

Hayes, J. J., and Hansen, J. C. (2001). Nucleosomes and the chromatin fiber. Curr Opin Genet Dev *11*, 124-129.

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

Imhof, A., and Becker, P. B. (2001). Modifications of the histone N-terminal domains. Evidence for an "epigenetic code"? Mol Biotechnol *17*, 1-13.

Ito, T., Bulger, M., Pazin, M. J., Kobayashi, R., and Kadonaga, J. T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. Cell *90*, 145-155.

Jaskelioff, M., Gavin, I. M., Peterson, C. L., and Logie, C. (2000). SWI-SNF-mediated nucleosome remodeling: role of histone octamer mobility in the persistence of the remodeled state. Mol Cell Biol *20*, 3058-3068.

Jaskelioff, M., Van Komen, S., Krebs, J. E., Sung, P., and Peterson, C. L. (2003). Rad54p is a chromatin remodeling enzyme required for heteroduplex DNA joint formation with chromatin. J Biol Chem *278*, 9212-9218.

Kassabov, S. R., Zhang, B., Persinger, J., and Bartholomew, B. (2003). SWI/SNF unwraps, slides, and rewraps the nucleosome. Mol Cell *11*, 391-403.

Kim, J. L., Morgenstern, K. A., Griffith, J. P., Dwyer, M. D., Thomson, J. A., Murcko, M. A., Lin, C., and Caron, P. R. (1998). Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. Structure *6*, 89-100.

Kimmins, S., and Sassone-Corsi, P. (2005). Chromatin remodelling and epigenetic features of germ cells. Nature *434*, 583-589.

Kingston, R. E., and Narlikar, G. J. (1999). ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. Genes Dev *13*, 2339-2352.

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

Langst, G., and Becker, P. B. (2001). ISWI induces nucleosome sliding on nicked DNA. Mol Cell *8*, 1085-1092.

Langst, G., and Becker, P. B. (2004). Nucleosome remodeling: one mechanism, many phenomena? Biochim Biophys Acta *1677*, 58-63.

Langst, G., Bonte, E. J., Corona, D. F., and Becker, P. B. (1999). Nucleosome movement by CHRAC and ISWI without disruption or trans-displacement of the histone octamer. Cell *97*, 843-852.

Laurent, B. C., Treich, I., and Carlson, M. (1993). The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. Genes Dev *7*, 583-591.

Lilley, D. M. (1980). The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. Proc Natl Acad Sci U S A *77*, 6468-6472.

Lilley, D. M. (1992). DNA--protein interactions. HMG has DNA wrapped up. Nature *357*, 282-283.

Lilley, D. M., and Kemper, B. (1984). Cruciform-resolvase interactions in supercoiled DNA. Cell *36*, 413-422.

Ljungman, M., and Zhang, F. (1996). Blockage of RNA polymerase as a possible trigger for u.v. light-induced apoptosis. Oncogene *13*, 823-831.

Lorch, Y., Zhang, M., and Kornberg, R. D. (1999). Histone octamer transfer by a chromatin-remodeling complex. Cell *96*, 389-392.

Luger, K., and Richmond, T. J. (1998). The histone tails of the nucleosome. Curr Opin Genet Dev *8*, 140-146.

Lusser, A., and Kadonaga, J. T. (2003). Chromatin remodeling by ATP-dependent molecular machines. Bioessays *25*, 1192-1200.

Manzan, A., Pfeiffer, G., Hefferin, M. L., Lang, C. E., Carney, J. P., and Hopfner, K. P. (2004). MlaA, a hexameric ATPase linked to the Mre11 complex in archaeal genomes. EMBO Rep *5*, 54-59.

Muchardt, C., and Yaniv, M. (2001). When the SWI/SNF complex remodels...the cell cycle. Oncogene *20*, 3067-3075.

Muller, C., and Leutz, A. (2001). Chromatin remodeling in development and differentiation. Curr Opin Genet Dev *11*, 167-174.

Narlikar, G. J., Phelan, M. L., and Kingston, R. E. (2001). Generation and interconversion of multiple distinct nucleosomal states as a mechanism for catalyzing chromatin fluidity. Mol Cell *8*, 1219-1230.

Osterod, M., Larsen, E., Le Page, F., Hengstler, J. G., Van Der Horst, G. T., Boiteux, S., Klungland, A., and Epe, B. (2002). A global DNA repair mechanism involving the Cockayne syndrome B (CSB) gene product can prevent the in vivo accumulation of endogenous oxidative DNA base damage. Oncogene *21*, 8232-8239.

Owen-Hughes, T. (2003). Colworth memorial lecture. Pathways for remodelling chromatin. Biochem Soc Trans *31*, 893-905.

Park, J. S., Marr, M. T., and Roberts, J. W. (2002). E. coli Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. Cell *109*, 757-767.

Paulsen, M., and Ferguson-Smith, A. C. (2005). DNA methylation in genomic imprinting, development, and disease. J Pathol *195*, 97-110.

Pazin, M. J., and Kadonaga, J. T. (1997). SWI2/SNF2 and related proteins: ATP-driven motors that disrupt protein-DNA interactions? Cell *88*, 737-740.

Peterson, C. L. (2002). Chromatin remodeling: nucleosomes bulging at the seams. Curr Biol *12*, R245-247.

Petukhova, G., Stratton, S., and Sung, P. (1998). Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. Nature *393*, 91-94.

Petukhova, G., Van Komen, S., Vergano, S., Klein, H., and Sung, P. (1999). Yeast Rad54 promotes Rad51-dependent homologous DNA pairing via ATP hydrolysis-driven change in DNA double helix conformation. J Biol Chem *274*, 29453-29462.

Picketts, D. J., Higgs, D. R., Bachoo, S., Blake, D. J., Quarrell, O. W., and Gibbons,R. J. (1996). ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome. Hum Mol Genet *5*, 1899-1907.

Poot, R. A., Dellaire, G., Hulsmann, B. B., Grimaldi, M. A., Corona, D. F., Becker, P. B., Bickmore, W. A., and Varga-Weisz, P. D. (2000). HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1 and two novel histone-fold proteins. Embo J *19*, 3377-3387.

Quinn, J., Fyrberg, A. M., Ganster, R. W., Schmidt, M. C., and Peterson, C. L. (1996). DNA-binding properties of the yeast SWI/SNF complex. Nature *379*, 844-847.

Raschle, M., Van Komen, S., Chi, P., Ellenberger, T., and Sung, P. (2004). Multiple interactions with the Rad51 recombinase govern the homologous recombination function of Rad54. J Biol Chem *279*, 51973-51980.

Richmond, E., and Peterson, C. L. (1996). Functional analysis of the DNA-stimulated ATPase domain of yeast SWI2/SNF2. Nucleic Acids Res *24*, 3685-3692.

Ristic, D., Wyman, C., Paulusma, C., and Kanaar, R. (2001). The architecture of the human Rad54-DNA complex provides evidence for protein translocation along DNA.
Proc Natl Acad Sci U S A *98*, 8454-8460.

Rouleau, N., Domans'kyi, A., Reeben, M., Moilanen, A. M., Havas, K., Kang, Z.,
Owen-Hughes, T., Palvimo, J. J., and Janne, O. A. (2002). Novel ATPase of SNF2-like protein family interacts with androgen receptor and modulates androgen-dependent transcription. Mol Biol Cell *13*, 2106-2119.

Saha, A., Wittmeyer, J., and Cairns, B. R. (2002). Chromatin remodeling by RSC involves ATP-dependent DNA translocation. Genes Dev *16*, 2120-2134.

Sawa, H., Kouike, H., and Okano, H. (2000). Components of the SWI/SNF complex are required for asymmetric cell division in C. elegans. Mol Cell *6*, 617-624.

Shen, X., Mizuguchi, G., Hamiche, A., and Wu, C. (2000). A chromatin remodelling complex involved in transcription and DNA processing. Nature *406*, 541-544.

Singleton, M. R., Scaife, S., and Wigley, D. B. (2001). Structural analysis of DNA replication fork reversal by RecG. Cell *107*, 79-89.

Singleton, M. R., and Wigley, D. B. (2002). Modularity and specialization in superfamily 1 and 2 helicases. J Bacteriol *184*, 1819-1826.

Smith, G. D., Nagar, B., Rini, J. M., Hauptman, H. A., and Blessing, R. H. (1998). The use of SnB to determine an anomalous scattering substructure. Acta Crystallogr D Biol Crystallogr *54 (Pt 5)*, 799-804.

Soejima, H., and Wagstaff, J. (2005). Imprinting centers, chromatin structure, and disease. J Cell Biochem *95*, 226-233.

Soultanas, P., Dillingham, M. S., Wiley, P., Webb, M. R., and Wigley, D. B. (2000). Uncoupling DNA translocation and helicase activity in PcrA: direct evidence for an active mechanism. Embo J *19*, 3799-3810.

Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Studitsky, V. M., Clark, D. J., and Felsenfeld, G. (1994). A histone octamer can step around a transcribing polymerase without leaving the template. Cell *76*, 371-382.

Subramanya, H. S., Bird, L. E., Brannigan, J. A., and Wigley, D. B. (1996). Crystal structure of a DExx box DNA helicase. Nature *384*, 379-383.

Sukhodolets, M. V., Cabrera, J. E., Zhi, H., and Jin, D. J. (2001). RapA, a bacterial homolog of SWI2/SNF2, stimulates RNA polymerase recycling in transcription. Genes Dev *15*, 3330-3341.

Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B., and Luger, K. (2003). Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. J Mol Biol *326*, 371-380.

Svejstrup, J. Q. (2003). Rescue of arrested RNA polymerase II complexes. J Cell Sci *116*, 447-451.

Swagemakers, S. M., Essers, J., de Wit, J., Hoeijmakers, J. H., and Kanaar, R. (1998). The human RAD54 recombinational DNA repair protein is a double-stranded DNA-dependent ATPase. J Biol Chem *273*, 28292-28297.

Thoma, N. H., Czyzewski, B. K., Alexeev, A. A., Mazin, A. V., Kowalczykowski, S. C., and Pavletich, N. P. (2005). Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54. Nat Struct Mol Biol *12*, 350-356.

Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D., and Hoeijmakers, J. H. (1992). ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. Cell *71*, 939-953.

Tsukiyama, T., Daniel, C., Tamkun, J., and Wu, C. (1995). ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. Cell *83*, 1021-1026.

Tsukiyama, T., and Wu, C. (1995). Purification and properties of an ATP-dependent nucleosome remodeling factor. Cell *83*, 1011-1020.

Turner, B. M. (2000). Histone acetylation and an epigenetic code. Bioessays *22*, 836-845.

Tuteja, N., and Tuteja, R. (2004). Unraveling DNA helicases. Motif, structure, mechanism and function. Eur J Biochem *271*, 1849-1863.

van den Boom, V., Jaspers, N. G., and Vermeulen, W. (2002). When machines get stuck--obstructed RNA polymerase II: displacement, degradation or suicide. Bioessays *24*, 780-784.

Varga-Weisz, P. D., and Becker, P. B. (1998). Chromatin-remodeling factors: machines that regulate? Curr Opin Cell Biol *10*, 346-353.

Varga-Weisz, P. D., Wilm, M., Bonte, E., Dumas, K., Mann, M., and Becker, P. B. (1997). Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. Nature *388*, 598-602.

Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S., and Wigley, D. B. (1999). Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. Cell *97*, 75-84.

Vignali, M., Hassan, A. H., Neely, K. E., and Workman, J. L. (2000). ATP-dependent chromatin-remodeling complexes. Mol Cell Biol *20*, 1899-1910.

Villard, L., and Fontes, M. (2002). Alpha-thalassemia/mental retardation syndrome, X-Linked (ATR-X, MIM #301040, ATR-X/XNP/XH2 gene MIM #300032). Eur J Hum Genet *10*, 223-225.

Whitehouse, I., Stockdale, C., Flaus, A., Szczelkun, M. D., and Owen-Hughes, T. (2003). Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme. Mol Cell Biol *23*, 1935-1945.

Wolffe, A. P. (1994). Transcriptional activation. Switched-on chromatin. Curr Biol *4*, 525-528.

Woodage, T., Basrai, M. A., Baxevanis, A. D., Hieter, P., and Collins, F. S. (1997). Characterization of the CHD family of proteins. Proc Natl Acad Sci U S A *94*, 11472-11477.

Xue, Y., Gibbons, R., Yan, Z., Yang, D., McDowell, T. L., Sechi, S., Qin, J., Zhou, S., Higgs, D., and Wang, W. (2003). The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies. Proc Natl Acad Sci U S A *100*, 10635-10640.

Yang, X. J., and Seto, E. (2003). Collaborative spirit of histone deacetylases in regulating chromatin structure and gene expression. Curr Opin Genet Dev *13*, 143-153.

Yu, S., Owen-Hughes, T., Friedberg, E. C., Waters, R., and Reed, S. H. (2004). The yeast Rad7/Rad16/Abf1 complex generates superhelical torsion in DNA that is required for nucleotide excision repair. DNA Repair (Amst) *3*, 277-287.

Zlatanova, J., Leuba, S. H., and van Holde, K. (1999). Chromatin structure revisited. Crit Rev Eukaryot Gene Expr *9*, 245-255.

Zofall, M., Persinger, J., and Bartholomew, B. (2004). Functional role of extranucleosomal DNA and the entry site of the nucleosome in chromatin remodeling by ISW2. Mol Cell Biol *24*, 10047-10057.

Appendix

Recipe for defined **LeMASTER MEDIUM**

AUTOCLAVE SOLUTION A

		1L	2L	4L
			(in g)	
1	Alanine	0.500	1	2.000
2	Arginine HCl	0.580	1.16	2.320
3	Aspartic Acid	0.400	0.8	1.600
4	L-Cystine	0.033	0.066	0.132
5	Glutamic Acid (Na)	0.750	1.5	3.000
6	Glutamine	0.333	0.666	1.332
7	Glycine	0.540	1.08	2.160
8	Histidine	0.060	0.12	0.240
9	Isoleucine	0.230	0.46	0.920
10	Leucine	0.230	0.46	0.920
11	Lysine HCl	0.420	0.84	1.680
12	Phenylalanine	0.133	0.266	0.532
13	Proline	0.100	0.2	0.400
14	L-Serine	2.083	4.166	8.332
15	Threonine	0.230	0.46	0.920
16	Tyrosine	0.170	0.34	0.680
17	Valine	0.230	0.46	0.920
18	Adenine	0.500	1	2.000
19	Guanosine	0.670	1.34	2.680
20	Thymine	0.170	0.34	0.680
21	Uracil	0.500	1	2.000
22	Sodium Acetate	1.500	3	6.000
23	Succinic Acid	1.500	3	6.000
24	Ammonium Cl	0.750	1.5	3.000
25	Sodium Hydroxide	0.850	1.7	3.400
26	Dibasic K_2HPO_4	10.500	21	42.000

→   let cool to 37°C, then add solution B (100 ml solution B / 1L autoclaved solutionA)

FILTERSTERILIZE SOLUTION B

		100 ml	200 ml	400 ml
			(in g)	
27	D-Glucose	10.00	20	40.00
28	$MgSO_4$-7$H_2$0	0.25	0.5	1.00
29	$FeSO_4$	4.2 mg	8.4 mg	16.8 mg
30	H_2SO_4 (liquid)	8 µL	16 µL	32 µL
31	Thiamine	0.005	0.01	0.020

→ disolve Seleno-Methionine in sterile H_2O and add 50 mg / 1L LeMaster Medium

Posener Str. 5
81929 München

Tel.: 0172/8623366
e-mail: duerr@lmb.uni-muenchen.de

Harald Dürr

Date of birth: 26.07.1974, Donauwörth

Marital status   married

Education: 1985-1994
Gymnasium Donauwörth, 7/94 Abitur

Military service: 1994 -1995
Gebirgsjägerbataillon 233 in Mittenwald

University studies: WS 95/96 – WS 00/01
Biology and Chemistry for mastership in secondary schools and Biology for Diploma at the LMU Munich,

August 2000
Diploma thesis with the title „Photoreaktion und Chromophorbindung der LOV-Domänen des Phototropins bei *Avena sativa*" at Botanical Institute Munich

July 2001
"erstes Staatsexamen" in Biology and Chemistry

September 2001
Degree of Diploma in Biology

since October 2001
PhD thesis in the laboratory of Prof. Dr. Karl-Peter Hopfner at the gene center Munich, LMU

August 2002 – August 2004
PhD fellowship of Boehringer Ingelheim Fonds

München 06.07.2005