Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Structural Framework for the Mechanism of Archaeal Exosomes in RNA processing

Structural Insights into DNA Duplex Separation by the Archaeal Superfamily 2 Helicase Hel308

Katharina Büttner

aus

Wolfratshausen

2007

The presented thesis was prepared in the time from December 2003 to March 2007 in the laboratory of Professor Dr. Karl-Peter Hopfner at the Gene Center of the Ludwig-Maximilians-University of Munich (LMU).

Parts of this PhD thesis have been published:

Büttner K., Nehring S., and Hopfner K.P., (2007).
Structural basis for DNA duplex separation by a Superfamily 2 Helicase.
NSMB 2007, Jul;14(7):647-52.

Büttner K. and Hopfner K.P., (2006).
Das Exosom.
Bioforum 2006; 29:19-21.

Büttner K., Wenig K., and Hopfner K.P., (2006).
The exosome: a macromolecular cage for controlled RNA degradation.
Mol. Microbiology 2006;61:1372-1379.

Büttner K., Wenig K.*.* and Hopfner K.P., (2005).
Structural Framework for the Mechanism of Archaeal Exosomes in RNA Processing. *Mol Cell.* 2005 Nov 11;20(3):461-71.

Other publications:

Karcher A., **Büttner K**., Märtens B., Jansen R.P. and Hopfner K.P., (2005).
X-ray structure of RLI, an essential twin cassette ABC ATPase involved in ribosome biogenesis and HIV capsid assembly. *Structure*. 2005 Apr. 13(4):649-59.

Presentations at international conferences:

Büttner K., Wenig K., and Hopfner KP.

Structural Framework for the Mechanism of Archaeal Exosomes in RNA Processing

Hybrid Methods; March 15-19, 2006, Lake Tahoe, CA, USA (Poster)

Our Father who art in heaven,
Hallowed be Thy name.
Thy kingdom come.
Thy will be done on Earth,
as it is in heaven.
Give us this day our daily bread.
And forgive us our debts,
as we forgive our debtors.
And lead us not into temptation.
But deliver us from evil.
For Thine is the kingdom,
and the power,
and the glory.
Forever,
Amen

Matthew, 6:8-13

PART I

Structural Framework for the Mechanism of Archaeal Exosomes in RNA processing

Table of contents

1 Introduction

The processing and degradation of cellular RNA is one of the most important processes in all living cells. Additional to the amount of RNA that is synthezised by a cell per minute (up to 2000 molecules of ribosomal (r)RNA as well as thousands of messenger (m)RNAs, transfer (t)RNAs, small nuclear (sn)RNAs and small nucleolar (sno)RNAs (Warner, 1999)), almost every RNA species undergoes several posttranscriptional processing reactions to produce functional RNA molecules. Furthermore, defective (pre-) mRNAs and nuclear RNA precursors rarely undergo repair processes as known for DNA, but are filtered out by quality control mechanisms and targeted for rapid degradation similar to proteins. In addition, posttranricriptionally regulated mRNA degradation modulate the celluar level of RNA, thereby influencing translation and gene expression. Thus, RNA processing and degradation events have to be constantely monitored by surveillance systems, which are able to distinguish between defective and non-defective RNA as well as between mature and precursor RNA.

In eukaryotes several enzymes which participate in the 3´$\rightarrow$ 5´ RNA surveillance machinery are organized into a ~ 400 kDA multisubunit complex, called the exosome (Mitchell et al., 1997). In the last years, this multiprotein complex has been shown to directly participate in the processing and degradation of almost every cellular RNA species (Allmang *et al.,* 1999; Allmang *et al.,* 2000; van Hoof *et al.,* 2000; Hilleren *et al.,* 2001; Suzuki *et al.,* 2001; Andrulis *et al.,* 2002; Torchet *et al.,* 2002; Orban and Izaurralde 2005). Thus, the exosome emerges as key component in the cellular 3´$\rightarrow$ 5´ RNA surveillance system in eukaryotes.

1.1 The exosome is conserved in eukaryotes and archaea

The human exosome was originally discovered in 1977 as an autoantigen in sera of patients suffering from polymyositis (PM) and scleroderma (Scl) overlap syndromes and was termed the "PM/Scl antigen" (Wolfe *et al.,* 1977; Reichlin *et al.,* 1984; Treadwell *et al.,* 1984). The PM/Scl antigen was shown to be a nuclear complex consisting of around 11 to 16 different proteins (Targoff and Reichlin 1985; Reimer *et al.,* 1986). The yeast exosome was discovered 20 years later by its role in the processing of 7S rRNA to mature 5.8S rRNA, an essential component of the ribosome (Mitchell *et al.,* 1997), and further analysis showed

that the PM/Scl complex is the human equivalent of this complex (Allmang *et al.,* 1999). The yeast complex, consisting of around ten different protein subunits, was designated "exosome" because of its exclusive 3´→ 5´ exoribonuclease activity. The core of human and yeast exosomes is assembled out of nine proteins, of which six have a molecular architecture similar to that of the eubacterial phosphodependent 3´→ 5´ exoribonucleases RNase PH and PNPase, (in yeast: Ribosomal RNA processing factor (Rrp)41p, Rrp42p, Rrp43p, Rrp45p, Rrp46p, mRNA transport regulator (Mtr)3p) (Symmons *et al.,* 2000; Koonin *et al.,* 2001; Harlow *et al.,* 2004) (Table 1). The remaining three subunits contain a putative S1 and KH RNA binding domain (Rrp4p, Rrp40p, cep1 synthetic lethality (Csl)4p) (Mitchell *et al.,* 1997; Allmang *et al.,* 1999; Mitchell and Tollervey 2000; Estevez *et al.,* 2001; Chekanova *et al.,* 2002). The tenth subunit in both human and yeast is the RNase R-like hydrolytic 3´→ 5´ exoribonuclease Dis3 (Rrp44 in human), which was shown to be responsible for RNase activity of the yeast exosome (Mitchell *et al.,* 1997; Dziembowski *et al.,* 2007). The eukaryotic exosome core is found in both the cytoplasm and the nucleus (Mitchell *et al.,* 1997; Allmang *et al.,* 1999; Mitchell and Tollervey 2000). The nuclear exosome interacts with an additional RNase D-like 3´→ 5´ exoribonuclease, Rrp6, that has a non-essential role in final 5.8S rRNA processing (Zhou and Deutscher 1997; Briggs *et al.,* 1998; Mitchell and Tollervey 2000) and Rrp47, which cooperates with Rrp6 (Mitchell and Tollervey 2003).

Table 1: Components of the yeast and archaeal exosome

Yeast	Archaea	Domains	Activity
Core proteins			
Rrp41p	aRrp41	RNase PH	Eukaryotes: 3´- 5´ phosphorolytic exoribonuclease[a]
			Archaea: 3´- 5´ phosphorolytic exoribonuclease and adenylase[b]
Rrp42p	aRrp42	RNase PH	
Rrp43p	aRrp42	RNase PH	
Rrp45p	aRrp42	RNase PH	
Rrp46p	aRrp41	RNase PH	Archaea: 3´-5´ phosphorolytic exoribonuclease and adenylase[b]
Mtr3p	aRrp41	RNase PH	Archaea: 3´-5´ phosphorolytic exoribonuclease and adenylase[b]
Csl4p	aCsl4	NT/S1/ZnR[c]	RNA binding[d]
Rrp4p	aRrp4	NT/S1/KH	Eukaryotes: 3´- 5´ hydrolytic exoribonuclease[e]
			RNA binding[f]
Rrp40p	aRrp4	NT/S1/KH	RNA binding[d]
Additional components			
Rrp6p		RNase D	3´- 5´ hydrolytic exoribonuclease[g]
Rrp44p		RNase R	3´- 5´ hydrolytic exoribonuclease[a]
	aDnaG	Toprim domain	unknown[h]

a. Mitchell *et al.* (1997).
b. Buttner *et al.* (2005); Lorentzen *et al.* (2005).
c. Eukaryotic Csl4 does not have the zinc co-ordinating cysteine residues of the aCsl4.
d. Predicted from the domain architecture.
e. Mitchell *et al.* (1997); Chekanova *et al.* (2002).
f. Chekanova *et al.* (2002); Buttner *et al.* (2005).
g. Burkard and Butler (2000).
h. Eubacterial DnaG synthesizes short RNA oligonucleotides on a ssDNA template (primase).

Homologs of yeast and human exosomal proteins were found in as diverse organisms as plants, flies and trypanosomes (Estevez *et al.,* 2001; Andrulis *et al.,* 2002; Chekanova *et al.,* 2002). Furthermore, homologs of four out of ten subunits of the complete eukaryotic complex were found in archaebacteria (Koonin *et al.,* 2001; Evguenieva-Hackenberg *et al.,* 2003) (Table 1). Archaeal exosomes exhibit a more simplified assembly. The core complex consists of only two Rnase PH-like proteins (Rrp41 and Rrp42) and only two putative RNA binding proteins with either S1 and KH (Rrp4) or S1 and Zn-Ribbon domains (Csl4) (Koonin *et al.,* 2001; Evguenieva-Hackenberg *et al.,* 2003).

However, the outstanding conservation of the exosomal subunits throughout different species suggests a conserved and fundamental role of the exosome in RNA maintenance in both eukaryotes and archaebacteria (Figure 1). In addition, the high similarity to the eubacterial 3´→ 5´ exoribonucleases RNase PH and PNPase implicate the presence of an evolutionary conserved RNA degradation mechanism (Koonin *et al.,* 2001).
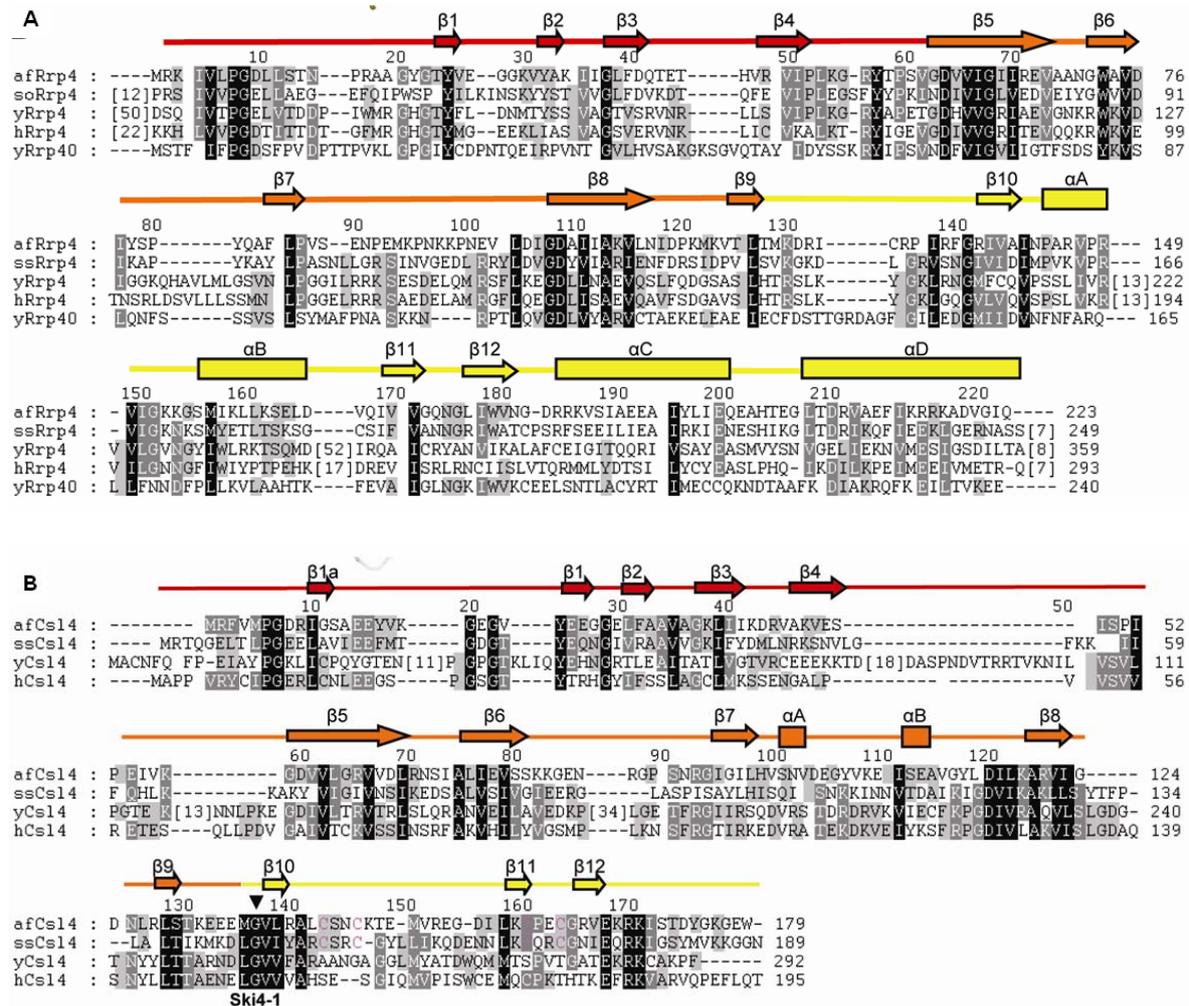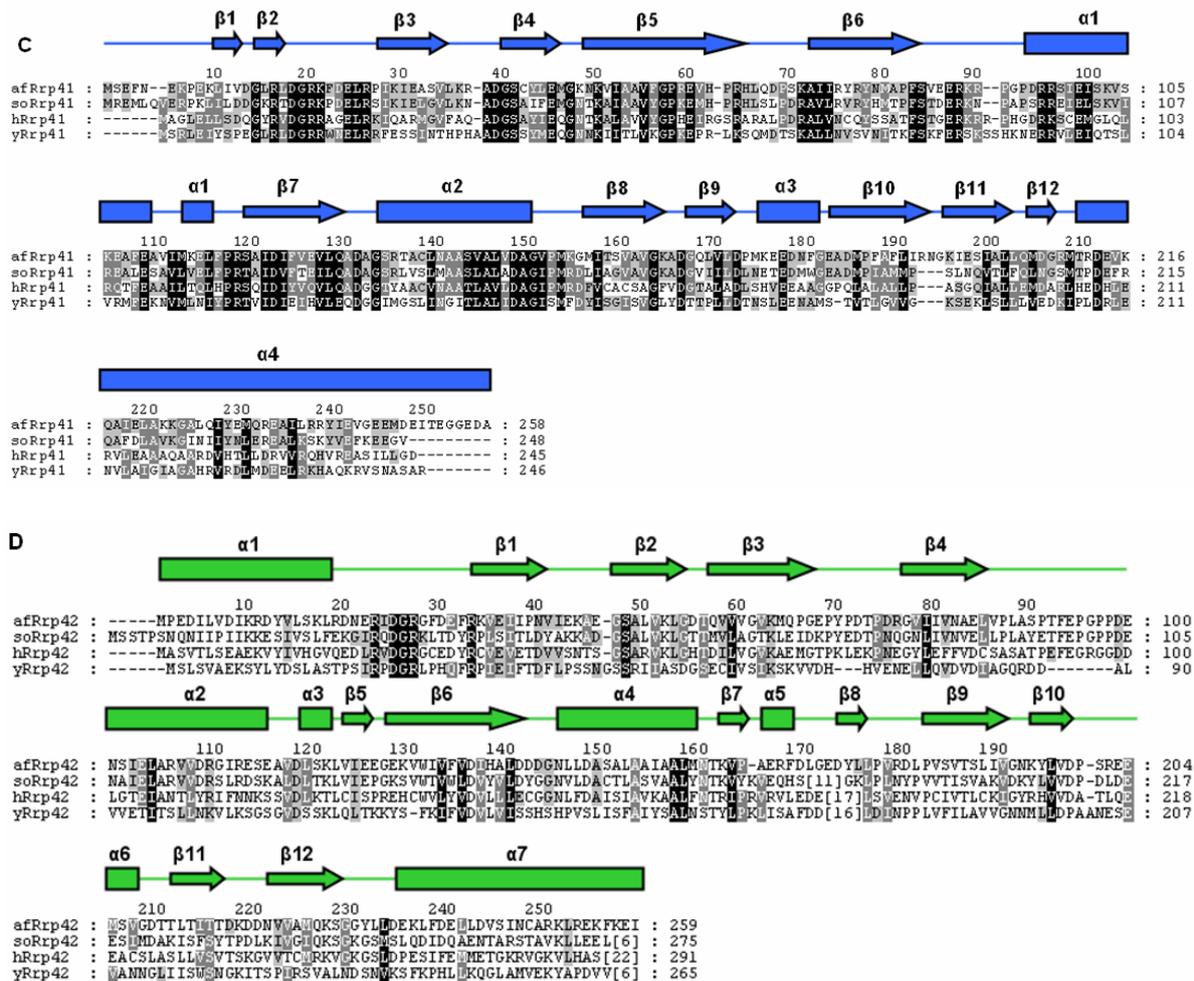
Figure 1: Structure-based sequence alignment of *Archaeoglobus fulgidus* (af) exosome subunits Rrp4 (A), Csl4 (B), Rrp41 (C) and Rrp42 (D) with the respective proteins from *Sulfolobus solfataricus* (so), human (h) and yeast (y). Conserved regions and residues are shaded light grey (low conservation), dark grey (moderate conservation) or black (high conservation). The annotated secondary structure of *A. fulgidus* proteins is shown on top of the alignments (arrows: β-strands, boxes: α-helices). Black triangle/Ski4-1 in B: relevant position of Csl4 point mutation (described in the text.).

1.2 Functions of the exosome and coactivator complexes

1.2.1 Cytosolic RNA decay and quality control

In *Saccharomyces cerevisiae* and human, mRNA is degraded via two possible pathways. Both pathways are initiated by shortening of the 3´-poly(A) tail by a deadenylase complex (Pan2-Pan3 and/or Ccr4-Caf1 in yeast). The major pathway in yeast (and the minor pathway in human cells) then follows subsequent decapping by Dcp1/Dcp2 and degradation by the 5´→ 3´ exonuclease Xrn1p

(Muhlrad *et al.,* 1994). All of the involved proteins are localized in cytoplasmic P-bodies, which are believed to be the active centres for 5´→ 3´ RNA decay (Sheth and Parker 2003). The major pathway in human cells (and the minor pathway in yeast) includes 3´→ 5´ exonucleolytic digestion by the exosome (Mitchell *et al.,* 1997; Decker 1998; Mitchell and Tollervey 2001), followed by degradation of the remaining cap structure by a "scavenger" decapping enzyme (Wang and Kiledjian 2001). However, it is not clear yet, which molecular factors target the mRNA for degradation and decide which of the two pathways is initialized.

The recruitment of the exosome to a different subset of cytosolic RNA species is mediated by sequence-and/or structure-specific RNA binding proteins, and the cytoplasmic activity of the yeast exosome has been shown to entirely depend on the presence of coactivators, like the SKI complex (Ski7p and Ski2p/Ski3p/Ski8p) (Anderson and Parker 1998; Brown *et al.,* 2000; van Hoof *et al.,* 2000; Araki *et al.,* 2001) (Figure 2). The interaction between the SKI complex and the exosome is thereby likely mediated via Ski7p, a GTP-binding protein (van Hoof *et al.,* 2000; Araki *et al.,* 2001), and the importance of the SKI complex is supported by the finding that mutations in SKI2, SKI3 and SKI8 inhibit 3´→ 5´ decay of mRNA *in vivo* (Anderson and Parker 1998; Brown *et al.,* 2000; van Hoof *et al.,* 2000). Furthermore, Ski2p is a member of the helicase superfamily 2 (SF2) and could be involved in the unwinding of RNA secondary structures prior to degradation.

The presence of the SKI complex is important for the activity of the exosome in normal mRNA turnover (Mitchell *et al.,* 1997). Furthermore, it has been shown that exosome and SKI complex function together with several other proteins like Upf1-Upf3 in the quality control and in the degradation of eukaryotic mRNAs containing a premature termination codon in a process termed nonsense-mediated decay (NMD) (Takahashi *et al.,* 2003) as well as in non-stop decay of mRNAs lacking a termination codon (Takahashi *et al.,* 2003).
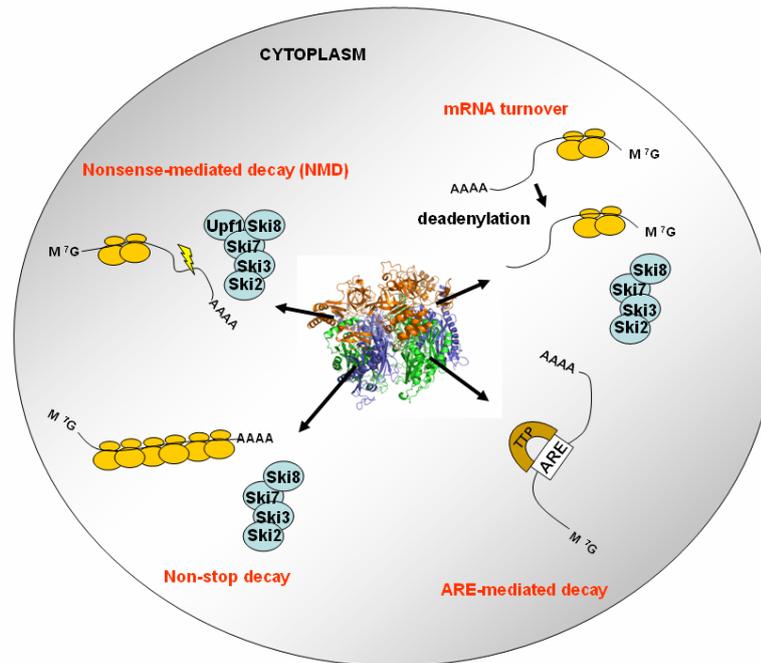
Figure 2: Functions and activation of the cytoplasmic exosome. M^{7}G: 7-methylguanosine. The yellow flash marks quality control pathways. Adapted from (Houseley *et al.,* 2006)

The detailed molecular mechanism of substrate specificity and exosome recruitment is still unclear. However, a subset of unstable mammalian mRNAs as well as some mRNAs in *Trypanosoma brucei* possesses AU-rich elements (AREs) in their 3´-untranslated region (UTR). By interacting with sequence-specific ARE-binding proteins, like Tristetraprolin (TTP), KH-type-splicing-regulatory-protein (KSRP), and RNA-helicase-associated-with-AU-rich-element (RHAU), the exosome is specifically recruited to the substrate RNA (Chen *et al.,* 2001; Wang and Kiledjian 2001; Mukherjee *et al.,* 2002; Haile *et al.,* 2003). This is in contrast to the sequence unspecific mRNA turnover and quality control pathways.

1.2.2 Nuclear RNA processing and degradation

Whereas the eukaryotic cytoplasmic exosome functions mainly in the normal turnover and quality control of (pre-) mRNA, the nuclear exosome takes part in the processing and degradation of several nuclear RNA species (Mitchell *et al.,* 1997; Allmang *et al.,* 1999) (Figure 3). According to the cytoplasmic SKI complex, the nuclear Trf4p/Air2p/Mtr4p polyadenylation (TRAMP) complex targets defective precursors of rRNAs, tRNAs, snRNAs, snoRNAs, and other transcripts for degradation by polyadenylation (de la Cruz *et al.,* 1998; Allmang *et al.,* 2000; Suzuki *et al.,* 2001; LaCava *et al.,* 2005; Wyers *et al.,* 2005), and thus functions

together with the exosome in sequence-independent RNA surveillance. Many snoRNAs function in rRNA processing and rRNA nucleotide modification like methylation and pseudouridinylation, whereas snRNAs are involved in pre-mRNA splicing as components of the splicesosome (Allmang *et al.,* 1999; van Hoof *et al.,* 2000).

Furthermore, the processing and maturation of pre-rRNAs like 7S rRNA to 5.8S rRNA, and the 3′ trimming of snoRNAs and snRNAs is performed by the nuclear exosome and involves participation of the nuclear cofactors Rrp6 and Rrp47 as well as Mtr4 (Allmang *et al.,* 1999; van Hoof *et al.,* 2000; Mitchell *et al.,* 2003).
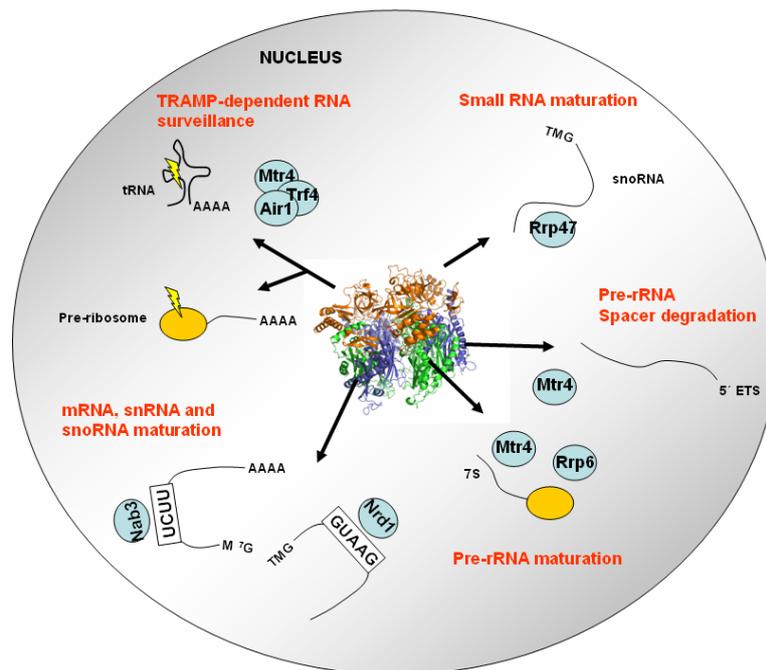
Figure 3: Functions and cofactors of the nuclear exosome. M^7G: 7-methylguanosine. The yellow flash marks quality control pathways. ETS: external transcribed spacer, TMG: trimethylguanosine. Adapted from (Houseley *et al.,* 2006)

The exosome has also been shown to degrade spacer elements that have been excised from rRNA precursors and during 3′ trimming of pre-rRNAs and snoRNAs, and there is emerging evidence that the exosome also degrades cryptic RNA PolII transcripts as well as RNA fragments arising at RNA interference processes (Orban and Izaurralde 2005; Wyers *et al.,* 2005). Sequence-dependent recruitment of the exosome to specific pre-mRNAs, snoRNAs and probably other RNAs that contain defined sequence motifs is likely mediated via interaction with sequence-specific RNA binding proteins like Nrd1 and Nab3 (Arigo *et al.,* 2006; Houalla *et al.,* 2006; Vasiljeva and Buratowski 2006).

1.3 Structure determination by X-Ray crystallography

The knowledge of the structure of a protein is one important step towards the knowledge of its detailed enzymatic function. Nowadays, there are well established techniques to determine the primary, secondary, and tertiary as well as the quaternary structure of a protein or protein complexes.

While the primary structure can be obtained via sequencing of the amino acid or the corresponding nucleic acid sequence, the determination of the secondary and tertiary structure requires detailed information about the arrangement of atoms within a protein. High resolution three dimensional structures of proteins are determined by nuclear magnetic resonance (NMR) or X-ray crystallography. While NMR is only applicable to smaller proteins with a maximum molecular weight of around 35 kDa, X-ray crystallography has no size limitation, but requires well-ordered protein crystals.

Low resolution quaternary structures of large protein complexes (MW > 250 kDa), organelles and cells can be obtained by (cryo-) electron microscopy (EM).

The following part briefly describes the theoretical background for structure determination by X-ray crystallography. For more detailed information, please see textbooks (Drenth 1999; McPherson 2001; Blow 2002; Massa 2002).

1.3.1 Crystallization

The prerequisite for X-ray crystallography is to grow protein crystals that diffract X-rays. Crystals, in where a protein molecule is packed in a repetitive and regular arrangement, form, when proteins are precipitated very slowly from supersaturated solutions. This thermodynamically driven process includes nucleation, growth of the crystal and growth termination, and is dependent on a variety of parameters, like pH, temperature, protein concentration, and nature of the precipitant. Many crystallization conditions have to be screened during a crystallization experiment to find the most useful crystallization condition. The most frequently used procedure for obtaining protein crystals is the vapour (sitting drop or hanging drop) diffusion method (McPherson, 1982).

In this thesis, the sitting drop vapour diffusion method was used. In this case, a buffered protein solution is mixed with precipitant solution. A reservoir contains a more concentrated precipitant solution. The reaction chamber is sealed and equilibrium between the drop and the container is slowly reached through vapour

diffusion. The precipitant concentration in the drop is increased by loss of water to the reservoir. Once the saturation point is reached, and other conditions such as pH and temperature are chosen correctly, protein crystals will occur in the drop, which may be suitable for X-ray diffraction experiments.

1.3.2 Theory of X-ray diffraction

X-rays are electromagnetic waves with a wavelength in the range of atomic distances (10^{-10} m = 1 Å). For X-ray diffraction experiments of protein crystals, usually X-rays between 0.8-1.6 Å are used. X-rays interact with electrons in the electron sheath of atoms in the protein crystal and cause dipole oscillation of the electrons at the X-ray frequency. The oscillating electrons emit X-rays with the same wavelength in every direction (coherent or elastic scattering).

As a crystal is composed of molecules located in a unit cell, which is periodically repeated in three dimensions, the emitted waves scattered from different atoms interfere, and normally cancel each other out. Only if the light path of waves differ by $n\lambda$ (n=integer), constructive interference is observed. The difference in the light path is thereby dependent on the distance between the scattering atoms. Thus, scattering can be described as reflection at imaginary lattice planes. The lattice planes pervade the crystal lattice and its lattice points. Intersections with the unit cell axes of the crystal lattice are termed Miller Indices (h,k,l), which describe the orientation of and the spacing between a set of parallel lattice planes. The conditions for constructive interference are given by Bragg´s Law (Bragg and Bragg. 1913):

$$n \cdot \lambda = 2 \cdot d \cdot \sin\theta$$

Only if the distance d between parallel lattice planes and the angle θ between the lattice planes and the incident beam follow Bragg´s Law, a reflection (h,k,l) is observed. The intensity of each reflection (h,k,l) represents the overall scattering from a particular set of parallel lattice planes (h,k,l). According to Friedel´s Law, reflections (h,k,l) and (-h,-k,-l) have the same magnitude, but opposite signs for phase angles (Friedel mates).

1.3.3 The Phase problem

X-rays are scattered at the electron sheath of atoms. The electrons are delocalized and their positions are described as a three dimensional distribution,

the so called electron density distribution ρ. As crystals are periodic assemblies of molecules, the electron density of a crystal resembles a periodic function. Thus, a Fourier Transformation (Fourier-Summation) can be applied to calculate the electron density for each point (x, y, z) in a crystal:

$$\rho(x,y,z) = \frac{1}{V} \sum_{hkl} F(h,k,l) \cdot \exp[-2\pi \cdot i(hx + ky + lz)]$$

This formula represents the summation over all structure factors F(h,k,l) for each position (x,y,z) in a normalized unit cell (1/V).

The structure factor F(h,k,l) (the Fourier coefficient) is the sum of scattering contribution of all atoms *j* in a unit cell, with (x,y,z) as fractional coordinates, to a reflection (h,k,l), and can be derived from the following equation.

$$F(h,k,l) = \sum_{j=1}^{N} f_j \cdot \exp[2\pi \cdot i(hx_j + ky_j + lz_j)] \cdot \exp[-B_i \cdot \sin^2 \theta / \lambda^2]$$

The atomic scattering factor (or form factor) f_j describes the scattering power of an atom *j*, which is dependent on the atom type and the diffraction angle (resolution) of the corresponding reflection. The Debye-Waller- or B-Factor, the last term in this equation, represents the contribution of thermal disorder or "vibration" (isotrop/anisotrop).

The electron density ρ is related to F(h,k,l) by a direct Fourier Transformation (Fourier-Integral):

$$F(h,k,l) = V \cdot \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(x,y,z) \cdot \exp[2\pi \cdot (hx + ky + lz)]\, dx\, dy\, dz$$

Thus, the structure factor F(h,k,l) can be easily calculated from a known structure. But *de novo* structure determination deals with the inverse problem.

F(h,k,l) is a complex number, which is formed by the amplitude IF(h,k,l)I and the phase αhkl of a scattered wave:

$$F(h,k,l) = |F(h,k,l)| \cdot \exp[i \cdot \alpha(h,k,l)]$$

Thus, the already mentioned inverse Fourier Transformation can be written as

$$\rho(x,y,z) = \frac{1}{V} \sum_{hkl} |F(h,k,l)| \cdot \exp[-2\pi \cdot i(hx + ky + lz - \alpha_{(h,k,l)})],$$

where the phase term αhkl becomes evident. Whereas the amplitude IF(h,k,l)I of a scattered wave is proportional to the square root of the measured intensity (I = $|F|^2$) for each reflection (h,k,l), and directly available from the diffraction pattern, the phase information is lost during measurement and thus accounts for the

unknown component of the complex number F(h,k,l). This is the so called "Phase problem" in crystallography.

1.3.4 Solving the phase problem

To overcome the phase problem in *de novo* structure determination, several approaches are applied including isomorphous replacement (MIR/SIR) and anomalous scattering techniques (MAD/SAD), or a combination of both methods (MIRAS/SIRAS), which require the attachment of a heavy atom and/or the presence of anomalous scattering atoms, respectively, to the protein in the crystal. If parts of the structure or the structure of a close homologue are known, phases can be obtained by molecular replacement.

In this thesis, single- and multiple anomalous scattering as well as molecular replacement techniques were used to obtain phases. Thus, these two methods will be described briefly in the following part.

1.3.4.1 Patterson Function

The Patterson function is an important tool to obtain phase angles in *de novo* structure determination. It is a Fourier Transformation of the measured intensities (its squared reflection amplitudes $|F|^2$), which do not depend on phases (phase angle = 0°):

$$P(u,v,w) = 1/V \sum_{hkl} |F_{hkl}|^2 \cos[2\pi(hu + kv + lw)]$$

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

1.3.4.2 Single- and multiple anomalous dispersion (SAD/MAD)

Diffraction is usually described as diffraction on "free" electrons, where the emitted beam undergoes neither changes in the intensity nor a phase delay. Near the absorption edge of an atom, if the energy of the incident beam is close to its eigenfrequency, resonance effects are observed. While some photons are absorbed and re-emitted at lower energy (fluorescence), some are emitted at the

same energy but with a phase shift. This phenomenon is described as anomalous dispersion. In case of anomalous dispersion, the atomic scattering factor, which is normally given for wavelengths far from the absorption edge of an atom, gains an anomalous contribution, which is composed of a real and an imaginary part (Figure 4).

$$f_{ano} = f + \Delta f + if'' = f' + if''$$
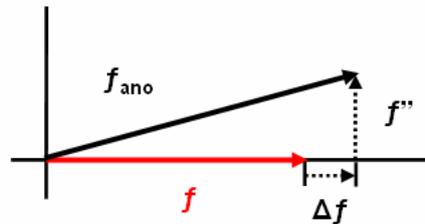
Figure 4: The anomalous contribution of the anomalous scattering factor f_{ano} is composed of a real part Δf and an imaginary part if''. The phase of the imaginary part is always shifted by 90°.

As normally occurring atoms like carbon, nitrogen and hydrogen do not scatter anomalously at wavelengths used in X-ray diffraction experiments, heavy atoms like selenium, platin or mercury, which show detectable anomalous scattering, are introduced into the protein.

The scattering is now a sum F_{PH} of the normal scattering from light atoms F_P and scattering from heavy atoms F_H with a normal (F_{HN}) and an anomalous (F_{HA}) contribution:

$$F_{PH} = F_P + F_{HN} + F_{HA} = F_P + F_H$$

While in normal scattering, the structure factors $F_P(h,k,l)$ and $F_P(-h,-k,-l)$ have the same amplitudes and opposite phases (Friedel Mates) according to Friedel´s Law, anomalous scattering causes violation of Friedel´s Law, and $F_{PH}(h,k,l)$ and $F_{PH}(-h,-k,-l)$ (now called Bijvoet pair) do not have the same magnitudes.

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

A Patterson map calculated with the Bijvoet differences $(\Delta|F|_{ano})^2$ results in a map showing only interatomic distance vectors between the anomalous scatterers, and allows localization of the heavy atom substructure in the unit cell. This allows the determination of the structure factor F_{HA} (amplitude and phase angle) of the heavy atoms and the calculation of its contribution to the structure factor F_H in SAD

experiments. The following equation allows calculation of F_P and the estimation of protein phase angles:

$$F_{PH} = \mid F_P \mid \cdot \alpha_P + \mid F_H \mid \cdot \alpha_H$$

As anomalous scattering depends on the wavelength, usually three wavelengths (peak, inflection and remote) are recorded in an MAD experiment. At the peak wavelength, $f"$ is maximal. f' is minimal at the inflection point and close to normal at the remote wavelength. The differences in the anomalous contribution (dispersive difference) between the three wavelengths can be used to define F_P more exactly from three different F_{HA} values. As the resonance wavelength of certain atoms are not only dependent on the atom itself but also on the chemical environment, the required wavelengths are usually determined experimentally by a fluorescence scan prior to data collection. Thereby $f"$ can be derived and f' can be calculated by the Kramer-Kronig equation for each given wavelength.

1.3.4.3 Molecular replacement

Molecular replacement is possible when parts of a protein structure or the structure of a close structural homologue of a protein are already known. The phase angles of the unknown structure are thereby obtained from the known structure. For this, the known model has to be rotated and translated into the electron density of the unknown structure. The replacement is a 6-dimensional search problem (or two 3 dimensional searches), which can be solved with the Patterson function. During translation and rotation search, the Patterson maps of the model structure and of the crystal diffraction data are compared. The intramolecular vectors depend only on the orientation of the molecule and are used for the rotation search. The radius of the unit cell thereby limits the length of the included vectors. The translation search uses intermolecular cross vectors, which depend on both the orientation and the position of the molecule. The derived coordinates of the molecules in the unit cell then allow the calculation of new structure factor amplitudes $|F_{calc}|$ and an estimation of the respective phase angles α_{calc}. The following equation than allows the calculation of an electron density for the new crystal structure ($|F_{obs}|$ = experimentally derived structure factor amplitudes):

$$F_{(h)} = (\mid F_{obs} \mid - \mid F_{calc}(h) \mid) \cdot \exp[i\alpha_{calc}(h)]$$

The calculated F_o-F_c density is useful for finding corrections between the new and the the known model. But it has always a bias towards the model structure, from which the phase angles were derived. To minimize model bias, usually the following equation is used:

$$F_{(h)} = (2 \mid F_{obs} \mid - \mid F_{calc}(h) \mid) \cdot \exp\left[i\alpha_{calc}(h)\right]$$

This $2F_o$-F_c electron density is used to build in the new structure or to change variations, to find a closer agreement between the calculated and observed structure factors. In the refinement procedure a reliability factor, the so called R-factor, is used to monitor the quality of the model.

1.4 Objectives

The exosome is a conserved multisubunit exoribonuclease complex, which was originally discovered as essential component in the 3´ maturation of ribosomal 5.8S rRNA in *Saccharomyces cerevisiae*. Since then, exosomes have been identified in humans, plants and flies as well as in archaebacteria. Functional studies showed that the exosome is not only essential for viability in yeast, but also involed in nearly all cellular RNA processing and degradation pathways. Thus, the exosome emerges as key component in the 3´→ 5´ RNA surveillance and quality control machinery. However, the detailed structural and molecular basis of regulated 3´ RNA degradation, substrate specificity and requirement of coactivators remain poorly understood.

The aim of this PhD thesis was to determine the structure of the nine subunit exosome from the archaeon *Archaeoglobus fulgidus* using X-ray crystallography. Crystal soaking experiments with the phosphate-mimic tungstate should locate phosphorolytic active centres and possible RNA binding sites. Furthermore, functional analysis of RNase and RNA binding activity of wild-type and structure-guided mutant exosomes should help to provide a model, which explains the underlying mechanism of the dual role of exosomes in processive 3´→ 5´ RNA degradation and precisely defined RNA processing, and provide a structural basis for its dependence on coactivator complexes. Due to the high conservation of exosome subunits in archaea and eukaryotes, the results may provide important insights, which can be extrapolated to the more complex eukaryotic exosome.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

All common chemicals were obtained from Merck (Darmstadt, Germany), Roth (Karlsruhe, Germany) and Sigma (Deisenhofen, Germany), unless otherwise stated. Enzymes and nucleotides for molecular biology were ordered by MBI Fermentas (St. Leon-Rot, Germany). Chromatographic material and columns as well as radioactive material (γ^{32}-P-ATP) were purchased from GE Healthcare (Freiburg, Germany). Crystallization screens, crystallization grade reagents and crystallization tools were obtained from Hampton Research (Aliso Viejo, USA), Nextal (Montreal, Canada; now Qiagen, Hilden, Germany) and Jena Bioscience (Jena, Germany). RP-HPLC purified DNA and RNA for cloning, crystallization and activity assays were ordered by Thermo Electron Corporation (Ulm, Germany) and Biomers (Ulm, Germany), respectively.

2.1.2 Bacterial strains

Table 2: *E. coli* strains

Strain	Description	Source of reference
XL-1 blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44relA1 *lac*[F´*proAB lacI^{q}ZΔdM15Tn10*(Tet^r)]	Stratagene, La Jolla, USA
Rosetta (DE3)	F- *ompT hsdS$_B$* (r$_B^-$ m$_B^-$) *gal dcm lacY1* (DE3) pRARE (Cm^R)	Novagen; Schwalbach/Ts., Germany
B834	transformed with pRARE (Cm^R) isolated from Rosetta (DE3) cells	Novagen; Schwalbach/Ts., Germany

2.1.3 Media and supplements

Table 3: Growth media

Media	Description	Source of reference
LB	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl (+ 1,5% (w/v) agar for selective media plates)	Miller, 1972
LeMaster´s		LeMaster, 1985

Autoclavable portion A (g/2000 ml)

Alanine	1.0	Serine	4.166
Arginine hydrochloride	1.16	Threonine	0.46
Aspartic acid	0.8	Tyrosine	0.34
Cystine	0.066	Valine	0.46
Glutamic acid	1.5	Adenine	1.0
Glutamine	0.666	Guanosine	1.34
Glycine	1.08	Thymine	0.34
Histidine	0.12	Uracil	1.0
Isoleucine	0.46	Sodium acetate	3.0
Leucine	0.46	Succinic acid	3.0
Lysine hydrochloride	0.84	Ammonium chloride	1.5
Phenylalanine	0.266	Sodium hydroxide	1.7
Proline	0.2	Dibasic potassium phosphate	21.0

Non-autoclavable portion B (200 ml)

Glucose	20.0 g
Magnesium sulphatex7H$_2$O	0.5 g
Iron sulphate	8.4 g
conc. Sulfuric acid	16.0 µl
Thiamine	10.0 mg

All amino acids were reagent-grade L-enantiomers purchased from Sigma (Deisenhofen, Germany). Solution A was autoclaved, cooled to 37°C and supplemented with filter-sterilized Solution B (200 ml B/2000 ml A) and

selenomethionine (dissolved in H_2O/HCl; 125 mg/2200 ml Solution A/B), Calbiochem, Schwalbach/Ts., Germany).

2.1.4 Buffers and solutions

The following table contains buffers and solutions which were used in this work. Buffers for specific protein purifications and biochemical assays are given separately in the corresponding part.

Table 4: Buffers, dyes and solutions

SDS-PAGE	
4x stacking gel buffer	0.5 M Tris, 0.4% (w/v) SDS, pH 6.8 (25°C)
4x separation gel buffer	3 M Tris, 0,4 % (w/v) SDS, pH 8.9 (25°C)
Electrophoresis buffer, 1x TGS	190 mM glycine, 50 mM Tris, 0.1 % (w/v) SDS
5x Loading Dye	50% glycerol, 250 mM Tris/HCl pH 6.8 (25°C), 7.5% SDS, 5 mM EDTA, 10 mM DTT, 0.5% bromphenolblue
Coomassie staining	
Staining solution	50% (v/v) ethanol, 7% (v/v) acetic acid, 0.2% Coomassie Brilliant blue R-250
Gel electrophoresis of nucleic acids	
Electrophoresis buffer, 1x TAE	40 mM Tris, 20 mM Acetic Acid, 1 mM Na_2EDTA
6x Loading Dye	1.5 g/L bromphenol blue, 1.5 g/L xylene cyanol, 50% (v/v) glycerol
Buffers for preparation of chemically competent cells	
TFB-1	30 mM KOAc, 50 mM $MnCl_2$, 100 mM KCl, 10 mM $CaCl_2$, 15% (v/v) glycerol, pH 5.8 (25°C)
TFB-2	10 mM MOPS, 10 mM KCl, 75 mM $CaCl_2$, 15% glycerol, pH 7.0 (25°C)

2.2 Methods

2.2.1 Bioinformatic Methods

2.2.1.1 Homology searches and alignments

DNA and protein sequences of archaeal proteins were found using the NCBI database (http://www.ncbi.nlm.nih.gov). Homology searches were performed using the NCBI Basic Local Alignment Search Tool (BLAST) server (http://www.ncbi.nlm.nih.gov/BLAST). Multiple sequence alignments were performed with ClustalW (http://align.genome.jp/) and edited manually using GeneDoc (Nicholas and Nicholas 1997).

2.2.1.2 Structural homology searches

Structural homology searches were performed with the DALI server (Holm and Sander 1995), http://www.ebi.ac.uk/dali). The submitted coordinates of a query protein are compared against protein coordinates deposited in the Protein Data Bank (www.rcsb.org). The comparison of 3D structures may reveal biologically interesting similarities that are not detectable by comparing protein sequences.

2.2.1.3 Calculation of protein parameters

Physical and chemical parameters of the recombinant proteins like molecular weight, (theoretical) isoelectric point (pI) and extinction coefficients were calculated with ProtParam (Gasteiger *et al.,* 2003) www.expasy.org/tools/protparam.html) from the ExPASy Proteomics Server.

2.2.1.4 Structure visualization and analyzation

Calculations of buried surface areas and molecule:molecule interactions were performed with CNS (Brunger *et al.,* 1998). Calculation of the electrostatic surface was done with PyMol (Baker *et al.,* 2001). Images of the crystal structures were prepared with PyMol (DeLano 2002). Superposition of two homologous structures was performed with COOT (Collaborative Computational Project 1994).

2.2.2 Molecular methods

2.2.2.1 Oligonucleotide design and Polymerase Chain Reaction (PCR)

PCR Primers for cloning of the genes of interest were designed using GeneRunner (see Table 5 for olignonucleotides). In general, oligonucleotides have a complementary part with a melting temperature (T^m) between 68°-72°C, an attached endonuclease restriction site and a 7-9 poly(A) overhang to assure efficiency of restriction endonuclease cleavage. *Archaeoglobus fulgidus* cell cultures were obtained from the DSMZ (Braunschweig, Germany), and genomic DNA was prepared with DNAzol® reagent (Molecular Research Center, Cincinnati, USA) according to the manufacturer's instructions. Genes of interest were amplified by PCR from *A. fulgidus* genomic DNA using Accuzyme® (Bioline, Luckenwalde, Germany) or *Pfx* Polymerase (Invitrogen, Karlsruhe, Germany). In general, a 50 µl PCR reaction contained 600 µM of each of the four dNTPs, 2.5 mM $MgCl_2$, 0.5 pM forward and reverse primer, and approximately 1 ng of template DNA. Each thermocycling program contained 25-30 cycles, whereas times and temperatures of denaturation, annealing and elongation were varied dependent on special requirements of the polymerase and primer-template pairs used in different amplifications. PCR products were purified from agarose gels with NucleoSpin® Extract II (Macherey and Nagel, Dueren, Germany).

2.2.2.2 Site-directed mutagenesis

Point mutations were introduced by PCR-based site-directed mutagenesis (Ho *et al.,* 1989). Here, two specially designed complementary oligonucleotides encoding the desired mutation are used to generate two DNA fragments with overlapping ends in the first amplification reaction (see Table 5 for oligonucleotides). In a second amplification reaction, these two DNA fragments serve as template to amplify the full-length gene with the incorporated nucleotide exchange. Primers for site-directed mutagenesis were designed with primerX (http://bioinformatics.org/primerx).

2.2.2.3 Restriction cleavage and ligation

DNA was digested using the corresponding restriction endonucleases and buffers (Fermentas, St. Leon-Rot, Germany) as recommended by the manufacturer.

Digested products were purified with NucleoSpin® Extract II (Macherey and Nagel, Dueren, Germany). Cleaved vector DNA (pET-21, pET-29, Novagen, Schwalbach/Ts., Germany) was additionally treated with calf intestine alkaline phosphatase (CIAP, Fermentas, St. Leon-Rot, Germany) to avoid religation. For ligation, a five- to tenfold excess of the digested DNA fragment was incubated with linearized vector and T4 ligase (Fermentas, St. Leon-Rot, Germany) in a 20 µl reaction following instructions of the manufacturer.

2.2.2.4 Transformation of *E. coli* and isolation of plasmid DNA

Transformation of plasmid DNA into competent *E. coli* cells was performed by mixing 50-100 µl of competent cells with 10-15 µl of the ligaton reaction or 1-2 µl of purified plasmid DNA (40 ng/µl) for 15 minutes on ice (Sambrook 1989). Cells were incubated at 42°C for 45 seconds and chilled on ice. 500 µl of fresh LB medium was added, followed by incubation at 37°C for 45 minutes in a thermo shaker. Cells were plated on LB agar plates containing the respective antibiotics and incubated at 37°C overnight. Plasmid DNA was isolated from a 5 ml overnight culture using the NucleoSpin®-Plasmid Quick Pure Kit (Macherey-Nagel, Dueren, Germany). DNA-sequencing of wild-type and mutant clones was performed by Medigenomix (Martinsried, Germany).

2.2.2.5 Preparation of competent cells

Preparation of transformation competent bacteria cells was performed according to (Hanahan 1983). 200 ml LB media was incubated with 2 ml of an overnight culture of the desired bacterial strain. The cells were grown for 2.5 – 3 hours at 37°C to an OD_{600} of 0.4-0.55, cooled down to 4°C on ice and centrifuged for 10 minutes at 3000 rpm and 4°C. The following steps were performed at 4°C. The pellet was washed with 30 ml of TFB-1 (Table 4), incubated for 10 minutes and centrifuged again. The washed pellet was resuspended in 4 ml of TFB-2 (Table 4), aliquoted and flash frozen in liquid nitrogen. Competent cells were stored at -80°C.

2.2.3 Electrophoretic methods

2.2.3.1 Electrophoretic separation of DNA

DNA was separated in horizontally poured 1 % Agarose/1x TAE gels containing 0.7 µg/ml ethidium bromide. Before loading, samples were mixed with 6X Loading

dye (Table 4). DNA was visualized using standard ultraviolet transilluminator (λ = 254 nm, Eagle Eye, Stratagene, LaJolla, USA).

2.2.3.2 Protein separation by SDS-PAGE

Protein samples were analyzed by SDS-PAGE with 15% polyacrylamide gels (acrylamide-bisacrylamide (37.5:1) (Laemmli 1970) using the vertical Mini-PROTEAN 3 System (Bio-Rad, Munich, Germany). Before loading, samples were mixed with 4x Loading Dye (Table 4) and boiled at 95°C for 2 minutes. Gels were routinely stained with Coomassie staining solution (Table 4) and destained with water.

2.2.4 Proteinchemical methods

2.2.4.1 Protein expression

For the over expression of recombinant proteins, competent *E. coli* Rosetta (DE3) cells (Novagen, Schwalbach/Ts., Germany) were transformed with plasmid DNA carrying the gene of interest. Cells were grown at 37°C in LB medium in the presence of the appropriate antibiotics to an OD_{600} of 0.6-0.8. Protein expression was induced with 0.2 mM IPTG (Roth, Karlsruhe, Germany). After further growth overnight at 18°C cells were harvested by centrifugation at 4°C. Cell pellets were flash frozen in liquid nitrogen and stored at -80°C.

2.2.4.2 Selenomethionine labelling

Selenomethionine incorporation was performed by co-transformation of the methionine auxotroph *E. coli* strain B834 (Novagen, Schwalbach/Ts., Germany) with plasmid DNA containing the gene of interest and the pRARE (Cm^r) plasmid isolated from Rosetta (DE3) cells. Cells were grown in 2000 ml LB medium supplemented with the appropriate antibiotics at 37°C to an OD_{600}=0.6. Cells were harvested and the pellet was resuspended in the same volume of LeMaster´s medium (Table 3) supplemented with selenomethionine (125 mg/2000 ml) and antibiotics. To deplete the medium of any residual methionine, cells were grown to an OD_{600} of 0.8 at 37°C. Protein expression was induced with 0.2 mM IPTG and cells were grown overnight at 18°C. Cells were harvested by centrifugation at 4°C. Cell pellets were flash frozen in liquid nitrogen and stored at -80°C.

2.2.4.3 Gelfiltration experiments and determination of molecular weight

The molecular weight of protein complexes was estimated by analytical gelfiltration chromatography using a Superose-6 10/300 HR gelfiltration column (GE Healthcare, Freiburg, Germany), equilibrated with the corresponding protein buffer. Prior to the experiment, the column was calibrated with Gel Filtration Standard (Bio-Rad, Munich, Germany). The stoichiometry of the complexes was additionally analyzed by SDS-PAGE.

2.2.5 Structure determination of the archaeal exosome

2.2.5.1 Cloning and expression of the exosome complexes from *Archaeoglobus fulgidus*

The coding sequences of the archaeal exosome subunits Rrp41, Rrp42, Rrp4 and Csl4 were amplified from *Archaeoglobus fulgidus* (af) genomic DNA by PCR using the corresponding PCR primers (Table 5). The purified PCR products of the two core subunits Rrp41 and Rrp42 were cloned into a modified bicistronic pET-21 vector (Novagen, Schwalbach/Ts., Germany; Lammens et al., 2004) using the corresponding restriction sites. The gene products of Rrp4-6xHis and Csl4 were cloned into a modified bicistronic pET-29 (Novagen, Schwalbach/Ts., Germany; Table 6). To obtain stoichiometric complexes of two exosome isoforms, the resulting expression plasmids containing Rrp41/Rrp42 and Rrp4 (denoted Rrp4-Exosome) or Rrp41/Rrp42 and Rrp4/Csl4 (denoted Csl4-exosome), were co-transformed into competent *E. coli* Rosetta (DE3). Proteins were co-expressed overnight at 18°C to allow association of the individual complexes and harvested by centrifugation. Point mutations in the core subunit Rrp41 were generated by site-directed mutagenesis (Table 5), and the mutated gene product was cloned with Rrp42 into the bicistronic pET-21. Mutant complexes of the Rrp4-exosome were expressed according to the wild-type protein. For SAD phasing protein was derivatizised with selenomethionine as described in 2.2.4.2.

2.2.5.2 Purification of the two exosome isoforms

Cells were resuspended in Lysis buffer (Table 7) supplemented with 200 µM PMSF (Roth, Karlsruhe, Germany), and disrupted by sonication. Cell debris was removed by centrifugation. *E. coli* proteins were heat denaturised for 25 minutes at

80°C and removed by centrifugation. To purify the Rrp4-exosome, the respective supernatant was loaded onto a Ni^{2+}-NTA column (5 ml, Qiagen, Hilden, Germany), equilibrated with the corresponding Equilibration buffer (Table 7). After washing with Wash buffer I and II (Table 7), the Rrp4-Exosome was step-eluted with Elution buffer (Table 7). Fractions were analyzed using Bradford protein assay (Bio-Rad, Munich, Germany). Protein containing fractions were pooled, diluted 1:4 with Dilution buffer (Table 7), and further purified by cation exchange chromatography (Resource™ S, GE healthcare, Freiburg, Germany) using the Äkta System (GE Healthcare, Freiburg, Germany). The heat-stable supernatant of the Csl4-exosome was filtered using a 0.45 µm sterile-filter before loading onto an anion exchange chromatography column (MonoQ HiTrap, GE Healthcare, Freiburg, Germany). Prior to the experiments, the ion exchange columns were equilibrated with the respective Buffer A (Table 7) and protein was eluted with a gradient of 20 column volumes from 50 mM NaCl (Buffer A) to 1 M NaCl (Buffer B, Table 7). The pooled peak fractions were concentrated and loaded onto a Superdex S200 16/60 size exclusion chromatography column equilibrated with the corresponding buffer (Table 7). Peak fractions were concentrated to 15 mg/ml for crystallization using centrifugal filter devices (Amicon® Ultra, Millipore, Billerica, MA, USA) with a 10 kDa nominal molecular weight cutt-off. Mutant Rrp4-exosome was purified in the same way. Selenomethionine-containing Rrp4- and Csl4-exosome were purified according to the wild-type proteins, whereas all buffers were degassed before use and contained 10 mM β-mercaptoethanol (Roth, Karlsruhe, Germany) or 2 mM DTT (Roth, Karlsruhe, Germany)

Table 5: Oligonucleotides

Oligo name	Sequence
afRrp41 for NdeI	aaaaaa<u>catatg</u>tcggaattcaatgaaaaaccagaa
afRrp41 rev Hind III	aaaaaa<u>aagctt</u>tcaggcatcttcaccaccctctg
afRrp42 for Nco I	aaaaaaa<u>ccatgg</u>gccctgaagacatccttgtggacatt
afRrp42 rev Not I	aaaa<u>gcggccgc</u> ttaaatttccttaaatttctccctcag
afRrp4 for Nco I	aaaaaaa<u>ccatgg</u>gcaggaagatagtactgccaggagat
afRrp4_cHis rev Not I	aaaaaaa<u>gcggccgc</u>ttgaattccgacatctgccttcct
afCsl4 for Nde I	aaaaaa<u>catatg</u>agattcgtaatgccgggagat
afCsl4 rev Hind III	aaaaaa<u>aagctt</u>ctaccactctcccttgccgtaat
Desired mutation	Sequences
afRrp41 R65E for	gaagtgcatcca**gaa**caccttcaggat
afRrp41 R65E rev	atcctgaaggtg**ttc**tggatgcacttc
afRrp41 D180A for	atgaaagaggag**ggc**aattttggtgag
afRrp41 D180A rev	ctcaccaaaatt**gcc**ctcctctttcat

Underlined regions mark recognition sites for restriction endonucleases.

Bold letters correspond to the mutated codons. All sequences are given in 5´- 3´ direction.

Table 6: Expression plasmids

#	Insert	Vector	Restriction sites	Tag	Remarks
1	afRrp41-afRrp42	pET-21	Nde I, Not I	-	Bicistronic
2	afRrp4	pET-29	Nco I, Not I	c-6xHis	-
3	afRrp4-afCsl4	pET-21	Nde I, Not I	-	Bicistronic
4	afRrp41 (R65E)-afRrp42	pET-21	Nde I, Not I	-	Bicistronic
5	afRrp41 (D180A)-afRrp42	pET-21	Nde I, Not I	-	bicistronic

Table 7: Buffers for the purification of the exosome complexes

Buffer	Description
Ni-NTA	
Lysis buffer/Equilibration buffer	20 mM NaH_2PO_4 pH 7.5 (25°C), 200 mM NaCl, 10 mM β-mercaptoethanol
Wash buffer I	20 mM NaH_2PO_4 pH 7.5 (25°C), 1.5 M NaCl, 10 mM Imidazol, 10 mM β-mercaptoethanol
Wash buffer II	20 mM NaH_2PO_4 pH 7.5 (25°C), 200 mM NaCl, 10 mM Imidazol, 10 mM β-mercaptoethanol
Elution buffer	20 mM NaH_2PO_4 pH 6.8 (25°C), 200 mM NaCl, 250 mM Imidazol, 10 mM β-mercaptoethanol
Dilution buffer	20 mM NaH_2PO_4 pH 6.8 (25°C), 10 mM β-mercaptoethanol
Resource S (Rrp4-exosome/Mono Q (Csl4-exosome)	
Buffer A	20 mM NaH_2PO_4 pH 6.8 (25°C), 50 mM NaCl, 2 mM DTT
Buffer B	20 mM NaH_2PO_4 pH 6.8 (25°C), 1 M NaCl, 2 mM DTT
Gelfitration (Rrp4-exosome)	20 mM Tris/HCl pH 7.8 (25°C), 200 mM NaCl, 2 mM DTT
Gelfiltration (Csl4-exosome)	20 mM Tris/HCl pH 7.8 (25°C), 200 mM NaCl, 10 % glycerol, 2 mM DTT

2.2.5.3 Crystallization and structure determination of the exosome complexes

Native Rrp4-exosome was crystallized by sitting drop vapour diffusion technique by mixing 1 µl protein (15 mg/ml in gelfiltration buffer, Table 7) and 1 µl of reservoir solution (0.1 M HEPES pH 6.2, 35% MPD, 10 mM $CaCl_2$ and 15% glycerol) at 20°C. Crystals of native Csl4-exosome (15 mg/ml in gelfiltration buffer, Table 7) were obtained accordingly by mixing 1 µl of protein with 1 µl of 0.1 M sodium acetate pH 5.6, 32% MPD and 10 mM NaCl as precipitant solution at 20°C. Prior to data collection, crystals were mounted in nylon loops and flash frozen in liquid nitrogen. Redundant single wavelength anomalous dispersion (SAD) experiments were recorded at ID29 (ESRF, Grenoble, France) with selenomethionine-containing crystals of both Rrp4- and Csl4-exosome at the K absorption edge to 3.0 Å and 3.2 Å, respectively. Data from 360 images (Rrp4-exosome, 1° rotation)

and 500 images (CsI4-exosome, 0.5° rotation) were integrated and scaled with XDS and XSCALE (Kabsch 1993). For the Rrp4-exosome, atomic positions for 57 Selenium atoms were located with SHELXD (Schneider and Sheldrick 2002). Phases to 3.0 Å were obtained with SHARP software (Global phasing), and used as input for automated and manual model building with RESOLVE/REFMAC (Terwilliger 2002) and MAIN (Turk 1992). Automated model building allowed partial tracing of poly-alanine chains into the electron density to simplify manual model building. Refinement to 3.0 Å was performed with CNS (Brunger et al., 1998). For the CsI4-exosome, the atomic postions of 54 Seleniums were found accordingly, and phases to 3.2 Å were obtained. As to this time point the initial 3.0 Å model of the Rrp4-exosome already existed, the core domain (Rrp41-Rrp42) was used as search model for molecular replacement. The additional CsI4 subunit was afterwards build manually using MAIN (Turk 1992). Refinement of the complete CsI4-exosome at 3.2 Å followed iterative cycles of manual model completion with MAIN (Turk 1992) and refinement with CNS (Brunger et al., 1998). The initial models were used to phase 2.7 Å native data sets recorded at ID14-1 (ESRF, Grenoble. France) and PX beamline (SLS, Villigen, Switzerland) using PHASER (Collaborative Computational Project 1994). Refinement of the native data were performed with CNS (Brunger et al., 1998), and included overall anisotropic B-factor and bulk solvent corrections, individual B-factor refinement, simulated annealing, and positional refinement. Initial three-fold noncrystallographic symmetry (NCS) within the exosome complexes was gradually removed at later refinement steps to account for subunit flexibility. Stereochemistry of the final models was analyzed using PROCHECK (Laskowski et al., 1993)

2.2.5.4 Tungstate soaking experiments

Crystals of the Rrp4-exosome were soaked with 50 mM $Na_2WO_4 x 2H_2O$ to locate potential phosphate or RNA binding sites (Lima *et al.,* 1997). For soaking experiments, tungstate was diluted into reservoir solution to keep crystallization conditions constant. The crystallization buffer was slowly exchanged with the tungstate solution and damage of crystals was monitored. After 10 minutes, 30 minutes and 1 hour, undamaged crystals were mounted in nylon loops and flash frozen in liquid nitrogen, respectively. Data were recorded at the tungstate peak wavelength to a resolution of 2.9 Å at PX beamline (SLS, Villigen, Switzerland).

Data from 180 images (1° oscillation) were integrated and scaled with XDS and XSCALE (Kabsch 1993). The data set was phased by molecular replacement using the refined Rrp4-exosome as search molecule for PHASER (Collaborative Computational Project 1994). Anomalous signals were visualized by calculation of an anomalous difference density ($|F(+)|-|F(-)|$).

2.2.6 Biochemical Methods

2.2.6.1 Radioactive Labelling of RNA

The radioactive labelling of RNA was carried out by mixing 30 pmol of a 30mer poly(rA)-RNA oligoribonucleotide with 60 µCi γ-^{32}P-ATP and 4.5 U T4 Polynucleotide Kinase (Fermentas, St. Leon-Rot, Germany) in a 30 µl reaction for 1 hour at 37°C following the provided instructions. The reaction contained additionally 0.8 U/µl RNasin (Promega, Mannheim, Germany), and was perfomed in DEPC-treated water. Tips and reaction tubes were RNase-free. The RNA was purified using MicroSpin G-25 columns (GE Healthcare, Freiburg, Germany).

2.2.6.2 RNA degradation and polyadenylation assays

RNA degradation and polyadenylation activity of the archaeal exosome isoforms were analyzed by mixing 5 or 50 pmol of protein with 50 fmol of RNA substrate in a 20 µl reaction containing the corresponding buffer (Table 8) for 10 minutes at 60°C. The reaction contained additionally 0.8 U/µl RNasin (Promega, Mannheim, Germany). Reactions were stopped by adding 0.5 volumes 4x loading dye (Table 8), and transferred onto a 20 % polyacrylamide/7 M Urea gel. Gels were run in 0.5 x TB-Buffer for 2 hours at 600 V. The resolved reaction products were analyzed by phosphorimaging (GE Healthcare, Freiburg, Germany) using a StormScanner (GE Healthcare, Freiburg, Germany). To analyze relative RNase activities of exosome isoforms, 5 pmol (250 nM) protein ws incubated with 50 fmol (2.5 nM) of the 30mer oligo(rA) substrate at 60°C for 5 minutes. Degradation products were resolved as described and quantified with ImageQuant® software (GE Healthcare, Freiburg, Germany). Calculated standard deviations represent means of three individual degradation experiments.

2.2.6.3 RNA binding assays

To test RNA binding abilities of wild-type and mutant exosome complexes, RNA binding assays were performed with radioactive labelled 30mer poly(rA)-oligoribonucleotide as substrate. For each reaction, 5, 2.5, or 0.25 µM protein was incubated with 50 fmol RNA substrate in the corresponding buffer (Table 8) in a 20 µl reaction for 5 minutes at room temperature. Reactions contained additionally 0.8 U/µl RNasin (Promega, Mannheim, Germany). The samples were mixed with 6x Loading Dye (Table 4), and RNA:protein complexes were resolved on a native 5% polyacrylamid gel. Gels were run in 0.5 x TB-Buffer for 2 hours at 80 V. The reactions products were analyzed by phosphorimaging (GE Healthcare, Freiburg, Germany).

Table 8: Buffers and solution for biochemical assays

Buffer	Description
RNA degradation assay	
5x Reaction buffer	120 mM Tris/HCl pH 7.8, 300 mM KCl, 50 mM MgCl$_2$, 50 % glycerol, 10 mM DTT, 0.5 mM EDTA, 500 µg/ml BSA, 50 mM NaH$_2$PO$_4$ (pH 7.8)
RNA polyadenylation assay	
5x Reaction buffer	120 mM Tris/HCl pH 7.8, 300 mM KCl, 50 mM MgCl$_2$, 50 % glycerol, 10 mM DTT, 0.5 mM EDTA, 500 µg/ml BSA, 100 mM ADP
4x Loading Dye	25% glycerol, 125 mM Tris/HCl pH 6.8 (25°C), 3.5% SDS, 2.5 mM EDTA, 5 mM DTT, 0.5% Bromphenolblue, 3 M Urea
10 x TB Buffer	0.89 M Tris, 0.89 M Boric acid
RNA binding assay	
5x Reaction buffer	120 mM Tris/HCl pH 7.8, 300 mM KCl, 50 mM MgCl$_2$, 50 % glycerol, 10 mM DTT, 0.5 mM EDTA
6x Loading Dye	1.5 g/L bromphenol blue, 1.5 g/L xylene blue, 50% (v/v) glycerol

3   Results

3.1   Purification, crystallization and structure determination of the archaeal exosomes

Exosomes emerge as central $3' \rightarrow 5'$ RNA processing and degradation machineries in eukaryotes and archaea. The structural mechanism of exosomes in fundamental processes like quality control and normal turnover of cytosolic (pre-) mRNA (Hilleren *et al.,* 2001) as well as in the processing and degradation of nuclear RNA species (Butler 2002; Vasudevan and Peltz 2003; Raijmakers *et al.,* 2004) is still poorly understood. To provide a framework for the architecture and mechanism of exosomes, I determined structures of two archaeal exosome isoforms.

The genes encoding the four archaeal exosome subunits from *Archaeaglobus fulgidus* were amplified by PCR from genomic DNA. Although individually expressed recombinant exosome subunits are all soluble, the whole complex was co-expressed and not reconstituted after expression. This co-expression strategy was already successfully used in the case of subunits of the Mediator complex (Baumli *et al.,* 2005; Hoeppner *et al.,* 2005). For co-expression, individual genes were cloned into mono- and bicistronic vectors. The resulting expression plasmids (Table 6) were co-transformed into *E. coli* Rosetta (DE3) cells, and complexes of the Rrp4-exosome and Csl4-exosome were co-expressed. The 250 kDa Rrp4-exosome contains the two core subunits Rrp41 and Rrp42 as well as the Rrp4 subunit with a C-terminal 6xHis-Tag. The 235 kDa Csl4-exosome contains the two core subunits Rrp41 and Rrp42, and both Rrp4 and Csl4 subunits (Figure 5).
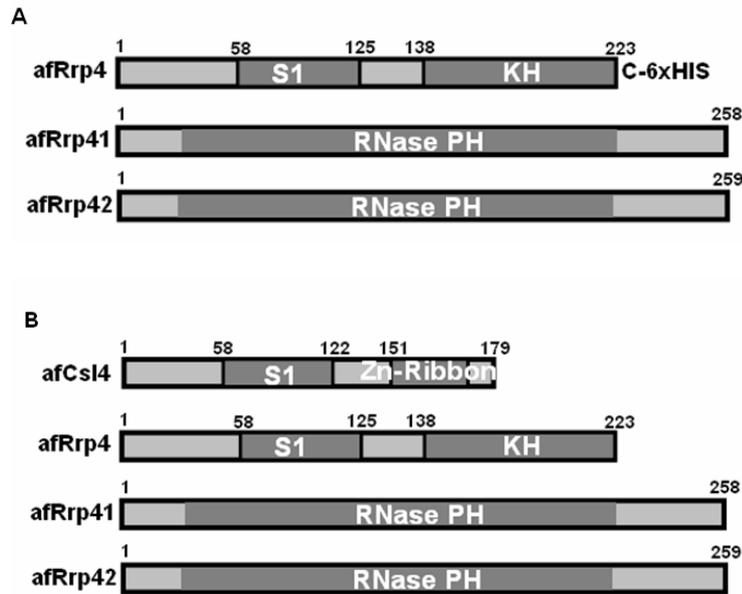
Figure 5: Schematic overview about subunit compostion of Rrp4-exosome (A) and Csl4-exosome (B) with indicated domain denotation.

3.1.1 Purification of the Rrp4-exosome

The Rrp4-exosome was purified to near homogeinity by an initial heat denaturation of *E. coli* proteins, immobilized Ni^{2+}-NTA affinity chromatography (utilizing the 6xHis-Tag on Rrp4), Resource™ S cation exchange chromatography, and HiLoad 16/60 Superdex S200 size exclusion chromatography as described in 2.2.5.2. Rrp4-exosome eluted in one distinct peak with an approximately molecular weight of 200-250 kDa (theoretical molecular mass: 250 kDa), and a 1:1:1 stoichiometry of all three subunits (Figure 6). Protein containing fractions were concentrated to a final concentration of 15 mg/ml and used for crystallization trials. From a 6 l expression culture, 15-20 mg of pure protein could be obtained in total (Figure 6). Selenomethionine-containing protein and mutant protein were purified accordingly.
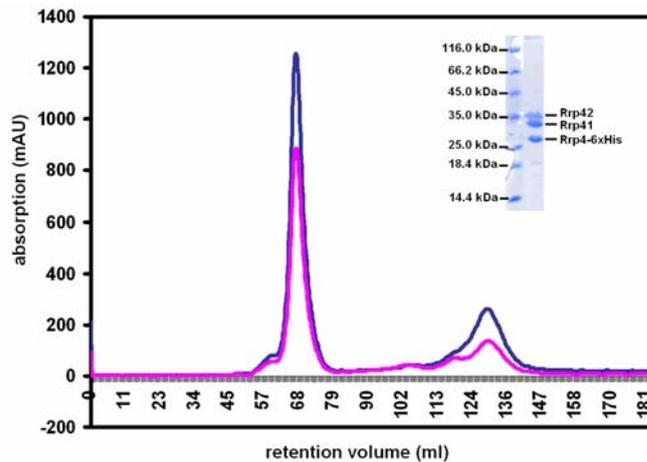
Figure 6: Elution profile of the Rrp4-exosome from HiLoad 16/60 S200 size exclusion column. (absorption at 280 nm, blue; absorption at 260 nm, magenta). Second peak may represent monomeric exosome subunits. Insert shows SDS-PAGE analysis of a representative peak fraction.

3.1.2 Purification of the Csl4-exosome

The Csl4-exosome was purified by heat denaturisation of *E. coli* proteins, Mono Q anion exchange chromatography and HiLoad 16/60 Superdex S200 size exclusion chromatography as described in 2.2.5.2. The Csl4-exosome complex eluted in one distinct peak with an appr. molecular weight of 200-250 kDa (theoretical MW 235 kDA) (Figure 7). The co-expression strategy resulted in a heterogeneous mixture of complexes that contain Rrp41 and Rrp42 in almost stoichiometric amounts, and Csl4 and Rrp4 in substoichiometric amounts (Figure 7). Protein containing fractions were pooled and concentrated to a final concentration of 15 mg/ml, and purified protein was used for crystallization set ups. In general, around 20-25 mg of protein could be obtained from 6 l expression culture.
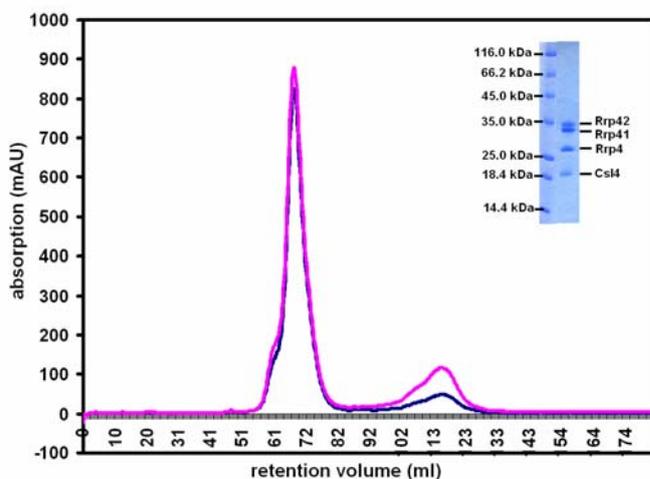
Figure 7: Elution profile of the Csl4-exosome from HiLoad 16/60 S200 size exclusion column. (absorption at 280 nm, blue; absorption at 260 nm, magenta). Insert shows SDS-PAGE analysis of a representative peak fraction.

3.1.3 Purification of mixture complexes

Given the defined stoichiometry of eukaryotic exosome complexes (Rrp4:Rrp40:Csl4 in 1:1:1 stoichiometry), the ability of archaeal exosomes to form different isoforms *in vitro* was surprising. As described, co-expression of all four Rrp41, Rrp42, Rrp4 and Csl4 subunits in *E. coli* resulted in a heterogeneous mixture of complexes that contain Rrp41 and Rrp42 in almost stoichiometric amounts, and Csl4 and Rrp4 in substoichiometric amounts (Figure 7).

As seen in Figure 8, the stoichiometry of Rrp4:Csl4 depends on the affinity purification strategy and can result in 2:1 or 1:2 stoichiometries (Figure 8B,D). Co-expression of Rrp41, Rrp42 and Csl4 resulted in 1:1:1 stoichiometry, consistent with the 1:1:1 stoichiometry of the Rrp4-exosome (Figure 8A,C).
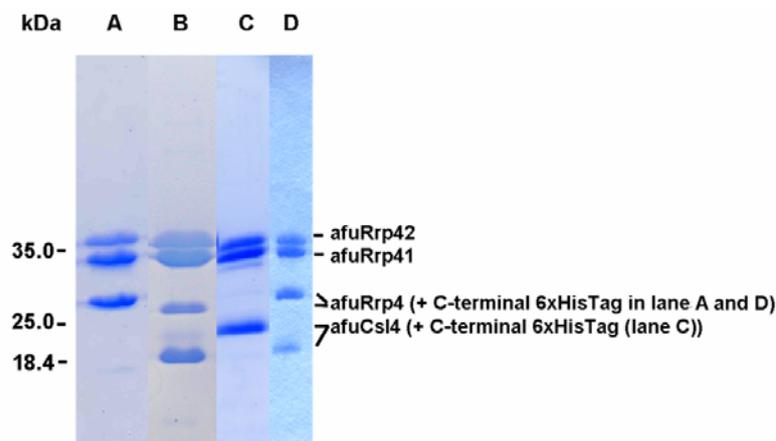
Figure 8: A) Co-expression and co-purification of Rrp4-exosomes (Rrp4-6xHis, Rrp41 and Rrp42) shows 1:1:1 stoichiometry. B) Co-expression and co-purification of Rrp41, Rrp42 and Rrp4 and Csl4 resulted in a mixture of complexes with Rrp4 and Csl4 subunits in varying substoichiometric amounts (depending on mutual expression levels). C) Co-expression and co-purification of Csl4-exosomes (6xHis-Csl4, Rrp41 and Rrp42) reveales 1:1:1 stoichiometry. D) Co-expression of Rrp4-6xHis, Rrp41, Rrp42 and Csl4 and purification by Ni^{2+}-NTA (utilizing the HisTag), Mono Q and size exclusion resulted in exosome isoforms containing both Rrp4 and Csl4 in different stoichiometric amounts ($Rrp4_3:Csl4_0$, $Rrp4_2:Csl4_1$, $Rrp4_1:Csl4_2$). Molecular weight markers are indicated left (in kDa). Figure shows Coomassie stained polyacrylamid gel electrophoresis from four individual co-purification experiments.

As stoichiometry fluctuations of Csl4 and Rrp4 subunits are also observed in the analysis of exosomes from archaeal cells (Evguenieva-Hackenberg *et al.,* 2003; Farhoud *et al.,* 2005), these data indicate that heterogenic Rrp4-Csl4 caps exist *in vitro* and thus are likely to exist also *in vivo*.

3.1.4 Crystallization of the Rrp4-exosome

For initial crystallization, trials were set by hand using commercial crystallization screens in 96-well sitting drop plates (Corning, Schiphol-Rijk, Netherlands). In general, the total reservoir volume was between 50 and 80 µl and 1 µl of protein was mixed with 1 µl of reservoir solution. Plates were incubated at 20°C. Small cubic crystals grew after 1-2 days in Hampton Screen I and II conditions as well as in conditions of the Jena Biosciences Screen 7 (Figure 9A,B). For first diffraction experiments, crystals were mounted directly and flash frozen in liquid nitrogen. These cubic crystals from different conditions diffracted to ~ 7 Å at PX beamline (SLS, Viligen, Switzerland). Nevertheless, no intense ice rings could be detected.

For structure determination, crystal size and quality had to be improved. Several conditions from the Hampton and Jena Biosciences Screens were chosen for extensive refinement and included variation of pH and MPD/glycerol concentration as well as addition of different salts (NaCl, LiCl, $MgCl_2$, $CaCl_2$). During the refinement procedure, a pH dependent change in the crystal morphology could be detected. By raising the pH, the crystals lost their cubic form and grew in a more cuboid like shape. Finally, the best crystal conditions could be achieved on the basis of Jena Biosciences Screen 7 #D4 by varying the pH and MPD/$CaCl_2$ concentration (Figure 9C) by which crystals with a maximum size of 100 µm x 100 µm x 200 µm could be obtained. These crystals belonged to space group $P2_1$ with unit cell constants a=101.6 Å, b=129.6 Å, c=102.3 Å, β=101.6° and exhibit one molecule (=one nine subunit exosome complex) per asymmetric unit. This results in a Matthews coefficient of 2.86 $Å^3$/dalton of protein (Matthews 1968; Kantardjieff and Rupp 2003) and a solvent content of 57 %. The crystals diffracted X-rays to 2.7 Å resolution at ID14-1 beamline (ESRF, Grenoble, France) and were used to record a native data set. Selenomethionine-containing crystals grew in the same refinement conditions.

Figure 9: Initial cubic crystals from Rrp4-exosome grown in two representative conditions. A) Hampton Screen I #40: 20% iso-Propanol, 0.1 M NaCitrate pH 5.6, 20% PEG4000. B) Jena Bioscience Screen 7 #D4: 44% MPD, 0.1 M NaHepes pH 7.5. C) Refined selenomethionine-containing cuboid crystal of the Rrp4-exosome in a refinement condition containing 30% MPD, 0.1 M NaHEPES pH 6.2 and 10 mM $CaCl_2$.

3.1.5 Crystallization of the Csl4-exosome

The Csl4-exosome was crystallized analogous to the Rrp4-exosome by initial screening with commercially available crystallization conditions. Crystal trials were set in 96-well sitting drop plates and 1 µl of protein was mixed with 1 µl of reservoir solution. Plates were incubated at 20°C. Interestingly, only the major species, the "Csl4-exosome" isoform crystallized with 1:1:1 (Rrp41:Rrp42:Csl4) stoichiometry. Similar looking crystals appeared after 2-4 days in only two different conditions (Figure 10A,B). However, these crystals were directly mounted and flash frozen in liquid nitrogen. Diffraction experiments on PX beamline (SLS, Villigen, Switzerland) already resulted in diffraction up to 4 Å. Refinement of Hampton Screen II #10 resulted in crystals with an particulate unusual shape (Figure 10C,D), which diffracted X-rays up to 2.7 Å. Crystals belonged to space group $P4_322$ with unit cell constants a=b=137.5 Å, c=261.0 Å and contained one molecule (=one nine subunit exosome complex) in the asymmetric unit. This results in a solvent content of 54 %, and a Matthews coefficient of 2.66 $Å^3$/dalton of protein (Matthews 1968; Kantardjieff and Rupp 2003). Selenomethionine-containing protein could be crystallized in the same refinement conditions and exhibited the same knobby outgrowths.
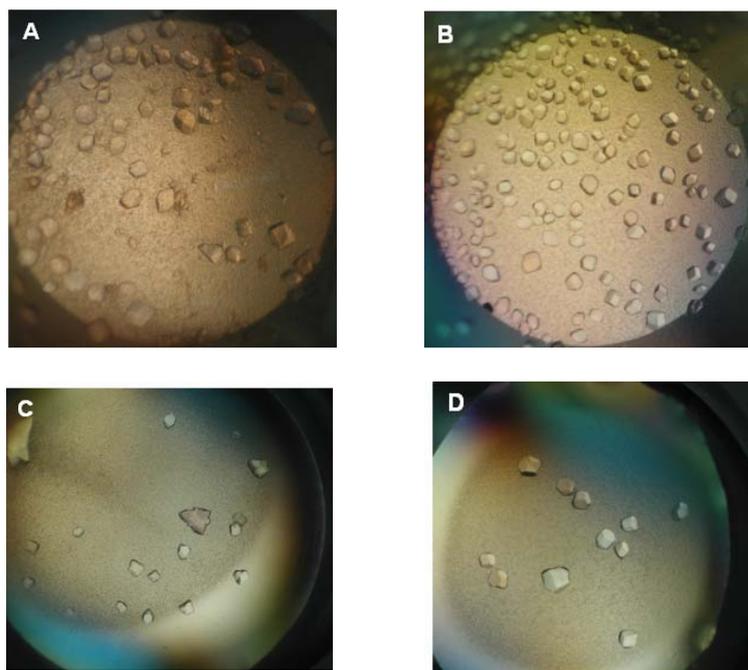
Figure 10: Crystals grown in initial screening conditions Hampton Screen II #10 (A, 30% MPD, 0.1 M NaAcetate pH 4.6, 0.2 M NaCl) and Nextal Classic Screen I (B, 0.2 M MgCl$_2$, 0.1 M NaHEPES pH 7.5, 30% iso-propanol). Refined crystals of selenomethionine-containing crystals (C, 35% MPD, 0.1 M NaAcetate pH 5.6, 20 mM NaCl) and native crystals of Csl4-exosome (D, 32% MPD, 0.1 M NaAcetate pH 5.6, 10 mM NaCl). Crystals grew to an averaged maximum size of 70 µm x 70 µm x 70 µm.

3.1.6 Data collection

Diffraction data of the selenomethionine containing crystals of both Rrp4- and Csl4-exosome were collected at ID29 (ESRF, Grenoble, France) with an ADSC Q315R CCD detector. Diffraction data of the native Rrp4-exosome were collected at ID14-1 with an ADSC Q210 CCD detector. Diffraction data of native Csl4-exosome were measured at PX beamline (SLS, Villigen, Switzerland) with a mar225 mosaic CCD detector. 360 frames (1° oscillation) were collected for each data set of the Rrp4-, whereas 500 frames (0.5° oscillation) were recorded for each Csl4-exosome data set.

Phase determination was carried out by a single wavelength anomalous dispersion (SAD) experiment at the selenium K edge for both Rrp4- and Csl4-exosome, using one selenomethionine containing crystal each. As measurements of SAD datasets depend on accurate measurement of the data set at the accurate wavelength, the optimal peak wavelength was determined experimentally by a

fluorescence scan for each selenomethionine containing crystal prior to data collection. Data for the peak wavelength at 0.9795 Å (12.66123 KeV, f´-8.60, f´´ 6.00) were collected for the Rrp4-exosome to 3.0 Å. A native data set of the Rrp4-exosome was collected at 1.003 A to 2.7 Å resolution. Data of selenomethionine containing crystals of the Csl4-exosome were measured at 0.9795 Å (12.661 KeV, f´-8.5, f´´ 5.6) to 3.2 Å. Data of native Csl4-exosome crystals were recorded at 1.01 Å to 2.7 Å.

3.1.7 Structure determination and refinement of the Rrp4-exosome

All data were processed with XDS and XSCALE (Kabsch 1993). Data were indexed and scaled in $P2_1$, and later analysis of systematic absences revealed the presence of the screw axis (Figure 11).
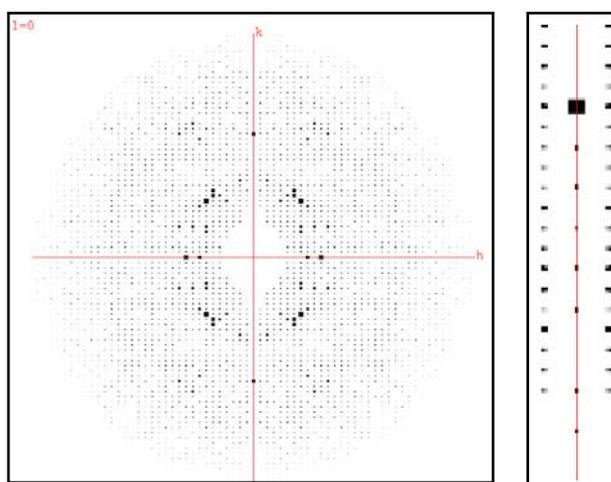
Figure 11: Image of the peak data set with hklview (CCP4, 1994) on section hk0 (left), and a zoomed section on the k-axis (h=0, l=0) (right). Only every second reflexion is present, which corresponds to a $P2_1$ screw axis.

The Rrp4-exosome contains 58 methionines (plus 9 N-terminal methionines) in total. Thus, it belongs to one of the biggest macromolecular protein complexes solved by selenomethionine SAD. In the case of the yeast RNA Polymerase II, *in vivo* selenomethionine incorporation was used to facilitate model building of a 3.1 Å resolution structure, whereby 94 out of 103 possible selenomethionine peaks were detected (Cramer *et al.,* 2001). However, 57 selenomethionine positions were located using SHELXD (Collaborative Computational Project 1994). Phases were obtained to 3.0 Å with SHARP. Phase improvement by solvent flattening with SOLOMON (Collaborative Computational Project 1994) resulted in an interpretable electron density (Figure 12).
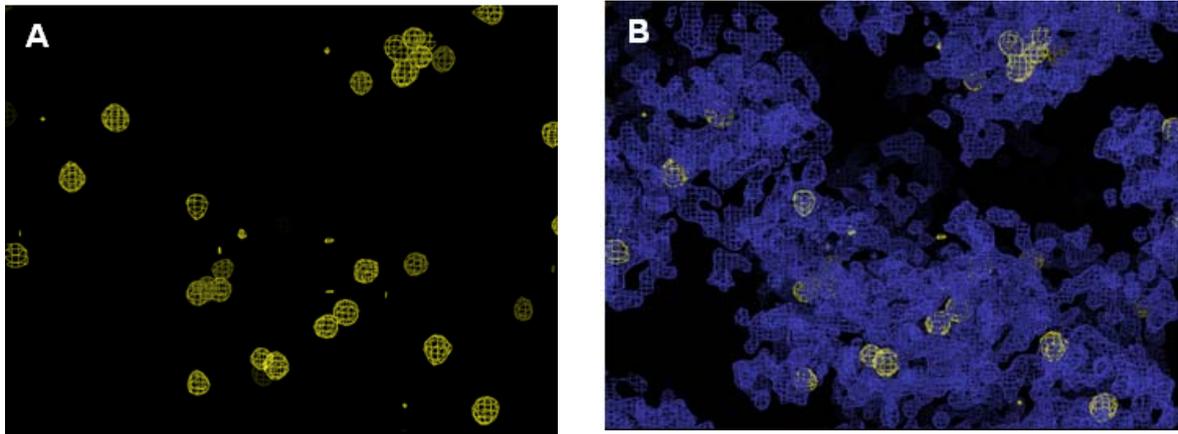
Figure 12. A) Anomalous difference fourier map (yellow mesh) contoured at 3σ obtained from Selenomethionine phasing. B) 1σ contoured single anomalous dispersion (SAD) map of the Rrp4-exosome at 3.0 Å resolution (blue mesh) with superimposed anomalous difference fourier map (yellow mesh) contoured at 3σ.

Phases to 3.0 Å allowed partial automated model building with RESOLVE/REFMAC (Terwilliger 2002), and most of the experimental electron density was traced as a poly-alanine model, which facilitates manual model building with MAIN (Turk 1992). The presence of at least five selenomethionies per single subunit, which could be used as sequence markers, additionally simplified the manual building process. One of each subunit was built manually. To obtain three copies of each subunit, the Rrp41-Rrp42-Rrp4 model was manifolded, followed by manual rotation into the two uninterpreted electron density regions of the exosome complex. The model was additionally fitted by rigid body refinement using CNS (Brunger *et al.,* 1998). The first refinement steps included rigid body refinement, bulk solvent correction and overall anisotropic B-factor correction as well as three-fold non-crystallographic symmetry (NCS) within the exosome. The refined model was later on used to phase the 2.7 Å resolution native data set by molecular replacement using PHASER (Collaborative Computational Project 1994). One single solution was refined at 2.7 Å by iterative cycles of bulk solvent correction, simulated annealing, positional minimization and individual B-factor refinement with CNS (Brunger *et al.,* 1998), and additional manual completion of the model with MAIN (Turk 1992). The initial three-fold NCS was gradually loosend during refinement to account for subunit flexibility. Solvent molecules were introduced using CNS (Brunger *et al.,* 1998) and verified manually. The final model of the Rrp4-exosome contains 2157 out of 2220 possible amino acids and 102 solvent molecules. Refinement resulted in an R-factor of 21.6 (R_{free}=26.4) and

good stereochemistry (Figure 13). Ramachandran plot shows 87 % of the residues in the most favored region. Refinement statistics are listed in Table 9.
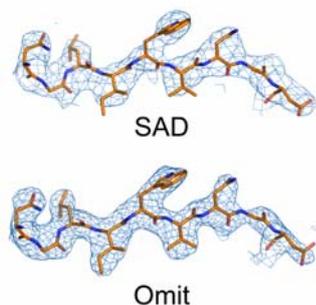
Figure 13: Representative portion of single anomalous dispersion (SAD) and composite omit maps (Omit, calculated after 500K-0K simulated annealing to remove model bias), shown at the Rrp4 KH domain. The maps are contoured at 1σ. The final model is overlaid and shown as color coded stick representation.

3.1.8 Structure determination and refinement of the Csl4-exosome

All data were initially indexed and scaled with XDS and XSCALE (Kabsch 1993) in space group P422. The presence of a screw axis was verified by analysis of systematic absences and could be limited to the two enantiomorphs $P4_122$ and $P4_322$ (International Tables Crystallography, Volume A: Space-group symmetry, 2002) (Figure 14). In this case, systematic absences alone allow no discrimination, and the correct space group can only be revealed after map calculation. Thus, both space groups had to be used for initial phase determination.
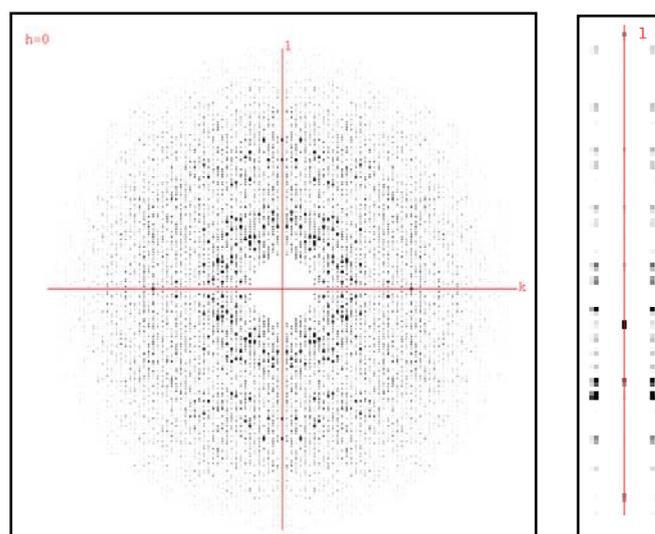
Figure 14: Image of the peak data set with hklview (CCP4, 1994) on section 0kl (left). A zoomed view on the l-axis (h=0, k=0) shows the presence of only every forth reflection, which corresponds to a 4_1 or 4_3 screw axis (right).

Comparable to the Rrp4-exosome, the Csl4-exosome contains 54 selenomethionines (plus 9 N-terminal selenomethionines) in total. With SHELXD

(Collaborative Computational Project 1994) atomic positions of 54 selenomethionines could be located. Phases to 3.2 Å could be obtained with SHARP (Global phasing), and an interpretable electron density was calculated in space group P4$_3$22 after phase improvement by solvent flattening with SOLOMON (Collaborative Computational Project 1994). For initial model building, the already build core domain of the Rrp4-exosome (Rrp41-Rrp42 subunits) was used as search molecule for molecular replacement with PHASER (Collaborative Computational Project 1994). One single search solution was fitted to the data by rigid body refinement using CNS (Brunger *et al.,* 1998). The additional Csl4 subunit and additional arising parts were afterwards build manually into the emerged difference fourier density. According to the Rrp4-exosome, the intrinsic three-fold NCS was used for ensuing refinement steps. After rigid body refinement, bulk solvent and overall anisotropic B-factor correction using CNS (Brunger *et al.,* 1998), the model was used to phase the 2.7 Å native data set by molecular replacement using PHASER (Collaborative Computational Project 1994). The final model was refined to 2.7 Å resolution by repeated rounds of bulk solvent correction, simulated annealing, positional minimization and restrained individual B-factor correction using CNS (Brunger *et al.,* 1998) and manual building with MAIN (Turk 1992). The internal three-fold NCS was gradually removed at later refinement steps as performed for the Rrp4-exosome. Water molecules were integrated with CNS (Brunger *et al.,* 1998) and verified manually. A zinc ion was modelled in each of the three zinc-binding domains of Csl4. The final model of the Csl4-exosome comprises 2019 out of 2088 possible amino acids, 91 solvent molecules, and three zinc ions. The model was refined to an R-factor of 22.4 (R$_{free}$=27.5) and exhibits good stereochemistry in Ramachandran plot with 84 % of the residues in the most favored region. Refinement statistics are listed in Table 9.

Table 9: Data collection and refinement statistics

Data collection

Data set	Rrp4 Exosome[1]			Csl4 Exosome[2]	
	Native	SeMet K Peak	WO4	Native	SeMet K Peak
X-ray source	ID14 (ESRF)	ID29 (ESRF)	ID29 (ESRF)	PX (SLS)	ID29 (ESRF)
Wavelength (Å)	0.933	0.9795	1.2143	1.01	0.9795
Data range (Å)	20-2.7	20-3.0	20-2.9	20-2.7	20-3.2
Observations (unique)	344114 (69958)	224914 (100738[d])	218821 (113308[d])	934134 (69124)	686258 (81959[d])
I/σI (last shell)	13.2 (3.7)	9.7 (3.0)	9.3 (2.4)	22.9 (7.0)	14.2 (6.3)
Completeness (last shell)	98.1 (84.4)	98.9 (95.7)	98.2 (92.5)	99.5 (99.9)	99.9 (99.8)
R_{symm}[a]	0.10 (0.59)	0.084 (0.45)	0.054 (0.39)	0.08 (0.39)	0.148 (0.49)

Refinement

Data set	Rrp4 Exosome	Csl4 Exosome
	Native	Native
Data range (Å)	20-2.7	20-2.7
Reflections F>0 (cross validation)	69955 (3589)	69041 (1929)
R_{work}[b](R_{free}[c])	0.216 (0.264)	0.224 (0.275)
RMS bond length (Å) (angles)	0.008 (1.38)	0.007 (1.4)

[1] Rrp4-exosome (P2$_1$) cell constants (Å): a=101.6, b=129.6, c=102.3, β=101.6°

[2] Csl4-exosome (P4$_3$22) cell constants (Å): a=b=137.5, c=261.0

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

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

[c]R_{free} = cross validation R-actor for 5% of reflections against which the model was not refined.

[d]Unique anomalous.

3.2 Crystal structure of the archaeal exosome

3.2.1 Structural overview of the exosome isoforms

The Rrp4- and Csl4-exosome isoforms of *Archaeoglobus fulgidus* form large globular molecules with overall dimension of 90 Å x 95 Å x 95 Å (Figure 15). Each exosome assembly consists of three copies of the RNase PH-type sununits Rrp41 and Rrp42. In addition, each complex contains three subunits of either Rrp4 (Rrp4-exosome) or Csl4 (Csl4-exosome).
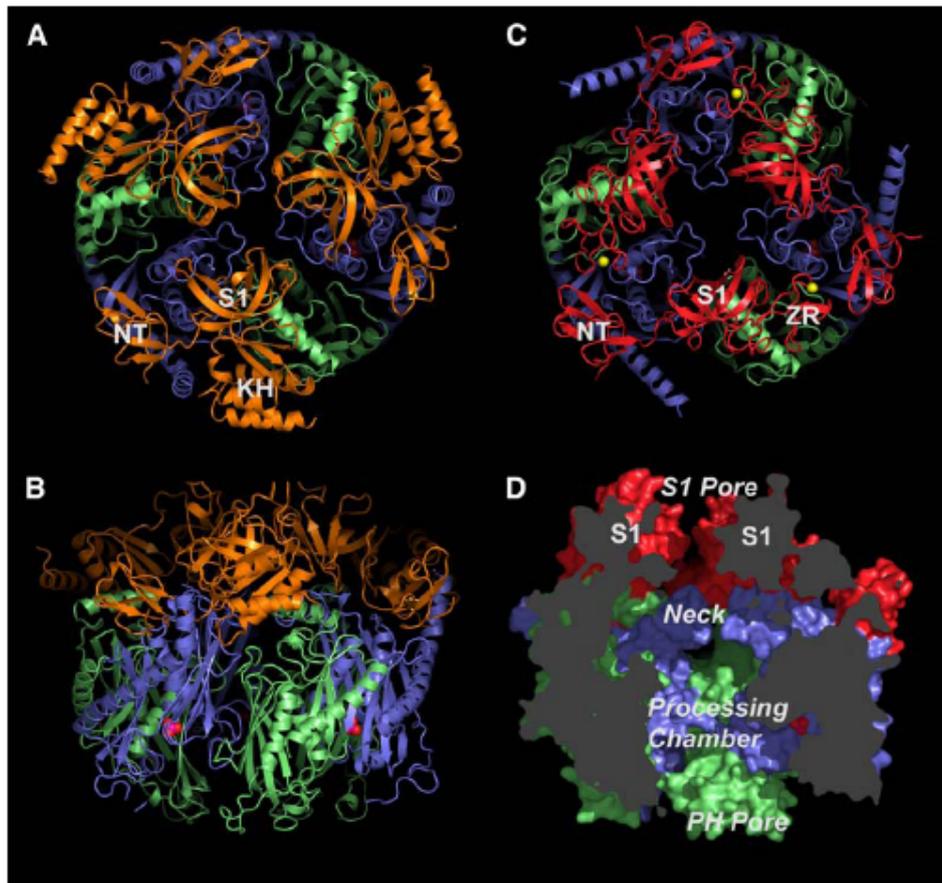
Figure 15: Ribbon models of the nine subunit archaeal Rrp4- (A: top view, B: side view) and Csl4-exosomes (C: top view). Both exosomes contain a ring of alternating Rrp41 (blue) and Rrp42 (green) subunits. Three tungstate moieties (magenta/red cpk model) indicate three active sites in the Rrp41:Rrp42 ring (see text). Three Rrp4 subunits (orange, with annotated NT, S1 and KH domains) or Csl4 subunits (red, with annotated N-terminal (NT), S1 and zinc ribbon (ZN) domains plus yellow zinc ions) bind to the top face of the Rrp41-Rrp42 ring in both exosome complexes, respectively. D) A slice view through the molecular surface of the Cls4-exosome reveals the channel that proceeds from S1 domains (S1 pore) via the central processing chamber to a pore at the RNase-PH domains (PH pore). A 10 Å narrow neck restricts the channel diameter between the S1 pore and the processing chamber.

3.2.1.1 The Rrp41 and Rrp42 subunits form the RNase PH-like core

Three tightly engaged Rrp41:Rrp42 subunits assemble the hexameric core particle of the exosome, which contain the phoshporolytic chamber (Figure 15). Both Rrp41 and Rrp42 subunits exhibit an RNase PH-like fold, with the canonical β-α-β-α layers (Figure 16).
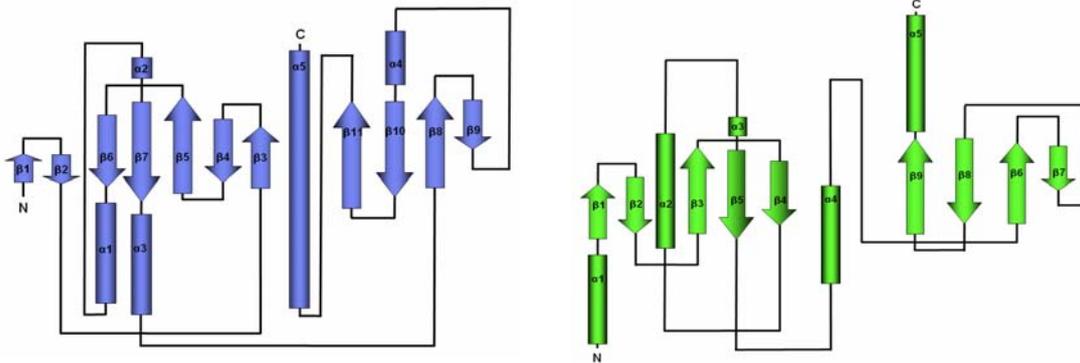
Figure 16: Topology Plot of Rrp41 (blue) and Rrp42 (green) with secondary structure annotation.

Rrp41 and Rrp42 superimpose with eubacterial RNase PH with an overall r.m.s. devation of 1.3 Å and 1.9 Å, respectively, and around 72% of the residues have a similar α-carbon position. Comparison and superimposing of the two archaeal proteins reveal that they are as similar to each other as they are to other RNase PH domains (Figure 15). Rrp41 and Rrp42 exhibit around 21% sequence identity, and their structures superpose with an overall r.m.s. deviation of 1.7 Å for around 74% of the α-carbon positions. The structural differences are limited to the length of individual helices at the N- and C-terminus and to the length of loop regions (Figure 17).

Figure 17: Structural similarity of the *A. fulgidus* Rrp41 (blue) and Rrp42 (green) exosome components with *Bacillus subtilis* RNase PH (wheat, PDB code 1OYS, Harlow et al., 2004).

As seen in Figure 18A, there are two different kinds of interactions between adjacent Rrp41 and Rrp42 subunits. Thus, the hexameric ring is better described as a trimer of heterodimers. In the heterodimers, the two RNase PH domains of Rrp41 and Rrp42 are antiparallel to each other, and the dimerization interface buries a surface area of around 3000 Å^2 (Figure 18B). This kind of up-down dimerization results in an extended β-sheet formed by the adjacent C-terminal β-sheets of the two subunits. In contrast, the trimerization interface is predominantely formed by interactions between the two N-terminal β-sheets and bury a surface area of only 2490 Å^2 (Figure 18B).
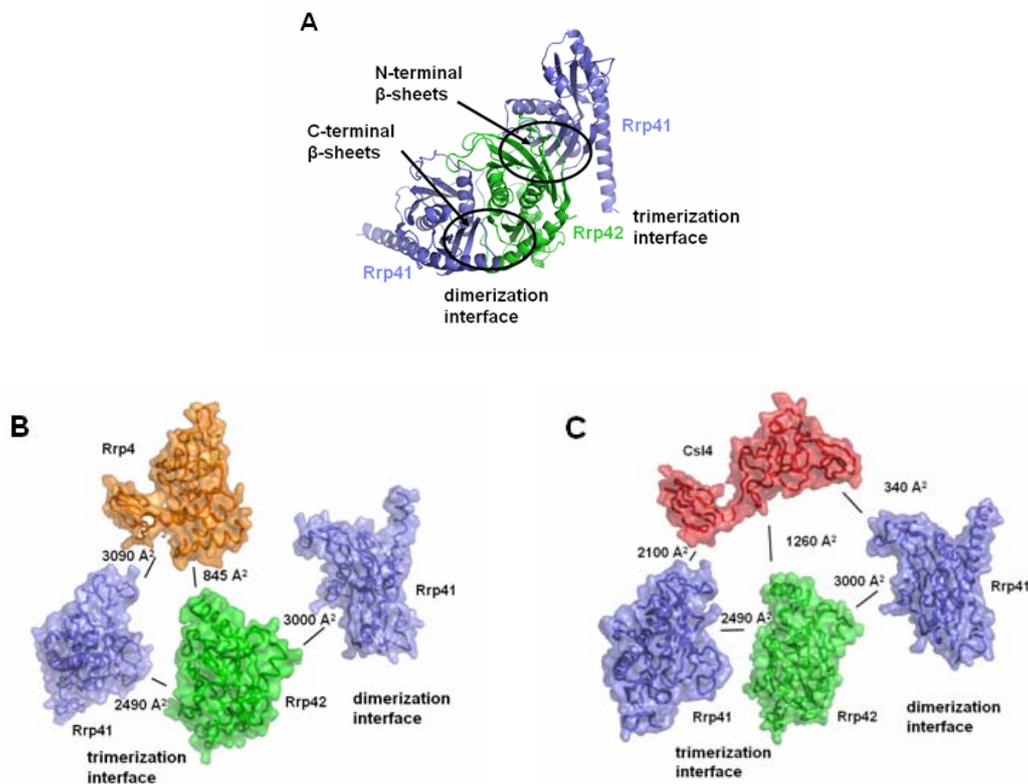
Figure 18: A) Ribbon presentation of *A. fulgidus* exosome RNase PH-domain heterodimers depicting intersubunit contacts to the adjacent Rrp41-domain. B) and C) Surface and ribbon presentation of Rrp41, Rrp42, Rrp4 (B) and Rrp41, Rrp42 and Csl4 (C) subunits "blown up" to visualize the dimerization and trimerization surface. Total buried surface areas between respective subunits indicated numerically and by lines. Color code as indicated in Figure 16.

As a consequence of the up-down dimerization the three Rrp41 subunits are parallel to each other and the three Rrp42 subunits have opposite orientations. A similar overall architecture has been observed in the homohexameric crystal structure of *Bacillus subtilits* RNase PH (Harlow *et al.,* 2004), in the crystal structure of the *Sulfolobus solfataricus* exosome (Lorentzen *et al.,* 2005; Liu *et al.,*

2006), and in bacterial PNPase, which is a trimer of two distinct RNase PH domains (Symmons *et al.,* 2000).

3.2.1.2 The Rrp4 and Csl4 subunits form multimeric caps

Rrp4 or Csl4 subunits are located on one side of the (Rrp41:Rrp42)$_3$ ring and form flat trimeric cap-like assemblies (Figure 15). Both Rrp4 and Csl4 subunits are composed of three domains (Figure 19). Both Rrp4 and Csl4 bind to the Rrp41 subunit with their N-terminal domain and to the adjacent Rrp42 subunit with their S1/KH or S1/zinc-ribbon domains, thereby stabilizing the exosome assembly (Figure 15 and 18B,C). An extended linker connects the N-terminal and the S1 domains and reaches across two adjacent Rrp41:Rrp42 subunits. While the KH domain of the Rrp4 subunit does not interact with the adjacent Rrp41 subunit, the Zn-ribbon domain of Csl4 makes a view interaction with the Rrp41 subunit of the adjacent heterodimer (Figure 18B,C). However, there is no significant interaction among the Rrp4 and Csl4 molecules itself, indicating that all Rrp4 and Csl4 subunits bind to the (Rrp41:Rrp42)$_3$ assembly independently. This is consistent with the heterogeneity of exosomes in *in vitro* reconstituted complexes (see 3.1.3) and previous analysis of exosomes from archaeal cells (Evguenieva-Hackenberg *et al.,* 2003; Farhoud *et al.,* 2005).
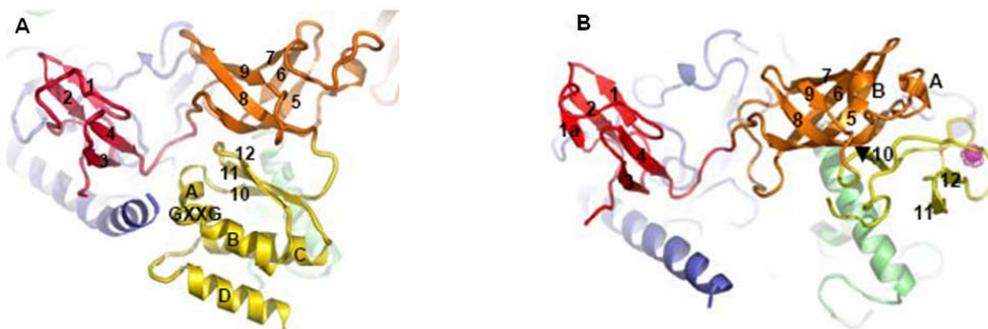
Figure 19: A) Domain organization of Rrp4, with highlighted and annotated secondary structure. Rrp4 possesses an N-terminal all-β domain (red), a middle S1 domain (orange) and a C-terminal KH-domain (yellow). The KH domain contains the conserved GXXG motif that is implicated in nucleic acid recognition. B) Cls4 possesses an N-terminal all-β domain (red), a middle S1 domain (orange) and a C-terminal zinc ribbon domain (yellow, with magenta zinc ion). The location of the Ski4-1 mutation in the interface of S1 and Zn-ribbon domains is indicated by a black triangle.

Both Rrp4 and Csl4 subunits contain an N-terminal six-stranded all-β domain, which binds into the "armpit" formed by the protruding C-terminal helix of Rrp41

(Figure 20). The structural data and the comparable high buried surface area between the N-terminal domains and the respective Rrp41 subunits indicate that this domain is the major anchor for the caps to the phosphorolytic ring (Figure 18B,C). The core of the interface is formed by a Pro-Gly loop that is conserved among archaeal and eukaryotic Rrp4, Rrp40 and Csl4 sununits, suggesting that archaeal and eukaryotic Csl4- and Rrp4-type subunits interact with the RNase PH domain ring in a similar manner (Figure 20 and below).
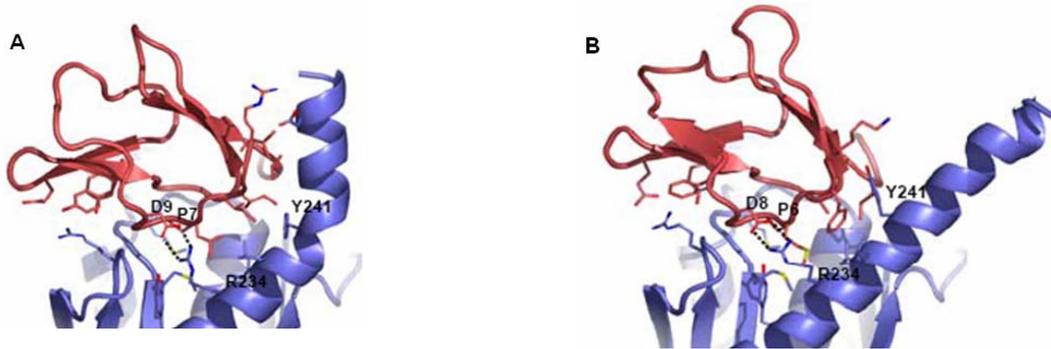
Figure 20: A) Detailed view on the interface of the N-terminal domain of Rrp4 (red) and the C-terminal helix of Rrp41 (blue). The backbone is shown as ribbon, interface side chains are displayed as color coded sticks and annotaded. B) Detailed view on the interface of the N-terminal domain of Cls4 (red) and the C-terminal helix of Rrp41 (blue), similar to panel B.

Apart from some additional conserved sidechain interactions, the remainder of the Csl4:Rrp41 and Rrp4:Rrp41 interface is surprisingly different. Rrp41 responds to the different N-terminal domains by conformational plasticity (e.g., Tyr241[Rrp41] in Figure 23). In addition, the C-terminal helix of Rrp41 is substantially kinked in the Rrp4-exosome as compared to the Csl4-exosome due to the presence of the additional KH domain (Figure 15A,C and 20).

The S1 domains belong to the oligonucleotide and/or oligosaccharide binding (OB) fold proteins and possess the typical kidney-shaped, five-stranded β-barrel structure (Figure 19). In both Rrp4 and Csl4, the S1 domain is situated at the center of the caps and interact with a radial helix on top of Rrp42. Thereby, the concave sides of the S1 domains of Rrp4 and Csl4 form an app. 15 Å and 18 Å wide pore (denoted "S1 pore"), respectively. The S1 domains as well as the N-terminal domain of both Rrp4 and Csl4 share not only the fold, but also the overall position.

The Rrp4 KH domain is situated on the outside of the cap (Figure 15 A). The KH domain possesses the "eukaryotic" KH fold and packs to the convex side of the S1 domain (Figure 19). The KH domain also interacts with the C-terminal helix of

Rrp41 and the radial helix on top of Rrp42 (Figure 18B). All these interfaces exhibit a substantial hydrophobic nature, suggesting that the overall domain arrangement of S1 and KH domains is fixed, apart from local mobility of the S1 domains (Figure 21).
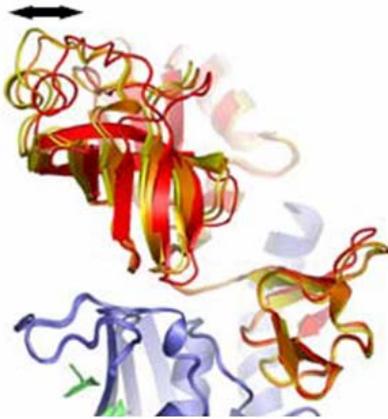
Figure 21: Local mobility of the S1 domain revealed by superposition of the three Rrp41:Rrp42:Rrp4 trimers in the structurally asymmetric Rrp4-exosome. The three Rrp4 subunits are shown in red, orange and yellow, while Rrp41 and Rrp42 are shown blue and green, respectively. The S1 domains can at least move by 5-6 Å (double headed arrow), a distance that could allow threading of RNA one base at a time (see text)

Cls4 contains a zinc-ribbon domain instead of a KH domain, in which four cysteine residues coordinate a central zinc ion (Figure 19B). The zinc-ribbon domain occupies the space between S1 and N-terminal domains of Csl4 and neither shares fold, nor position with the Rrp4 KH domain. Like the KH domain, the zinc-ribbon domain shares a considerable hydrophobic interface with the S1 domain. A mutation in a conserved glycine in this interface gives rise to the Ski4-1 phenotype in yeast (Figure 1 and 19B). Ski4-1 strains have normal rRNA maturation activity but are defective in mRNA degradation (van Hoof *et al.,* 2000). The interface location of the mutation suggests that the mutual arrangement of S1 and the zinc-ribbon domain is important for mRNA degradation. In support of this, the zinc-ribbon domain is positioned between two adjacent S1 domains, toward the outside of the cap, and additionally interacts with the adjacent Rrp41 molecule (Figure 15C and 18C). This orientation is ideally suited to provide additional macromolecular interaction sites for recognition of RNA or protein cofactors.

3.2.1.3 Entry into and Exit from the inner processing chamber

The "S1 pore" in both complexes represents the entrance opening to a central channel formed by the (Rrp41:Rrp42)$_3$ ring. Directly beneath the "S1 pore", an Rrp41 loop (residues 57 to 74) protrudes into the central channel. Stacking interactions between the three Rrp41 loops are mediated via Pro64^{Rrp41} and His66^{Rrp41} and result in narrowing of this "neck" to app. 10 Å diameter (Figure 22).
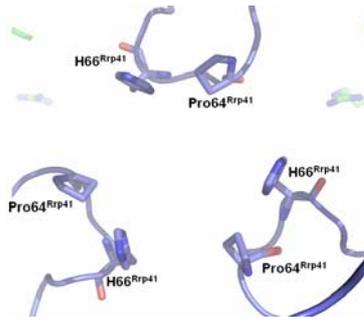
Figure 22: Close up view on the "neck" loop of the three Rrp41 molecules. Stacking interaction between Residue 64 (Proline) and 66 (Histidine) stabilize the loops in a conformation that closes most of the central channel beneath the S1 pore, leaving a central hole of app. 10 Å.

Beneath the neck, the central widens to form the central chamber of the archaeal exosome. This processing chamber contains the phosphorolytic active sites and is additionally connected to the solvent by a second ~ 18 Å wide opening, denoted "PH pore". The "PH pore" is located opposite to the "S1 pore" and represents the exit (Figure 15D).

3.2.2 Active sites and the processing chamber of the exosome core

RNase PH family RNases degrade RNA by adding inorganic phosphate across the $5´\rightarrow 3´$ phosphodiester bond between the ultimate and penultimate nucleotides (Figure 23) (Symmons *et al.,* 2000). This reaction releases a nucleotide diphosphate from the 3´-end of the RNA substrate ($[pN]_n$ + Pi $\rightarrow$ $[pN]_{n-1}$ + NDP). In the presence of ADP and low amounts of phosphate, RNase PH family enzymes are capable of catalyzing the reverse reaction and polyadenylate the 3´ end.
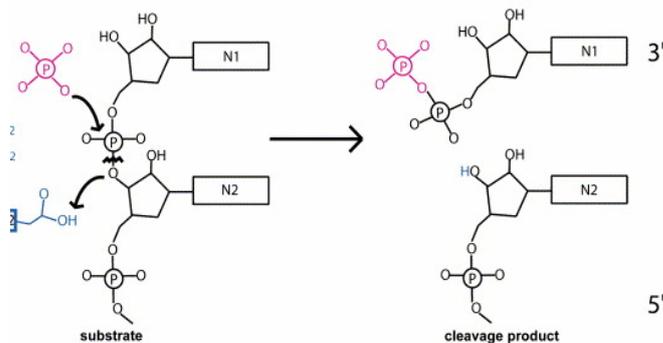
Figure 23: Schematic drawing of the mechanism in phosphorolytic cleavage reactions. The attacking inorganic phosphate is shown in magenta (adapted from (Lorentzen and Conti 2005)).

In order to identify the phosphorolytic active sites in the archaeal exosome, crystals were soaked with tungstate. Tungstate mimics phosphate and is useful to locate phosphorolytic sites and possible nucleic acid binding domains of RNase PH-type enzymes (Symmons *et al.,* 2000; Harlow *et al.,* 2004). Crystals of the Rrp4-exosome were incubated in a solution containing 50 mM tungstate. Anomalous dispersion data at the tungsten LIII absorption edge were recorded. Strong anomalous difference electron density could be observed in all three active site pockets in the Rrp41 subunit in the central chamber (Figure 24). In contrast,

no anomalous density in the equivalent positions of the Rrp42 subunits could be observed. These data indicate that the archaeal exosome contains three phosphorolytic active sites at the interface of Rrp41 and Rrp42, which is consistent with mutagenesis results derived in this thesis (see below) and from others (Lorentzen *et al.,* 2005). In these Rrp41 pockets, N-termini from two Rrp41 helices form two suitable phosphate binding moieties on which strong tungstate anomalous difference density can be observed (Figure 24A, 1, 1a). The highest tungstate occupancy (>8σ) is found near the end of each pocket (Figure 27A, 1, 2 and 3). To obtain a detailed insight into the residues involved in binding of the tungstate, a tungstate/phosphate was modelled into one of the active site pockets (Figure 24C, pocket 1, 1a). The tungstate/phosphate is coordinated by $G135^{Rrp41}$, $S136^{Rrp41}$ and $R97^{Rrp41}$. Two acidic residues, $D180^{Rrp41}$ and $D186^{Rrp41}$, are suitably positioned to participate in the phosphorolytic attack. The functional relevance of these sites was later on refirmed by a $D180^{Rrp41}A$ mutation, which resulted in complete loss of RNase activity (see below). Additional, a less occupied tungstate density in form of an elongated stretch between the pocket and the central channel can be observed in the shown active site (Figure 24A,C, 1a). Here, three arginines, $R91^{Rrp41}$, $R96^{Rrp41}$ and $R107^{Rrp41}$, form a positively charged rim beneath the density and are suitably positioned to bind the backbone of RNA in a sequence-independent manner (Figure 24C). Thus, the second observed density could mark a path for RNA that enters the phosphorolytic active sites.
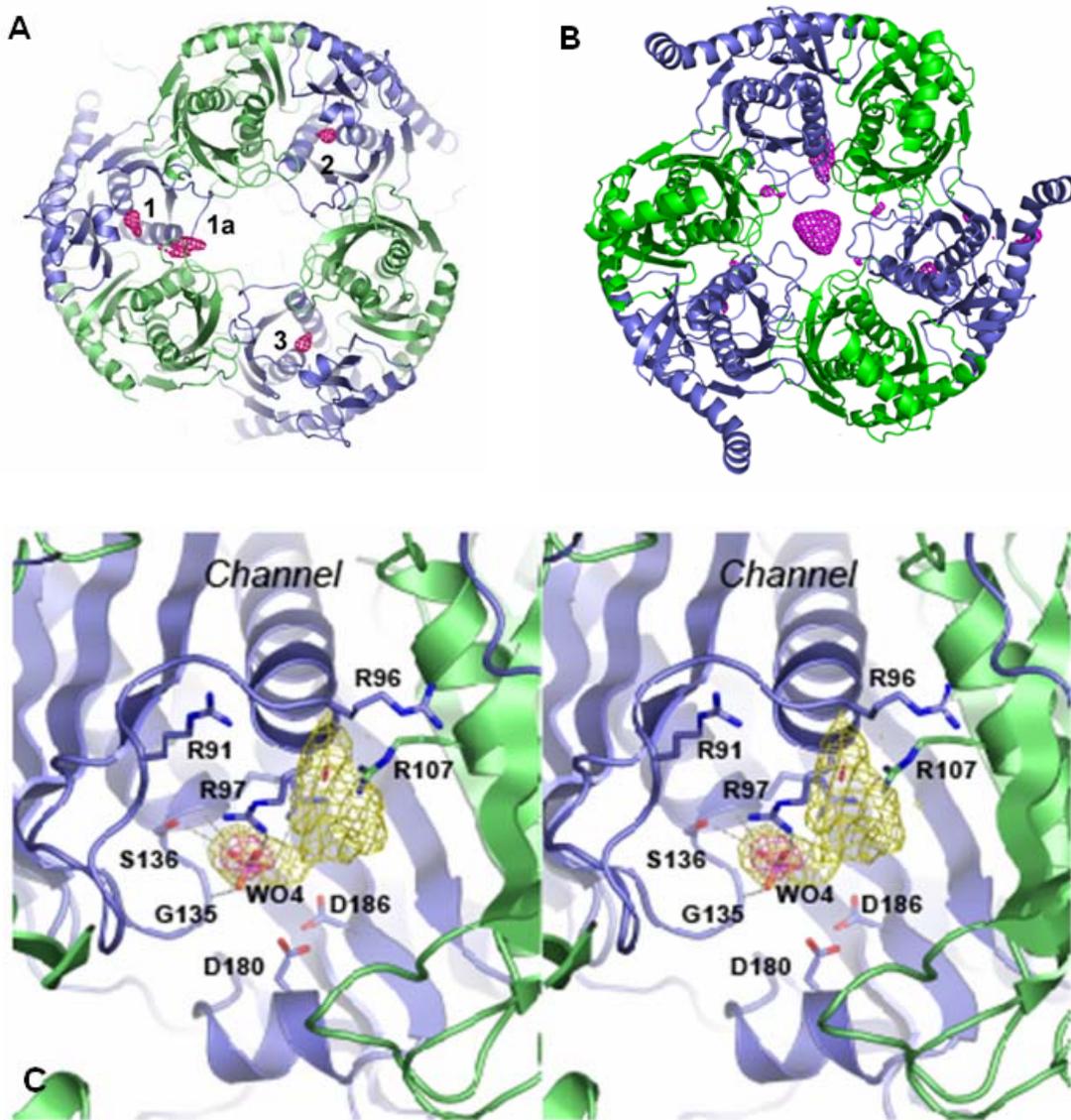
Figure 24: A) Ribbon model of the Rrp4-exosome (bottom view), superimposed with anomalous difference density (magenta) derived from tungstate binding experiments. B) Ribbon model of the Rrp4-exosome (top view, without Rrp4 subunits), similar to A). C) Stereo plot of a close-up view of the active site pocket in the interface of Rrp41 (blue) and Rrp42 (green). The Rrp4-exosome is shown as ribbon model, using the color code of Fig. 1. The side chains of notable conserved active side residues are shown as annotated color coded sticks. Tungsten anomalous difference density is shown with 4σ (yellow) and 8σ (magenta) contour. A tungstate moiety (WO_4) is modeled into the 8σ contour.

Additional tungsten anomalous difference density could be observed in the opening between the three neck loops (Figure 24B and 15D). This remarkably broad and strong density may represent binding of a tungstate cluster and supports the idea that the neck region interact with the phosphate backbone of RNA. Furthermore, the location of the tungsten difference densities may represent the possible pathway of substrate RNA (see below). However, the presented

results are consistent with recent results on the exosome from *S. solfataricus* in which the neck loops directly interacts with the backbone of a trapped RNA oligonucleotide (Lorentzen *et al.,* 2007).

3.3 The archaeal exosome shows RNA dependent activities

3.3.1 RNA degradation and polyadenylation

To directly assess the RNA degradation activity of the *Archaeoglobus fulgidus* exosome, RNase experiments with wild-type and mutant protein were carried out as described. As seen in Figure 25A, both Rrp4- and Csl4-exosomes possess processive RNase activity and degrade a single-stranded poly(A) RNA-substrate *in vitro*. Under the given conditions the Csl4-exosome (CEx, Rrp41, Rrp42, Csl4) and the Csl4+Rrp4 containing exosome complexes (CREx, Rrp41, Rrp42, Rrp4, Csl4) show the highest degradation activity, whereas the Rrp4 exosome (REx, Rrp41, Rrp42, Rrp4) exhibits lowest RNase activity (Figure 25B), suggesting that the cap proteins might function differently in RNA recognition and binding. Both complexes are also capable to catalyze the reverse reaction in which a single-stranded RNA substrate is polyadenylated in the presence of ADP (Figure 25A). This polyadenylation activity has also been shown for eubacterial PNPase, which catalyzes the addition of poly(A) tails to RNAs (Mohanty and Kushner 2000), and for the *S. solfataricus* exosome (Lorentzen *et al.,* 2005). As expected, mutation of the conserved D180 to an uncharged alanine resulted in complete loss of both RNase and polyadenylation activity.
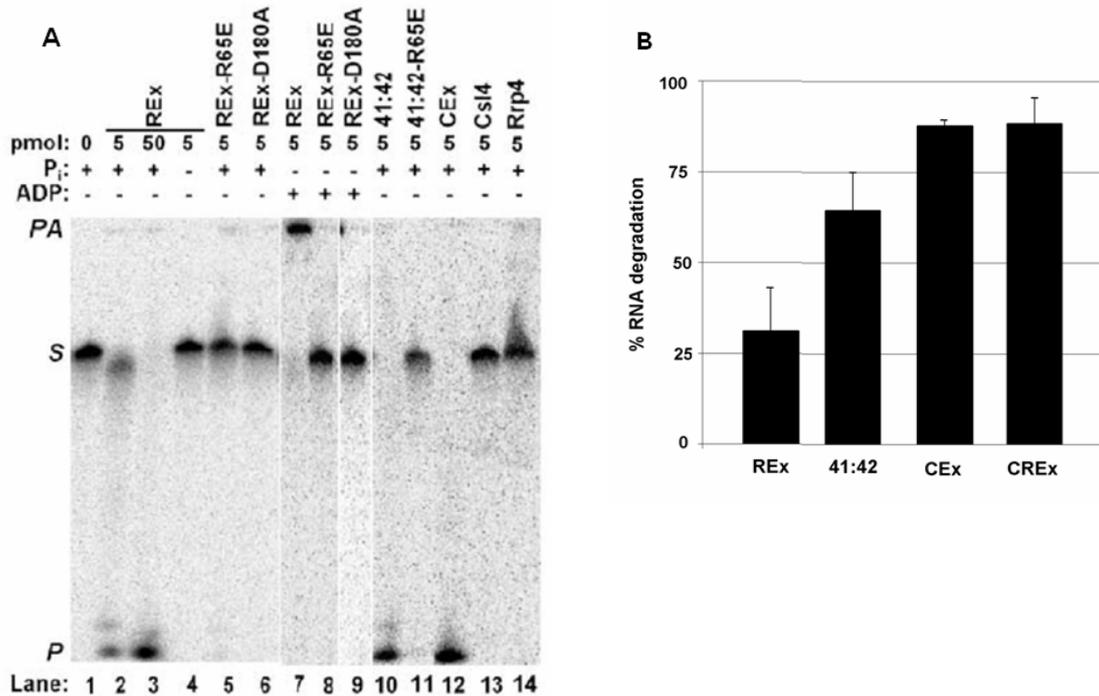
Figure 25: A) RNase activity and polyadenylation assays. Protein, phosphate or ADP was added as indicated and substrate oligo(rA) 30mer (S) and polyadenylated (PA) or phosphorolytically degraded (P) nucleotide products were resolved by denaturing polyacrylamide gel electrophoresis. The Rrp4-exosome (REx) shows processive degradation (lanes 2, 3) and processive polyadenylation (lane 7) of a 30mer oligo(rA), dependend on phosphate (lane 4) or ADP. Both activities are abolished or severely inhibited by the D180A mutation in the active site (lane 6, 9) or the R65E mutation in the neck loop (lane 5, 8). Controls show that Csl4-exosome (CEx, lane 12) and the Rrp41:Rrp42 complex (41:42, lane 10, 11) possess a similar activity, while Csl4 (lane 13) and Rrp4 (lane 14) subunits alone do not possess RNase activity *in vitro*. **B)** Relative RNase activities of exosome isoforms.

To more directly assess the functional relevance of the "neck", RNase and polyadenylation activity of a R65^{Rrp41}E mutant protein was analyzed. As seen in Figure 25A, no residual RNase activity can be observed. In addition, the R65^{Rrp41}E mutant enzymes do not possess any polyadenylation activity.

3.3.2 RNA binding

As the loss of RNase activity in the R65^{Rrp41}E mutant indicates that RNA substrates enter the exosome via S1 pore and neck, the role of the "neck" was further addressed by analyzing RNA binding abilities of wild-type and mutant R65E^{Rrp41} Rrp4-exosome and Rrp41:Rrp42 core complex as described. As shown in Figure 26, both Rrp4-exosome and its R65E mutation bind RNA with

comparable affinities. However, the obvious differences in the band-shift patterns may indicate a somehow altered RNA binding mode due to the presence of the neck mutation. This data is not unexpected, as S1 and KH domain may bind RNA independently from the central chamber and the neck.
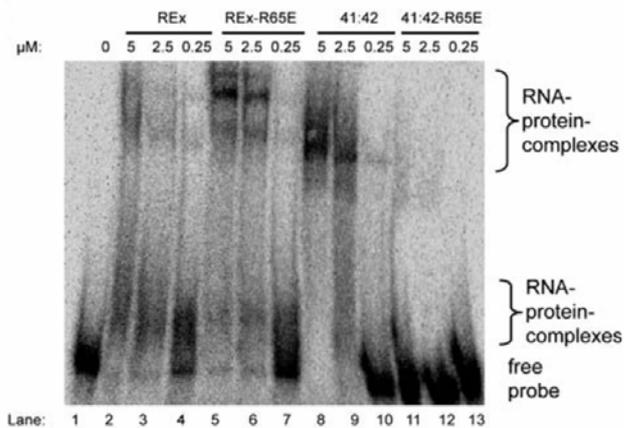
Figure 26: Gel shift assays of the wild-type and R65E containing Rrp4-exosome (REx) and Rrp41:Rrp42 ring (41:42). R65E changes the shift pattern of the Rrp4-exosome (lanes 2-4 and 5-7), but abolishes RNA binding by the Rrp41:Rrp42 ring (lanes 8-10 and 11-13).

Furthermore, binding of RNA to the Rrp41:Rrp42 ring and its R65E mutation was analyzed. As seen in Figure 26, the R65E mutation completely abolishes RNA binding by the Rrp41:Rrp42 ring. The most plausible explanation for these results is that RNA initially binds to S1 and/or KH domains and then enters the processing chamber via the narrow neck region.

All in all, these functional data support the idea that the conserved arginine 65 and the neck motif are important elements in processive RNA degradation, although they are more than 50 Å away from the phosphorolytic active sites. Furthermore, these functional observations strongly support a model, in which RNA passes through the neck into the central chamber to the active sites. Consistently, bacterial PNPase that carries a mutation in an equivalent neck (FFRR-loop) shows decreased phosphorylase activity *in vivo* (Jarrige *et al.,* 2002).

4 Discussion

4.1 The Cap proteins form a macromolecular recognition surface

4.1.1 Both Rrp4 and Csl4 domains share structural homology to ribosomal proteins

The exosome of *Archaeoglobus fulgidus* is a globular, double-donut like structure with a central hole (Figure 15). One ring of this donut is assembled by three heterodimers of the RNase PH-like Rrp41:Rrp42 proteins. The second ring is formed by three copies of Rrp4 or Csl4 and is located on one side of the RNase PH ring. As both Rrp4 and Csl4 subunits possess well-known protein-nucleic acid recognition motifs (KH and/or S1 domain) the structural data suggest that the caps form a large RNA recruitment surface.

Structural and functional data for a number of S1 and/or OB fold domains indicate an involvement in the recognition and binding of both RNA and DNA substrates. The S1 and KH domains of *Arabidopsis thaliana* (At) Rrp4 were already shown to bind RNA *in vitro*, and the KH domains of both AtRrp4 and *A. fulgidus* Rrp4 also possess the typical nucleic-acid recognizing helix-turn-helix motif with the central GXXG sequence (Figure 19A) (Chekanova *et al.*, 2002). In addition, both S1 and KH domains of eubacterial PNPase are required for efficient RNA degradation (Stickney *et al.*, 2005), and the S1 domain of PNPase contains a high affinity poly(A) binding site (Yehudai-Resheff *et al.*, 2003).

Comparison of the Rrp4/Csl4 S1 domain with DNA oligonucleotide bound S1 domains reveals that the archaeal S1 domains probably bind the nucleic acid at the side, which frames the S1 pore in the crystal structure of the archaeal exosome (Figure 15A,C and 27) (Cavarelli *et al.*, 1994; Shuman and Lima 2004). The KH domain of Rrp4 is located at the periphery of the ring. However, comparison with nucleic acid bound KH domains show that the proposed RNA binding region is fully accessible to substrate RNA (Figure 15A and 27).

Figure 27: Tube representation of the S1 and KH domains of Rrp4 (orange), superimposed with nucleic acid complexes of OB (PDB code 1ASZ, (Cavarelli *et al.,* 1994)), KH (PDB code 1J5K, (Braddock *et al.,* 2002) and N-terminal (NT, PDB code 1V8Q, (Wang *et al.,* 2004); without nucleic acid) domains (blue with brown nucleic acids).

In contrast, the Csl4 subunit contains a zinc-ribbon instead of a KH domain. Zn-ribbon domains are structurally divers, but typically form protein-protein or protein-nucleic acid recognition modules (Bernstein *et al.,* 2003; Naryshkina *et al.,* 2003). In addition, both Rrp4 and Csl4 subunits share a similar N-terminal domain, which seems to be conserved also in the human exsome Rrp4, Rrp40 and Csl4 counterparts (Figure 1) (Liu *et al.,* 2006).

To look for structural neighbors of the zinc-ribbon and N-terminal domains, a DALI search was performed. Surprisingly, the only significant structural neighbors of zinc-ribbon and N-terminal domains turned out to be ribosomal proteins. The N-terminal domain possesses a fold similar to ribosomal protein L27 (PDB entry 1V8Q, Figure 27), whereas the zinc-ribbon domain most closely matches ribosomal protein L37ae (PDB entry 1FFK chain W). Both ribosomal proteins are implicated in binding to ribosomal RNA (Ban *et al.,* 2000; Wang *et al.,* 2004). From an evolutionary point of view, it is interesting that ribosome architecture (L27 and L37ae and S1 domain proteins) and ribosome maturation (exosome cap domains) share structural features.

4.1.2 Electrostatic conservation reveals possible RNA binding sites

Because all nine domains of the Csl4 and Rrp4 caps belong to domain families that are typically involved in macromolecular interactions, the caps could form a large platform that recruits macromolecular substrates and cofactors. To further characterize the molecular surface, the electrostatic potential was mapped onto the Rrp4- and Csl4-exosome complexes. Consistent with the possibility that the caps form a major RNA binding site, Rrp4 and Csl4 caps contain a positive

surface potential, whereas the remaining surface is highly negatively charged (Figure 28). The strongest positive potential is found in the S1 pores. Significantly, the locations of these positive patches superimpose well with characterized nucleic acid binding sites of numerous S1 and/or OB fold domains. Evidently, the typical location of RNA binding sites on S1 domains and the matching surface charge suggest that the S1 pore is a major RNA binding site on the caps. The proposed RNA binding site on the KH domain, which also exhibits positive surface potential, is quite remote from the S1 domains and suggests that RNA binding to these two sites, if any, occurs independently. This structural implication is supported by independent RNA binding of S1 and KH domains of *A. thaliana* Rrp4 *in vitro* (Chekanova *et al.,* 2002). However, a possible interplay between both KH and S1 domain in RNA substrate recognition and distinction can not be ruled out. This idea is supported by recent small angle X-ray scattering (SAXS) studies of the *Pyrococcus furiosus* exosome, which suggest that the S1 and KH domains are quite dynamic in solution (Ramos *et al.,* 2006) and is consistent with the high flexibility observed for S1 and KH domains in the crystal structure of PNPase (Symmons *et al.,* 2000).
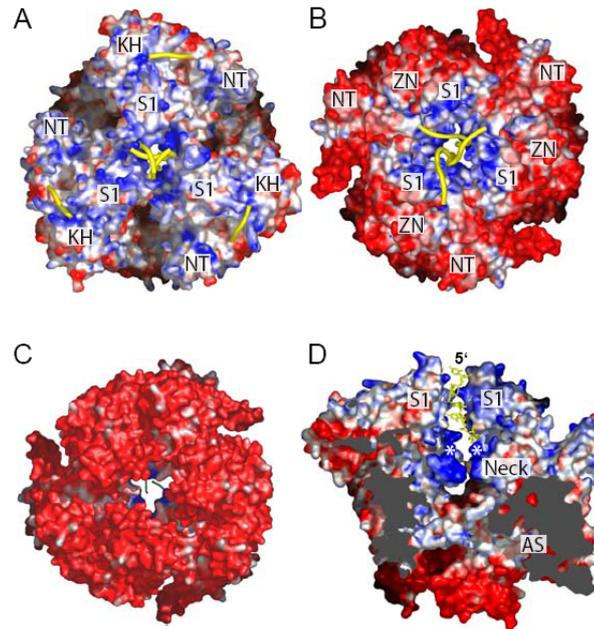
Figure 28: Molecular surfaces of Rrp4- and Csl4 exosomes in different views, colored according to electrostatic potential (red -7kt/e⁻ to blue +7kT/e⁻). Rrp4- (panel A, top view) and Csl4-exosome (panel B, top view) contain strong positive patches at S1 and KH domains. In contrast, the exterior of the RNase-PH domain ring is highly negatively charged (panel C, bottom view of the Csl4 exosome). The strongest positive potential reaches from "S1 pore" via the neck loops into the processing chamber (panel D: slice view of the Rrp4 exosome), indicating a likely path for RNA. The location of the functionally important neck residue R65^{rrp41} is indicated by (*). The positive patches on S1 and KH domains match RNA binding sites of related OB fold and KH domains. RNA derived from a superposition of exosome S1 and KH domains with nucleic acid complexes of OB fold or KH domains are shown as yellow tube.

Taken together, the orientation of S1, KH, and zinc-ribbon domains suggests that the caps form a macromolecular interaction surface with two zones. The inner zone, formed by the S1 domain, would be ideally suited to regulate RNA entry through the S1 pore and neck (see below). The function of the peripheral zone (KH and zinc-ribbon domains) is less clear, its peripheral location might indicate that these domains are more involved in the specific recognition of RNA targets such as rRNA, mRNA, and poly(A) tails, but could also be involved in the recognition of cofactors. A role in substrate and/or cofactor selection would be consistent with the Ski4-1 phenotype, where a separation-of-function mutation in the S1-zinc-ribbon interface of yeast Csl4 specifically abolishes mRNA degradation, but does not affect rRNA maturation (van Hoof *et al.,* 2000). In support of different substrate specificities of S1 and KH domains, the functional

results reported here showed that the tested oligo(rA) 30-mer is more efficiently degraded by the Csl4-exosome and Csl4/Rrp4-hybrid complexes than by the Rrp4-exosome or the Rrp41:Rrp42 ring alone (Figure 25B). Thus, the Rrp4/Csl4 subunits are probably involed in specificity and regulation of RNA processing.

4.2 Central channel and proposed RNA path

How does substrate RNA now reach the active sites inside the phosphorolytic chamber? As seen in the crystal structure of the archaeal exosome, the processing chamber contains two suitable openings, denoted S1 and PH pores. The PH pore at the bottom of the phosphorolytic ring is closer to the active sites and slightly wider than the S1 pore. However, it does not contain recognizable RNA binding elements, it carries a strong negative surface potential, and it is far away from the putative RNA binding S1/KH domain layer (Figure 28D). It is therefore unlikely that RNA enters through the PH pore. A more intriguing possibility for RNA entry is the S1 pore at the Rrp4/Csl4 cap structures. Because of the highly postitve electrostatic potential and the favorable location of the RNA binding S1 and KH/Zn-ribbon domains, this pore is well suited to interact with RNA (Figure 31). Consistently, the path from the putative RNA binding S1/KH domains to the active sites through the neck carries the strongest positive electrostatic potential in the archaeal exosome. In addition, strong tungsten anomalous difference density was observed in the opening between the three neck loops, which supports the idea that the neck loop directly interacts with the phosphate backbone of RNA (Figure 24B and 29).
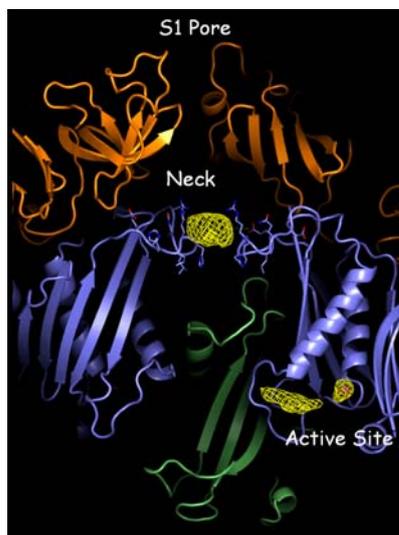
Figure 29: Functional analysis of the neck. Shown is a slice through the Rrp4-exosome (in color coded ribbon representation) with highlighted neck side chains and tungstate difference density (yellow mesh). Tungstate density and the R65 ^{Rrp41}E mutation (annotated) indicate that the neck region is important for RNA processing and presumably relays RNA from the S1 domains to the active sites (AS).

The functional importance of the neck could be further assessed by analyzing the capability of the R65E^{Rrp41} mutant to bind and to degrade a 30mer poly(A) RNA substrate. As mentioned, this R65 is located in a loop in Rrp41 which directly lines the neck. Furthermore, the positive charge at this position is highly conserved throughout archaeal and eukaryotic Rrp41-like subunits (Lorentzen *et al.,* 2005; Liu *et al.,* 2006), and a mutation in an equivalent neck in bacterial PNPase (FFRR-loop) results in a protein with decreased acitivty *in vivo* (Jarrige *et al.,* 2002). The mutation itself did not influence the assembly of the nine subunit exosome and appears to have no evident architectural function (data not shown). The fact, that exosomes carrying the R65E^{Rrp41} mutation are defect in degrading and binding as well as polyadenylating the RNA substrate indicates that the RNA is no longer able to access the active sites. All in all, the structural and functional observations strongly support a model in which RNA passes through the neck into the central chamber, and the "neck" directly interacts with RNA (Figure 28D and 29). In addition, this model is supported by recent structural results on the *S. solfataricus* exosome, which could be co-crystallized in the presence of an RNA molecule containing a stable stem-loop structure at the 5′-end and a 3′-poly(A) tail (Lorentzen *et al.,* 2007). Although no well-ordered density was observed for the 5′ stem-loop structure, Lorentzen *et al*. could detect well-ordered electron density for one nucleotide. This nucleotide is bound in the central channel at the narrow constriction formed by the "neck" loop, which clearly shows, that the "neck" interacts with RNA.

4.3 Implication for exosome degradation mechanism

The crystal structure of the *Archaeoglobus fulgidus* exosome revealed that the archaeal exosome possesses three active sites buried in a cage-like structure, which are connected to the putative RNA binding domains by a central channel and pore. The architecture of the archaeal exosome now provides a first basis for understanding controlled RNA degradation and suggests a likely mechanism for processive RNA degradation which may be also applicable, at least to some extent, to the eukaryotic and eubacterial counterparts (Figure 30A). The putative RNA binding subunits Rrp4 and Csl4 form a suitable multidomain surface which is probably involved in substrate and/or cofactor recruitment and regulation. A direct role for these subunits in both RNA substrate recognition and distinction is

supported by a separation-of-function mutation in yeast Csl4 (Ski4-1), which affected mRNA decay but not rRNA processing (van Hoof *et al.,* 2000) and by the observation that both KH and S1 domains of eubacterial PNPase are required for efficient RNA degradation *in vivo* (Stickney *et al.,* 2005). Thus, binding of substrate and/or cofactors may be mediated by the peripheral KH and Zn-ribbon domains, whereas the central S1 domains are ideally suited to guide RNA via the S1 pore to the "neck" and into the central chamber. As the S1 domains exhibit intrinsic flexibility, they appear sufficiently mobile to promote RNA translocation of one base at a time (Figure 24). In archaeal exosomes, the 3′-tail of single-stranded RNA that enters the central chamber has a choice of three active sites (Lorentzen and Conti 2005). This high concentration of active sites in the archaeal exosome, in combination with tight RNA binding at the neck (Lorentzen *et al.,* 2007) and S1 domains, may explain the high processivity of these enzymes on unstructured RNA. At present, it is still not clear whether RNA is processively degraded only in one active site, or alternates between all three. However, structural and functional data from the archaeal exosome are in stark contrast to the eukaryotic exosome. The human exosome has obviously retained only one phosphorolytically active site (Rrp41), whereas the yeast exosome appears to have lost all phosphorolytically activity in the RNase PH core or at least depend entirely on the aid of coactivator complexes (Liu *et al.,* 2006; Dziembowski *et al.,* 2007).
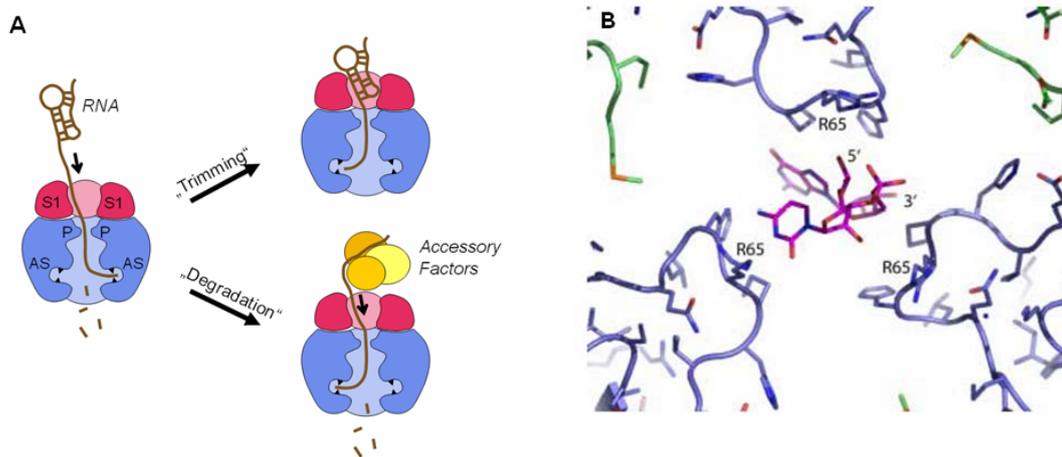
Figure 30: A) Proposed model for RNA degradation and trimming by exosomes. For simplicity, only two out of three active sites (AS) and S1 domains are shown. Secondary structures can stall at the S1 pore (S1) and neck (P) and abort processive degradation. Accessory factors with helicase activity could stimulate processive degradation. Sufficiently stable secondary structures of protein:RNA complexes are only trimmed. B) Figure: Model of a RNA dinucleotide (magenta) in the neck (blue Rrp41, green Rrp42) suggests that only unfolded RNA can pass the narrow 10 Å restriction below the S1 pore. The RNA was taken from PDB 1ASZ and manually fitted into the neck.

Both S1 pore and "neck" are only 8-12 Å wide, a feature that suggests that only single-stranded RNA without any secondary structures can enter the processing chamber. This idea is supported by molecular modelling (Figure 30B) and recent data on the *S. solfataricus* exosome in complex with an RNA stem-loop (Lorentzen *et al.,* 2007). In addition, the positively charged "neck" is also a conserved feature in eukaryotic exosomes (Liu *et al.,* 2006), supporting that the model presented here is also applicable to the eukaryotic RNA processing machinery. However, RNA molecules in a cellular environment are typically bound to proteins or form stable secondary structures, and as that are protected to degradation. Thus, the relative width of the entry pore could be important to prevent uncontrolled decay of cellular RNA. In support of this, the eukaryotic exosome is more or less inactive on long RNA substrates *in vitro* and requires the presence of coactivators like the yeast SKI and TRAMP complexes (Brown *et al.,* 2000; van Hoof *et al.,* 2000; Araki *et al.,* 2001; LaCava *et al.,* 2005; Wyers *et al.,* 2005). *In vivo*, these coactivator complexes are probably involved in the efficient unfolding of RNA secondary structures and/or in the removal of bound proteins. Both yeast SKI and TRAMP complexes contain subunits that possess or are predicted to possess RNA helicase activity. Thus, the coactivator complexes may target and unfold substrate

RNA, thereby providing sufficiently unstructured RNA for degradation by the exosome. In addition, the model predicts that processive degradation stalls at stable secondary structures or protein roadblocks that are not unwound or removed by additional factors (Figure 30A). This idea is consistent with recent data on the *S. solfataricus* exosome and eubacterial PNPase, which were both shown to stop processive degradation eight to nine nucleotides before an engineered double-stranded RNA stem-loops (Carpousis *et al.,* 1999; Lorentzen and Conti 2005). Furthermore, the structures presented here suggest that initial threading of RNA substrates into the central channel should be facilitated by a highly unstructured 3′ tail of several nucleotides in length. Although the role and the source of polyadenylation in archaeal cells remains to be elucidated (Portnoy *et al.,* 2005; Portnoy and Schuster 2006), the poly(rA) tails of eukaryotic mRNA and the 3′ poly(rA) tails added by the TRAMP complex on nuclear RNA targets of the exosome are ideally suited for such an initial threading. All in all, the structural data of the archaeal exosome explain how sequence-unspecific RNA degradation is coupled to simple but tight regulatory strategies like limitation of access to the active sites by narrow constrictions, and interaction with coactivator complexes. Furthermore, these strategies, which prevent uncontrolled decay of cellular RNA, can be extrapolated to both eukaryotic and eubacterial RNA degradation machineries.

4.4 Common features in 3′→ 5′ RNA degrading enzymes

4.4.1 The structure of the exosome is highly conserved in both archaea and eukaryotes

The archaeal exosome consists of a hexameric core of three phosphorolytically active RNase PH subuntis (Rrp41) and three inactive RNase PH subunits (Rrp42). A trimeric ring of subunits with putative RNA-binding domains (Rrp4 and Csl4) is positioned on top of the hexamer opposite to the RNA degrading sites. According to its simpler archaeal counterpart, the eukaryotic core exosome consists of three Rrp41- (Rrp41, Rrp46, Mtr3) and three Rrp42- (Rrp42, Rrp45, Rrp43) like subunits as well as Rrp4, Rrp40 and Csl4 subunits (Mitchell *et al.,* 1997). Electron microscopy as well as yeast two hybrid data already suggested the association of the six eukaryotic RNase PH-type subunits into a hexameric, PNPase-like structure (Aloy *et al.,* 2002; Raijmakers *et al.,* 2004). Now, the recently solved

crystal structure of the human exosome allows detailed atomic insights into the complex structure of the eukaryotic exosome (Liu *et al.,* 2006) and comparison of the structure with its archaeal counterpart.

The overall structure of the human exosome is quite similar to that of the *Archaeoglobus fulgidus* exosome (Liu *et al.,* 2006), with a hexameric core of three RNase PH-like heterodimers, capped by a trimer of S1-domain-containing proteins (Figure 31).
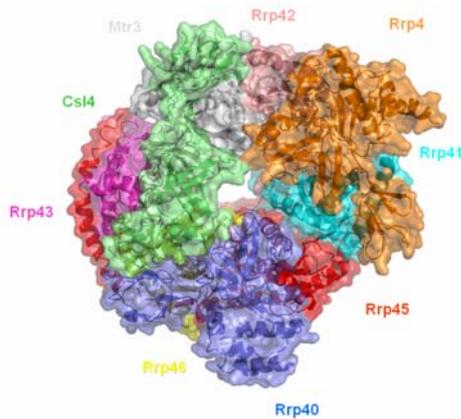
Figure 31: Crystal structure of the human exosome core complex reveals a similar overall architecture compared to the archaeal exosome (PDB code 2NN6, (Liu *et al.,* 2006)). As shown, the C-terminal extension of the human Rrp45 subunit (red) wraps around the Rrp46 and Rrp43 subunit.

The highest similarity between both exosome structures is found for the subunits of the RNase PH ring. Each Rrp41-and Rrp42-like subunit of the human exosome can be superimposed with an r.m.s.d. between 1.3 Å to 2.0 Å for over 70 % of all cα-positions to the corresponding *A. fulgidus* subunits (Figure 32A,B). Differences are limited to the length of single loop regions or helices, as in the case of the human Rrp45 subunit, which possesses a 180 aa long C-terminal extension that wraps around the adjacent Rrp46 and Rrp43 subunit and appears to be conserved in higher eukaryotes (Figure 31 and 32B) (Liu *et al.,* 2006).

The human S1-domain containing cap proteins Rrp4 and Rrp40 superimpose with an r.m.s.d. of 2.4 Å and 2.6 Å, respectively, and the human Csl4 subunit with an r.m.s.d. of 2.3 Å for below 70 % of all cα-positions to the archaeal Rrp4 and Csl4 proteins (Figure 32C,D,E).
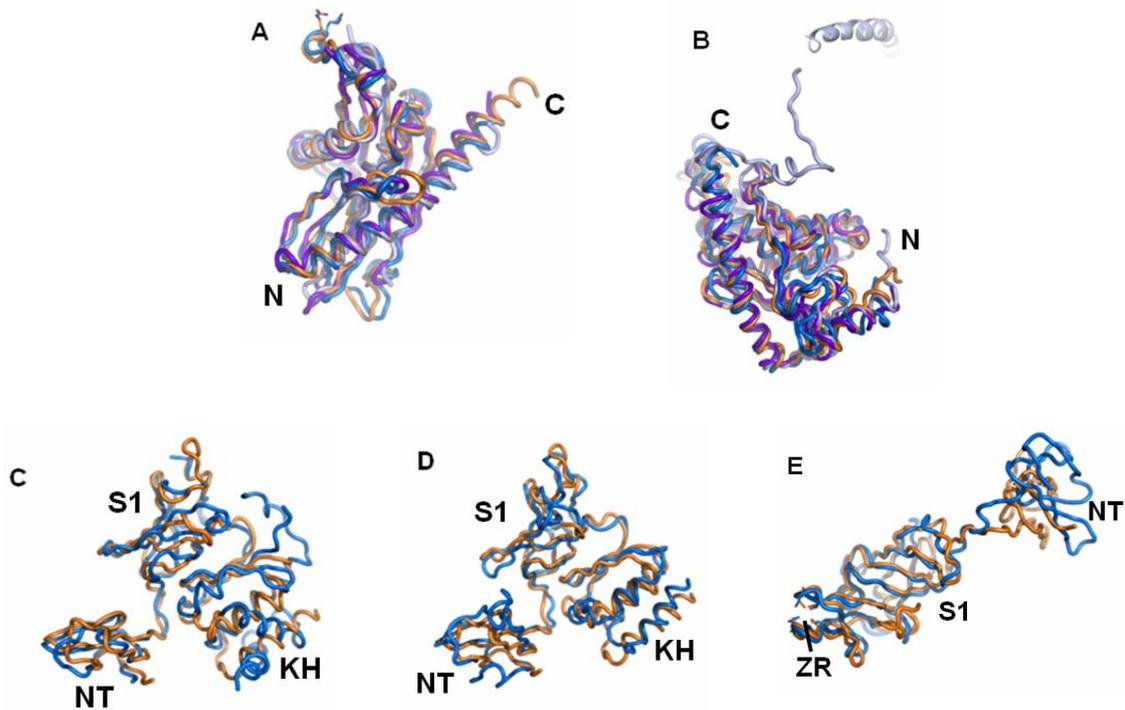
Figure 32: Structural overlays of the RNase PH like and S1-containing subunits of archaeal and human exosome (PDB code 2NN6; (Liu *et al.,* 2006)). A) archaeal Rrp41 (orange) and human Rrp41 (blue), Rrp46 (lightblue) and Mtr3 (purple). B) archaeal Rrp42 (orange) and human Rrp42 (blue), Rrp45 (lightblue) and Rrp43 (purple). C) archaeal (orange) and human (blue) Rrp4. D) archaeal Rrp4 (orange) and human (blue) Rrp40. E) archaeal (orange) and human (blue) Csl4.

The overall fold topology of human Rrp4 and Rrp40 N-terminal and S1 domain is remarkably similar to archaeal Rrp4. However, while the KH domain of human Rrp40 matches well with archaeal Rrp4, human Rrp4 exhibits some differences in the orientation of the KH domain, and possesses an additional C-terminal helix (Figure 32C,D) (Liu *et al.,* 2006). Human Csl4 exhibits a similar topology to archaeal Csl4 in the S1 and C-terminal domain. Differences in the N-terminal domain are visible in the orientation, but not in the overall fold (Liu *et al.,* 2006). Although eukaryotic Csl4 lost the C-terminal Zn-ribbon domain, it still adopts the same conformation as the C-terminal domain of archaeal Csl4 (Figure 32E). Thus, the zinc ion probably has only a structural role, for instance to maintain the fold of this small domain in the harsh growth einvironment of archaea. However, the subunit diversification of the RNase PH-like domains and S1/KH-like domains in eukaryotic exosomes may result in different substrate preferences or binding of specific protein partners.

The presence of heterotrimeric caps in the human exosome, which are necessary to form a stable assembly (Liu *et al.,* 2006 143), now raises the question if there are also heterotrimeric caps in archaeabacteria *in vivo*. To see if archaeal homotrimeric Rrp4 and Csl4 caps are compatible with heterotrimeric caps, one Rrp4 subunit was substituted by its corresponding Csl4 subunit *in silico*, after structurally superimposing the two Rrp41-Rrp42 rings of the Csl4- and Rrp4-exosomes (Figure 33). The Csl4 subunit fits well into the space between the two remaining Rrp4 subunits without evident steric clashes and the necessity of structural changes. Comparison with the heterotrimeric cap structure of the human exosome indeed reveals striking similarities. In conjunction with the presented biochemical findings (see 3.1.3) these data support the idea that heterotrimeric caps form also *in vivo* in archaeabacteria.
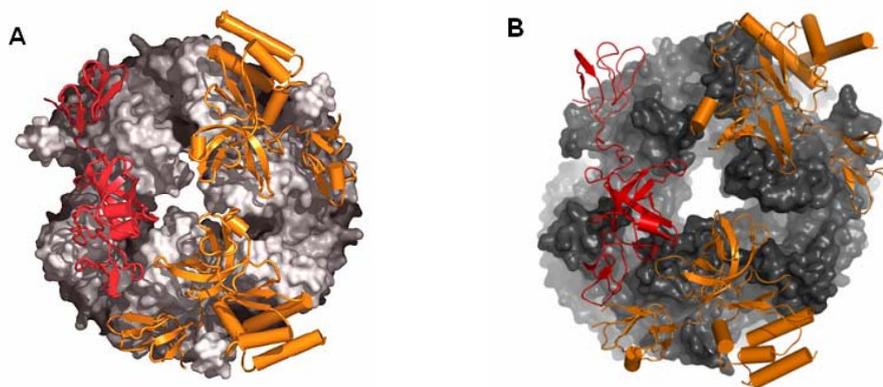
Figure 33: A) Model for a heterotrimeric Csl4-Rrp4-Rrp4 cap on the molecular surface of the RNase-PH domain (grey). Placement of a Csl4 subunit (red ribbon model) into an Rrp4 cap (orange ribbon models, one of the subunits is substituted by Csl4). The placement was guided by superimposing the RNase PH domains of Rrp4- and Csl4-exosomes *in silico*. B) Structure of the human heterotrimeric cap on the surface of the RNase PH ring (grey). Colors as indicated in A). (Rrp40: bottom; Rrp4: up).

Besides the S1 pore, an additional highly conserved feature of both archaeal and eukaryotic exosome is the presence of the positively charged "neck" loop (residues 60-70 in archaeal Rrp41 and residues 56-64 in human Rrp41). In the human exosome, only the Rrp41 "neck" loop is well ordered, but clearly adopts a conformation similar to that in archaeal Rrp41 (Liu *et al.,* 2006).

Although RNA processing by eukaryotic exosomes is probably more complex due to additional RNase activities and cofactors, the conservation of S1 pore and "neck" loop, and the structural and functional results presented here support a

unifying mechanistic framework for the RNA degradation mechanism in both archaeal and eukaryotic exosomes.

4.4.2 The structure of the archael exosome reveals similarities to eubacterial 3´→ 5´ RNA degrading enzymes

The identification of the exosome 30 years ago already showed that its subunits share high sequence similarity with eubacterial RNases (Wolfe *et al.,* 1977). From an evolutionary point of view the exosome appears to have evolved from homo-oligomeric bacterial proteins like RNase PH (Wolfe *et al.,* 1977; Mitchell *et al.,* 1997; Koonin *et al.,* 2001). RNase PH from *E. coli* consists of a hexameric ring of three identical RNase PH-homodimers with up-down orientations (Harlow et al., 2004). The assembly into a ring like structure is important for activity in the maturation of tRNA precursors (Deutscher *et al.,* 1988; Zhou and Deutscher 1997; Harlow *et al.,* 2004). RNase PH-like domains are also present in eubacterial polynucleotide phosphorylase (PNPase). PNPase is a central factor in the 3´→ 5´ degradation of unstructured RNA substrates in bacterial cells and in eukaryotic organelles such as chloroplasts and mitochondria (Crofton and Dennis 1983; Haag and Lewis 1994; Zhou and Deutscher 1997; Yehudai-Resheff *et al.,* 2003; Bollenbach *et al.,* 2004). Bacterial PNPase consists of a tandem repeat of RNase PH-like domains, followed by a KH and a S1 domain on a single polypeptide (Figure 34A) (Symmons *et al.,* 2000). Three PNPase molecules assemble in a hexameric ring which has a similar architecture not only to the one described for eubacterial RNase PH, but also to the RNase PH ring of the archaeal and eukaryotic exosome (Figure 34B) (Symmons *et al.,* 2000; Symmons *et al.,* 2002; Harlow *et al.,* 2004; Lorentzen *et al.,* 2005; Liu *et al.,* 2006). The RNase PH-type ring of PNPase is topped on one side by three KH and S1 domains (Figure 34C). The location of S1 and KH domains in PNPase is remarkably similar (although not identical) to the location of S1 and KH domains of the Rrp4 subunits bound to the exosome core particle in respect to the active sites (Figure 34A), suggesting a common mechanistic basis for RNA recognition in both processing machineries. Only one of the RNase PH-like domains of PNPase, denoted RNase PH domain 2, is catalytically active (Symmons *et al.,* 2000). According to that, only one of the archaeal RNase PH-like proteins possesses a functional characterized active site. In addition, the experimentally defined tungstate binding sites of the *A. fulgidus*

Rrp4-exosome closely matches those derived for PNPase (Figure 34A) (Symmons *et al.,* 2000).
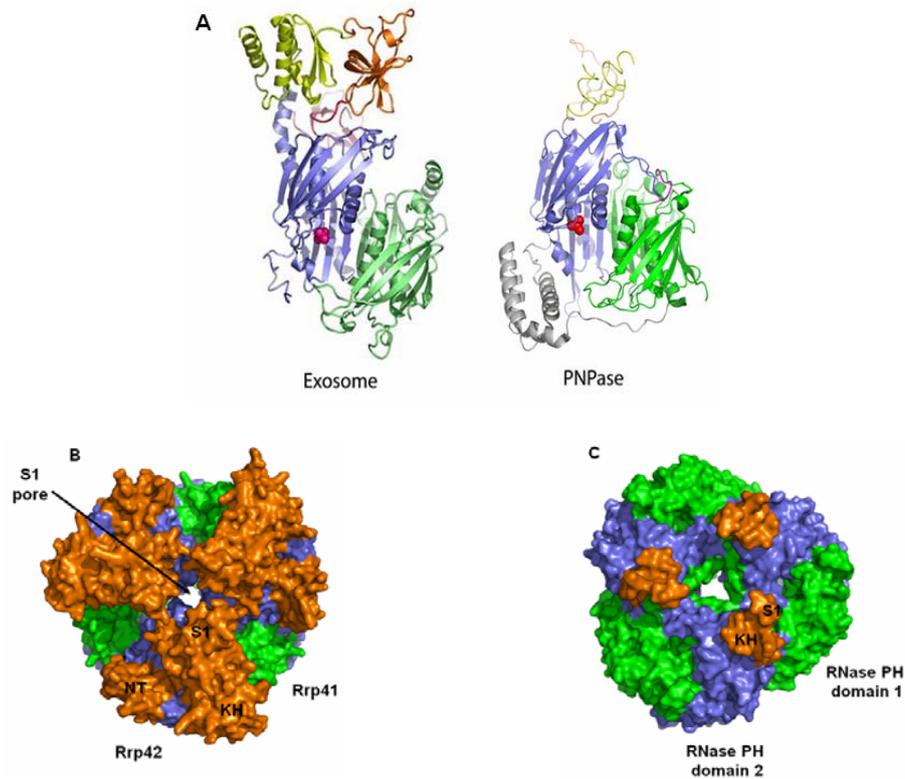
Figure 34: A) Ribbon representations of the Rrp4 exosome complex (only three out of nine subunits are shown) and bacterial poly-nucleotide phosphorylase (PNPase, only one out of three subunits is shown, PDB code 1E3H, (Symmons *et al.,* 2000)). Active sites shown with magenta tungstate moieties. PNPase FFRR-loop is shown in magenta. Color code of Figure 16, S1 domain: orange, KH domain: yellow. B) and C) Surface representation of the nine subunit *A. fulgidus* exosome (left) and the structurally related bacterial PNPase trimer (right). Color code of Figure 16. Green PNPase RNase PH domain 1: inactive, blue PNPase RNase PH domain 2: active.

Another remarkable similarity is the presence of a postivitely charged loop in both exosome ("neck"-loop, residues 60-70 in Rrp41) and PNPase ("FFRR"-loop, residues 78-92 in RNase PH domain 1). Although both loops are conserved neither in sequence nor in structure, they both occupy an ideal position to constrict access to the central channel and to interact with RNA substrates passing through the channel (Symmons *et al.,* 2000). Consistent with that, degradation activity of the archaeal exosome is abolished by the "neck" mutation R65E^{Rrp41} (Figure 25A), whereas a mutation in the "FFRR-loop" of PNPase significantly diminishes RNA binding and degradation abilities *in vivo* (Jarrige *et al.,* 2002). The structural and functional data on the archaeal exosome and eubacterial PNPase, together with

data from RNA protection assays on PNPase (Spickler and Mackie 2000), indicate that both RNA degradation machineries recruit substrates through the central channel, although the details of the substrate recognition are likely to differ. However, the remarkable overall equivalence suggests a common evolutionary basis of RNA degradation by archaeal/eukaryotic exosomes and bacterial/organelle PNPases. In addition, it raises the possibility that the mechanism of 3´ RNA decay by phosphorolytic degradation complexes is more or less conserved across the kingdoms of life.

4.5 Similar principle in protein degradation by the proteasome

The proposed structural mechanism for RNA degradation by the exosome reveals a remarkably similar principle to the degradation of proteins, a feature that has long been suspected (van Hoof and Parker 1999). In eukaryotes and archaea, proteins are degraded by the proteasome, or in eubacteria by the related HslU and ClpP proteases. They all share a similar cage-like structure in which unfolded substrate biopolymers are fed through narrow entrance pores to reach the active sites (Figure 35). The active sites are located in a central chamber, which shields the reaction from the outside. Once the reaction has started, the substrate is trapped in the processing chamber to prevent release between consecutive rounds of cleavage. An additional high local concentration of active sites as in the case of the archaeal exosome or archaeal proteasome could thereby be important for processivity.
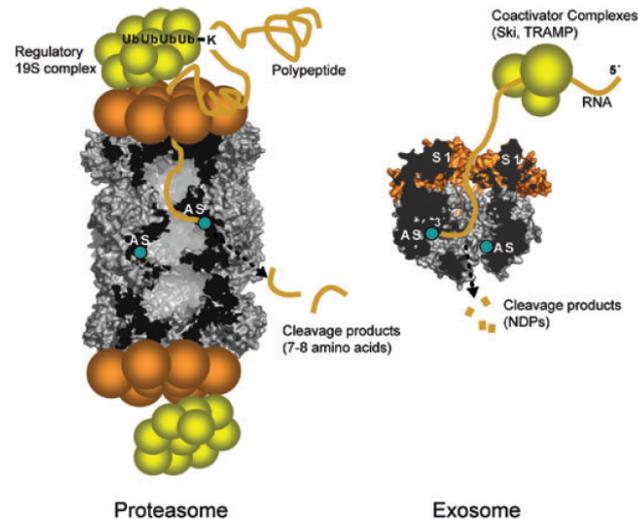
Figure 35: Comparison of the archaeal exosome (RNA degradation) with the proteasome (Protein degradation). Both degradation multiprotein complexes are assembled into a chamber like structure. Substrate molecules enter the processing chambers through narrow entrance pores. This allows degradation of only unfolded protein (Proteasome) or single-stranded RNA (exosome) molecules. Accessory factors are needed to unfold protein or RNA secondary and tertiary structure, respectively. AS: active sites; S1: S1 domains

In addition, targeting and unfolding of the substrates is facilitated by the presence of coactivators. In the case of the eukaryotic proteasome, regulatory complexes bind to the outside of the entry pores and thus are ideally positioned to specifically recognize ubiquitinylated proteins marked for degradation. The regulatory subunits unfold the substrate proteins in an ATP-dependent manner and promote threading of the poylpeptides into the processing chamber (Voges *et al.,* 1999). Although the regulatory mechanisms of the exosome are less understood, recent results could identify several coactivator complexes of the eukaryotic exosome. The cytoplasmic activity of the yeast exosome depends on the presence of the SKI complex (Anderson and Parker 1998; Brown *et al.,* 2000; Araki *et al.,* 2001), whereas the degradation of several nuclear RNAs requires the Trf4p-Air1/2p-Mtr4p and/or Trf5p-Air1p-Mtr4p (TRAMP) complexes (Haracska *et al.,* 2005; LaCava *et al.,* 2005; Vanacova *et al.,* 2005; Wyers *et al.,* 2005; Houseley *et al.,* 2006). Both coactivator complexes contain a SF2 helicase (Ski2p or Mtr4p), which could be involved in the unwinding of RNA secondary structures prior to degradation by the core exosome. Thus, the structural mechanism and the molecular regulation of protein and RNA degradation exert at least three striking similarities: the first is based on a simple size exclusion strategy, which limits access to only unfolded

substrate molecules. The second is based on the presence of multiple active sites located in a protected environment, which allows processive degradation, and which prevents dissociation of the substrate. The third similarity is based on the interaction with several coactivator complexes, which are involved in substrate targeting, and which provide unfolded substrate molecules. And although the subunits of the exosome and the proteasome share no structural homology, the similar principle in the controlled degradation of RNA and polypeptides is quite fascinating.

4.6 Future persepectives and outlook

The structural data on the archaeal exosome presented here not only suggest a likely mechanism of how RNA is bound and degraded by the exosome, but also reveals an interesting analogy between controlled RNA and protein degradation by nano-compartmentalized enzymes.

However, there are still many unanswered questions regarding substrate specificity and activation of the exosome in both eukaryotes and archaea. One of the most outstanding questions in understanding RNA surveillance and quality-control pathways affects the ability of the exosome to distinguish on the one hand between RNA that needs to be fully degraded and RNA that needs to be processed, and on the other hand between defective RNA and normal RNA processing intermediates. Although we can not answer this question at the moment due to the lack of detailed structural and functional information, it is possible that the S1, KH and zinc-ribbon domains of archaeal Rrp4 and Csl4 have an important role in substrate recognition. The presence of three different putative RNA binding subunits in the eukaryotic exosome may indicate that their functional and biochemical difference is responsible for the recognition of different types of RNA molecules (e.g. mRNA vs. rRNA). In addition, as the gene for archaeal Csl4 is not located in the same operon as Rrp41, Rrp42 and Rrp4 (Koonin *et al.,* 2001), an cell-cycle dependent regulated expression of Csl4 may allow the archaeal exosome to degrade or process other RNA substrates as in the presence of Rrp4 alone. Furthermore, the presence of exoribonucleolytic subunits associated with the eukaryotic core exosome may suggest an additional possibility to specifically target substrate RNA, although it remains unclear by which mechanism specific substrates are directed to specific exonuclease active sites.

Another important thing that has to be investigated in future work is the mechanistic interplay between the exosome and its coactivator complexes like the cytoplasmatic SKI and the nuclear TRAMP complex. The biochemical basis for the activation of the exosome is still unclear, but it is possible that these complexes target and provide unfolded RNA and/or modulate the activity of the core exosome by specifically binding to it. In archaea, the possible presence of coactivator complexes needs to be investigated. Although archaeal exosome complexes were shown to co-purify with a protein that exhibits sequence homology to eubacterial DNA primase DnaG, the function of this interaction and the possible role of this primase homologue remains unclear (Evguenieva-Hackenberg *et al.,* 2003; Farhoud *et al.,* 2005).

Finally, it is unclear if RNA substrates are marked for degradation equivalent to the poly-ubiquitylation of proteins and thus facilitate recognition by the exosome. A general "degradation signal" could be e.g. a 3´-polyadenylated tail, as observed for many exosome substrates. This idea is supported by the polyadenylation of nuclear exosome RNA substrates by the TRAMP complex prior to degradation (LaCava *et al.,* 2005), and the high processivity of the archaeal exosome on oligo(rA) RNA substrate. Interestingly, archaea do not possess homologues of known polyadenylation factors, but the *S. solfataricus* exosome itself catalyses the polyadenylation of RNA *in vivo* (Portnoy *et al.,* 2005; Portnoy and Schuster 2006), consistent with its polyadenylation activity *in vitro*. Polyadenylation activity has also been shown for eubacterial PNPase (Grunberg-Manago *et al.,* 1955). In contrast to eukaryotes, 3´-polyadenylation of mRNA in eubacteria was shown to serve as signal for degradation (Dreyfus and Regnier 2002). Although structural and functional data on both exosome and PNPase suggest that polyadenylated single-stranded RNA 3´-tails facilitate entry in the respective processing chamber (Symmons *et al.,* 2000), the direct biochemical role of the polyadenylated 3´-tails in RNA degradation by the exosome in eukaryotes and archaea needs to be elucidated.

These and many other questions will have to be answered in future work on the archaeal and eukaryotic exosome using a combination of structural and molecular biology.

5 Summary

The exosome is a conserved 3´→ 5´ exoribonuclease complex involved in cellular RNA metabolic processes in eukaryotes and archaea. Its involvement in the accurate processing of nuclear RNA precursors and in the degradation of RNA in both nucleus and cytoplasm implies a central function in the eukaryotic RNA surveillance machinery. This widespread function implies the ability of the exosome to distinguish between RNA substrates that should be matured by the removal of nucleotides to a precisely defined end point, and defective RNAs that undergo rapid and complete degradation. However, the structural and molecular mechanisms of processive 3´→ 5´ RNA degradation and substrate specificity remain unclear.

To obtain insights into the structural and functional organization of the exosome, I determined crystal structures of two 230 kDa nine subunit exosome isoforms from *Archaeoglobus fulgidus*. Both exosome isoforms contain a hexameric ring of RNase PH-like domain subunits Rrp41 and Rrp42 with a central chamber. A trimer of Rrp4 or Csl4 subunits is situated on one side of the RNase PH domain ring and forms a multidomain macromolecular interaction surface with central S1 domains and peripheral KH and zinc-ribbon domains. Tungstate soaks identified three phosphorolytic active sites inside the central processing chamber. Additional structural and functional results suggest that the S1 domains of Rrp4 or Csl4 subunits and a subsequent neck in the RNase PH domain ring form an RNA entry pore that only allows access of unstructured RNA to the active sites. The structural results presented here can not only mechanistically unify observed features of exosomes, including processive 3´ RNA degradation of unstructured RNA, the requirement for regulatory factors and coactivators to degrade structured RNA, and the precision in processing RNA species to a defined length. But the high conservation of the archaeal exosome to the eukaryotic exosome and its additional high structural similarity to bacterial mRNA-degrading PNPase suggest a common basis for 3´ RNA-degradation in all kingdoms of life. Furthermore, the structure of the archaeal exosome reveals remarkable architectural and functional similarities to the protein degrading proteasome.

6 References

Allmang, C., J. Kufel, G. Chanfreau, P. Mitchell, E. Petfalski and D. Tollervey (1999). "Functions of the exosome in rRNA, snoRNA and snRNA synthesis." Embo J **18**(19): 5399-410.

Allmang, C., P. Mitchell, E. Petfalski and D. Tollervey (2000). "Degradation of ribosomal RNA precursors by the exosome." Nucleic Acids Res **28**(8): 1684-91.

Allmang, C., E. Petfalski, A. Podtelejnikov, M. Mann, D. Tollervey and P. Mitchell (1999). "The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases." Genes Dev **13**(16): 2148-58.

Aloy, P., F. D. Ciccarelli, C. Leutwein, A. C. Gavin, G. Superti-Furga, P. Bork, B. Bottcher and R. B. Russell (2002). "A complex prediction: three-dimensional model of the yeast exosome." EMBO Rep **3**(7): 628-35.

Anderson, J. S. and R. P. Parker (1998). "The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex." Embo J **17**(5): 1497-506.

Andrulis, E. D., J. Werner, A. Nazarian, H. Erdjument-Bromage, P. Tempst and J. T. Lis (2002). "The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila." Nature **420**(6917): 837-41.

Araki, Y., S. Takahashi, T. Kobayashi, H. Kajiho, S. Hoshino and T. Katada (2001). "Ski7p G protein interacts with the exosome and the Ski complex for 3'-to-5' mRNA decay in yeast." Embo J **20**(17): 4684-93.

Arigo, J. T., D. E. Eyler, K. L. Carroll and J. L. Corden (2006). "Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3." Mol Cell **23**(6): 841-51.

Baker, N. A., D. Sept, S. Joseph, M. J. Holst and J. A. McCammon (2001). "Electrostatics of nanosystems: application to microtubules and the ribosome." Proc Natl Acad Sci U S A **98**(18): 10037-41.

Ban, N., P. Nissen, J. Hansen, P. B. Moore and T. A. Steitz (2000). "The complete atomic structure of the large ribosomal subunit at 2.4 A resolution." Science **289**(5481): 905-20.

Baumli, S., S. Hoeppner and P. Cramer (2005). "A conserved mediator hinge revealed in the structure of the MED7.MED21 (Med7.Srb7) heterodimer." J Biol Chem **280**(18): 18171-8.

Bernstein, D. A., M. C. Zittel and J. L. Keck (2003). "High-resolution structure of the E.coli RecQ helicase catalytic core." Embo J **22**(19): 4910-21.

Blow, D. (2002). Outline of Crystallography for Biologists. New York, Oxford University Press.

Bollenbach, T. J., G. Schuster and D. B. Stern (2004). "Cooperation of endo- and exoribonucleases in chloroplast mRNA turnover." Prog Nucleic Acid Res Mol Biol **78**: 305-37.

Braddock, D. T., J. L. Baber, D. Levens and G. M. Clore (2002). "Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: solution structure of a complex between hnRNP K KH3 and single-stranded DNA." Embo J **21**(13): 3476-85.

Briggs, M. W., K. T. Burkard and J. S. Butler (1998). "Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation." J Biol Chem **273**(21): 13255-63.

Brown, J. T., X. Bai and A. W. Johnson (2000). "The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex in vivo." Rna **6**(3): 449-57.

Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54 (Pt 5)**: 905-21.

Butler, J. S. (2002). "The yin and yang of the exosome." Trends Cell Biol **12**(2): 90-6.

Carpousis, A. J., N. F. Vanzo and L. C. Raynal (1999). "mRNA degradation. A tale of poly(A) and multiprotein machines." Trends Genet **15**(1): 24-8.

Cavarelli, J., G. Eriani, B. Rees, M. Ruff, M. Boeglin, A. Mitschler, F. Martin, J. Gangloff, J. C. Thierry and D. Moras (1994). "The active site of yeast aspartyl-tRNA synthetase: structural and functional aspects of the aminoacylation reaction." Embo J **13**(2): 327-37.

Chekanova, J. A., J. A. Dutko, I. S. Mian and D. A. Belostotsky (2002). "Arabidopsis thaliana exosome subunit AtRrp4p is a hydrolytic 3'-->5' exonuclease containing S1 and KH RNA-binding domains." Nucleic Acids Res **30**(3): 695-700.

Chen, C. Y., R. Gherzi, S. E. Ong, E. L. Chan, R. Raijmakers, G. J. Pruijn, G. Stoecklin, C. Moroni, M. Mann and M. Karin (2001). "AU binding proteins recruit the exosome to degrade ARE-containing mRNAs." Cell **107**(4): 451-64.

Collaborative Computational Project, N. (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr. **50**: 760-763.

Cramer, P., D. A. Bushnell and R. D. Kornberg (2001). "Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution." Science **292**(5523): 1863-76.

Crofton, S. and P. P. Dennis (1983). "Cloning and orientation of the gene encoding polynucleotide phosphorylase in Escherichia coli." J Bacteriol **154**(1): 58-64.

de la Cruz, J., D. Kressler, D. Tollervey and P. Linder (1998). "Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in Saccharomyces cerevisiae." Embo J **17**(4): 1128-40.

Decker, C. J. (1998). "The exosome: a versatile RNA processing machine." Curr Biol **8**(7): R238-40.

Deutscher, M. P., G. T. Marshall and H. Cudny (1988). "RNase PH: an Escherichia coli phosphate-dependent nuclease distinct from polynucleotide phosphorylase." Proc Natl Acad Sci U S A **85**(13): 4710-4.

Drenth, J. (1999). Principles of Protein X-Ray Crystallography. Heidelberg, Springer-Verlag.

Dreyfus, M. and P. Regnier (2002). "The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria." Cell **111**(5): 611-3.

Dziembowski, A., E. Lorentzen, E. Conti and B. Seraphin (2007). "A single subunit, Dis3, is essentially responsible for yeast exosome core activity." Nat Struct Mol Biol **14**(1): 15-22.

Estevez, A. M., T. Kempf and C. Clayton (2001). "The exosome of Trypanosoma brucei." Embo J **20**(14): 3831-9.

Evguenieva-Hackenberg, E., P. Walter, E. Hochleitner, F. Lottspeich and G. Klug (2003). "An exosome-like complex in Sulfolobus solfataricus." EMBO Rep **4**(9): 889-93.

Farhoud, M. H., H. J. Wessels, P. J. Steenbakkers, S. Mattijssen, R. A. Wevers, B. G. van Engelen, M. S. Jetten, J. A. Smeitink, L. P. van den Heuvel and J. T. Keltjens (2005). "Protein complexes in the archaeon methanothermobacter thermautotrophicus analyzed by blue native/SDS-PAGE and mass spectrometry." Mol Cell Proteomics.

Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch (2003). "ExPASy: The proteomics server for in-depth protein knowledge and analysis." Nucleic Acids Res **31**(13): 3784-3788.

Grunberg-Manago, M., P. J. Oritz and S. Ochoa (1955). "Enzymatic synthesis of nucleic acidlike polynucleotides." Science **122**(3176): 907-10.

Haag, R. and R. A. Lewis (1994). "The partial purification and characterization of purine nucleoside phosphorylase from mammalian mitochondria." Mol Cell Biochem **135**(2): 129-36.

Haile, S., A. M. Estevez and C. Clayton (2003). "A role for the exosome in the in vivo degradation of unstable mRNAs." Rna **9**(12): 1491-501.

Hanahan, D. (1983). "Studies on transformation of Escherichia coli with plasmids." J Mol Biol **166**(4): 557-580.

Haracska, L., R. E. Johnson, L. Prakash and S. Prakash (2005). "Trf4 and Trf5 proteins of Saccharomyces cerevisiae exhibit poly(A) RNA polymerase activity but no DNA polymerase activity." Mol Cell Biol **25**(22): 10183-9.

Harlow, L. S., A. Kadziola, K. F. Jensen and S. Larsen (2004). "Crystal structure of the phosphorolytic exoribonuclease RNase PH from Bacillus subtilis and implications for its quaternary structure and tRNA binding." Protein Sci **13**(3): 668-77.

Hilleren, P., T. McCarthy, M. Rosbash, R. Parker and T. H. Jensen (2001). "Quality control of mRNA 3'-end processing is linked to the nuclear exosome." Nature **413**(6855): 538-42.

Ho, S. N., H. D. Hunt, R. M. Horton, J. K. Pullen and L. R. Pease (1989). "Site-directed mutagenesis by overlap extension using the polymerase chain reaction." Gene **77**(1): 51-59.

Hoeppner, S., S. Baumli and P. Cramer (2005). "Structure of the mediator subunit cyclin C and its implications for CDK8 function." J Mol Biol **350**(5): 833-42.

Holm, L. and C. Sander (1995). "Dali: a network tool for protein structure comparison." Trends Biochem Sci **20**(11): 478-80.

Houalla, R., F. Devaux, A. Fatica, J. Kufel, D. Barrass, C. Torchet and D. Tollervey (2006). "Microarray detection of novel nuclear RNA substrates for the exosome." Yeast **23**(6): 439-54.

Houseley, J., J. LaCava and D. Tollervey (2006). "RNA-quality control by the exosome." Nat Rev Mol Cell Biol **7**(7): 529-39.

Inners, L. D. and G. Felsenfeld (1970). "Conformation of polyribouridylic acid in solution." J Mol Biol **50**(2): 373-89.

Jarrige, A., D. Brechemier-Baey, N. Mathy, O. Duche and C. Portier (2002). "Mutational analysis of polynucleotide phosphorylase from Escherichia coli." J Mol Biol **321**(3): 397-409.

Kabsch, W. (1993). "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants." J Appl Cryst **26**(6): 795-800.

Kantardjieff, K. A. and B. Rupp (2003). "Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals." Protein Sci **12**(9): 1865-1871.

Koonin, E. V., Y. I. Wolf and L. Aravind (2001). "Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach." Genome Res **11**(2): 240-52.

LaCava, J., J. Houseley, C. Saveanu, E. Petfalski, E. Thompson, A. Jacquier and D. Tollervey (2005). "RNA degradation by the exosome is promoted by a nuclear polyadenylation complex." Cell **121**(5): 713-24.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(5259): 680-5.

Laskowski, R. A., M. W. MacArthur, D. S. Moss and J. M. Thornton (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." J Appl Cryst **26**(2): 283-291.

Lima, C. D., M. G. Klein and W. A. Hendrickson (1997). "Structure-based analysis of catalysis and substrate definition in the HIT protein family." Science **278**(5336): 286-90.

Liu, Q., J. C. Greimann and C. D. Lima (2006). "Reconstitution, activities, and structure of the eukaryotic RNA exosome." Cell **127**(6): 1223-37.

Lorentzen, E. and E. Conti (2005). "Structural basis of 3' end RNA recognition and exoribonucleolytic cleavage by an exosome RNase PH core." Mol Cell **20**(3): 473-81.

Lorentzen, E., A. Dziembowski, D. Lindner, B. Seraphin and E. Conti (2007). "RNA channelling by the archaeal exosome." EMBO Rep.

Lorentzen, E., P. Walter, S. Fribourg, E. Evguenieva-Hackenberg, G. Klug and E. Conti (2005). "The archaeal exosome core is a hexameric ring structure with three catalytic subunits." Nat Struct Mol Biol.

Massa, W. (2002). Kristallstrukturbestimmung. Stuttgart, B.G. Teubner.

Matthews, B. W. (1968). "Solvent content of protein crystals." J Mol Biol **33**(2): 491-497.

McPherson, A. (2001). Crystallization of Biological Macromolecules, Cold Spring Harbor Laboratory Press.

Mitchell, P., E. Petfalski, R. Houalla, A. Podtelejnikov, M. Mann and D. Tollervey (2003). "Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs." Mol Cell Biol **23**(19): 6982-92.

Mitchell, P., E. Petfalski, A. Shevchenko, M. Mann and D. Tollervey (1997). "The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases." Cell **91**(4): 457-66.

Mitchell, P. and D. Tollervey (2000). "Musing on the structural organization of the exosome complex." Nat Struct Biol **7**(10): 843-6.

Mitchell, P. and D. Tollervey (2001). "mRNA turnover." Curr Opin Cell Biol **13**(3): 320-5.

Mitchell, P. and D. Tollervey (2003). "An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'-->5' degradation." Mol Cell **11**(5): 1405-13.

Mohanty, B. K. and S. R. Kushner (2000). "Polynucleotide phosphorylase, RNase II and RNase E play different roles in the in vivo modulation of polyadenylation in Escherichia coli." Mol Microbiol **36**(4): 982-94.

Muhlrad, D., C. J. Decker and R. Parker (1994). "Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript." Genes Dev **8**(7): 855-66.

Mukherjee, D., M. Gao, J. P. O'Connor, R. Raijmakers, G. Pruijn, C. S. Lutz and J. Wilusz (2002). "The mammalian exosome mediates the efficient

degradation of mRNAs that contain AU-rich elements." Embo J **21**(1-2): 165-74.

Naryshkina, T., A. Bruning, O. Gadal and K. Severinov (2003). "Role of second-largest RNA polymerase I subunit Zn-binding domain in enzyme assembly." Eukaryot Cell **2**(5): 1046-52.

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

Orban, T. I. and E. Izaurralde (2005). "Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome." Rna **11**(4): 459-69.

Portnoy, V., E. Evguenieva-Hackenberg, F. Klein, P. Walter, E. Lorentzen, G. Klug and G. Schuster (2005). "RNA polyadenylation in Archaea: not observed in Haloferax while the exosome polynucleotidylates RNA in Sulfolobus." EMBO Rep **6**(12): 1188-93.

Portnoy, V. and G. Schuster (2006). "RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R." Nucleic Acids Res **34**(20): 5923-31.

Raijmakers, R., G. Schilders and G. J. Pruijn (2004). "The exosome, a molecular machine for controlled RNA degradation in both nucleus and cytoplasm." Eur J Cell Biol **83**(5): 175-83.

Ramos, C. R., C. L. Oliveira, I. L. Torriani and C. C. Oliveira (2006). "The Pyrococcus exosome complex: structural and functional characterization." J Biol Chem **281**(10): 6751-9.

Reichlin, M., P. J. Maddison, I. Targoff, T. Bunch, F. Arnett, G. Sharp, E. Treadwell and E. M. Tan (1984). "Antibodies to a nuclear/nucleolar antigen in patients with polymyositis overlap syndromes." J Clin Immunol **4**(1): 40-4.

Reimer, G., U. Scheer, J. M. Peters and E. M. Tan (1986). "Immunolocalization and partial characterization of a nucleolar autoantigen (PM-Scl) associated with polymyositis/scleroderma overlap syndromes." J Immunol **137**(12): 3802-8.

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

Schneider, T. R. and G. M. Sheldrick (2002). "Substructure solution with SHELXD." Acta Crystallogr D Biol Crystallogr **58**(Pt 10 Pt 2): 1772-9.

Sheth, U. and R. Parker (2003). "Decapping and decay of messenger RNA occur in cytoplasmic processing bodies." Science **300**(5620): 805-8.

Shuman, S. and C. D. Lima (2004). "The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases." Curr Opin Struct Biol **14**(6): 757-64.

Spickler, C. and G. A. Mackie (2000). "Action of RNase II and polynucleotide phosphorylase against RNAs containing stem-loops of defined structure." J Bacteriol **182**(9): 2422-7.

Stickney, L. M., J. S. Hankins, X. Miao and G. A. Mackie (2005). "Function of the conserved S1 and KH domains in polynucleotide phosphorylase." J Bacteriol **187**(21): 7214-21.

Suzuki, N., E. Noguchi, N. Nakashima, M. Oki, T. Ohba, A. Tartakoff, M. Ohishi and T. Nishimoto (2001). "The Saccharomyces cerevisiae small GTPase, Gsp1p/Ran, is involved in 3' processing of 7S-to-5.8S rRNA and in degradation of the excised 5'-A0 fragment of 35S pre-rRNA, both of which are carried out by the exosome." Genetics **158**(2): 613-25.

Symmons, M. F., G. H. Jones and B. F. Luisi (2000). "A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation." Structure Fold Des **8**(11): 1215-26.

Symmons, M. F., M. G. Williams, B. F. Luisi, G. H. Jones and A. J. Carpousis (2002). "Running rings around RNA: a superfamily of phosphate-dependent RNases." Trends Biochem Sci **27**(1): 11-8.

Takahashi, S., Y. Araki, T. Sakuno and T. Katada (2003). "Interaction between Ski7p and Upf1p is required for nonsense-mediated 3'-to-5' mRNA decay in yeast." Embo J **22**(15): 3951-9.

Targoff, I. N. and M. Reichlin (1985). "Nucleolar localization of the PM-Scl antigen." Arthritis Rheum **28**(2): 226-30.

Terwilliger, T. C. (2002). "Automated structure solution, density modification and model building." Acta Crystallogr. **58**(Pt 11): 1937-40.

Torchet, C., C. Bousquet-Antonelli, L. Milligan, E. Thompson, J. Kufel and D. Tollervey (2002). "Processing of 3'-extended read-through transcripts by the exosome can generate functional mRNAs." Mol Cell **9**(6): 1285-96.

Treadwell, E. L., M. A. Alspaugh, J. F. Wolfe and G. C. Sharp (1984). "Clinical relevance of PM-1 antibody and physiochemical characterization of PM-1 antigen." J Rheumatol **11**(5): 658-62.

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

van Hoof, A., P. Lennertz and R. Parker (2000). "Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs." Mol Cell Biol **20**(2): 441-52.

van Hoof, A. and R. Parker (1999). "The exosome: a proteasome for RNA?" Cell **99**(4): 347-50.

van Hoof, A., R. R. Staples, R. E. Baker and R. Parker (2000). "Function of the ski4p (Csl4p) and Ski7p proteins in 3'-to-5' degradation of mRNA." Mol Cell Biol **20**(21): 8230-43.

Vanacova, S., J. Wolf, G. Martin, D. Blank, S. Dettwiler, A. Friedlein, H. Langen, G. Keith and W. Keller (2005). "A new yeast poly(A) polymerase complex involved in RNA quality control." PLoS Biol **3**(6): e189.

Vasiljeva, L. and S. Buratowski (2006). "Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts." Mol Cell **21**(2): 239-48.

Vasudevan, S. and S. W. Peltz (2003). "Nuclear mRNA surveillance." Curr Opin Cell Biol **15**(3): 332-7.

Voges, D., P. Zwickl and W. Baumeister (1999). "The 26S proteasome: a molecular machine designed for controlled proteolysis." Annu Rev Biochem **68**: 1015-68.

Wang, H., C. H. Takemoto, K. Murayama, H. Sakai, A. Tatsuguchi, T. Terada, M. Shirouzu, S. Kuramitsu and S. Yokoyama (2004). "Crystal structure of ribosomal protein L27 from Thermus thermophilus HB8." Protein Sci **13**(10): 2806-10.

Wang, Z. and M. Kiledjian (2001). "Functional link between the mammalian exosome and mRNA decapping." Cell **107**(6): 751-62.

Wolfe, J. F., E. Adelstein and G. C. Sharp (1977). "Antinuclear antibody with distinct specificity for polymyositis." J Clin Invest **59**(1): 176-8.

Wyers, F., M. Rougemaille, G. Badis, J. C. Rousselle, M. E. Dufour, J. Boulay, B. Regnault, F. Devaux, A. Namane, B. Seraphin, D. Libri and A. Jacquier

(2005). "Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase." <u>Cell</u> **121**(5): 725-37.

Yehudai-Resheff, S., V. Portnoy, S. Yogev, N. Adir and G. Schuster (2003). "Domain analysis of the chloroplast polynucleotide phosphorylase reveals discrete functions in RNA degradation, polyadenylation, and sequence homology with exosome proteins." <u>Plant Cell</u> **15**(9): 2003-19.

Zhou, Z. and M. P. Deutscher (1997). "An essential function for the phosphate-dependent exoribonucleases RNase PH and polynucleotide phosphorylase." <u>J Bacteriol</u> **179**(13): 4391-5.

PART II

Structural Insights into DNA Duplex Separation by the Archaeal Superfamily 2 Helicase Hel308

Table of contents

1 Introduction

1.1 Helicases – Classification and nucleic acid unwinding mechanisms

1.1.1 Classification of helicases into Superfamilies

Helicases are ubiquitous enzymes which use the free energy of adenosine triphosphate (ATP) hydrolysis to catalyze the thermodynamically unfavourable separation of nucleic acid duplexes (DNA:DNA, DNA:RNA, RNA:RNA) into the corresponding single strands (Bjornson *et al.,* 1996; Hall and Matson 1999). Since the discovery of the first helicase more than 30 years ago (Abdel-Monem and Hoffmann-Berling 1976), dozens of these essential proteins have been identified and characterized in all kingdoms of life. DNA-acting helicases are found to play essential roles in nearly all DNA metabolizing processes, including DNA replication, recombination, repair and transcription. RNA helicases are found to have essential functions in translation and RNA processing/degradation pathways (Staley and Guthrie 1999; Venema and Tollervey 1999; Linder and Stutz 2001; Tuteja and Tuteja 2004) as well as in the mammalian innate immune response (Meylan *et al.,* 2006).

Helicases are classified according to amino acid sequence homology, oligomeric state and unwinding polarity (3´$\rightarrow$ 5´ or 5´$\rightarrow$ 3´), and are grouped into five Superfamilies (SF1-SF5) (Gorbalenya *et al.,* 1989; Singleton and Wigley 2002) (Table 1). SF1 and SF2 are the two largest groups. They are closely related and typically characterized by a set of conserved sequence motifs called Q, I, Ia, and II-VI (Figure 1). Regarding differences in length, spacing between motifs and the presence of unique motifs, SF1 and SF2 helicases are divided into different subfamilies, like DEAD- and DEXH-box helicases (Gorbalenya *et al.,* 1989; Linder and Daugeron 2000; Caruthers and McKay 2002).

Table 1: Characteristics of the five Helicase Superfamilies

Superfamily	Conserved Motifs	Examples	Source	Oligomeric state	Unwinding polarity
SF1	I, Ia, Ib, II, III, IV, V, VI	PcrA, Rep, UvrD	all kingdoms	monomer, dimer	$3´ \rightarrow 5´, 5´ \rightarrow 3´$
SF2	Q, I, Ia, Ib, II, III, IV, V, VI	HCV NS3, RecG, RecQ	all kingdoms	monomer, oligomer	$3´ \rightarrow 5´, 5´ \rightarrow 3´$
SF3	A, B, C	SV40-LTag	RNA and DNA viruses	hexamer	$5´ \rightarrow 3´$
SF4	1, 1a, 2, 3, 4	DnaB	eubacteria, bacteriophages	hexamer	$5´ \rightarrow 3´$
SF5	n.d.	Rho	eubacteria	hexamer	$5´ \rightarrow 3´$

Members of SF3-SF5 also share conserved motifs (Table 1), but are structurally and functionally less characterized (Gorbalenya *et al.,* 1989; Hall and Matson 1999). In contrast to members of SF1 and SF2, which function as monomers (e.g. PcrA, HCV NS3, RecG), dimers (e.g. Rep), and/or higher oligomers (e.g. HCV NS3) (Cheng *et al.,* 2001; Dillingham *et al.,* 2003; Maluf *et al.,* 2003; Levin *et al.,* 2004; Tackett *et al.,* 2005), members of SF3-SF5 are usually hexamers (Picha *et al.,* 2000). However, the three-dimensional fold of the ATP binding domains (RecA fold) is similar in helicases of different families, whereas additional amino (N)- and/or carboxy (C)-terminal domains function in substrate recognition and specificity, or interaction with other proteins (Subramanya *et al.,* 1996; Singleton *et al.,* 2000).

1.1.2 Role of the seven helicase motifs in SF1 and SF2 helicases

High resolution X-ray crystallographic studies have revealed that the conserved helicase signature motifs of SF1 and SF2 helicases (Q, I, Ia, II-VI) form a large functional domain whereby they act together in ATP- and/or nucleic acid binding (Hall and Matson 1999) (Figure 1). Extensive structure-function analyses of the motfs have been used to discover their roles in helicase mechanism.
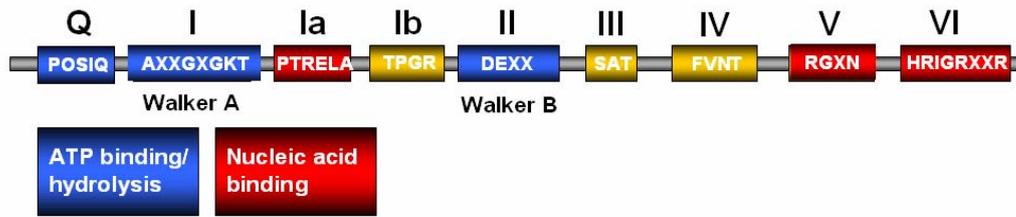
Figure 1: Diagram of the seven classical conserved helicase motifs defined by Gorbalenya and Koonin (annotated I to VI). The consensus sequence is adapted to DEXX box related enzymes.

Motif I (Walker A motif) and motif II (Walker B motif) are directly involved in the binding of the nucleotide, and are generally found in nucleotide trisphosphate hydrolysing enzymes (Gorbalenya *et al.,* 1989). While the invariant lysine (G**K**T) of motif I has been shown to bind to the β- and γ-phosphate of the ATP molecule and to stabilize the transition state during catalysis, the aspartate of the DE-motif has been shown to interact with the ATP-associated Mg^{2+}-ion and to activate the attacking water molecule required for ATP hydrolysis (Story and Steitz 1992). Mutation of either lysine or aspartate to uncharged amino acids abolishes ATP binding and ATP dependent nucleic acid unwinding (Pause and Sonenberg 1992; Hall and Matson 1999). The Q motif as well as motif IV have also an implicated function in ATP recognition/binding and hydrolysis (Tanner 2003; Tanner *et al.,* 2003)(Korolev *et al.,* 1997; Velankar *et al.,* 1999).

Motifs Ia and V have been shown to interact with the nucleic acid substrate as mutations in these motifs diminish ssDNA-stimulated ATPase activity (Marintcheva and Weller 2003). Motif VI is suggested to be involved in the binding of the nucleic acid. By binding of the invariant arginies to the β- and γ-phosphates of ATP after a nucleotide binding-induced conformational closure of domains 1 and 2, the enzyme additionally may mediate conformational changes and couple ATP hydrolysis to DNA unwinding (Pause and Sonenberg 1992; Kim *et al.,* 1998).

Motif III and Ib as well as motif IV have also been shown to play essential roles in coupling ATP hydrolysis to DNA unwinding activity as they are implicated in binding to the nucleotide and to the nucleic acid substrate (Pause and Sonenberg 1992; Graves-Woodward *et al.,* 1997).

However, not every protein which possesses the mentioned conserved helicase motifs exhibits helicase acitivty. The seven motifs are also present in *E. coli* MfD protein (Selby and Sancar 1993; Assenmacher *et al.,* 2006; Deaconescu *et al.,*

2006) and the chromatin remodelling factor SWI2/SNF (Richmond and Peterson 1996; Swagemakers *et al.,* 1998; Durr *et al.,* 2005). Both do not contain helicase activity, but displace stalled RNA Polymerase II from a site of DNA damage and dissociate/replace nucleosomes during chromatin remodelling events, respectively. Many other helicases have been found to additionally function as molecular motors that displace proteins from the nucleic acid substrate instead of unwinding duplex regions (Jankowsky *et al.,* 2001; Flores *et al.,* 2005; Macris and Sung 2005).

1.1.3 Helicase unwinding mechanism

The basic activity of helicases is to couple ATP binding and hydrolysis to conformational changes that result in base pair separation and/or translocation along a nucleic acid substrate. By this, each helicase goes through defined ATP ligation states (empty, ATP, ADP*P_i and ADP). One or more ATP ligation states can cause a modulation of the binding affinity to the nucleic acid substrate and a conformational change. Changes in the ATPase cycle that are associated with an observable energy change are likely to force a so called power stroke that leads to processive or non-processive translocation and/or duplex separation.

The detailed structural and molecular mechanism of nucleic acid duplex separation by helicases still remains unclear. However, there are some features that seem to be common to all helicases studied to date. Resulting from these similarities, three models became popular.

1.1.3.1 Inchworm and modified inchworm mechanism

The inchworm mechanism is proposed for monomeric helicases, which possess two nucleic acid binding sites located on the same polypeptide. Both binding sites function independently in binding and release of the nucleic acid substrate in response to the signals received from a single ATPase site (Yarranton *et al.,* 1979; Wong and Lohman 1992; Velankar *et al.,* 1999). A cycle of nucleic acid binding, release and translocation starts with one binding site bound tightly (tight site), and the second binding site bound weakly (weak site) to the nucleic acid substrate (Figure 2). Changes in the ATP ligation state causes the weak site to dissociate from the nucleic acid. An ensuing power stroke leads to a movement away from the tight site and to rebinding of the weak site at a position ahead. The tight site is

then proposed to become weak and to dissociate from the nucleic acid. A second power stroke leads to a directed movement of the original tight site to the site ahead.
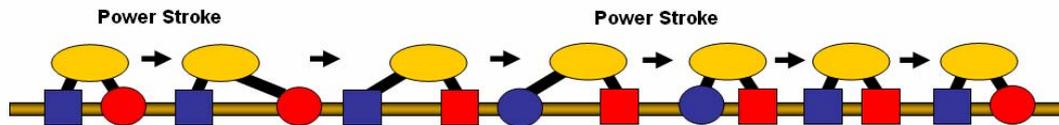
Figure 2: Schematic drawing of the stepping inchworm mechanism. A helicase monomer with a tight (quader) and a weak (circle) nucleic acid binding site undergoes translocation (power stroke) and nucleic acid affinity changes (tight to weak binding) (adapted from (Patel and Donmez 2006)). The red module can be seen as "leading" site, the blue module as "trailing" site. Changes in the NTP ligation state and the resulting modulations in the substrate affinity causes the helicase to undergo up to six conformational changes in one stepping cycle.

This typical inchworm model is proposed for translocation of PcrA helicase along ssDNA, whereas domains 1A and 2A of PcrA alternate in weak and tight binding of the single-stranded substrate (Velankar *et al.,* 1999) (Figure 3A).

A modified inchworm model has been proposed for a monomeric UvrD mutant (UvrDd40C), which failed to dimerize but showed unaltered ATPase and DNA unwinding acitivty (Mechanic *et al.,* 1999). Here, the leading nucleic acid binding site exhibits double-stranded and single-stranded DNA binding ability, whereas the trailing site shows only single strand binding affinity (Figure 3B). The modified inchworm mechanism is also applicable to hetero- and homodimeric helicases. Here, each subunit would represent an individual nucleic acid binding site that is controlled by the NTP binding site of the respective molecule. This type of model is proposed for heterodimeric RecBC DNA helicase (Bianco and Kowalczykowski 2000; Bianco *et al.,* 2001). RecB, the leading domain, anchors the enzyme to one strand of the duplex DNA and translocates along it. RecC, the trailing domain, is responsible for ATP hydrolysis dependent unwinding. The leading domain can bind up to 23 nt ahead of the trailing domain and thereby translocate over DNA gaps.

1.1.3.2 Active rolling mechanism

The active rolling model has been proposed for dimeric helicases like the Rep protein (Wong and Lohman 1992; Delagoutte and von Hippel 2003). In this model,

the NTP ligation state of each subunit influences and alternate the binding to single-stranded and duplex DNA. Thereby both subunits leave their relative postion on the nucleic acid substrate depending on the NTP ligation state, resulting in being the trailing or the leading subunit (Figure 3C).
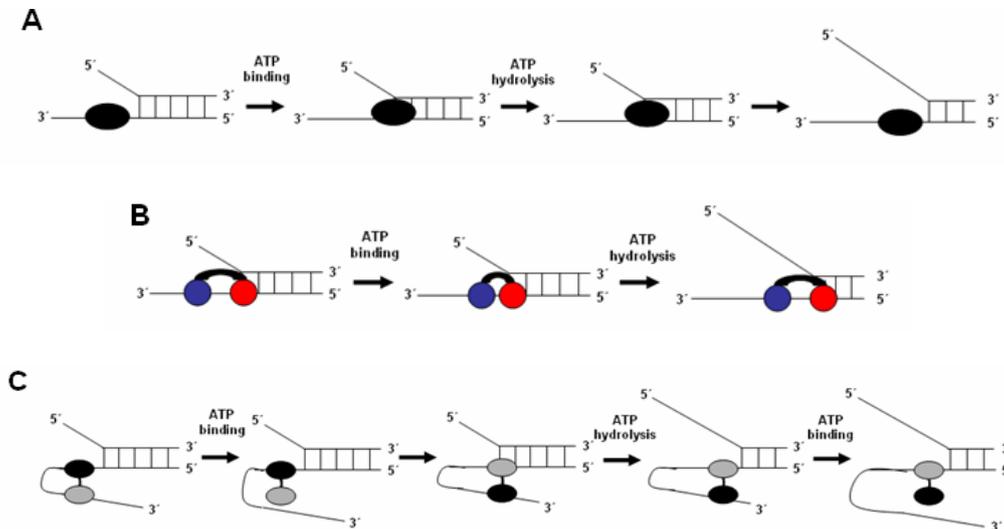
Figure 3: A) In the Inchworm model for monomeric helicases the enzyme is bound to single-stranded nucleic acid and translocates along the single-strand to the fork region, possibly upon binding of ATP. Destabilization of the duplex region takes place upon ATP hydrolysis (Yarranton *et al.,* 1979). B) In the modified inchworm model for monomeric helicases the enzyme contains two individual nucleic acid binding sites. The leading site (red) exhibits affinity for both double-stranded and single-stranded nucleic acids, whereas the trailing site (blue) shows affinity for sinlge-stranded nucleic acid. Starting from an extended conformation, the enzyme undergoes a conformational change upon ATP binding. This results in movement of the trailing site towards the leading site. On ATP hydrolysis, the leading site moves forward in the duplex region and unwinds the double-strand (Mechanic *et al.,* 1999). To apply this model to dimeric helicases, one of the molecules would represent the leading subunit, while the second molecule would represent the trailing subunit. C) In the active rolling model both subunits of the dimer are initially bound to single-stranded nucleic acid. On binding of ATP, the respective subunit releases the nucleic acid and binds to the duplex region at the fork. This results in helix destabilization, accompanied by ATP hydrolysis (Bjornson *et al.,* 1996).

1.1.3.3 Brownian Motor mechanism

The Brownian motor model was recently proposed for HCV NS3 helicase and T7 g4 helicase and includes Brownian motion and a power stroke (Betterton and Julicher 2005; Levin *et al.,* 2005; Stano *et al.,* 2005). In contrast to the inchworm

mechanism, it does not require two binding sites for nucleic acid within the active site of the enzyme. Here, the different ATP ligation states result in two distinct conformational states and nucleic acid binding modes (weak (+ATP) and tight (-ATP)). Changes are associated with alternations in the energy profile, which allow or disallow movement of the helicase on the nucleic acid substrate. Whereas tight binding traps the enzyme on the nucleic acid, the weak state is associated with thermal fluctuations, which allow the helicase to dissociate from the substrate and to translocate in either direction (Brownian motion). The transition from the weak to the tight state (probably upon ATP hydrolysis) results in a power stroke and translocation of the molecule. The Brownian motion model lets the helicase translocate in either direction. Those molecules that have fluctuated in the forward direction move ahead. Those that have fluctuated in the opposite direction return to the original position. Only the repetition of these steps leads to net forward movement along the nucleic acid substrate.

1.2 Helicases and the thread of stalled replication forks

Helicases have crucial roles in nearly all nucleic acid metabolic processes. One of the most important processes in all living cells is the accurate and processive DNA replication. During normal cell growth, exogenous (chemicals, radiation) and endogenous (oxygen radicals, replication errors) sources of damage can inactivate a large portion of replication forks (Heller and Marians 2005; Patel and Donmez 2006). Apparently, the interplay between DNA replication and repair of stalled replication forks is a basic phenomenon in all living cells, and is associated with the conversion of DNA strands into a variety of different structures (McGlynn and Lloyd 2002; Kreuzer 2005). A multitude of DNA helicases has been found to be involved in the processing and breakdown of these structures in both prokaryotes (UvrD, Rep, RecQ) and eukaryotes (BLM, WRN, Mus308, PolΘ), allowing restart of replication and DNA repair in an origin-independent manner (Michel *et al.,* 2004).

1.2.1 RecQ and its role in the repair of stalled replication forks in prokaryotes

In prokaryotes, stalled or collapsed replication forks are primary repaired via recombinational processes (Cox 2002). The repair pathways are varied and

redundant, reflecting the multitude of DNA structures that might be found at a stalled replication fork (Michel *et al.,* 2004).

Whereas collapsed forks at a site of a DNA strand break are mainly repaired via the RecBCD pathway, a stalled replication fork caused by a DNA lesion is repaired via the RecF pathway (Michel *et al.,* 2004). The RecF pathway includes proteins like RecA, RecF, RecG, RecJ, RecO, RecR and RuvABC as well as the DNA helicase RecQ (Chow and Courcelle 2004; Michel *et al.,* 2004).

The prototype of the RecQ DNA helicase family is the *Escherichia coli* RecQ protein (Figure 4) (Nakayama *et al.,* 1985), which is a member of the DEAH subgroup of 3´$\rightarrow$ 5´ SF2 helicases (Gorbalenya *et al.,* 1989; Harmon and Kowalczykowski 2001). Homologs of *E. coli* RecQ are found in yeast (Rqh1, Sgs1) and metazoans (BLM, WRN, RecQ4, RecQ5, RecQ1), and have implicated roles in the maintenance of genome stability (Ellis 1997). Archaea typically lack orthologues of RecQ.
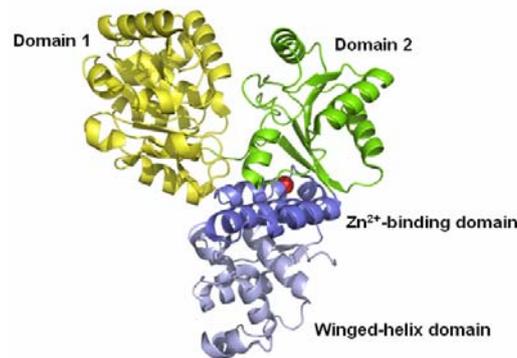
Figure 4: Crystal structure of the catalytic core of *E. coli* RecQ (PDB code 1OYW) (Bernstein *et al.,* 2003). The ribbon model shows two RecA-like domains 1 (yellow) and 2 (green), and the C-terminal Zn^{2+}-binding (slate) and winged-helix domains (lightblue). A bound Zn^{2+} is shown as a red sphere. The C-terminal HRDC domain is not present in the crystal structure. The C-terminal domains of RecQ have implicated functions in substrate specificity, substrate binding and helicase activity (Liu *et al.,* 1999; Gajiwala and Burley 2000; Killoran and Keck 2006; Killoran and Keck 2006).

RecQ proteins are suggested to play different roles in maintaining genome stability. As shown in Figure 5, prokaryotic RecQ catalyzes the unwinding of DNA at gaps or double strand breaks, consistent with its *in vitro* ability to unwind double-stranded blunt-ended DNA substrates as well as duplexes with a 3´-tail overhang (Umezu *et al.,* 1990; Bennett *et al.,* 1999; Harmon and Kowalczykowski 2001; Mohaghegh *et al.,* 2001; Cui *et al.,* 2003; Sharma *et al.,* 2005). The interplay

with the 5´-exonuclease RecJ allows generation of 3´ ssDNA. The assembly of RecA filaments on the ssDNA then initiates strand exchange and D-Loop formation (Harmon and Kowalczykowski 1998).
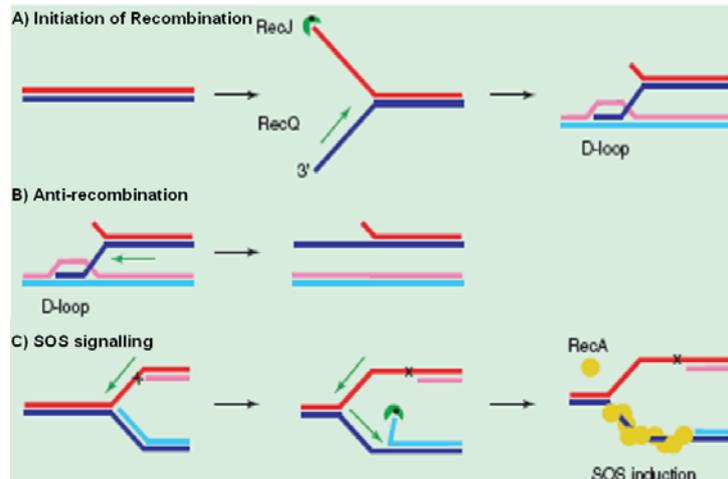
Figure 5: Roles of RecQ DNA helicase in genomic maintenance. A) RecQ was found to initiate recombination by unwinding DNA at gaps or double strand breaks (only in recBC-cells). B) RecQ has been shown to disrupt D-Loop structures, resulting in an anti-recombinational function. C) RecQ SOS signaling model (Hishida *et al.,* 2004). DNA strands: red: leading strand, blue: lagging strand, green arrow: 3´→ 5´unwinding polarity of RecQ. (adapted from (Heyer, 2004))

RecQ was also found to disrupt synthetic D-Loop structures and Holliday junction intermediates (Hanada *et al.,* 1997; Harmon and Kowalczykowski 1998; Bennett *et al.,* 1999; Wu and Hickson 2003). This anti-recombination function may be relevant for the suppression of illegitimate recombination, which may result in DNA rearrangements such as deletions, translocations, duplications or inversions *in vivo* (Gangloff *et al.,* 1994). In the past years, an additional role of RecQ in the induction of the prokaryotic SOS response has been revealed. In the SOS-model proposed by Hishida *et al.*, RecQ binds a gap on the leading strand of a stalled replication fork. After unwinding the duplex template ahead of the fork in 3´→ 5´ direction, RecQ switches to the lagging strand template. Displacement of the nascent lagging strand by RecQ and the concerted action of the exonuclease RecJ generates a ssDNA gap on the lagging strand on which RecA can be assembled. RecA filament formation subsequently leads to the induction of the SOS response and repair of the DNA lesion, trans-lesion synthesis and/or recombinational restart (Hishida *et al.,* 2004).

The molecular mechanism of RecQ in DNA repair is not fully clear and seems to depend mainly on the regulated interaction between RecQ helicases and other replication/repair proteins in both prokaryotes and eukaryotes (Hickson 2003). However, the capability of prokaryotic and eukaryotic RecQ helicases to resolve such a variety of DNA structures is consistent with a role in suppressing illegitimate recombination by disrupting aberrant intermediates at sites of stalled replication forks and during homologous recombination.

1.2.2 The Mus308 family of DNA helicases and their role in sensitivity to DNA cross-linking agents in higher eukaryotes

A special DNA damage, which can cause the stall of a replication fork is a DNA interstrand cross-link. DNA interstrand cross-links (ICLs) are highly cytotoxic to dividing cells as they covalently connect the two complementary strands of the DNA double helix (Magana-Schwencke *et al.,* 1982; Lawley and Phillips 1996). DNA cross-linking agents like cisplatin and mytocmycin C are commonly used in cancer therapy as chemotherapeutics in order to inhibit DNA replication and to stop cell division (Jamieson *et al.,* 1999; Kow *et al.,* 2005).

ICLs cause damages to both DNA strands at the same or very close nucleotide position, and repair of an ICL itself and the repair of ICL induced stalled replication forks is a combination of nucleotide excision repair (NER), homologous recombination and/or translesion DNA synthesis, as studied in *E. coli* and *S. cerevisiae* (Yoakum and Cole 1978; Sladek *et al.,* 1989; Dronkert and Kanaar 2001; McHugh and Sarkar 2006).

Whereas the repair of ICL in bacteria and yeast is well studied, the removal of ICLs and the restart of ICL-induced stalled replication forks in mammalian cells is not well understood, and several mechanisms for repair of or tolerance to DNA cross-links have been proposed (Dronkert and Kanaar, 2001; Grossmann et al., 2001; De Silva et al., 2000). However, genetic data and recent studies with mutant cell lines sensitive to DNA cross-linking agents suggest that proteins of both excision repair and homologous recombination pathways are involved in the repair of ICLs (Hoy *et al.,* 1985; Brookman *et al.,* 1996; Busch *et al.,* 1997; Dronkert and Kanaar 2001; Medhurst *et al.,* 2006).

One of the identified proteins, whose *in vivo* role could be directly linked to the repair of ICLs and/or the repair of ICL-induced stalled replication forks is the

Drosophila menlanogaster Mus308 and its orthologs from *Homo sapiens*. The Mus308 family of helicases and putative helicases is implicated in supporting genome stability in metazoans (Boyd *et al.,* 1990; Shima *et al.,* 2003; Laurencon *et al.,* 2004). In contrast to RecQ-like helicases, no sequence homologs of Mus308-like helicases are found in bacteria and yeast. Recently, the archaeal Mus308 homolog could be identified (Fujikane *et al.,* 2005; Guy and Bolt 2005).

1.2.2.1 *Drosophila melanogaster* Mus308

The prototype of the Mus308 family of helicases is the *Drosophila melanogaster* Mus308, which was identified more than 30 years ago and was found to be required for resistance to DNA cross-linking agents. A direct role of Mus308 in the repair of interstrand cross-links could be further verified in a variety of genetic studies (Boyd *et al.,* 1990).

Mus308 exhibits a unique domain architecture. It contains a C-terminal region, which shows homology to bacterial A-family DNA polymerases like *E. coli* DNA Polymerase I. In addition, the N-terminal region shows homology to SF2 helicases (Boyd *et al.,* 1990; Harris *et al.,* 1996). The detailed cellular function of Mus308 is not known, due to a lack of functional information about polymerase and helicase activity. According to the nucleotide excision repair (NER) model in prokaryotes, Mus308 may play a role analogous to bacterial Pol I in gap fill-in during recombinational repair of DNA cross-links in which the helicase domain could function similarly to UvrD (Harris *et al.,* 1996; Marini and Wood 2002).

1.2.2.2 Human PolΘ

The human Mus308 ortholog PolΘ comprises a similar overall architecture as *Drosophila* Mus308 (Seki *et al.,* 2004). The polymerase activity of purified PolΘ has been shown to be highly efficient at bypassing abasic sites in DNA templates, preferentially inserting adenine opposite an AP-site (Seki *et al.,* 2004). PolΘ additionally exhibits an unexpected tendency to misincorporate bases and to extend from mispaired termini, a feature that is untypical for an A-family polymerase (Seki *et al.,* 2003). It may have a function in maintaining normal genomic integrity by allowing bypass of lesions, when a DNA replication fork encounters an unrepaired lesion. Lack of PolΘ bypass activity therefore could cause replication fork collapse or breakage, leading to genomic instability.

As PolΘ is preferentially expressed in lymphoid tissues, an additional function during somatic hypermutation of antibody genes, by which AP sites are generated as a mutagenic intermediate, may be possible (Kawamura *et al.,* 2004). However, similar to *Drosophila* Mus308, nothing is known of its DNA helicase activity, although ATPase activity of PolΘ has been shown to be dependent on ssDNA (Seki *et al.,* 2003).

1.2.2.3 Human Hel308

The human Hel308 protein (hHel308) exhibits significant sequence homology to the N-terminal helicase domain of *Drosophila* Mus308 (Marini and Wood 2002). Although hHel308 is lacking the C-terminal polymerase domain it has been found that mutations in the *Drosophila* Hel308 ortholog cause increased sensitivity to nitrogen mustard and other DNA cross-linking agents (Laurencon *et al.,* 2004) as shown for *Drosophila* Mus308. hHel308 has been shown to exhibit single-stranded DNA-dependent ATPase and RPA (replication protein A) stimulated 3´$\rightarrow$ 5´ DNA unwinding acitivty (Marini and Wood 2002).

1.2.2.4 Human PolN (Polν)

Human polymerase ν (Polν) shows homology to the Mus308 polymerase domain (Marini *et al.,* 2003). Polν shows a low fidelity and incorporates T opposite template G with half the yield of incorporating the correct base and has a processivity similar to exonuclease-deficient Klenow fragment of *E. coli* Pol I (Takata *et al.,* 2006). Polν can also perform error-free translesion synthesis past thymine glycol (Tg), a common endogenous and radiation-induced product of reactive oxygen species DNA damage, and efficiently extends from the lesion. Error-free bypass results in low mutation frequency and the possibility to repair the Tg by base excision repair (BER). As hHel308 and Polν were shown to be expressed in the same tissues, a coupling between helicase and polymerase activity of these two enzymes could be possible (Marini *et al.,* 2003). Thus, different expression patterns of hHel308/Polν and PolΘ could be indicative for similar functions in different tissues (Marini *et al.,* 2003; Seki *et al.,* 2003).

1.3 Archaeal Hel308 belongs to Mus308 family of DNA helicases and exhibit RecQ like functions

1.3.1 Classification of archaeal Hel308 into the Mus308 helicase family

The presence of an archaeal Mus308 homolog was recently discovered by two independent research groups. Fujikane et al. discovered archaeal Hel308 (pfuHel308) in the hyperthermophilic archaeon *Pyrococcus furiosus* in an *in vitro* experiment to identify proteins responsible for Holliday junction (HJ) branch migration in archaea (Fujikane *et al.,* 2005). The gene, denoted *Hjm* (holliday junction migration), encodes the Hjm protein, which exhibits ATP-dependent unwinding activity of synthetic HJs.

At the same time, Guy and Bolt identified the homolog from *Methanothermobacter thermautotrophicus* (mthHel308) in an *in vivo* screen to investigate potential analogous activities to bacterial RecQ using putative archaeal helicases with unknown function (Guy and Bolt 2005). Thereby, they specifically looked for archaeal helicases which, similar to bacterial RecQ, act upon DNA structures formed at sites of stalled DNA replication using a standard heterologous genetic system (Wechsler and Gross 1971; Vandewiele *et al.,* 2002; Fujikane *et al.,* 2005; Guy and Bolt 2005).

Both groups found a significant conservation of the protein in both major branches of archaea, and additionally a significant sequence similarity to the N-terminal region of Mus308 helicase family members. No significant overall sequence homology to eukaryotic or bacterial RecQ helicases could be detected. Based on the amino acid sequence, both pfuHel308 and mthHel308 can be grouped into the helicase superfamily 2 of DNA and RNA helicases. In addition to the seven conserved helicase motifs they exhibit residues which are unique to the Mus308 family of helicases, and which are lacking or altered in RecQ-like proteins (Table 2). In the Mus308 helicase family, the threonine within motif V is usually a hydrophobic residue in other members of SF2, whereas the methionine in motif VI is usually a charged residue (Marini and Wood 2002; Guy and Bolt 2005).

Table 2: Sequence details in helicase motifs IVa, V and VI

	IVa	V	VI
afuHel308	GA**AFHHAGL**	ATP**T**L	**M**AGRAG
pfuHel308	GV**AFHHAGL**	ATP**T**L	**M**LGRAG
mthHel308	GI**AFHHAGL**	ATP**S**L	**M**SGRAG
hsHel308	GV**AYHHSGL**	CTS**T**L	**M**IGRAG
hsPolΘ	GV**AFHHAGL**	ATS**T**L	**M**VGRAG
hsBLM	AL**A**Y-**HAGL**	ATIAF	ESGRAG

Bold letters: invariant residues in Hel308/Mus308

afuHel308: *Archaeoglobus fulgidus;* pfuHel308: *Pyrococcus furiosus;* mthHel308: *Methanothermobacter thermautotrophicus;* hsHel308: *Homo sapiens* Hel308; hsPolΘ: *Homo sapiens* PolΘ, hsBLM: *Homo sapiens* BLM. (adapted from Guy and Bolt, 2005).

However, the Mus308 helicase familiy seems to combine motifs from various subfamilies. For example, motif II, the so called Walker B motif, is similar to the DEXH family of helicases (Figure 6). This family includes many "repair/recombination" helicases such as RecQ-like proteins as well as the chromatin remodelling factor Rad54, although no overall sequence conservation can be detected (Durr *et al.,* 2005; Guy and Bolt 2005). Based on recent analysis, the helicase domain of members of the Mus308 family is most closely related to the Ski2 family of RNA helicases as they all contain a conserved glutamine residue in the so called Q-Loop about 20 amino acids upstream of motif I (Figure 6) (Tanner 2003; Tanner *et al.,* 2003). Also motif IVa shows characteristics of the yeast RNA processing helicases Ski2p/Mtr4p (Marini and Wood 2002; Tanner *et al.,* 2003), and additional similarity to motif IVa of RecQ like helicases (Harris *et al.,* 1996; Bernstein *et al.,* 2003) (Table 2 and Figure 6).

AfHEL308
Pa0592
DmMus308
HsPolQ
HsHel308
yMtr4
HsMtr4
ySki2
HsSki2

Row 1 (α1–β3, residues 1–100):

```
                α1                    α2      ß1      α3        ß2      α4        ß3
              10        20        30        40        50        60        70        80        90       100
AfHEL308 : -----MKVEELAESISSYAVGILKEEGIEELFIPQAEAVEK-VFSGKNLILAMPTAIGKTLLAEMAIVREAIK-GGESLYVVPLRAIAGEKYESFKKWEKIGLRIGI : 100
Pa0592   : -----MKVGELN--VSEKIKEILRERGIEEELYIPQAEALTSGVLEGENLIVAIPTASGKTLIAEIAIANKLLEEGGIAVYIVPLKAIAEKFREPKDWERLGLKVAM : 100
DmMus308 : [206]TLRSISAWNLPMSIQAEYKKKGVVDMFDWQVECL[5]IPEHCNLVYSAPTSAGKTLVSEILLKTVLERGK-VLLILEPISVVREKMFYMQDLLTPAGYRVE : 310
HsPolQ   : [66]-DKLLLANWGIPKAVLEKYHSFGVKKMFEWQAECL[4]IVLEGKNLVYSAPTSAGKTLVAELLILKRVLEMRQIALEILEFVSVAKEKKYYLQSLFQEVGIKVD : 169
HsHel308 : [310]NDLGPFYSLPSKVRDLYAQFKGIEKLYEWQHTCL[4]IVQERKNLVLVAEILMLQELLCCRKDVLMILEPYVIVQEKISGLSSFGIELGFFVE : 413
yMtr4    : [98]-LTNSETLQVEQDGKVRLSHQVRHQVALIPNYDY[29]IDRGESLVVSAHTSAGKTVVAEYAIAQSLKN-KQRVITSLIKAISNQKYRELL-AEFGDVG--L : 221
HsMtr4   : [89]-LSLADLMPRIKVQSVETVEGCTHEVALIAEEDY[28]VDNNQSVLVSAHTSAGKTVCAEYAIAMAHRN-MTIIYTSLIKAISNQKYREMY-EEFQDVG--L : 211
ySki2    : [269]NADDAEIDEILPIGIDFGRTKPVSKSVIVKKEW[38]IEQGDSLVFVAAHTSAGKTVVAEYAIAMAHRN-MTIIYTSLIKAISNQKFRDFK-ETFDDVNIGL : 403
HsSki2   : [250]LARASSLEDIVLKEASTAVSTPEAPEPISQEQW[38]IERHDSVFVAAHTSAGKTVVAEYAIALAQKH-MTIILYTSLIKAISNQKFRDFR-NTFGDV--GL : 382
```

Q-Loop   Motif I   Motif Ia

Row 2 (β4–β7, residues 101–199):

```
         ß4      α5  α6        ß5    α7      α8        ß6      α9        ß7
              110       120       130       140       150       160       170       180       190
AfHEL308 : -STGDYE--SRDEHLGDCDIIVTTSEKADSLIRNRAS--WIKAVSGLVVDEIHLLDSEKRGATLEILVTKMRRMNKALRVIGLSATAPNVTEIABWIDD---ADYYVS : 199
Pa0592   : -ATGDYD--SKDEWLGKYDIILTAAEKFDSLIRHGSS--WIRDVKMLVVDEIHLIGSRDRGATLEFIITQMLGR---AQLIGLSATIGNPEELAEWLN---AKLIRS : 196
DmMus308 : GFYCGY---TPPGGFESLHVAICTIEKANSIVNKLMEBQGRLETIGMVVVDEVHLLSDKGRGYILELLAKILYM[7]IQVIMSATLBNVQLLQSWLD---ADLYIT : 415
HsPolQ   : GYMCST----SPSRHFSSLDIAVCTIERANCLINRLIEENKMDLIGMVVVDELEMLIGEGSRGATLEMTLAKILYTSKTTQILIGMSATLNVEDLQKFIQ---AEYYTS : 285
HsHel308 : EYACSKGRFPPTKRREKKSLYIATIEKGHSLVNSLIETGRIDSLGLVVVDELEMIGEGSRGATLEMTLAKILYTSKTTQILIGMSATLNVEDLQKFIQ---AEYYTS : 517
yMtr4    : -MTGDIT-INP-----DAGCLVMTTEILRSMIYRGSE--VMREVANWIFDEVHYMRDKBRGVVWEETIILLPDK---VRYVFLSATVPNAMEFAEWIC[6]CHIVYT : 319
HsMtr4   : -MTGDVT-INP-----TASCLVMTTEILRSMIYRGSE--VMREVANWIFDEIHYMRDSBRGVVWEETIILLPDN---VHYVFLSATVPNARQFAEWIC[6]CHVIYT : 309
ySki2    : -ITGDVQ-INP-----DANCLIMTTEILRSMIYRGAD--LIRDVEFVIFDEVHYMNDQDRGVVWEEVIIMLPQH---VKIILLSATVPNTYEBANWIG[6]IYVIST : 501
HsSki2   : -LTGDVQ-LHP-----EASCLIMTTEILRSMIYSGSD--VIRDIEWVIFDEVHYMDVBRGVVWEEVIIMLPDH---VSIILLSATVPNALEFADWIG[6]IYVIST : 480
```

Motif Ib (part I)   Motif Ib (part 2)   Motif II   Motif III

Row 3 (β8–α13, residues 200–300):

```
       ß8  ß9  ß10       α10         ß11     α11         α12            α13
       200       210       220       230       240       250       260       270       280       290
AfHEL308 : DWRPVEVEGVLCEGTTELFDGAFSTSRRVKFEBIVEECVAENGGIVLVESTGRAEKIFVKTSAITAKYVENEGLEKAI--PEENEGEMSRK----IAECVRKCAA : 300
Pa0592   : DWRPVAPLRKGVFYQGPIFWEDG--SERYNSWEEIVYDAVKKGKGALVRVNMREKAEKIFILPAKKVKNYLDRKEL-REIRBIPAESLEENITNEK--LAKAILSCVA : 298
DmMus308 : NYRPVALPKEMIKVGTVTYDH[17]LENDSDDVALICIETLLEGCSIVNFCPSKDWCEKLAVQTATAIHV[16]NP--RALAELKQQLRDII-TGLDGVWSKAITYACA : 545
HsPolQ   : DFRRPVITLESVKVGNSTYDS[13]VKGDEDHVVSIICYETICDNHSVLLFCPSKKWCEKLADIIAREFYN[19]EQ--RELBLVMDQLRRLLSGLDSVLQKTVPWCVA : 414
HsHel308 : QFRRPVBIKEYLKINDTIYEY[21]LKKMDPDHLVALVTEVIPNYSCNVAEMTCPSKKDWCEKLKHKEKEKCNVAEMTCKFLSKEYLHKEKEKCKNLKNIGNGNLCPNYKRTIPFGVA : 641
yMtr4    : NFRPTPIQHYIFP[8]IVDE[43]GDAKGDIYKIVKMIWKKKYNPIVVFSFSKRDCEBIALKMSKLDFN[10]IFN-NAIAIIPETDRELIQIKH--ILPLIRRGIG : 473
HsMtr4   : DYRPTPIQHYIFP[8]IVDE[35]GPSN--VFKIVKMIMERNFQPIVIFSFSKRDCEAYALQMTKLDFN[9]EVFS-NAIIDCDSDEDKKLKQVEH--ILPLIRRGIG : 453
ySki2    : PKRPVPTEINIW[38]TDNG[52]GPSKKTWPEIVNYLRRRELLPMVVRVFSKRKCEBYDWIEGINFC[9]MFIE-KSIHTRKKEDRDLIQILK--TRSLLERGIA : 693
HsSki2   : VTRPVPLEHYLF[12]LDS[38]AQDRGVYLSLIASLRTRAQLPVVVFTFSRGRCEDEQASGFITSLDLT[9]LFLQ-RCIARLRGSDRQLIQVLH--MSELLNRGIG : 633
```

Motif IV

Row 4 (β12–α16, residues 300–405):

```
       ß12     α14       ß13       ß14        α15         ß15    α16
       300       310       320       330       340       350       360       370       380       390       400
AfHEL308 : FHHAGLLNGORRVBDAFRRGNIKVVVATPTLAAGVNLPARRVIVRSLIRFDGY-SRIKVSEYKQMAGRAGRPGMDERGEAIIIVG-KDREIAVKRYIFCEPERI : 405
Pa0592   : PHHAGLGRDERIVVEDNFRKGLIKVVVATPTLSAGINTPAFRVIIRDTFEYSEFGMEIPDVLVCQMGRAGRPRYDEVGEAIIVST-TEEPSLVIDHYIKCKPEKI : 404
DmMus308 : FHHAGLTTEBRDIIEASFKACALKVVVATSTLSSGVNLPARRVIIRTPIFGG---RIDITYRQMVGRIGRAGMKGCESILICN--EINARMGRDLLVSELQPI : 646
HsPolQ   : FHHAGITFEBRDIIEGAFRQGLIRVLAATSTLSSGVNLPARRVIIRTPIFGG---RLDILTYRQMVGRIGRAGMKGIDTVGESILICK--NSEKSKGIALLQCSLKPV : 515
HsHel308 : MHHSGITSDERKULPEAVSRGLITCTSLAPARRVIFTCDSNEMRNQYKQMAGRIGRAGDIDTVGESILILAP-YVAK--EIKRNQYKQMAGRIGRAGDIDTVGESILILQ--EKDKQQVLELITKPLENC : 742
yMtr4    : IHHSGLLPILKETIEILFQECBLKVLPATPFSILCLMNPAKTVVFTSVIKWDGQQFRIVSGGEYIQMSGRAGRRGLDDRGIVIMMD-EKMEPQVAKGMVKLQADRI : 579
HsMtr4   : IHHSGGLLPILKETIEILFSECBLKVIKAFPAMGINMPARRVLFTNAIKFDGKDFRILSSGEYIQMSGRAGRRGLDDRGIVILMVD-EKMSPTIGKQLLKCSADDI : 559
ySki2    : VHHSGILPILKEVIVEMLFSKGFIKVLPATPAMGINMPARRVIVMPSGEYIQMSGRAGRRGLDSTGTVIVMAYNSPLSIATFKEVTMCVPTRI : 800
HsSki2   : VHHSGILPILKEVMELFSRGFIKVLIPATPAMGINMPARRVVFDSMFKHDGSTFRILSSGEYIQMSGRAGRRGLDPTGTVIILCKGRVPEMADLHRMMGKPSQI : 740
```

Motif IVa   Motif IVb   Motif V   ß-Hairpin   Motif VI

Row 5 (α17–α22, residues 405–506):

```
            α17            α18   α19       α20          ß16 ß17  α21         α22
       410       420       430       440       450       460       470       480       490       500
AfHEL308 : TSKLGVETHIRFHSISIICDGYAKTLEELEDFFADIFFFKQNEISLSYE--LERVVRQLENWGM--VVEAAHLAPTYLGSLVSRLYIDPLTGF-IPHDVLSRMELS- : 506
Pa0592   : PSQLSNESIERSQLIAIAITFGYSEFKEIYDPLER[6]EYAHQGKDPYMLEEKIRRIIYFLLENEFIEVTLEDEIKPLPLGVRTTKLYIDPMTAK-IPKDTLPRIEKNP : 509
DmMus308 : TSC[8]STHIKRATIEVISSGVANTKEDIDFFVNCILLSAQKAPHAKEKPPDEESDANYINDALDFLV[17]YVATLGAACLASSMPPTDGLIIPAELQKSRRSFV : 770
HsPolQ   : RSC[9]TGSMIRAILEIIVGGVASTSQDMHTYAACIFLAASMKEGKQGIQRNQESVQLGAIEACVMWLLE[19]PTLGSATLSSSLSPADTLDIFADLQRAMKGPV : 642
HsHel308 : YSHL[8]KGIQTLFLSIIGLKIATNLDDIYHFMNGIFFGVQQKVLLKEKS-LWEITVESLRYLTEKGLLQK[16]TLGRASF-KGTIDLAYCDILYRDLKKGLEGL : 863
yMtr4    : DSAPHLGYNMILNLMIVEGISPEFMLEH-------SFFQFQFHYRAIPGVVEKVKNSEEQYNK----IVIPNEESVVIYYKIRQQLAK--LGKE-----IEEYIHK : 665
HsMtr4   : NSAPHLTYNIVLNLIIRVEEINPEYMLEK-------SFYQFQFHYRAIPGVVEKVKNSEEQYNK----IVIPNEESVVIYYKIRQQLAK--LGKE-----IEEYIHK : 645
ySki2    : QSQFRLTYNMILNLIIREALRVEEMIKY------SFSENAKETLQPEHBKQIKVLQEELQT----IBYKSCEICDNDIEKFLELMLA-----YKBATVNLMQEMVK : 890
HsSki2   : QSQFRLTYMILNLIIRVDALRVEDMMKR------SFSEFPSRKDSKAHBQALABELTKRLGA----LBEEPDMTGQLVDLPEYYS-----WGEBITETQHMIQRRIME : 830
```

Row 6 (α23–α27, residues 510–613):

```
           α23            α24              α25            α26              α27
       510       520       530       540       550       560       570       580       590       600       610
AfHEL308 : -DIGALHLICRTPDMERLTVRKTDSNVEEEAFRLRKELSIYPSDFSVEYDWFLSEVKTALCLKDWIEBKDEDEICAIKYGIAPGDLRRIEVETABWLSNAMNRIAEE : 610
Pa0592   : NPLGIBHMISLAPDITPLSYSKRETSMLEDEYYSLMDRLIYFELDYENERKFFR-AFKTALVLNAWINEVPEGEIVERNVEPGDIYRIEVETABWLIYSLGEIAKV : 613
DmMus308 : LESE-LIPAYLVTTPMFEDWTTI--DIYRFFC[43]IHKRIFYIALALEELVNET-------------PINVVVHLBYLBQLQQMASTEAGIVTAFCNSL : 894
HsPolQ   : LEND-LIIIYLVTPMFEDWTTI--DIYRFFC[43]IHKRIFTSLVLLDLISEV---------------PLRBINQIXYGCNRGOIQSLIQQSAAVMAGMITVFSNRL : 768
HsHel308 : VLESLLHLILYLTTPYDLVSQCNPD--IMIYFRQFSQLSPABQNVAAILGVSESF[22]YLSFVLYTLLKETNWTVSEIBNMPRGYIQNLITGTASBSSCVLHFCEEL : 986
yMtr4    : -PANALSFLIQPGRLVEISVNGKDNYGIGAVVD----------------SFFQFQFHYRAIPGVVEKVKNSEEQYNK-------------------------[377] : 771
HsMtr4   : -PKYQLPFLIQPGRLVKVKNEG-DDFGIGVVVN----------------SFYQFQFHYRAIPGVVEKVKNSEEQYNK----------------------[367] : 750
ySki2    : -SPSILHIIKEGRLVAFRDPN-DCLKLGFVFK----------------SFSENAKETLQPEHBKQIKVLQEELQT-------------------------[367] : 1015
HsSki2   : -SVNGLKSLSAGRVVVVKNQE-HHNALGVILQ----------------SFSEFPSRKDSKAHBQALABELTKRLGA----------------------[386] : 935
```

Row 7 (α28–α33, residues 620–691):

```
          α28      α29      α30     α31    α32      α33
       620       630       640       650       660       670       680       690
AfHEL308 : VGNTSVGGLTERIKHGYVKBLIIBLVRIRHICRVRARKLYNAGIRNAEDIVRHREKVASLIGRGIAERVVEGISVKSLNPES----------------- : 691
Pa0592   : -[9]YVNTLRLRVKHGIRBLIPLIMBPMVSRKRARALYNAGPKDIESIRNAKPSEILRIEGIGAKIVEGIFKYLGKEVKITERPRKGTLDYFLNP----- : 714
DmMus308 : -[7]IVSQFKDIPIFGIHRDLIDIMTPDISQKRARALBDAGITSIVBLAGADFVEIEKVLYNSISPFDSAKQHDHENADEAAKRNVVRNFYITGKAGMTVS[1061] : 2059
HsPolQ   : -[7]LISQFQKIPTFGIQRELCDLVRVSLDNAQRARVLYASGFHTVADILARANIVEIEVILKNAVPFKSARKAVDEEBEAVERRNMRTIWVTGRKGLTER[1718] : 2590
HsHel308 : -[8]LVELTKKLITYQVKABLIPLMBYTGVLEGRAKQLYSACYKSLMHIANANPEVLVRTIDHLSRRQAKQIVSSAKMLLHEKAEALQEEVEBLLRLPSDFPGA[8] : 822
yMtr4    : -------------------------------------------------------------------------------------------------- : 771
HsMtr4   : -------------------------------------------------------------------------------------------------- : 750
ySki2    : -------------------------------------------------------------------------------------------------- : 1015
HsSki2   : -------------------------------------------------------------------------------------------------- : 935
```

RAR

Figure 6: Sequence alignment of *Archaeoglobus fulgidus* Hel308 with selected SF2/Ski2 family helicases. Conserved regions and residues are shaded light grey (low conservation), dark grey (moderate conservation) or black (high conservation). The annotated secondary structure of *Archaeoglobus fulgidus* Hel308 is shown on top of the alignment (arrows: ß-strands, boxes: α-helices). Residues implicated in DNA binding are highlighted with triangles: dark red: 3´- 5´ single strand, magenta: interactions with ss-ds junction on ß-hairpin loop, yellow: double strand. The positions of sequence motifs are indicated. Abbreviations: AfHel308: *Archaeoglobus fulgidus* Hel308; Pa0592: *Pyrococcus abyssii* Hel308; DmMus308: *Drosophila melanogaster* Mus308; HsPolQ: *Homo sapiens* PolΘ; HsHel308: *Homo sapiens* Hel308; yMtr4: *Saccharomyces cerevisiae* Mtr4; hMtr4: *Homo sapiens* Mtr4; ySki2: *Saccharomyces cerevisiae* Ski2; hSki2: *Homo sapiens* Ski2

1.3.2 Functional characteristics of the archaeal Hel308 family members

The *in vitro* properties observed for archaeal Hel308 are all remarkably similar to the *in vitro* properties of eubacterial RecQ helicase (Fujikane *et al.,* 2005; Guy and Bolt 2005). Besides single-stranded DNA dependent ATPase activity, archaeal Hel308 binds and unwinds a variety of different branched nucleic acid substrates in 3´→ 5´ direction, with a significant preference for binding to and unwinding of a fork with lagging strand only (Table 3). In addition, pfuHel308 and mthHel308 have been shown to unwind synthetic Holliday junctions and D-Loop structures (only shown for mthHel308) (Guy and Bolt 2005).

Table 3: Prefered DNA structures of archaeal Hel308

| Y-Fork | Lagging strand only fork (FD) | Leading strand only fork | Fork with leading and lagging strand | Holliday junction (HJ) |

The functional complementarity between archaeal Hel308 and eubacterial RecQ was also shown in genetic complementation experiments using *E. coli* cells (Fujikane *et al.,* 2005; Guy and Bolt 2005), and the results implicate that archaeal Hel308 targets DNA structures, which form at stalled replication forks (Figure 5). Analogous to RecQ, archaeal Hel308 may clear the lagging strand, thereby

allowing the generation of an Okazaki fragment initiation zone, and the loading of archaeal primosome components or translesion polymerases to allow replication over damage (Guy and Bolt 2005). Interaction with components of the primosome, replisome or with polymerases could be mediated by the C-terminal domain of archaeal Hel308, which have been shown to be important for the interaction with archaeal PCNA (Fujikane *et al.,* 2005).

The ability of archaeal Hel308 to dissolve synthetic HJ and to dissociate RecA-mediated recombination intermediates produced by a plasmid DNA (Fujikane *et al.,* 2005) could be important for an anti-recombination function (similar to RecQ) or in the late stage of homologous recombination (similar to RuvABC). Many proteins involved in the onset of homologous recombination like the recombinase Rad51, the single-stranded binding protein RPA, and the DNA repair proteins Rad50/Mre11 are highly conserved between archaea and eukarya (Uemori *et al.,* 1995; Ishino *et al.,* 1998; Hopfner *et al.,* 2000), which is consistent with the general finding that archaea share many similarities with eukarya in their genetic information processing pathways, in terms of sequences of the involved protein factors (Olsen and Woese 1997). This similarity makes it likely that they use similar mechanisms to repair and restart replication forks in response to damage signalling mechanisms at sites remote from replication origins.

1.4 Objectives

Helicases play essential roles in nearly all nucleic acid metabolic pathways in which they use the free energy of ATP hydrolysis to catalyze the thermodynamically unfavourable separation of nucleic acid duplex separation.

However, the detailed structural and molecular mechanism of nucleic acid duplex separation by SF2 helicases still remains unclear, but is critical for understanding the molecular mechanisms of nucleic acid metabolic pathways in living cells.

The aim of this PhD thesis was to reveal the structure of the Superfamily 2 (SF2) Mus308 helicase homolog from the archaeon *Archaeoglobus fulgidus* (afHel308) by X-ray crystallography. The eukaryotic and archaeal Mus308 family of helicases has an implicated function in DNA repair and was shown to preferentially unwind lagging strands at stalled replication forks. Comparison of the structures of afHel308 in the absence and presence of a DNA oligonucleotide should reveal important insights into the structural interplay between the enzyme and its DNA

substrate during processive 3´→ 5´ DNA duplex separation, and may provide an DNA unwinding model which is also applicable to other SF2 family members. Biochemical analysis of structure-guided mutant afHel308 should support the model obtained from the structures and may provide an explanation of how ATP binding and hydrolysis is coupled to DNA translocation and DNA base pair separation. In addition, the interaction between the enzyme and its DNA substrate may allow extrapolation of how afHel308 is targeted to possible *in vivo* substrates and interacts with components of the DNA repair machinery.

2 Materials and Methods

2.1 Materials

See Part I

2.2 Methods

General methods are described in Part I. The specific purification procedures and crystallization of the *Archaeoglobus fulgidus* Hel308 protein as well as the biochemical assays are listed in the following chapter.

2.2.1 Structure determination of the archaeal SF2 Helicase

2.2.1.1 Cloning, expression and purification of *Archaeoglobus fulgidus* Hel308

The coding sequence of archaeal Hel308 was amplified by PCR from *Archaeoglobus fulgidus* (af) genomic DNA (see Table 4 for oligonucleotides). The PCR product was cloned into a pET-29 vector (Novagen, Schwalbach/Ts., Germany) using the corresponding restriction sites (Table 5), and transformed into competent *E. coli* Rosetta (DE3) cells. The protein was expressed for 15 h at 18°C. Cells were harvested by centrifugation, resuspended in Lysis buffer (Table 6) supplemented with 200 µM PMSF, and disrupted by sonication. Cell debris was removed by centrifugation. *E. coli* proteins were denatured by a heat step for 10 minutes at 80°C and removed by centrifugation. The supernatant was loaded onto an equilibrated Ni^{2+}-NTA column (5 ml, Qiagen, Hilden, Germany; Table 6). After washing with Wash buffer I and II, the protein was step-eluted with Elution buffer (Table 6). Protein containing fractions were determined using Bradford Protein Assay (Bio-Rad, Munich, Germany), pooled and diluted 1:4 with Dilution buffer (Table 6). afHel308 was further purified by anion exchange chromatography (Resource Q, GE Healthcare, Freiburg, Germany). The column was equilibrated with Buffer A (Table 6) prior to the experiment. The protein was eluted with a gradient of 20 columns volumes from 50 mM NaCl (Buffer A) to 1 M NaCl (Buffer B, Table 6). Peak fractions were concentrated and loaded onto a Sephadex S200 16/60 size exclusion column (GE Healthcare, Freiburg, Germany) equilibrated with the corresponding buffer (Table 6). Pooled peak fractions were concentrated for

crystallization to 10-12 mg/ml using centrifugal filter devices (Amicon® Ultra, Millipore, Billerica, MA, USA). For MAD phasing selenomethionine was incorporated as described (Part I, 2.2.2.4). Derivatized protein was purified according to wild-type protein, whereas all buffers were degassed before use and contained 10 mM β-mercaptoethanol or 2 mM DTT (Roth, Karlsruhe, Germany). Mutations were introduced by overlap PCR (Part I, 2.2.2.2) and cloned into pET-29 vector (Novagen, Schwalbach/Ts., Germany). Expression and purification of the mutant protein were carried out as described for the wild-type protein. The C-terminal truncated proteins (Δ626 and Δ400) were obtained by PCR amplification and cloned into pET-29 vector (Novagen, Schwalbach/Ts., Germany). Expression of the truncated constructs was performed according to the full-length protein. The truncated protein Δ626 was purified according to the full-length protein. Δ400 was purified by Ni^{2+}-NTA affinity chromatography, HighTrap SP cation exchange chromatography and Sephadex S200 16/60 size exclusion chromatography using the same buffers as for the wild-type protein (Table 6).

Table 4: Oligonucleotides

Oligo name	Sequence
afHel308 for Nde I	aaaaaaa<u>catatg</u>aaggtggaggagcttgctg
afHEL308_c rev Not I	aaaaaaaa<u>gcggccgc</u>tgactccgggtttaaagacttaa
Desired deletion	Sequences
afHEL308_Δ400_c rev Not I	aaaa<u>gcggccgc</u>tccaaaaatatacctttcacagcaat
afHEL308_Δ626_c rev Not I	aaaa<u>gcggccgc</u>atgctttatcctctccgtaagccc
afHEL308_E146Q_for	gccttgtcgttgac<u>caa</u>atacacctactc
afHEL308_E146Q_rev	cgagtaggtgtat<u>ttg</u>gtcaacgacaaggc

Underlined regions mark the recognition sites for restriction endonucleases and mutated codons.

All sequences are given in 5´- 3´ direction.

Table 5: Expression plasmids

#	Insert	Vector	Restriction sites	Tag
1	afHEL308	pET-29	Nde I, Not I	c-6xHis
2	afHEL308 Δ400	pET-29	Nde I, Not I	c-6xHis
3	afHEL308 Δ626	pET-29	Nde I, Not I	c-6xHis
4	afHEL308 E146Q	pET-29	Nde I, Not I	c-6xHis

Table 6: Buffers for the purification of archaeal Hel308

Buffer	Description
Ni-NTA	
Lysis buffer/Equilibration buffer	20 mM NaH_2PO_4 pH 7.5 (25°C), 200 mM NaCl, 10 mM β-mercaptoethanol
Wash buffer I	20 mM NaH_2PO_4 pH 7.5 (25°C), 1.5 M NaCl, 10 mM Imidazol, 10 mM β-mercaptoethanol
Wash buffer II	20 mM NaH_2PO_4 pH 7.5 (25°C), 200 mM NaCl, 10 mM Imidazol, 10 mM β-mercaptoethanol
Elution buffer	20 mM NaH_2PO_4 pH 6.8 (25°C), 200 mM NaCl, 250 mM Imidazol, 10 mM β-mercaptoethanol
Dilution buffer	20 mM NaH_2PO_4 pH 6.8 (25°C), 10 mM β-mercaptoethanol
Resource Q (full length/mutant Hel308)/HiTrap SP (Δ400)	
Buffer A	20 mM NaH_2PO_4 pH 7.5 (25°C), 50 mM NaCl, 2 mM DTT
Buffer B	20 mM NaH_2PO_4 pH 7.5 (25°C), 1 M NaCl, 2 mM DTT
Gelfiltration	20 mM Tris/HCl pH 7.8 (25°C), 200 mM NaCl, 10 % glycerol, 2 mM DTT

2.2.1.2 Crystallization and structure determination of apo-afHel308

Crystals of apo afHEL308 in space group P6$_1$22 with one molecule per asymmetric unit were obtained by sitting drop vapor diffusion by mixing 1 µl protein (10 mg/ml in 20 mM Tris pH 7.8, 200 mM NaCl, 10 % glycerol) and 1 µl of reservoir solution (0.1 M NaCitrate pH 5.6, 1.0 M $(NH_4)_2PO_4$, 15 % glycerol) at 20°C. Prior to data collection, crystals were mounted in nylon loops and flash frozen in liquid nitrogen. To determine the structure of apo afHEL308, a multiple-wavelength anomalous

diffraction (MAD) experiment was recorded at the K absorption edge to 3.7 Å at beamline ID29 (ESRF, Grenoble, France). The optimal wavelengths were determined experimentally with a fluorescence scan. Synchrotron data from 90 images (1° rotation per image) were processed and scaled with XDS and XSCALE (Kabsch 1993). Atomic positions of 13 Selenium atoms were located with SHELXD (Sheldrick and Schneider 1997). Phases were obtained with SHARP software (Global phasing) and improved with SOLOMON (Collaborative Computational Project 1994), resulting in an interpretable electron density. The model was build manually with MAIN (Turck 1992). Refinement was performed with CNS (Brunger *et al.,* 1998), and included overall anisotropic B-factor and bulk solvent corrections, individual B-factor refinement, simulated annealing, and positional refinement. The 3.7 Å resolution structure was used as a search model to obtain phases for a native data set at 3.1 Å resolution (recorded at PX (SLS, Villigen, Switzerland)) by molecular replacement using PHASER (Collaborative Computational Project 1994). The stereochemistry of the final model of apo afHel308 was analyzed with PROCHECK (Laskowski *et al.,* 1993).

2.2.1.3 Crystallization and structure determination of DNA:afHel308 complex

To obtain crystals of afHel308 in complex with DNA, the protein was incubated with a 15mer double-stranded DNA with a 10mer 3′-overhang (5′-CTAGCAAGCCAGAATTCGGCAGCGT-3′, 5′-ACGCTGCCGAATTCT-3′) in size exclusion buffer (Table 6), and a protein:DNA ratio of 1:1.2 for 20 minutes on ice. Crystals of the DNA:afHEL308 complex were obtained according to apo afHEL308 in space group P6$_1$22 with one molecule per asymmetric unit, using 0.1 M HEPES pH 6.4, 20 % MPD, 0.1 M MgAcetate, 15 % glycerol as precipitant solution. Data from 90 images (1° rotation per image) diffracting to 3.0 Å resolution were collected at PX beamline (SLS, Villigen, Switzerland) and processed with XDS (Kabsch 1993). The structure of the DNA:afHEL308 complex was solved by molecular replacement with PHASER (Collaborative Computational Project 1994), using the refined model of apo afHEL308 as search molecule. A single molecular replacement solution was refined at 3.0 Å by using the procedure described above. The stereochemistry of the final model of the DNA:afHel308 complex was analyzed with PROCHECK (Laskowski *et al.,* 1993).

HPLC purified and lyophilized oligonucleotides used for crystallization were obtained from Thermo Electron Corporation (Ulm, Germany) and dissolved in 10 mM Tris/HCl [pH 8.5]) at a concentration of 3 nm/µl. Adjustment of the pH of the DNA solution to 7-8 was performed to avoid protein denaturation after DNA addition due to acidification. Double-stranded DNA substrate was generated by mixing equimolar amounts of the two complementary oligonucleotides, heating to 95°C for 5 minutes, followed by slow cooling to room temperature. The protein:DNA complex was directly used for crystallization without further purification procedures.

2.2.2 Biochemical methods

2.2.2.1 Radioactive labelling of DNA substrates

DNA strands were end-labeled using T4 polynucleotide kinase (PNK) (New England Biolabs, Frankfurt, Germany) and γ-^{32}P-ATP (GE Healthcare, Freiburg, Germany) for 60 minutes at 37°C in a 20 µl reaction following the supplied instructions. DNA was separated from unincorporated γ-^{32}P-ATP using the Nucleotide Removal Kit (Qiagen, Hilden, Germany). For generating double-stranded DNA the ^{32}P-labeled strand was mixed with an appropriate amount of unlabeled complementary DNA strands. Annealing was performed in buffer containing 40 mM Tris/HCl pH 7.5, 100 mM NaCl, 10 mM MgCl$_2$ and 1 mM DTT for 6–18 h by reducing the reaction temperature from 95 to 20°C in a heating block. Success of annealing was controlled by 12% TAE-PAGE, and annealed holliday junction (HJ) substrate were used to assay helicase acitivty. DNA substrate used for ATPase activity assays was prepared in the same way without radioactive labeling. Here, success of annealing was monitored by UV-shadowing. The sequence of the oligonucleotides used for biochemical assays are listed in Table 7.

Table 7: Oligonucleotides for biochemical assays

Name	Sequence
HJ1	atcgatagtctctagacagcatgtcctagcaagccagaattcggcagcgt
HJ2	gacgctgccgaattctggcttgctaggacattctttgcccacgttgaccc
HJ3	gggtcaacgtggcaaagaatgtcctacgtccgatacggataatcgccat
HJ4	atggcgattatccgtatcggacgtcggacatgctgtctagagactatcga
FD1	atcgatagtctctagacagcatgtcctagcaagccagaattcggcagcgt
FD2	gacgctgccgaattctggcttgctatgtaactctttgcccacgttgaccc
FD3	ggacatgctgtctagagactatcgat
ssDNA	tttttttttttttt

All sequences are given in 5′- 3′ direction. HJ: Holliday junction, FD: Lagging strand only fork (see also Table 3)

2.2.2.2 ATPase assays

For assaying ATPase activity of wild-type and mutant afHel308, 100 nM of each afHel308, Hel308a^{E146Q}, afHel308$^{\Delta400}$, and afHel308$^{\Delta626}$ were incubated with and without 4 nM DNA substrate in 20 mM Tris/HCl pH 7.5, 50 mM KCl, 1mM MgCl$_2$, 10% glycerol, 2 mM DTT, 1 mM cold ATP, 90 µg/ml bovine serum albumin (BSA), 0.1% PEG 8000 and 20 nM γ-^{32}P-ATP in a 20 µl reaction mixture at 37°C for 30 minutes. Released phosphate was analyzed by thin layer chromatography on polyethyleneimine cellulose using 1 M Formic Acid/0.5 M LiCl as running buffer and phosphorimaging (GE Healthcare; Freiburg, Germany). The amount of hydrolyzed ATP was quantified with ImageQuant® (GE Healthcare; Freiburg, Germany), and is given in mol ATP hydrolyzed/mol protein per second. Standard deviations represent calculated means ± standard error of three independent experiments.

2.2.2.3 DNA unwinding assays

Helicase activity was assayed with 100 nM of each protein in a 20 µl reaction containing 20 mM Tris/HCl pH 7.5, 50 mM KCl, 1mM MgCl$_2$, 10% glycerol, 2 mM DTT, 90 µg/ml bovine serum albumin (BSA), 0.1% PEG 8000 and 1 mM ATP in the absence or presence of 4 nM DNA substrate at 37°C for 15 min. The reaction

was terminated by the addition of 0.3% SDS, 10 mM EDTA, 5% glycerol and 0.1% bromphenol blue. The products of the reaction were fractionated by 12% non-denaturing PAGE in TAE buffer. Gels were dried and DNA unwinding was quantified by phosphorimaging (GE Healthcare; Freiburg, Germany). Values and error bars represent means ± standard error of three independent experiments, and are given in percent DNA unwound.

2.2.3 Limited proteolysis

Limited proteolysis was performed in order to reveal conformational changes in the protein upon binding of ATP and/or DNA. The reaction was carried out in size exclusion buffer (Table 6) in a total volume of 20 µl. For each reaction, 14 µg of afHel308 wild-type protein were incubated with different amounts of proteinase K, trypsin or chymotrypsin (Sigma, Deisenhofen, Germany) (0.1, 1, 10 µg) in the absence and presence of 500 µM AMP-PNP (Sigma, Deisenhofen, Germany), 1 mM $MgCl_2$ and 20 µM DNA (15mer duplex with 10mer 3´ tail overhang) for 25 minutes at room temperature. The reaction was stopped by addition of 2 µl PMSF (saturated solution in 2-propanol). The reaction products were analyzed by SDS-PAGE and protein bands were stained with Coomassie Brilliant Blue R250 (Roth, Karlsruhe, Germany).

3 Results

3.1 Purification, crystallization and structure determination of the archaeal helicase Hel308 from *Archaeoglobus fulgidus*

The archaeal helicase afHel308 is grouped into the Superfamily 2 (SF2) of DNA and RNA helicases and shares high sequence similarity to human PolΘ (Marini *et al.,* 2003; Seki *et al.,* 2003; Yoshimura *et al.,* 2006) and Hel308 (Marini and Wood 2002) as well as to *Drosophila* Mus308 (Oshige *et al.,* 1999) and the yeast RNA decay factors Ski2p and Mtr4p (Koonin *et al.,* 2001). Whereas the ATP-dependent DNA unwinding mechanism of SF1 helicases like PcrA (Velankar *et al.,* 1999), Rep (Korolev *et al.,* 1997) and UvrD (Lee and Yang 2006) is structurally well characterized, processive nucleic acid unwinding by SF2 helicases is much less understood. To provide a structural framework of how processive SF2 $3' \rightarrow 5'$ helicases actually separate the two strands of a nucleic acid, and to obtain insights into substrate recognition mechanisms by Mus308 family members, archaeal Hel308 was crystallized in the absence and presence of DNA.

The gene from the archaeal helicase Hel308 was amplified from *Archaeoglobus fulgidus* (af) genomic DNA by PCR as described in 2.2.1.1. The PCR product was cloned into a pET-29 vector with a C-terminal 6xHis-Tag. The resulting expression plasmid was transformed into *E. coli* Rosetta (DE3) cells and the protein was expressed overnight at 18°C.

3.1.1 Purification of afHel308

afHel308 was purified by an initial heat denaturization of *E. coli* proteins, immobilized Ni^{2+}-NTA affinity chromatography (utilizing the C-terminal 6xHis-Tag), Resource™ Q anion exchange chromatography, and HiLoad 16/60 Sephadex S200 size exclusion chromatography, following the procedure described in 2.2.1.1. The protein eluted in one distinct peak (Figure 7), which corresponded to a molecular weight of around 78 kDa (theoretical MW: 78.7 kDa). Protein containing fractions were pooled and concentrated to a final concentration of 10-12 mg/ml. The purified protein was used for crystallization trials and biochemical characterization. From 6 l of expression culture, around 10-15 mg of purified protein could be obtained. (Figure 7) The selenomethionine-containing protein as

well as the mutant constructs were purified according to the wild-type protein. Mutant proteins showed expression levels similar to the wild-type protein and were used for biochemical studies.
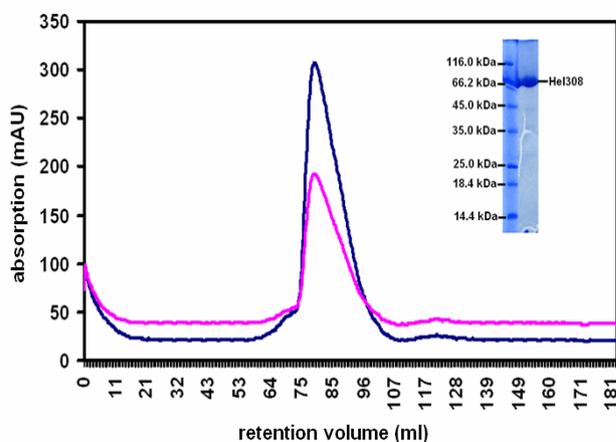

Figure 7: Elution profile of afHel308 from HiLoad 16/60 S200 size exclusion column. (absorption at 280 nm, blue; absorption at 260 nm, magenta). Insert shows SDS-PAGE analysis of a representative peak fraction.

3.1.2 Crystallization of apo afHel308

The afHel308 protein was crystallized by initial screening using commercial crystallization screens in 96-well sitting drop plates by mixing 1 μl of protein with 1 μl of reservoir solution. Small hexagonal shaped crystals appeared within 2-3 days in a variety of different conditions of the Nextal Classic Screen I. For first diffraction experiments, some hexagonal shaped crystals were incubated in mother liquor supplemented with 10-15 % glycerol, 2,3-Butandiol or PEG400. Crystals were mounted in nylon loops and flash frozen in liquid nitrogen. Crystals diffracted to around 6 Å at PX beamline (SLS, Viligen, Switzerland). Diffraction power seemed to depend mainly on the crystal size. No ice rings were formed in crystals incubated in glycerol, whereas 2,3-Butandiol and PEG400 did not seem to avoid ice formation, and therefore were not recommended as cryo-protectants in this case. For structure determination, the initial crystals had to be improved and bigger crystals had to be obtained. First refinement trials of Nextal Classic Screen I #26 resulted in bigger crystals, but only slight improvement of X-ray diffraction capacity. Therefore, an excessive additive screen was performed. The initial crystallization condition Nextal Classic I #26 was mixed in different ratios with other Nextal Screens (Cryo´s, PEG´s, Classic II). By this, big crystals could be obtained in many cryo compatible conditions from which crystals could be directly mounted and flash frozen (Figure 8). Thereby, native diffraction data to 3.1 Å could

be measured. Crystals belonged to space group $P6_122$ with unit cell constants of a=b= 136.15 Å, c=230.82 Å, $\alpha=\beta$= 90°, γ=120° and one molecule per asymmetric unit. The overall weak scattering power may be due to a high solvent content of almost 68 %, and a Matthews volume of 3.84 Å^3/dalton of protein (Matthews, 1968; Kantadjieff and Rupp, 2003). Almost isomorphous selenomethionine-containing crystals in space group $P6_122$ (unit cell constants: a=b=135.69 Å, c=227.56 Å) could be obtained with the same procedure and exhibited also a hexagonal shape with maximum dimensions of 60 μm x 60 μm x 300 μm.
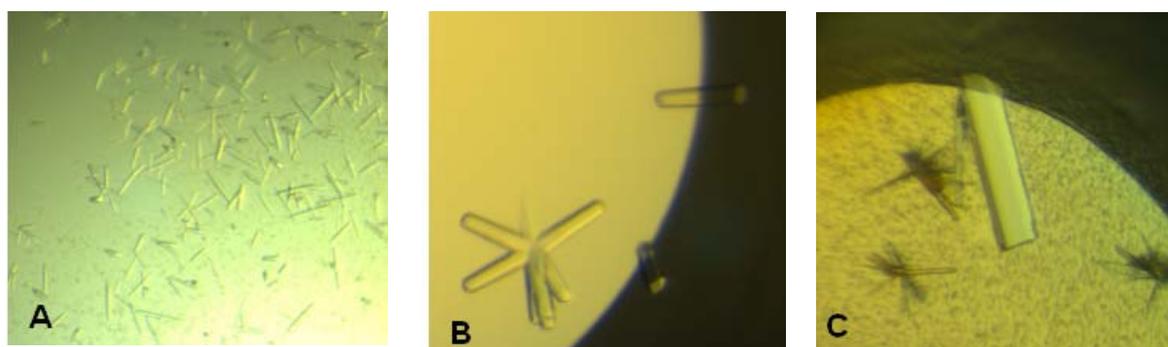

Figure 8: Crystals of apo afHel308 A) First small hexagonal crystals were obtained in Nextal Classic Screen I #26 (0.1 M NaCitrate pH 5.6, 1 M $(NH_4)_2PO_4$). B) Crystal size could be increased by variation of pH and addition of glycerol (0.1 M NaCitrate pH 5.4, 1 M $(NH_4)_2PO_4$, 5% glycerol). C) representative crystal of the refinement procedure using different ratios of Nextal Classic I #26 and Nextal PEG´s or Nextal Cryo´s, respectively (here: 80% Nextal Classic I #26 and 20% Nextal PEG´s #14).

3.1.3 Crystallization of the binary afHel308:DNA complex

In order to obtain structural information about how afHel308 interacts with DNA, and how ATP hydrolysis is coupled to DNA translocation und unwinding, afHel308 was crystallized with its DNA substrate. The apo afHel308 protein crystallized in conditions with high salt concentrations. These conditions are unlikely optimal for crystallization of protein:DNA complexes. A high salt concentration may be usefull for keeping DNA binding apo proteins in solution, as high ionic strength weakens charge-charge interactions between protein molecules. But it would also negatively influence charge dependent interactions between protein and nucleic acids. Therefore, new crystallization conditions had to be found. Initial screenings also included the use of different DNA substrates, like single-stranded DNA as well as double-stranded DNA with 3´-tail overhangs of different lengths (Table 8).

Table 8: DNA substrate for co-crystallization trials

#	Name	Sequence	Remarks
1	25mer ds with 3′ 10mer overhang for	ATCGATAGTCTCTAGACAGCATG TC	diffracting binary crystals
	25mer ds with 3′ 10mer overhang rev	CTAGAGACTATCGAT	
2	20mer ds with 3′ T and 5′ A 5mer overhang for	AAAAAATCGATAGTCTCTAGACA GC	protein aggregation
	20mer ds with 3′ T and 5′ A 5mer overhang rev	GCTGTCTAGAGACTATCGATTTT TT	
3	Fork 20mer for	GCAGTGCTCGCATGGAGCTG	no crystals
	Fork 12mer rev	CAGCTCCATGAT	
	Fork 9mer rev	GAGCACTGC	
4	15mer ds with 3′ 10mer overhang for	GCAGTGCTCGTTTTT	no binary crystals
	15mer ds with 3′ 10mer overhang rev	CGAGCACTGC	
5	25mer ss	ACGCTGCCGAATTCTGGCTTGC TAG	non diffracting crystals

All Sequences are given in 5′- 3′ direction. ds = double strand; ss = single strand

As binding of afHel308 to DNA is not sequence specific, the use of double-stranded DNA with 3′-tail overhangs and the resulting presence of a double-strand:single-strand DNA junction was thought to provide an optimal binding platform for the enzyme to the DNA. This strategy was also successfully used to capture a DNA polymerase:DNA substrate binary complex (Li *et al.,* 1998) as well as to crystallize the SF1 helicase UvrD in complex with its DNA substrate (Lee and Yang 2006), and is consistent with the finding that most helicases need a single-stranded nucleic acid region to bind and to initiate their action of strand separation. Double-stranded DNA was generated as described in 2.2.1.3. Prior to crystallization, protein was mixed with DNA in a 1:1.2 molar ratio and incubated on ice for 20 minutes. The behaviour of the protein after DNA addition was monitored carefully. Some of the DNA substrates were completely useless, as addition resulted in precipitation of the protein, which may be caused by pH changes or

DNA dependent aggregation. The protein:DNA crystallization trials were set up by mixing 1 µl of protein:DNA mixture with 1 µl of reservoir solution in 96-well sitting drop plates (Corning, Schiphol-Rijk, Netherlands). Plates were incubated at 20°C. After 2-3 days, hexagonal crystals appeared in a variety of different conditions. Condition #A3 of the Hampton Natrix Screen (Figure 9, with DNA substrate #3) resulted in small hexagonal crystals. The crystallization condition was further refined by variation of MPD and glycerol concentration, and yielded in crystals with maximum dimensions of 50 µm x 50 µm x 250 µm (Figure 9). Crystals could be directly mounted in nylon loops and flash frozen in liquid nitrogen for data collection. Crystals belonged to space group $P6_122$ and exhibited unit cell constants of a=b=138.38 Å, c=252.97 Å. Like in the apo protein crystals, one molecule per asymmetric unit is found. This results in a Matthews volume of 3.78 $Å^3$/Dalton (Matthews 1968; Kantardjieff and Rupp 2003), and a high solvent content of 67.5 %.
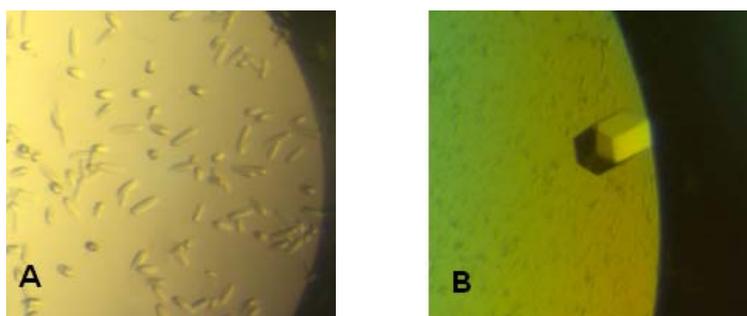

Figure 9: A) Initial crystals of afHel308 in complex with a 15 bp double-stranded DNA with an additional 10 base 3′-tail in Hampton Natrix Screen #A3 (0.05 M MES pH 5.6, 20% MPD, 0.1 M MgAcetate). Crystals exhibited maximum dimension of around 10 µm x 10 µm x 50 µm. B) refined crystal of afHel308:DNA complex in 0.1 M MES pH 6.0, 22% MPD, 0.1 M MgAcetate, 10% glycerol.

3.1.4 Data collection

Diffraction data of the selenomethionine-containing crystals were collected at ID29 (ESRF, Grenoble, France) with an ADSC Q315R CCD detector. Diffraction data of native crystals as well as data of the binary DNA:protein crystals were recorded at PX beamline (SLS, Villigen,Switzerland) with a mar225 mosaic CCD detector. For each data set 90 images (1° oscillation each) were collected.

For phase determination, a two wavelength anomalous dispersion experiment was carried out at the selenium K edge using one selenomethionine containing crystal.

The optimal wavelengths were determined experimentally with a fluorescence scan prior to data collection. Data for the peak wavelength at 0.9795 Å (12.65766 kEv, f´ -8.30, f´´ 5.90) and the inflection point at 0.9797 Å (12.65541 kEv, f´-11, f´´ 3.00) were collected to 3.7 Å, respectively. A remote data set was additionally recorded, but was not included for phase determination due to extensive radiation damage and decrease in scattering capacity of the crystal to 4.5 Å.

A native data set was recorded at 1.0 Å to 3.1 Å. Data of the DNA:afHel308 complex were collected at 0.920 Å to 3.0 Å (Figure 10). Due to the extreme radiation sensitivity of the crystals, MOSFLM (Powell 1999) was used to calculate an optimal collection strategy prior to data collection to obtain a reasonable redundancy and overall completeness with a minimum number of images.
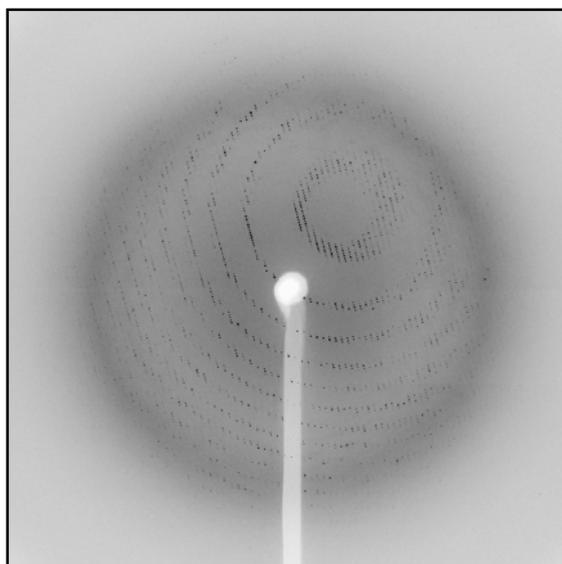

Figure 10: Diffraction image of afHel308 in complex with DNA recorded at PX beamline (SLS, Villigen, Switzerland) to 3.0 Å.

3.1.5 Structure determination and refinement of apo afHel308

All data were processed with XDS and XSCALE (Kabsch 1993). Data were first indexed and scaled in P622 as preliminary analysis allowed no adjugdement of a present screw axis. After scaling, it was possible to identify a screw axis by looking at systematic absences and reflexion conditions, respectively (Figure 11), and to limit potential space groups to $P6_122$ and $P6_522$ (International Tables Crystallography, Volume A: Space-group symmetry, 2002). As systematic absences alone allow no discrimination between these two enantiomorphs, both space groups were used for initial phase determination.
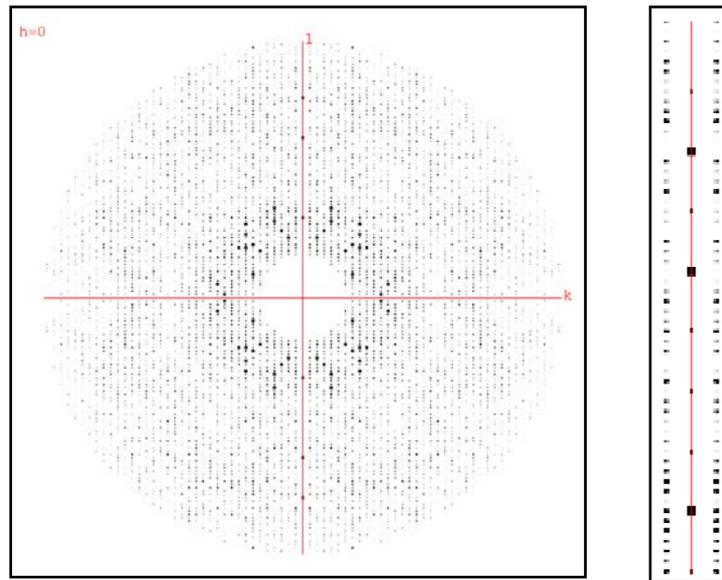
Figure 11: Image of the peak data set with hklview (CCP4, 1994) on section 0kl (left). A zoomed view on the l-axis (h=0, k=0) shows the presence of only every sixth reflection, which corresponds to a 6_1 or 6_5 screw axis (right).

All 13 selenomethionine sites could be located using SHELXD (Collaborative Computational Project 1994). Phases were calculated to 3.7 Å with SHARP (Global phasing), followed by solvent improvement with SOLOMON (Collaborative Computational Project 1994). An interpretable electron density could be calculated with data scaled in $P6_122$ (Figure 12).


Figure 12: 1σ contoured multiple anomalous dispersion (MAD) map of afHel308 at 3.7 Å resolution (blue mesh) with superimposed selenomethionines (red balls). Note that secondary structures are interpretable, but sidechains are difficult to identify.

However, as already mentioned, phases could only be obtained to 3.7 Å. This resolution allowed no automated model building with programs like ARP/wARP

(Morris *et al.,* 2003) or REFMAC/RESOLVE (Terwilliger 2002), and even manual model building was far from easy. However, the presence of 13 selenomethionines as sequence markers allowed manual building of 665 out of 691 possible amino acids in the asymmetric unit with MAIN (Turck 1992). After bulk solvent correction and overall anisotropic B-factor correction, the refined model could be used to phase a 3.1 Å native data set by rigid body refinement with CNS (Brunger *et al.,* 1998). As the protein consists of five more or less individual domains (see below), it was helpful to define individual "rigid groups", with each group containing only residues of one of the five domains. The ensuing refinement of the native data set included repeated cycles of simulated annealing, positional minimization, and restrained individual B-factor refinement with CNS (Brunger et al., 1998), and additional manual completion of the model with MAIN (Turck 1992).

The final model of apo afHel308 contains 672 out of 691 possible amino acids. Although the resolution did not allow the detection of water molecules, ten inorganic phosphates could be built into the model. Refinement resulted in an R-factor of 22.9 (R_{free}=27.5) with reasonable stereochemistry (Figure 13). In Ramachandran plot 82.7% of the residues are in allowed regions. Refinement statistics are listed in Table 9


MAD
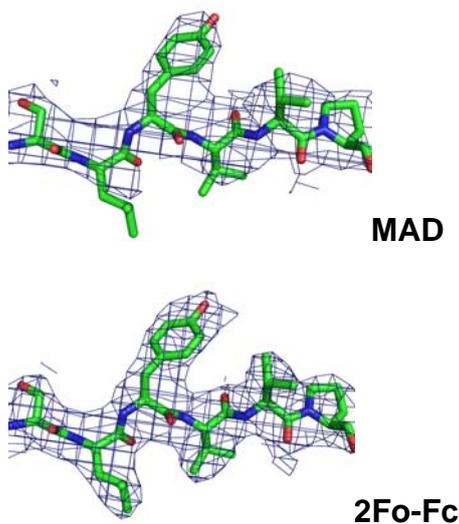
2Fo-Fc

Figure 13: Representative portion of 1σ contoured multiple anomalous dispersion (MAD) and $2F_o$-F_c electron density maps of afHel308.

3.1.6 Structure determination and refinement of afHel308 in complex with a DNA substrate

The 3.0 Å data from the DNA:afHel308 crystal were indexed and scaled with XDS and XSCALE (Kabsch 1993). Data were scaled in $P6_{1}22$. Although the space group did not change, the length of the c-axis of the unit cell increased around 20 Å. The structure of the afHel308:DNA complex was determined by molecular replacement with PHASER (Collaborative Computational Project 1994), using the whole uncomplexed structure as search molecule. An almost clearly defined electron difference density of the bound nucleic acid was obtained from a single MR solution. As most of the DNA is in direct contact with the protein or involved in crystal lattice formation and thus stabilized, the complete DNA could be modelled into the electron density. Only the three terminal base pairs at the duplex end are not visible in the electron density. They are not stabilized by protein or crystal lattice contacts and may exhibit an increased flexibility. Refinement at 3.0 Å followed the procedure described above and included iterative rounds of manual model completion with MAIN (Turck 1992), and refinement with CNS (Brunger *et al.,* 1998). The definition of individual "rigid groups", as in the case of the apo enzyme, helped fitting the model into the electron density. The refinement resulted in an R-factor of 23.1 (R_{free} 27.4) and good stereochemistry (Table 9). The Ramachandran plot shows 85.7% of the residues in allowed regions The final model comprises 682 out of 691 possible amino acids and one 15 base pair duplex with an additional 10mer 3´-tail overhang.

Table 9: Data collection and refinement statistics

Data collection

Data set	*apo* afHel308[1]			DNA:afHel308a[2]
	SeMet K Peak	SeMet K Inf	Native	DNA complex
x-ray source	ID29 (ESRF)	ID29 (ESRF)	PX (SLS)	PX (SLS)
Wavelength (Å)	0.9797	0.9795	1:00	0.920
Data range (Å)	20-3.7	20-3.7	20-3.1	20-3.0
Observations (unique)	78956 (24771[d])	97600 (14094)	102897 (22283)	315440 (54172)
I/σI (last shell)	10.37 (3.53)	15.19 (6.20)	15.57 (1.96)	16.93 (8.97)
Completeness %/(last shell)	98.7 (94.8)	99.0 (99.9)	94.0 (46.5)	99.9 (100)
R_{symm}[a]	11.4 (44.0)	11.1 (39.1)	7.2 (49.7)	9.0 (19.4)

Refinement

Data set	*apo* afHel308	DNA:afHel308a
	Native	DNA complex
Data range (Å)	20-3.1	20-3.0
Reflections F/σF>2.0 (cross validation)	21877 (1062)	28659 (2883)
R_{work}[b] (R_{free}[c])	22.9 (27.5)	23.1 (27.4)
RMS bond length (Å) (angles)	0.007 (1.62)	0.008 (1.25)

[1] apo afHel308 (P6$_1$22), cell constants (Å): a=b=136.15, c=230.82

[2] afHel308:DNA complex (P6$_1$22) cell constants (Å): a=b=138.38, c=252.97

[a] to [d]: see Part I, Table 9.

Model contents: apo Hel308: 1 Hel308 (1-17, 24-331, 335-350, 356-686), 10 inorganic phosphates; DNA:Hel308: 1 Hel308 (1-21, 25-686), 1 Hel308 (1-21, 25-686).

3.2 Structure of archaeal afHel308

The archaeal helicase afHel308 consists of five structural domains (domain 1 to domain 5) that are arranged to form a rectangular molecule of approximately 70 Å x 55 Å x 65 Å (Figure 14A). Domain 1 and domain 2 exhibit a characteristic RecA-type α/β fold (Figure 14B). Both domains comprise the seven conserved helicase motifs in their interface cleft, which are implicated in ATP-Hydrolysis (I, II, III, IV, VI) and/or DNA binding (Ia, Ib, IV, V, VI) (Gorbalenya and Koonin, 1993, Velankar et al., 1999). Both, the RecA-fold domains and the conserved motifs are lined in an arrangement similar to that observed in other members of the SF2 of helicases like RecQ (Bernstein *et al.,* 2003) and Hepatitis C virus NS3 (Kim *et al.,* 1998), which implicate a related ATP-hydrolyis mechanism (Caruthers and McKay 2002).
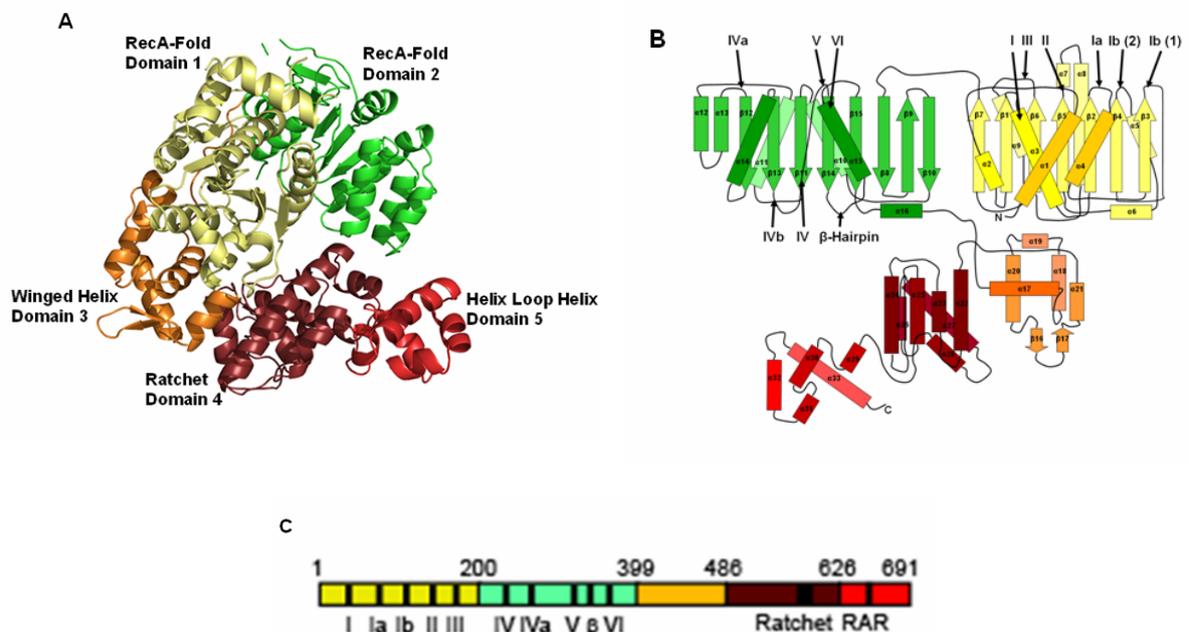

Figure 14: A) Ribbon presentation of the crystal structure of afHel308 in the absence of DNA. afHel308 consists of two RecA-fold domains (1, yellow, and 2, green) that harbor the ATPase active site in the interface, a winged-helix domain (3, orange), a ratchet domain (4, deep red) and a helix-loop-helilx domain (5, red). B) Topology diagram of afHel308 with secondary structure annotation and depicted conserved helicase motifs.C) Scheme of the five domains of afHel308. Domain boundaries are indicated on top, sequence motifs beneath the scheme. Roman numbers: SF2 helicase motifs; β-hairpin loop; R: ratchet helix; RAR: the Arg-Ala-Arg motif.

Domain 3 possesses a winged-helix (WH) motif. It is connected via a long flexible linker to domain 2, which allows tight packing against domain 1. The WH fold of afHel308 comprises five α-helices and two β-strands in the canonical order α17-

α18-α19-α20-βx16-β17-α21 (Figure 14B). Residues in α17-α18-α20 thereby form the tightly packed hydrophobic core of the winged-helix domain. WH domains have been found to act as DNA binding motifs in many proteins (Gajiwala and Burley 2000), and are frequently found in general transcription factors like TFE/TFIIEα (Brennan 1993; Meinhart *et al.,* 2003) and TFIIF (Groft *et al.,* 1998; Kamada *et al.,* 2001). Interestingly, a WH domain is also found in the SF2 helicase RecQ (Bernstein et al., 1998). Most DNA binding WH domains interact with DNA via the "recognition" helix and via wing 1, and generally show a continous positively charged surface on the DNA binding site (Gajiwala and Burley 2000). In aHEl308, both "recognition" helix and wing 1 are exposed on the surface. In contrast to other known WH domains, the side of the afHel308 domain that corresponds to the DNA binding face of typical WH domains shows a dominant distribution of negatively charged amino acids. This makes it unlikely that the afHel308 WH domain binds nucleic acids like other typical DNA binding winged-helix proteins (Figure 15). A comparable distribution of negatively charged amino acids along the WH domain is also found in archaeal TFE, consistent with the finding that archaeal TFE does not bind DNA *in vitro* (Meinhart *et al.,* 2003). Due to the location of domain 3, and the amount of surface exposed hydrophobic amino acids, a more relevant function in protein-protein interaction may be possible.

afHel308 WH   **RecQ WH**   **E2F-4 WH**   **TFE/TFIIEα**
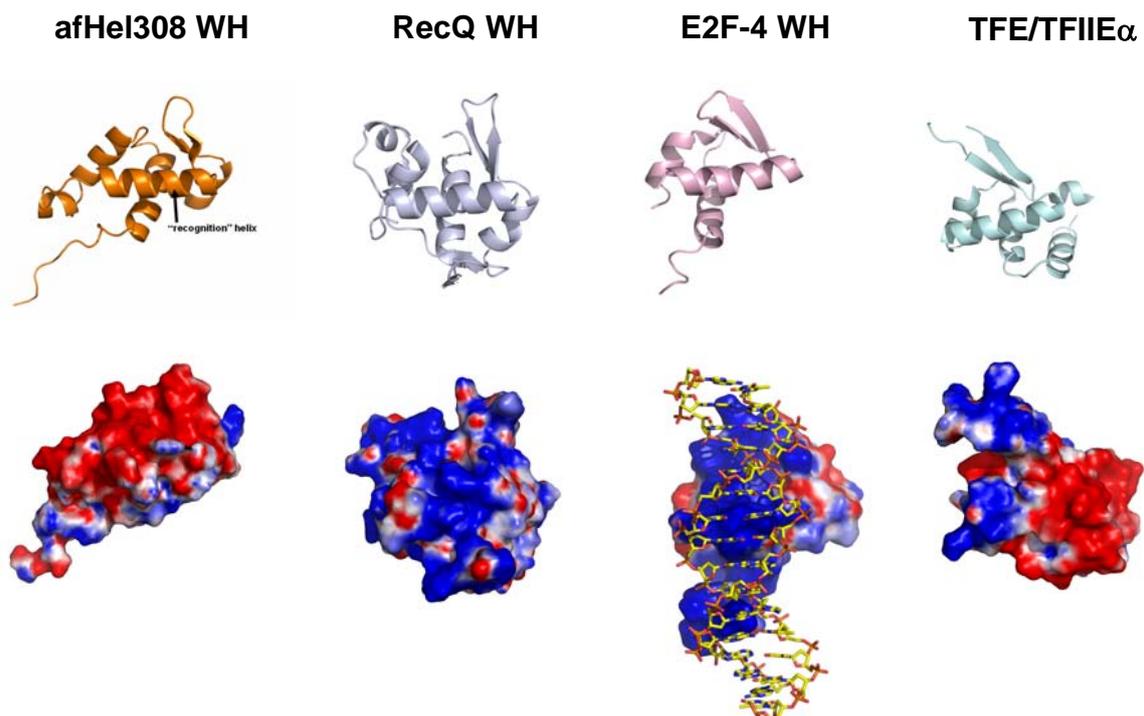
Figure 15: Comparison of winged-helix domains. At the top, a ribbon representation is shown. At the bottom, the molecular surface potential is shown (blue: +10kT/e⁻, red: -10kT/e⁻). E2F-4 is shown with bound DNA (yellow sticks). From left to right: afHel308 (this study), RecQ WH domain (PDB code 1OYW (Bernstein *et al.,* 2003), E2F-4 (PDB code 1CF7 (Zheng *et al.,* 1999), TFE/TFIIEα (PDB code 1Q1H (Meinhart *et al.,* 2003).

Domain 4 consists of a seven-helix bundle. A DALI search with the whole domain 4 and individual parts of domain 4 revealed no significant structural neighbors. Domain 4 is packed between domain 2 and domain 3 mainly via polar contacts and hydrophobic interactions. Together with domain 1, the three domains form a ring like structure, which gives the molecule its peculiar shape (Figure 15A). Because of the polar nature of contacts between domain 2 and domain 4 the interaction may be only transient. However, data from limited proteolysis experiments performed with three different proteases show similar digestion patterns of DNA-free and DNA-bound afHel308, indicating a comparable conformation in solution (see below). Domain 5 of afHel308 possesses a helix-loop-helix (HLH) motif and is situated at the periphery of the ring (Figure 14A). The HLH motif is a DNA binding domain, which is frequently found in DNA interacting proteins. It is also found in the archaeal XPF/Rad1 homolog Hef in which the HLH motif has been shown to be important for specific recognition of branched DNA structures (Nishino *et al.,* 2005; Nishino *et al.,* 2005). An additional DALI search revealed relevant structural similiarity to the thumb domain of eukaryotic polymerases of the Y-family like Pol κ (Uljon *et al.,* 2004), Pol ι (Nair *et al.,* 2006) and Rev1 (Nair *et al.,* 2005), and to the thumb domain of archaeal error-prone DNA polymerases of the DinB family (Silvian *et al.,* 2001) (Figure 16).
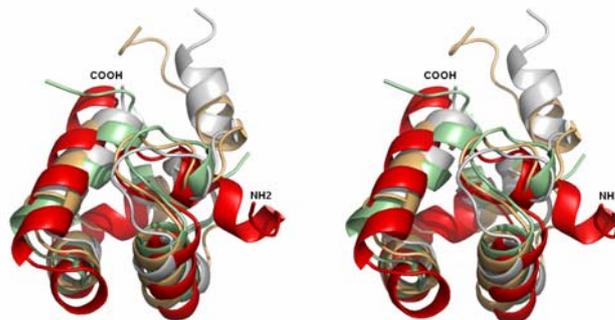

Figure 16: Stereo view of the superposition of "thumb" domain 5 (red, this study) of afHel308 with equivalent regions of Pol ι (lightgreen, PDB code 2ALZ, (Nair *et al.,* 2006)), Rev1 (wheat, PDB code 2AQ4, (Nair *et al.,* 2005)) and archaeal DinB (grey, PDB code 1K1Q, (Silvian *et al.,* 2001)

These polymerases are involved in translesion DNA synthesis (TLS) and are thought to act transiently at stalled replication forks (Friedberg and Gerlach 1999). The thumb domain is hereby involved in DNA binding and contributes to processivity (Zhou *et al.,* 2001). Although the sequence of the thumb domain between these different polymerases is conserved to a certain degree, there is no significant sequence conservation to the afHel308 thumb domain (Silvian *et al.,* 2001). However, the thumb domain seems to be conserved between afHel308 and *Drosophila melanogaster* Mus308 and human PolΘ (Figure 6) and may therefore be a unique feature of these helicases/polymerases.

3.3 Structure of the archaeal afHel308:DNA complex

3.3.1 Overall architecture

The overall structure of afHel308 in complex with DNA as well as the orientation of the five domains is very similar to that described for the DNA unbound afHel308 structure (Figure 17A). The exceptions are some slight movements of domain 5 and the ordering of some DNA and ATP binding motifs in the presence of DNA (Figure 17B). The single-stranded portion of the DNA substrate is threaded into the ring, which is formed by domains 1, 2, 3 and 4, whereas domain 5 binds the remote end of the 3′-tail. The single strand:double strand junction as well as some regions of the DNA duplex interact mainly with domain 2, whereas some contacts are contributed by domain 4. The surface of domain 2 contains a prominent β-hairpin loop that separates the strands of the substrate duplex. Remarkably, this β-hairpin loop already promoted melting of two base pairs in the absence of ATP. All five domains bind to the DNA substrate at various positions, whereby most of the nucleotides become visible in the electron density (Figure 17C). Additional contacts between the DNA duplex and a symmetry related molecule stabilize the double-stranded portion of the nucleic acid substrate (Figure 17D).
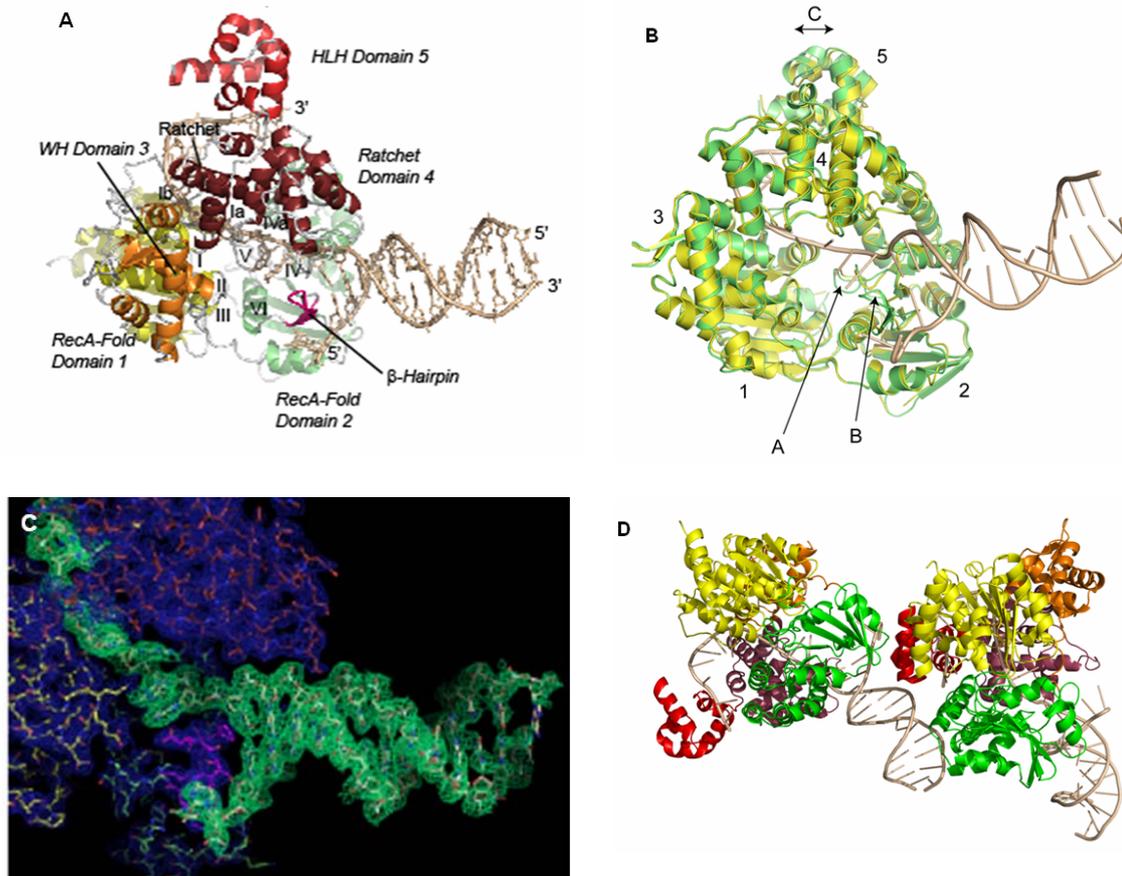
Figure 17: Ribbon model afHel308a in complex with the 15mer DNA duplex containing a 10mer single-strand 3' tail. Domains are colored according to Figure 15 and indicated. DNA is shown as brown stick model. Locations of sequence motifs are indicated. B) Comparison of apo and DNA bound afHel308. Overlay of apo (yellow) and DNA bound (green, with brown DNA) afHel308a, shown as ribbon models. R.m.s.d values between both structures were 0.88 Å for 684 Cα atoms. Domain numbers are indicated. DNA binding results in slight structural changes, including ordering of residues 331-335 (A: motif IVb), ordering of residues 350-356 (B: β-hairpin loop) and slight movement of domain 5. C) Composite simulated annealed omit electron density (0.8σ contour) shows how strands of the DNA duplex (green) are separated by afHel308 (blue). The final model is superimposed as color codes sticks. D) Two asymmetric units of afHel308:DNA complex. The double-stranded region of the DNA substrate interacts additionally with a symmetry related molecule. Interactions between DNA duplex of one afHel308 (left side) and a symmetry related Hel308 molecule are promoted mainly by Lys277 of domain 2 (green) and Arg672 of domain 5 (red) to the phosphate backbone of the product strand of the DNA duplex.

3.3.2 Interactions between afHel308 and its DNA substrate

All five domains of afHel308 bind to the DNA substrate at various positions (Figure 18A). The contacts with the substrate duplex and the single strand:double strand

junction are thereby predominantely mediated by interactions between domain 2 and the phosphate backbone. Additional contacts to the double-stranded region are contributed by domain 4. An important feature of domain 2 is the so called β-hairpin loop, which is located between motif V and VI of afHel308. Residues of this β-hairpin loop (R350, F351, Y354 and R357) intercalate between the last paired bases (-1) and the first unpaired bases (+1) on both strands of the duplex (Figure 18B). These interactions result in loss of base stacking and force separation of the two DNA strands in opposite directions, which prevents reannealing of the duplex. Two other residues of the β-hairpin loop, D352 and G353 bind the base at position +1 or facilitate formation of the hairpin turn, respectively. The structural results already argue that the energy required for duplex separation is provided by the free energy released by binding of the nucleic acid substrate to the enzyme and not by binding of ATP. This is consistent with thermodynamic data on the HCV NS3 helicase (Levin et al., 2005). ATP independent strand separation of several base pairs has also been observed for RecBCD, albeit at a structurally distinct region, and could be a more widespread feature of helicases (Singleton et al., 2004).
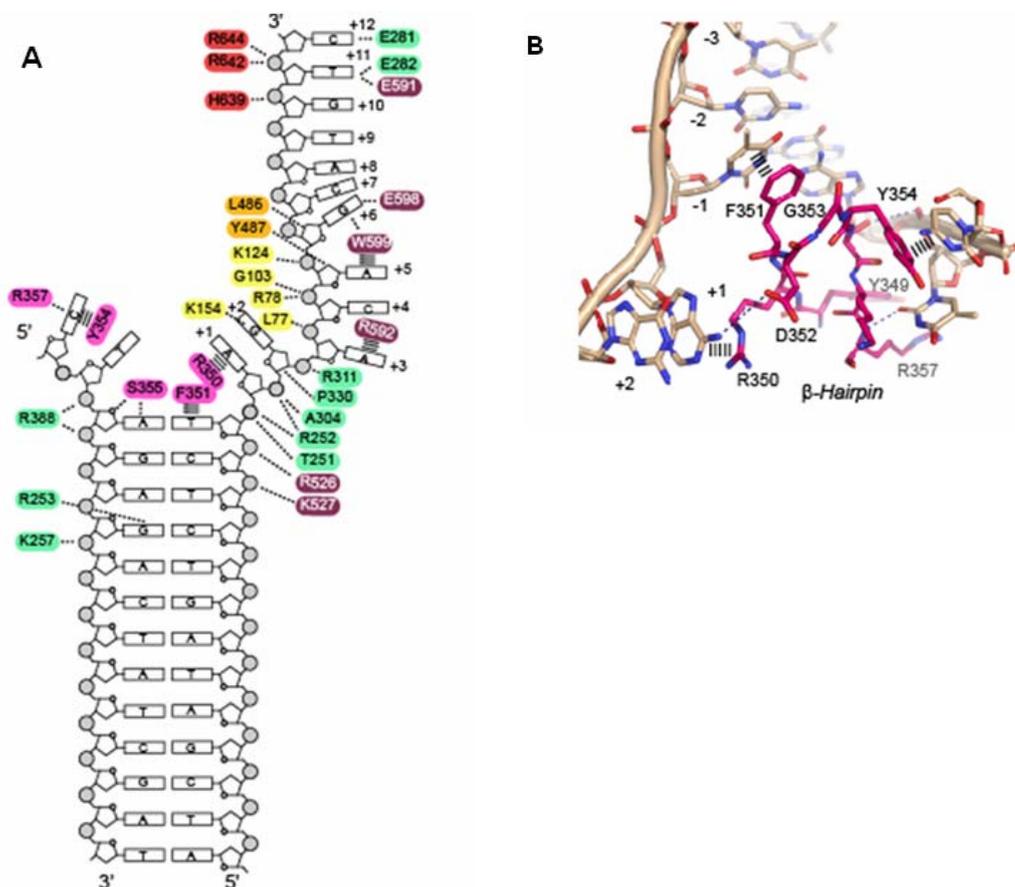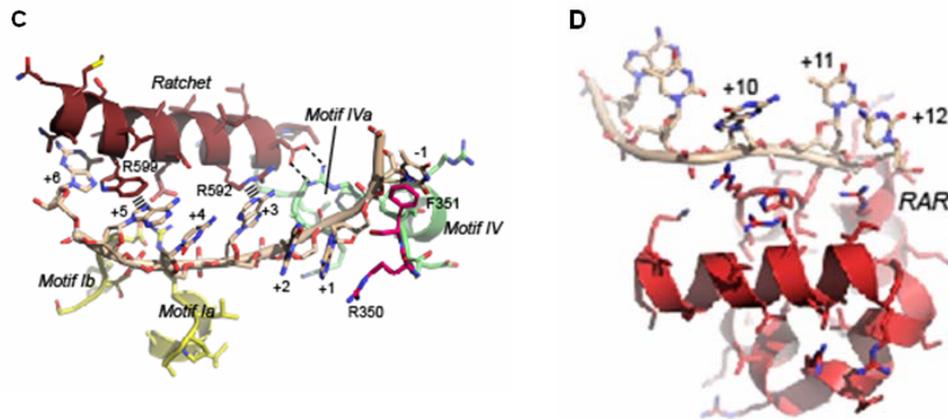
Figure 18: A) Scheme showing key interactions of afHel308 with the partially unwound DNA substrate. Residues are colored according to domain affiliation (magenta: β-hairpin motif). Interactions are shown by dashed lines. Key stacking interactions of side chains with bases are shown by stripes. B) The β-hairpin motif (magenta stick model with blue nitrogens and red oxygens) stabilizes the unwound duplex. DNA strands (brown stick model) are separated by disruption of base stacking in both strands. Key residues and interactions are annotated and highlighted. C) 3' tail binding by the translocation module. Domain 4 (dark red) stacks with bases at position +3 (R592) and +5 (R599). Two residues of the β-hairpin motif (magenta) and motif IV (green) are indicated. D) The helix-loop-helix domain 5 binds the backbone of the base positions +10, +11, +12 by a cluster of positively charged residues, including the Hel308/PolΘ/Mus308 conserved RAR motif.

The backbone of the unwound 3'-tail is bound at motifs IV (+1), IVa and IVb (+2) on domain 2 and motifs Ia (+4) and Ib (+5) on domain 1 (Figure 18C), consistent with an inchworm like transport of DNA by ATP-dependent conformational changes between domains 1 and 2. In this model, both domain 1 and domain 2 would function in the role of two independently acting nucleic acid binding regions, allowing translocation on DNA like an inchworm (Figure 3A). The binding of the single-stranded tail to the mentioned motifs is additionally stabilized by opposing base interactions with domain 4. The position of domain 4 and the obvious interactions with the single-stranded region indicate an involvement in the directional transport of the product tail across domains 1 and 2 (Figure 18C). A central helix in domain 4 possesses two residues, R592 and W599, which stack with base moieties at positions +3 (R592) and +5 (W599), opposing to motifs Ia and Ib on domain 1. The N-terminal end of this helix additionally interacts with motif IVa on domain 2. Thus, the central helix of domain 4 provides an ideal ratchet, which position is structurally linked to ATP dependent conformational changes in domain 2, and may directly influence the geometry of the single strand

binding between domains 1 and 4. In addition, the base stacking interactions could prevent backslipping of the enzyme on the DNA substrate. After emerging from the tunnel formed by domains 1, 3 and 4, the 3'-tail bends around domain 4 (bases +6 to +9) with apparently nonspecific contacts, and binds to the HLH domain 5 via the phosphate-sugar backbone of base positions +10, +11 and +12 (Figure 18D). A RAR (Arg-Ala-Arg) motif in domain 5 is conserved among several related helicases (POLΘ, Mus308, hHEL308), implicating that similar types of interactions are relevant for the functions of the eukaryotic enzymes (Figure 6). Altogether, the encircling of DNA between domain 1, 3 and 4 is likely important for translocation processivity, because the non-translocating DEAD box SF2 lack equivalents of domains 3 and 4 helicases (Andersen *et al.,* 2006; Bono *et al.,* 2006; Sengoku *et al.,* 2006). Consistent with the presented structural data, removal of domains 3 to 5 results in uncoupling of ATPase and helicase activities (see below).

3.3.3 Truncation of the C-terminal part results in loss of DNA unwinding activity but does not affect DNA dependent ATPase activity

To investigate a possible role of the C-terminal domains in coupling ATP hydrolysis to translocation and in substrate recognition, wild-type and mutant afHel308 were tested for ATPase and helicase activity.

ATPase activity of afHel308 is greatly stimulated by single-stranded (ss) DNA and to a much lesser extent by double-stranded (ds) DNA. All branched DNA substrates tested stimulate ATPase activity of wild-type afHel308 to a similar extent (Figure 19A). No significant activity is observed without DNA. This is consistent with observations made for pfuHel308 and mthHel308, hHel308 and also RecQ, assuming that ATP hydrolysis is in general stimulated by ssDNA and ss:ds transition occurring in branched DNA substrates (Bernstein *et al.,* 2003; Hishida *et al.,* 2004; Fujikane *et al.,* 2005; Guy and Bolt 2005). The ATPase activity of purified afHel308 Walker B variant E146Q is reduced to below 5 % of wild-type activity under the given conditions (Figure 19B,C,D). The truncated protein Δ626 exhibits ATPase activity comparable to the wild-type protein on all tested branched DNA substrates. In contrast, the truncated construct Δ400 is significantly more stimulated by the branched substrates (Figure 19C,D). On the other side, the stimulation of ATPase activity by ssDNA alone is reduced to around

40% of wild-type activity in the Δ626 mutant (Figure 20B), while the activity of truncated Δ400 is affected even worse. The reduction of activity in the truncated Δ400 and Δ626 proteins in the presence of ssDNA compared to the wild-type can be explained by the lack of additional ssDNA binding regions. This finding is consistent with the structural data, in which the C-terminal domains are mainly involved in the binding of the ssDNA portion, while the ds region is bound predominantely by domain 2 (Figure 18A).
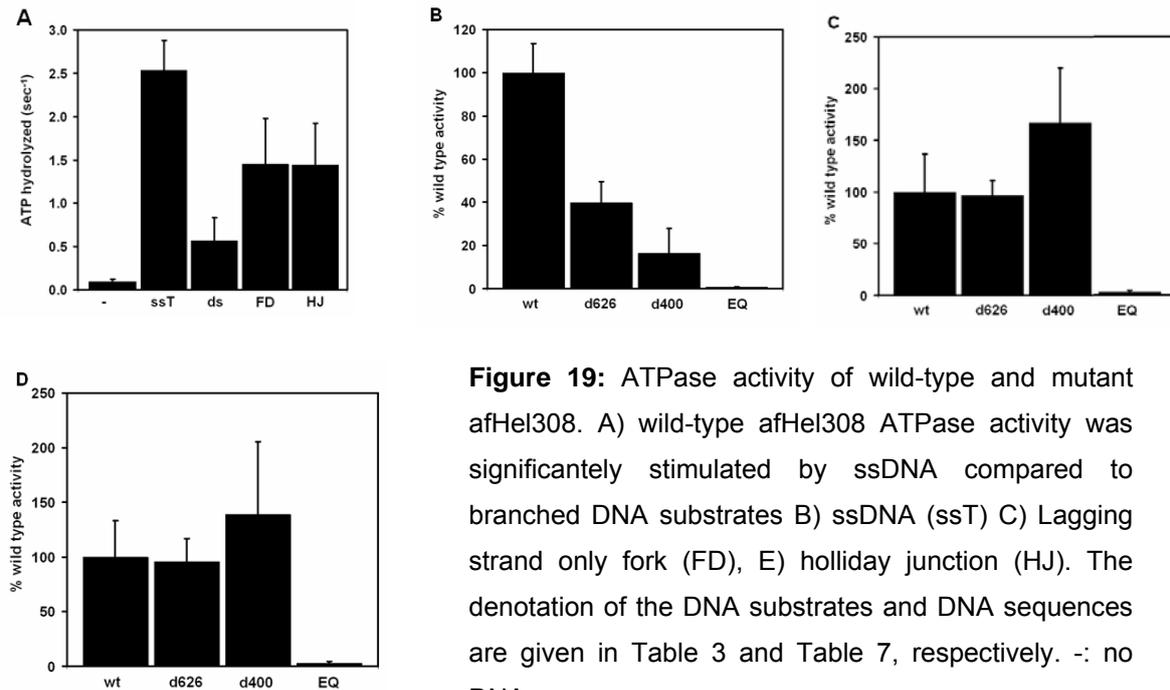

Figure 19: ATPase activity of wild-type and mutant afHel308. A) wild-type afHel308 ATPase activity was significantely stimulated by ssDNA compared to branched DNA substrates B) ssDNA (ssT) C) Lagging strand only fork (FD), E) holliday junction (HJ). The denotation of the DNA substrates and DNA sequences are given in Table 3 and Table 7, respectively. -: no DNA.

Helicase activity of afHel308 wild-type and mutant proteins was analyzed using a synthetic Holliday junction (Figure 20). As expected, the E146Q mutant exhibits no helicase activity, consisting with ATP-binding and -hydrolysis dependent unwinding activity. The Δ626 protein possesses helicase activity, although the activity is reduced compared to wild-type activity. HLH domains are generally implicated in the binding of dsDNA and in the recognition of branched molecules (Nishino *et al.,* 2005). Thus, deletion might influence specificity to as well as unwinding processivity of specific branched DNA substrates. As shown, the Δ400 protein exhibits no helicase activity (Figure 20). The truncated protein is still able to interact with DNA, which stimulates ATPase activity. Because of the lacking domains (especially domain 4), which are implicated in coupling the ATP hydrolysis to translocation and thereby promoting strand separation, no unwinding is observed. The ATPase activity of Δ400 is therefore uncoupled from possibly occurring unwinding reactions, which could result in enhanced ATPase activity.
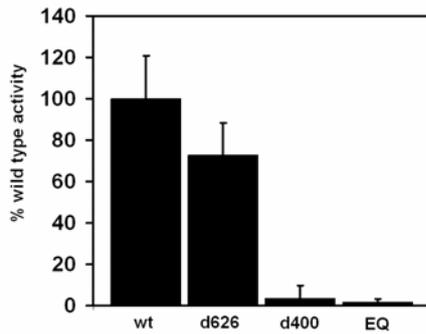
Figure 20: Comparison of unwinding activities of holliday junction between wild-type and mutant enzymes.

3.3.4 Limited proteolysis experiments reveal only small conformational changes during ATP- and DNA-binding

DNA unwinding and translocation activity of helicases require binding of ATP, which is accompanied by a closure of domains 1 and 2. This closure implies rotation of one domain towards the other, and a conformational change that in addition drives translocation. Based on the presented crystal structure of afHel308, the rotation of domain 2 towards domain 1 upon ATP binding would lead to only minor conformational changes in the protein on the whole. In order to support this idea without the presence of an ATP-bound structure, limited proteolysis experiments of afHel308 were performed in the presence and absence of the non-hydrolyzable ATP-analogue AMP-PNP and the DNA used for crystallization. Proteolysis was carried out with different amounts of three different proteases as described in 2.2.3. As seen in Figure 21, no significant changes in the proteolysis pattern could be observed in the presence of DNA as a function of AMP-PNP. However, small changes in the relative strength of single bands in the pattern of the proteinase K digestion could be observed in the presence of DNA (Figure 21A). Here, the binding of DNA may lead to protection, ordering or rearrangements of protein loops. These slight alterations in the relative strength of bands can also be observed in the presence of AMP-PNP. All in all, these results support at model in which only slight conformational and structural changes occur between domains 1 and 2 upon ATP-binding.
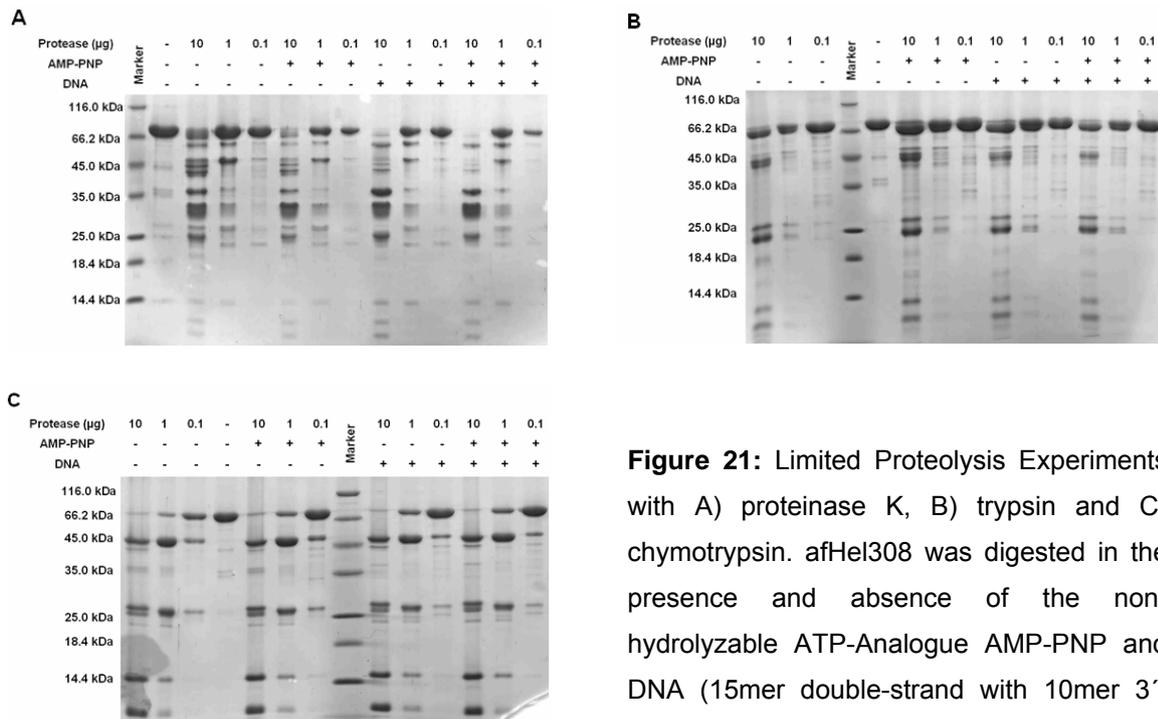
Figure 21: Limited Proteolysis Experiments with A) proteinase K, B) trypsin and C) chymotrypsin. afHel308 was digested in the presence and absence of the non-hydrolyzable ATP-Analogue AMP-PNP and DNA (15mer double-strand with 10mer 3´-overhang).

At the same time, it supports the similarity of apo and DNA-bound structure. Any flexibility in the C-terminal domain in solution would be altered upon DNA binding, as seen in the crystal structure. This would result in dramatic changes in the digestion pattern as more amino acids are accessible to the proteases. As this is not the case, the apo crystal structure seems to exhibit a conformation also adapted in solution in the absence of DNA.

4 Discussion

4.1 The crystal structure of archaeal afHel308 gives insights into the unwinding mechanism of SF2 helicases

The crystal structure of oligonucleotide bound archaeal afHel308 is the first crystal structure of a SF2 helicase in complex with a duplex DNA containing an additional single-stranded 3′-overhang. It reveals distinct interactions between the enzyme and the nucleic acid, which are contributed by the conserved helicase motifs found also in other helicases of the SF2 family. Thus, the structure provides a model for DNA duplex separation, which is applicable to a broader range of SF2 helicases, like the related Mus308 and Ski2 family. In addition, the structural data reveal how afHel308 specifically might interact with the lagging strand on a stalled replication fork.

4.1.1 Comparison of afHel308 with ATP-bound structures of SF2 helicases implicates translocation of the DNA substrate by ATP-dependent closure of the two RecA domains

afHel308 was crystallized in the absence of a bound nucleotide in both DNA bound and apo protein. Although melting of two base pairs without binding of ATP was observed, processive unwinding by 3′→ 5′ translocation requires ATP binding and hydrolysis. According to an inchworm like base pair separation mechanism, the binding of ATP is believed to induce a closure of the two RecA-like domains, resulting in conformational changes in the nucleic acid binding domains and base separation at the ss:ds junction (Figure 3A). ATP hydrolysis and release of the products drive the enzyme back into the previous conformation, a process which is also associated with changes in the affinity to the nucleic acid substrate and translocation.

The ATP bound form of a helicase is generally designated as "closed" form. Thus, our DNA bound structure should reflect the ATP-free state during processive translocation, which can be regarded as "open" conformation. As no ATP-bound structure of archaeal Hel308 is available, afHel308 was compared with HCV NS3 helicase (Kim *et al.,* 1998) and other recently determined ATP-bound structures of SF2 DEAD box RNA helicases (Andersen *et al.,* 2006; Bono *et al.,* 2006; Sengoku

et al., 2006) to learn more about ATP-dependent translocation. In the structure presented in this work, the 3´-tail of the nucleic acid is embedded between domain 1 and domain 2 on one side, and domain 4 on the other side. The phosphates of the backbone bound at motif Ia (position +4) on domain 1 and motif IV on domain 2 (position +1) are separated by two intermediary bases (Figure 22). This two-base-separation is also observed in the ATP-free structure of NS3 helicase bound to a RNA single-strand (Kim *et al.,* 1998).
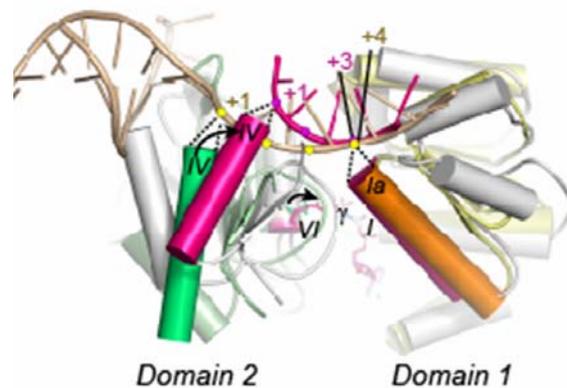

Figure 22: Comparison of afHel308 with the ATP bound structure of the DEAD box RNA helicase VASA. Superposition of domains 1 (afHel308: yellow with orange motif 1a helix; VASA: grey with magenta 1a helix) shows that domains 2 (afHel308: green with dark green motif IV helix; NS3: grey with magenta motif IV helix) rotate app. 20° (arrows) between apo and ATP bound SF2 enzymes. A candidate for this ATP induced structural change is an invariant arginine in motif VI (shown as green (afHel308) and magenta (VASA) sticks) that binds to the γ-phosphate of ATP (γ) in VASA and is located close to the expected ATP binding site in afHel308. In the afHel308 structure, motif IV binds phosphate position +1, while motif Ia binds phosphate position +4. In the ATP bound structure of VASA, motifs IV and Ia bind phosphates at corresponding position +1 and +3, respectively. This difference argues that ATP induced conformational change slips the 3' tail across motif Ia and in conjunction with ATP hydrolysis translocates the enzymes by one base.

The comparison between the described ATP-free structure with three independent structures of SF2 DEAD box enzymes in complex with ATP analogs and single-strand nucleic acids reveals that domain 2 is shifted appr. 20° towards domain 1 (Figure 22). This movement implicates an ATP-binding induced closure of domains 1 and 2. A closer look at the interactions with the nucleic acid further reveals, that the phosphates bound at motifs Ia and IV are now separated by only

one intermediary base (Andersen *et al.,* 2006; Bono *et al.,* 2006; Sengoku *et al.,* 2006).

Because of the similarity of the three dimensional fold of the compared structures in either ATP-free or ATP-bound state, and due to the conservation of the helicase motifs in SF2 members, it is expected that the observed movement of domain 2 and the pairwise match of phosphate distances is not coincidental. Thus, the difference of one base between ATP-bound and ATP-free states could be due to an ATP-binding induced conformational change that causes slippage of a single nucleotide across domain 1. As seen in the AMP-PNP bound structure of VASA, an invariant arginine in motif VI is involved in the binding of the γ-phosphate of the nucleotide (Sengoku *et al.,* 2006). Interestingly, mutation of the equivalent arginine in eIF4A and NS3 helicase strongly decreases ATPase activity (Pause and Sonenberg 1992; Sampath *et al.,* 2006). This arginine is supposed to contribute to ATP hydrolysis in a manner similar to that of the "arginine finger" of several GTPase activation proteins (Rittinger *et al.,* 1997; Scheffzek *et al.,* 1997) by stabilizing the transition state of the reaction and by triggering conformational changes after NTP hydrolysis. In the structure of afHel308, this arginine (R369) is located close to the expected ATP binding site in domain 2. Thus, the ATP driven conformational closure of domains 1 and 2 could be triggered by binding of the conserved arginine to the γ-phosphate of ATP (Figure 22). Based on the crystal structure of apo and DNA bound afHel308, the closing of domain 1 and 2 upon ATP binding should result in only small conformational changes. This is supported by the proteolysis experiments in the presence and absence of the non-hydrolyzable ATP-analogue AMP-PNP and DNA (Figure 22).

4.1.2 The suggested interplay between domain 1 and domain 2 and the C-terminal part upon binding and hydrolysis of ATP is consistent with the Brownian ratchet model proposed for HCV NS3 helicase

How does an ATP-induced conformational change lead to directed nucleic acid translocation and processive unwinding? Based on the described structural data a simple model for unwinding by SF2 helicase afHel308 can be predicted, which is applicable to other SF2 enzymes as well. As shown in Figure 23, DNA translocation is initiated when the protein binds to ssDNA via interactions in both

domain 1 and 2. ATP binding at the interface cleft of domain 1 and domain 2 may result in a structural change and closing of the cleft. Domain 2 thereby pushes the DNA via motif IV towards domain 1. This movement is promoted by a push of domain 2 via direct contacts with the ratchet helix. Since the ratchet helix in domain 4 base stacks with bases of the 3'-tail strand (+3 and +5), the combined movement of domains 2 and 4 may loosen the nucleic acid binding at domains 1, likely facilitating slippage of the 3' tail backbone across motif Ia. The modulation of stacking interactions between the ratchet helix and the nucleic acid upon changes in the nucleotide ligation state is consistent with biophysical data showing that single-strand binding of the related HCV NS3 helicase is weakend upon ATP binding (Levin *et al.,* 2003). ATP hydrolysis and product release reset the orientation of domain 2, thereby enabling the corresponding β-hairpin loop to melt another base pair of the duplex. In the ATP-free state domain 1 and domain 4 can form new tight interactions with the 3´-tail.
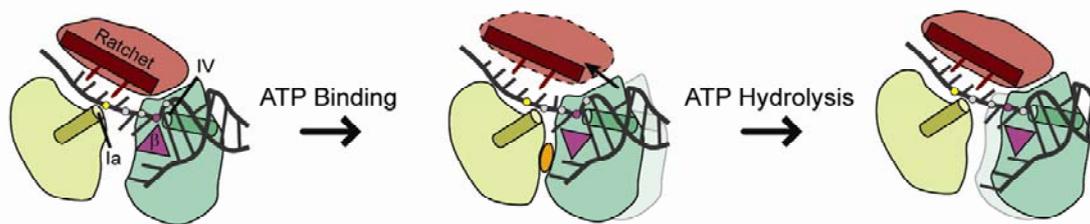

Figure 23: Schematic two step model for processive unwinding by Hel308a. Left and right are simplified schemes of the crystallographically observed conformations of Hel308. The middle scheme is modeled analog to ATP-bound SF2 enzymes (see Fig. 3a). Domains 3 and 5 are omitted for simplicity. Motifs Ia (yellow), IV (green), β-hairpin (magenta triangle) and the ratchet helix (red) are highlighted.

Since domain 2 is capable to melt at least two base pairs without ATP binding or hydrolysis, an additional ATP binding/hydrolysis cycle may in principle lead to rebinding of domain 2 farther upstream and therefore to larger step sizes. This idea may explain why 2-3 base step sizes are observed in single molecule experiments to probe helicase activity of NS3 helicase (Levin *et al.,* 2005). However, these structural observations are also in excellent agreement with thermodynamic data on HCV NS3, which suggest the transport of nucleic acids via a Brownian ratchet (Kim *et al.,* 1998; Levin *et al.,* 2005). HCV NS3 helicase also contains two conserved residues, V432 and W501, which are base stacking with the last and intermediary base of the bound DNA oligonucleotide. These residues

are also proposed to function as a ratchet and to modulate translocation upon ATP binding induced closure of domains 1 and 2, whereby W501 exhibit additional function in substrate specificity (Kim *et al.,* 2003).

The model predicts that removal of domains 3 to 5 might still allow DNA dependent stimulation of ATP-hydrolysis (via interactions with motifs on domain 1 and 2), but should lead to a loss of processive 3' → 5' translocation and therefore processive unwinding. Consistently, truncated afHel308 Δ400 (deletion of domains 3 to 5) still possesses DNA stimulated ATPase activity, but lacks helicase activity (see 3.3.3).

4.1.3 The β-hairpin loop is a widespread feature among SF2 helicases and implicates a conserved unwinding mechanism

The DNA bound structure of afHel308 suggests that separation of the DNA duplex is promoted by the prominent β-hairpin loop between motifs V and VI, which was not ordered in the crystal structure of apo afHel308 (Figure 19B) Two aromatic residues, F351 and Y354, of this β-hairpin loop wedge between the first base pairs at the ss-ds junction and promote melting of two base pairs in the absence of ATP (Figure 18C). The position and the obvious interaction with the unwound DNA strand thereby implicate a direct role in DNA unwinding. As this newly identified β-hairpin loop would be an elegant tool for the enzyme to destabilize the duplex DNA by stabilizing the nascent single strand during processive translocation, it could be a more widespread feature among SF2 helicases. A comparison of the primary sequences of archaeal afHel308 and other relevant SF2 helicases with focus on the presence of similar insertions between motifs V and VI is shown in Table 10. DEAD box helicases like VASA, SWI2/SNF2 enyzmes and RecG evidently lack such a β-hairpin loop. This is not only seen in the three dimensional structures of these enzymes, which do not reveal such insertion at this position (Singleton *et al.,* 2001; Caruthers and McKay 2002; Durr *et al.,* 2005), but is also consistent with the idea that latter enzymes either locally bend RNA duplexes, or translocate on DNA duplexes without strand separation. RecQ helicases also do not possess an equivalent β-hairpin loop and since no DNA-bound structure available, the unwinding mechanism of the SF2 RecQ like helicases remains unclear.

Table 10: Comparison of the sequence between motif V and Vi in a subset of SF2 helicases

	Motif V		**β-hairpin motif**		**Motif VI**
afHel308	GVNLPARRV	6	RFDGYSKE	7	QMAGRAGRR
yMtr4	GLNMPAKTV	6	KWDGQQFR	8	QMSGRAGRR
ySki2	GLNLPTRTVI	5	KHDGNGLR	8	QMAGRAGRR
HCV NS3	GFTGDFDSVI	3	TXVTQTVDFSLDPTFTIETTTLP	7	QRRGRTGRG
yeIF4A	GIDVQQVSLVI	5		6	HRIGRGGRFG
dmVASA	GLDIKNIKHVI	5		7	HRIGRTGRVGN
EcRecG	ELHLLVATTVI	12		14	QLRGRVGRG
SsoRad54	GINLTSANRVI	6-		6	QATDRVYR

As mentioned previously, afHel308 shares sequence similarity not only with the Mus308 family of helicases, but also with the RNA decay factors Ski2p and Mtr4p (Koonin *et al.,* 2001; Houseley *et al.,* 2006). Comparison of the region between motifs V and VI revealed that members of the Ski2 helicase family could possess an equivalent β-hairpin loop (Figure 6 and Table 10). This conservation argues that this peculiar appendix could separate RNA secondary structures or RNA:protein complexes in RNA decay processes in a way similar to that observed for archaeal afHel308. In addition, the HCV NS3 helicase and its homolog from yellow fever virus contains an equivalent β-hairpin loop between motifs V and VI (Kim *et al.,* 1998; Wu *et al.,* 2005) (Table 10 and Figure 24B).

Although HCV NS3 is one of the most intensely studied SF2 helicases, the mechanism of strand separation by NS3 is still unresolved. As no structure of NS3 bound to a ss:ds junction DNA is available, it is still not fully clear which motifs are involved in the separation of the two strands of the duplex. However, mutations in the NS3 β-hairpin loop (also denoted "Phe-loop") either deprive NS3 from all activities or specifically abolish duplex unwinding but not nucleic acid stimulated ATPase activity (Lam *et al.,* 2003). Furthermore, recent biochemical data showed that HCV NS3 also melts several base pairs of the duplex in the absence of ATP (Levin *et al.,* 2005). Thus, the evident structural and biochemical similarity of NS3 and afHel308 suggest that the β-hairpin is the sought after strand separation element of NS3. The data presented here now suggest that the afHel308 derived

mechanism for duplex destabilization is applicable to a broader range of SF2 helicases.

4.1.4 The proposed unwinding mechanism of archaeal afHel308 differs from unwinding by SF1 helicases and RecG

Interestingly, the presence of a "separation pin", albeit at different positions, is also seen in SF1 helicases like PcrA, UvrD and the RecBCD helicase (Velankar *et al.,* 1999; Singleton *et al.,* 2004; Lee and Yang 2006). The function of this "separation pin" is to some extent comparable to the function of the "wedge" domain found in RecG (Singleton *et al.,* 2001). However, the unwinding mechanism in these enzymes is different from the one proposed for archaeal afHel308.

As shown in Figure 24D, in the structure of the UvrD:DNA complex, the product 3´-tail and the substrate duplex region are bound at a right angle (Lee and Yang 2006). As proposed for the wrench-and-inchworm model for UvrD (Lee and Yang 2006) (Figure 3), ATP binding induces closure of the cleft between domains 1A and 2A and a conformational change in domains 1B and 2B, which is associated with an alteration in the relative DNA binding affinities. Rotation of the bound DNA helix is promoted by domains 1B and 2B and results in unwinding of one base pair at the ss:ds junction, forced by the presence of a separation pin after motif VI. This β-hairpin buttresses the end of the DNA duplex, thereby preventing reannealing. ATP hydrolysis results in resetting of the domains and unidirectional translocation along DNA. By this, the 3´-tail bound to UvrD appears to be peeled off the duplex (Velankar *et al.,* 1999; Lee and Yang 2006; Xie 2006).

Compared to other SF1 and SF2 helicases, the fork reversal enzyme RecG exhibits a unique domain orientation. Additional to the two RecA-like folds it possesses a large N-terminal domain (Figure 24C). This domain is denoted as "wedge" domain, and participates in the separation of the nascent DNA strands on both lagging and leading strand template at a stalled replication fork (Singleton *et al.,* 2001). The ATPase acitivity of RecG has been shown to depend mainly on dsDNA, promoting the idea that RecG is a ds translocase (Whitby and Lloyd 1998). Upon ATP binding and hydrolysis, the two RecA-like domains 1 and 2 are closing and opening, thereby changing their affinity to dsDNA. This allows translocation of the enzyme on the template duplex. The N-terminal wedge domain functions as steric hindrance on which the duplex strands on both leading and

lagging strand are separated. The unwound complementary nascent DNA strands subsequently anneal because of their proximity to each other, and form the "chicken foot" structure, which is a characteristic feature of fork reversal (Singleton *et al.,* 2001).
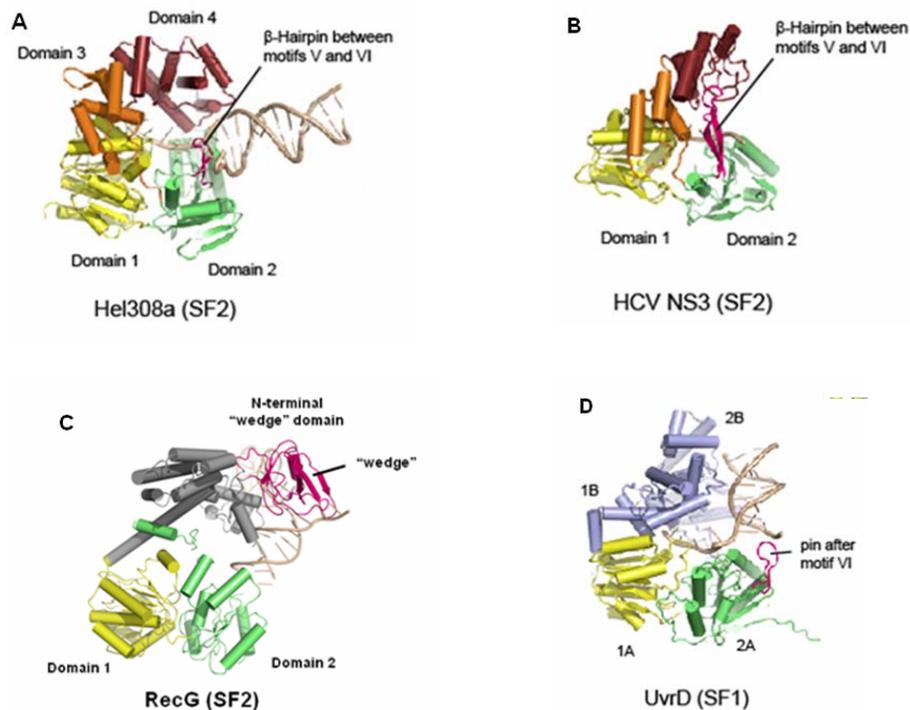

Figure 24: Comparison of SF2 helicase A) afHel308 (in complex with an unwound tailed duplex) to the SF2 helicase B) NS3 of the hepatitis C virus (Kim *et al.,* 1998) (in complex with a single strand) and C) RecG (Singleton *et al.,* 2001) (in complex with a three-way junction) and to the SF1 helicases D) UvrD (Lee and Yang 2006) (in complex with a tailed duplex). afHel308 and NS3 share related domains 3 (orange) and 4 (dark red) as well as a β-hairpin motif between motifs V and IV. In contrast, the double strand translocase RecG exhibit an N-terminal "wedge" domain, which functions in the unwinding of the leading and lagging strand duplex at stalled replication forks. The SF1 helicases UvrD possesses a different domain structure with 1B and 2B insertions, and lack the β-hairpin motif between motif V and VI (Singleton *et al.,* 2004; Lee and Yang 2006).

In contrast to the DNA bound structure of UvrD, the crystal structure of DNA bound afHel308 shows that the duplex portion of the DNA and the 3´- 5´ substrate single-strand are almost collinear (Figure 24A). The rotation of domain 2 relative to domain 1 upon ATP binding seems to be the driving force of translocation along the single strand. The β-hairpin thereby appears to be rather pushed through the duplex to displace the 5´- 3´ strand, resulting from a conformational change and

cleft opening during ATP hydrolysis and product release. Due to striking structural and biochemical similarity, the proposed mechanism may be also applicable to HCV NS3 helicase (Figure 24B). Furthermore, this mechanism is very versatile and could also be used to strip proteins from single-strands (Jankowsky *et al.,* 2001).

4.2 The interaction of the DNA substrate with afHel308 shows how the enzyme might interact with branched nucleic acids

Archaeal Hel308 is proposed to act on stalled replication forks in a manner similar to RecQ (Fujikane *et al.,* 2005; Guy and Bolt 2005). Consistent with that, archaeal Hel308 was shown to unwind a variety of different branched DNA structures in $3´ \rightarrow 5´$ direction. Based on the DNA bound crystal structure presented here in which the 3´-tail of the DNA is unexpectedly bound to domain 5, one can propose a simple model of how afHel308 recognizes branched nucleic acids, in particular its preferred *in vitro* substrates in form of replication forks or other branched DNA structures (Figure 25). In this model, domain 5 may bind the single strand in front of the leading strand duplex, positioning afHel308 to unwind the lagging strand. After loading and translocation along the lagging strand template in $3´ \rightarrow 5´$ direction, the nascent unwound lagging strand may replace the leading strand, resulting in the crystallographically observed structure in complex with DNA. The displacement of the nascent lagging strand would allow the loading of proteins required to restart DNA replication like primase, replicative helicase or replication polymerases and in an alternative pathway the induction of an archaeal SOS response due to the exposed ssDNA.
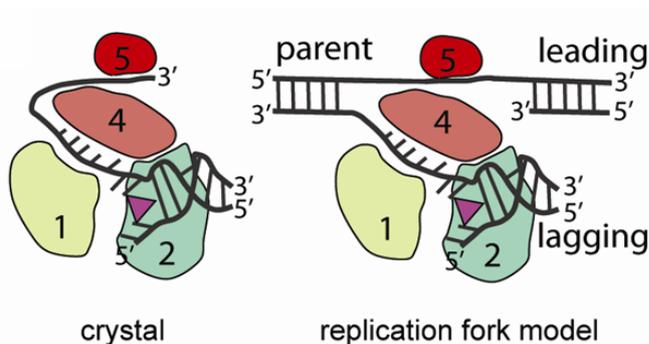
**Figure 25:** Model for the interaction of afHel308 with a replication fork. Left panel**:** schematic drawing of DNA bound afHel308 (color code of Fig. 16; domain 3 is omitted for simplicity). Right panel: model for the interaction of afHel308 with a replication fork.

In this model, domain 5 could have an important function in substrate recognition and binding. This is consistent with biochemical results showing that the helicase

activity of the truncated Δ626 protein is reduced on HJ substrate (see 3.3.3). The idea is supported to some extent by the unexpected interaction between the HLH domain of one afHel308 molecule and the dsDNA portion of a symmetry related afHel308 molecule as observed in the crystal packing (Figure 19D). To further analyze the possible interaction between the HLH domain of afHel308 with dsDNA, domain 5 was superimposed to the thumb domain of the translesion polymerase Rev1 (Nair *et al.,* 2005) (Figure 26C). As seen in the crystal structure of Rev1, the dsDNA is bound mainly via interactions with the helices corresponding to helix α30, α31 and α33 in afHel308 (Figure 26A,B). These helices interact with the ssDNA in afHel308, supporting the finding that the interaction is mainly backbone mediated and unspecific. However, comparison of the two DNA bound HLH domains with DNA derived from a symmetry related molecule in the crystal packing of afHel308 show that it would be possible, that the HLH domain additionally interacts with dsDNA of a branched DNA substrate like a Holliday junction intermediate (Figure 26C).
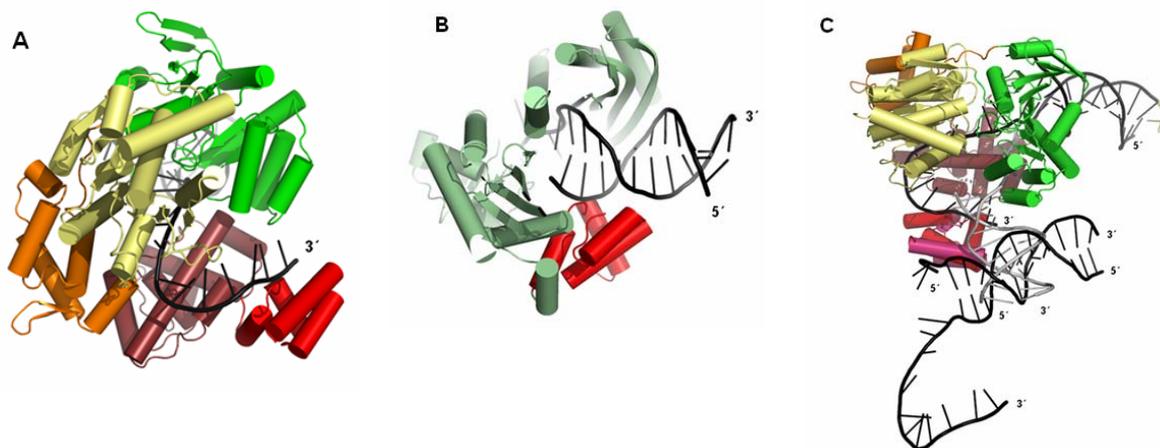
Figure 26: Comparison of the HLH domain of afHel308 (A) and Polymerase Rev1 ((B) PDB code 2AQ4, without residues 600-620). Overlay of the two HLH domains (B) red: afHel308; magenta: 2AQ4, (Nair *et al.,* 2005)) and the respective oligonucleotide raise the possibility that the afHel308 HLH domain is also involved in the recognition and binding of double-stranded or branched DNA structures (C).

All in all, the structural and biochemical data presented here exhibit high consistency with the biochemical properties observed for the archaeal homologs from *Pyrococcus furiosus* and *Methanothermobacter thermoautotrophicus*, which were found to partially complement RecQ in *E. coli* (Fujikane *et al.,* 2005; Guy and Bolt 2005). The DNA:bound structure of afHel308 can explain the interaction with branched DNA substrates, confirming the RecQ-like role of Mus308 family

helicases in the repair of stalled replication forks. The position of the HLH domain 5 thereby would be ideal to allow interaction with PCNA and/or translesion/replicative polymerases. Of course, the interaction with other important proteins found to be involved in these processes like endo- and exonucleases, primase and polymerase has to be elucidated.

Due to the overall homology between the helicase domain of eukaryotic and archaeal Hel308 members it is possible, that they recognize similar DNA structures *in vivo*, consistent with their proposed role in DNA repair and/or recombination.

5 Summary

Adenosine triphosphate (ATP) dependent nucleic acid unwinding by superfamily 2 (SF2) helicases is required for numerous biological processes, including DNA recombination, RNA decay and viral replication. The structural and molecular mechanism for processive duplex unwinding of SF2 helicases is still unclear, in part due to a lack of structural insights into the actual strand separation reaction.

Archaeal SF2 helicase Hel308 preferentially unwinds lagging strands at replication forks and is closely sequence related to human PolΘ and Hel308 as well as *Drosophila* Mus308. Furthermore, the RecA ATPase-core of archaeal Hel308 shares high sequence conservation to the SF2 RNA decay factors Ski2p and Mtr4p. Thus, archaeal Hel308 appears as representative model to understand processive 3´→ 5´ DNA unwinding by SF2 helicases.

During this PhD thesis crystal structures of *Archaeogloubs fulgidus* Hel308 (afHel308) in the absence and presence of a 15mer duplex DNA containing a 10mer 3´-overhang were determinded using X-ray crystallography. afHel308 exhibits two typical SF2 RecA-like domains at the N-terminus. The C-terminus comprises a winged-helix (WH) domain, followed by a unique seven-helix-bundle domain and a helix-loop-helix (HLH) domain. The DNA bound structure captures the initial duplex separation and argues that initial strand separation does not require ATP binding. Comparison with ATP bound SF2 enzymes suggests that ATP binding and hydrolysis promotes processive unwinding of one base pair by a ratchet like transport of the 3' product strand. In addition, the structure suggests that unwinding is promoted by a prominent β-hairpin loop. The identification of similar β-hairpin loops in Hepatitis C virus (HCV) NS3 helicase and RNA decay factors Ski2p and Mtr4p, and consistency of the results with biochemical data on HCV NS3 helicase argue that the observed duplex unwinding mechanism is applicable to a broader subset of processive SF2 helicases. Furthermore, the interaction between afHel308 and its DNA substrate also may explain how afHel308 is targeted to branched nucleic acid substrates.The presented results provide a first structural framework for duplex unwinding by processive SF2 helicases and reveal important mechanistic differences to SF1 helicases and the SF2 helicase RecG.

6 References

Abdel-Monem, M. and H. Hoffmann-Berling (1976). "Enzymic unwinding of DNA. 1. Purification and characterization of a DNA-dependent ATPase from Escherichia coli." Eur J Biochem **65**(2): 431-40.

Andersen, C. B., L. Ballut, J. S. Johansen, H. Chamieh, K. H. Nielsen, C. L. Oliveira, J. S. Pedersen, B. Seraphin, H. Le Hir and G. R. Andersen (2006). "Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA." Science **313**(5795): 1968-72.

Assenmacher, N., K. Wenig, A. Lammens and K. P. Hopfner (2006). "Structural basis for transcription-coupled repair: the N terminus of Mfd resembles UvrB with degenerate ATPase motifs." J Mol Biol **355**(4): 675-83.

Bennett, R. J., J. L. Keck and J. C. Wang (1999). "Binding specificity determines polarity of DNA unwinding by the Sgs1 protein of S. cerevisiae." J Mol Biol **289**(2): 235-48.

Bernstein, D. A., M. C. Zittel and J. L. Keck (2003). "High-resolution structure of the E.coli RecQ helicase catalytic core." Embo J **22**(19): 4910-21.

Betterton, M. D. and F. Julicher (2005). "Opening of nucleic-acid double strands by helicases: active versus passive opening." Phys Rev E Stat Nonlin Soft Matter Phys **71**(1 Pt 1): 011904.

Bianco, P. R., L. R. Brewer, M. Corzett, R. Balhorn, Y. Yeh, S. C. Kowalczykowski and R. J. Baskin (2001). "Processive translocation and DNA unwinding by individual RecBCD enzyme molecules." Nature **409**(6818): 374-8.

Bianco, P. R. and S. C. Kowalczykowski (2000). "Translocation step size and mechanism of the RecBC DNA helicase." Nature **405**(6784): 368-72.

Bjornson, K. P., I. Wong and T. M. Lohman (1996). "ATP hydrolysis stimulates binding and release of single stranded DNA from alternating subunits of the dimeric E. coli Rep helicase: implications for ATP-driven helicase translocation." J Mol Biol **263**(3): 411-22.

Bono, F., J. Ebert, E. Lorentzen and E. Conti (2006). "The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA." Cell **126**(4): 713-25.

Boyd, J. B., K. Sakaguchi and P. V. Harris (1990). "mus308 mutants of Drosophila exhibit hypersensitivity to DNA cross-linking agents and are defective in a deoxyribonuclease." Genetics **125**(4): 813-9.

Brennan, R. G. (1993). "The winged-helix DNA-binding motif: another helix-turn-helix takeoff." Cell **74**(5): 773-6.

Brookman, K. W., J. E. Lamerdin, M. P. Thelen, M. Hwang, J. T. Reardon, A. Sancar, Z. Q. Zhou, C. A. Walter, C. N. Parris and L. H. Thompson (1996). "ERCC4 (XPF) encodes a human nucleotide excision repair protein with eukaryotic recombination homologs." Mol Cell Biol **16**(11): 6553-62.

Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr., Sect. D: Biol. Crystallogr. **54**(Pt 5): 905-21.

Busch, D. B., H. van Vuuren, J. de Wit, A. Collins, M. Z. Zdzienicka, D. L. Mitchell, K. W. Brookman, M. Stefanini, R. Riboni, L. H. Thompson, R. B. Albert, A. J. van Gool and J. Hoeijmakers (1997). "Phenotypic heterogeneity in

nucleotide excision repair mutants of rodent complementation groups 1 and 4." Mutat Res **383**(2): 91-106.

Caruthers, J. M. and D. B. McKay (2002). "Helicase structure and mechanism." Curr Opin Struct Biol **12**(1): 123-33.

Cheng, W., J. Hsieh, K. M. Brendza and T. M. Lohman (2001). "E. coli Rep oligomers are required to initiate DNA unwinding in vitro." J Mol Biol **310**(2): 327-50.

Chow, K. H. and J. Courcelle (2004). "RecO acts with RecF and RecR to protect and maintain replication forks blocked by UV-induced DNA damage in Escherichia coli." J Biol Chem **279**(5): 3492-6.

Collaborative Computational Project, N. (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr. **50**: 760-763.

Cox, M. M. (2002). "The nonmutagenic repair of broken replication forks via recombination." Mutat Res **510**(1-2): 107-20.

Cui, S., R. Klima, A. Ochem, D. Arosio, A. Falaschi and A. Vindigni (2003). "Characterization of the DNA-unwinding activity of human RECQ1, a helicase specifically stimulated by human replication protein A." J Biol Chem **278**(3): 1424-32.

Deaconescu, A. M., A. L. Chambers, A. J. Smith, B. E. Nickels, A. Hochschild, N. J. Savery and S. A. Darst (2006). "Structural basis for bacterial transcription-coupled DNA repair." Cell **124**(3): 507-20.

Delagoutte, E. and P. H. von Hippel (2003). "Helicase mechanisms and the coupling of helicases within macromolecular machines. Part II: Integration of helicases into cellular processes." Q Rev Biophys **36**(1): 1-69.

Dillingham, M. S., M. Spies and S. C. Kowalczykowski (2003). "RecBCD enzyme is a bipolar DNA helicase." Nature **423**(6942): 893-7.

Dronkert, M. L. and R. Kanaar (2001). "Repair of DNA interstrand cross-links." Mutat Res **486**(4): 217-47.

Durr, H., C. Korner, M. Muller, V. Hickmann and K. P. Hopfner (2005). "X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA." Cell **121**(3): 363-73.

Ellis, N. A. (1997). "DNA helicases in inherited human disorders." Curr Opin Genet Dev **7**(3): 354-63.

Flores, M. J., N. Sanchez and B. Michel (2005). "A fork-clearing role for UvrD." Mol Microbiol **57**(6): 1664-75.

Friedberg, E. C. and V. L. Gerlach (1999). "Novel DNA polymerases offer clues to the molecular basis of mutagenesis." Cell **98**(4): 413-6.

Fujikane, R., K. Komori, H. Shinagawa and Y. Ishino (2005). "Identification of a novel helicase activity unwinding branched DNAs from the hyperthermophilic archaeon, Pyrococcus furiosus." J Biol Chem **280**(13): 12351-8.

Gajiwala, K. S. and S. K. Burley (2000). "Winged helix proteins." Curr Opin Struct Biol **10**(1): 110-6.

Gangloff, S., J. P. McDonald, C. Bendixen, L. Arthur and R. Rothstein (1994). "The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase." Mol Cell Biol **14**(12): 8391-8.

Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko and V. M. Blinov (1989). "Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes." Nucleic Acids Res **17**(12): 4713-30.

Graves-Woodward, K. L., J. Gottlieb, M. D. Challberg and S. K. Weller (1997). "Biochemical analyses of mutations in the HSV-1 helicase-primase that alter ATP hydrolysis, DNA unwinding, and coupling between hydrolysis and unwinding." J Biol Chem **272**(7): 4623-30.

Groft, C. M., S. N. Uljon, R. Wang and M. H. Werner (1998). "Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly." Proc Natl Acad Sci U S A **95**(16): 9117-22.

Guy, C. P. and E. L. Bolt (2005). "Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands." Nucleic Acids Res **33**(11): 3678-90.

Hall, M. C. and S. W. Matson (1999). "Helicase motifs: the engine that powers DNA unwinding." Mol Microbiol **34**(5): 867-77.

Hanada, K., T. Ukita, Y. Kohno, K. Saito, J. Kato and H. Ikeda (1997). "RecQ DNA helicase is a suppressor of illegitimate recombination in Escherichia coli." Proc Natl Acad Sci U S A **94**(8): 3860-5.

Harmon, F. G. and S. C. Kowalczykowski (1998). "RecQ helicase, in concert with RecA and SSB proteins, initiates and disrupts DNA recombination." Genes Dev **12**(8): 1134-44.

Harmon, F. G. and S. C. Kowalczykowski (2001). "Biochemical characterization of the DNA helicase activity of the escherichia coli RecQ helicase." J Biol Chem **276**(1): 232-43.

Harris, P. V., O. M. Mazina, E. A. Leonhardt, R. B. Case, J. B. Boyd and K. C. Burtis (1996). "Molecular cloning of Drosophila mus308, a gene involved in DNA cross-link repair with homology to prokaryotic DNA polymerase I genes." Mol Cell Biol **16**(10): 5764-71.

Heller, R. C. and K. J. Marians (2005). "Unwinding of the nascent lagging strand by Rep and PriA enables the direct restart of stalled replication forks." J Biol Chem **280**(40): 34143-51.

Hickson, I. D. (2003). "RecQ helicases: caretakers of the genome." Nat Rev Cancer **3**(3): 169-78.

Hishida, T., Y. W. Han, T. Shibata, Y. Kubota, Y. Ishino, H. Iwasaki and H. Shinagawa (2004). "Role of the Escherichia coli RecQ DNA helicase in SOS signaling and genome stabilization at stalled replication forks." Genes Dev **18**(15): 1886-97.

Hopfner, K. P., A. Karcher, D. Shin, C. Fairley, J. A. Tainer and J. P. Carney (2000). "Mre11 and Rad50 from Pyrococcus furiosus: cloning and biochemical characterization reveal an evolutionarily conserved multiprotein machine." J Bacteriol **182**(21): 6036-41.

Houseley, J., J. LaCava and D. Tollervey (2006). "RNA-quality control by the exosome." Nat Rev Mol Cell Biol **7**(7): 529-39.

Hoy, C. A., L. H. Thompson, C. L. Mooney and E. P. Salazar (1985). "Defective DNA cross-link removal in Chinese hamster cell mutants hypersensitive to bifunctional alkylating agents." Cancer Res **45**(4): 1737-43.

Ishino, Y., K. Komori, I. K. Cann and Y. Koga (1998). "A novel DNA polymerase family found in Archaea." J Bacteriol **180**(8): 2232-6.

Jamieson, E. R., M. P. Jacobson, C. M. Barnes, C. S. Chow and S. J. Lippard (1999). "Structural and kinetic studies of a cisplatin-modified DNA icosamer binding to HMG1 domain B." J Biol Chem **274**(18): 12346-54.

Jankowsky, E., C. H. Gross, S. Shuman and A. M. Pyle (2001). "Active disruption of an RNA-protein interaction by a DExH/D RNA helicase." Science **291**(5501): 121-5.

Kabsch, W. (1993). "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants." Journal of Appl. Crystallogr. **26**(6): 795-800.

Kamada, K., J. De Angelis, R. G. Roeder and S. K. Burley (2001). "Crystal structure of the C-terminal domain of the RAP74 subunit of human transcription factor IIF." Proc Natl Acad Sci U S A **98**(6): 3115-20.

Kantardjieff, K. A. and B. Rupp (2003). "Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals." Protein Sci **12**(9): 1865-1871.

Kawamura, K., R. Bahar, M. Seimiya, M. Chiyo, A. Wada, S. Okada, M. Hatano, T. Tokuhisa, H. Kimura, S. Watanabe, I. Honda, S. Sakiyama, M. Tagawa and O. W. J (2004). "DNA polymerase theta is preferentially expressed in lymphoid tissues and upregulated in human cancers." Int J Cancer **109**(1): 9-16.

Killoran, M. P. and J. L. Keck (2006). "Sit down, relax and unwind: structural insights into RecQ helicase mechanisms." Nucleic Acids Res **34**(15): 4098-105.

Killoran, M. P. and J. L. Keck (2006). "Three HRDC domains differentially modulate Deinococcus radiodurans RecQ DNA helicase biochemical activity." J Biol Chem **281**(18): 12849-57.

Kim, J. L., K. A. Morgenstern, J. P. Griffith, M. D. Dwyer, J. A. Thomson, M. A. Murcko, C. Lin and P. R. Caron (1998). "Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding." Structure **6**(1): 89-100.

Kim, J. W., M. Y. Seo, A. Shelat, C. S. Kim, T. W. Kwon, H. H. Lu, D. T. Moustakas, J. Sun and J. H. Han (2003). "Structurally conserved amino Acid w501 is required for RNA helicase activity but is not essential for DNA helicase activity of hepatitis C virus NS3 protein." J Virol **77**(1): 571-82.

Koonin, E. V., Y. I. Wolf and L. Aravind (2001). "Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach." Genome Res **11**(2): 240-52.

Korolev, S., J. Hsieh, G. H. Gauss, T. M. Lohman and G. Waksman (1997). "Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP." Cell **90**(4): 635-47.

Kow, Y. W., G. Bao, B. Minesinger, S. Jinks-Robertson, W. Siede, Y. L. Jiang and M. M. Greenberg (2005). "Mutagenic effects of abasic and oxidized abasic lesions in Saccharomyces cerevisiae." Nucleic Acids Res **33**(19): 6196-202.

Kreuzer, K. N. (2005). "Interplay between DNA replication and recombination in prokaryotes." Annu Rev Microbiol **59**: 43-67.

Lam, A. M., D. Keeney and D. N. Frick (2003). "Two novel conserved motifs in the hepatitis C virus NS3 protein critical for helicase action." J Biol Chem **278**(45): 44514-24.

Laskowski, R. A., M. W. MacArthur, D. S. Moss and J. M. Thornton (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." J Appl Cryst **26**(2): 283-291.

Laurencon, A., C. M. Orme, H. K. Peters, C. L. Boulton, E. K. Vladar, S. A. Langley, E. P. Bakis, D. T. Harris, N. J. Harris, S. M. Wayson, R. S. Hawley

and K. C. Burtis (2004). "A large-scale screen for mutagen-sensitive loci in Drosophila." Genetics **167**(1): 217-31.

Lawley, P. D. and D. H. Phillips (1996). "DNA adducts from chemotherapeutic agents." Mutat Res **355**(1-2): 13-40.

Lee, J. Y. and W. Yang (2006). "UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke." Cell **127**(7): 1349-60.

Levin, M. K., M. Gurjar and S. S. Patel (2005). "A Brownian motor mechanism of translocation and strand separation by hepatitis C virus helicase." Nat Struct Mol Biol **12**(5): 429-35.

Levin, M. K., M. M. Gurjar and S. S. Patel (2003). "ATP binding modulates the nucleic acid affinity of hepatitis C virus helicase." J Biol Chem **278**(26): 23311-6.

Levin, M. K., Y. H. Wang and S. S. Patel (2004). "The functional interaction of the hepatitis C virus helicase molecules is responsible for unwinding processivity." J Biol Chem **279**(25): 26005-12.

Li, Y., Y. Kong, S. Korolev and G. Waksman (1998). "Crystal structures of the Klenow fragment of Thermus aquaticus DNA polymerase I complexed with deoxyribonucleoside triphosphates." Protein Sci **7**(5): 1116-23.

Linder, P. and M. C. Daugeron (2000). "Are DEAD-box proteins becoming respectable helicases?" Nat Struct Biol **7**(2): 97-9.

Linder, P. and F. Stutz (2001). "mRNA export: travelling with DEAD box proteins." Curr Biol **11**(23): R961-3.

Liu, Z., M. J. Macias, M. J. Bottomley, G. Stier, J. P. Linge, M. Nilges, P. Bork and M. Sattler (1999). "The three-dimensional structure of the HRDC domain and implications for the Werner and Bloom syndrome proteins." Structure **7**(12): 1557-66.

Macris, M. A. and P. Sung (2005). "Multifaceted role of the Saccharomyces cerevisiae Srs2 helicase in homologous recombination regulation." Biochem Soc Trans **33**(Pt 6): 1447-50.

Magana-Schwencke, N., J. A. Henriques, R. Chanet and E. Moustacchi (1982). "The fate of 8-methoxypsoralen photoinduced crosslinks in nuclear and mitochondrial yeast DNA: comparison of wild-type and repair-deficient strains." Proc Natl Acad Sci U S A **79**(6): 1722-6.

Maluf, N. K., J. A. Ali and T. M. Lohman (2003). "Kinetic mechanism for formation of the active, dimeric UvrD helicase-DNA complex." J Biol Chem **278**(34): 31930-40.

Marini, F., N. Kim, A. Schuffert and R. D. Wood (2003). "POLN, a nuclear PolA family DNA polymerase homologous to the DNA cross-link sensitivity protein Mus308." J Biol Chem **278**(34): 32014-9.

Marini, F. and R. D. Wood (2002). "A human DNA helicase homologous to the DNA cross-link sensitivity protein Mus308." J Biol Chem **277**(10): 8716-23.

Marintcheva, B. and S. K. Weller (2003). "Helicase motif Ia is involved in single-strand DNA-binding and helicase activities of the herpes simplex virus type 1 origin-binding protein, UL9." J Virol **77**(4): 2477-88.

Matthews, B. W. (1968). "Solvent content of protein crystals." J Mol Biol **33**(2): 491-497.

McGlynn, P. and R. G. Lloyd (2002). "Recombinational repair and restart of damaged replication forks." Nat Rev Mol Cell Biol **3**(11): 859-70.

McHugh, P. J. and S. Sarkar (2006). "DNA interstrand cross-link repair in the cell cycle: a critical role for polymerase zeta in G1 phase." Cell Cycle **5**(10): 1044-7.

Mechanic, L. E., M. E. Latta and S. W. Matson (1999). "A region near the C-terminal end of Escherichia coli DNA helicase II is required for single-stranded DNA binding." J Bacteriol **181**(8): 2519-26.

Medhurst, A. L., H. Laghmani el, J. Steltenpool, M. Ferrer, C. Fontaine, J. de Groot, M. A. Rooimans, R. J. Scheper, A. R. Meetei, W. Wang, H. Joenje and J. P. de Winter (2006). "Evidence for subcomplexes in the Fanconi anemia pathway." Blood **108**(6): 2072-80.

Meinhart, A., J. Blobel and P. Cramer (2003). "An extended winged helix domain in general transcription factor E/IIE alpha." J Biol Chem **278**(48): 48267-74.

Meylan, E., J. Tschopp and M. Karin (2006). "Intracellular pattern recognition receptors in the host response." Nature **442**(7098): 39-44.

Michel, B., G. Grompone, M. J. Flores and V. Bidnenko (2004). "Multiple pathways process stalled replication forks." Proc Natl Acad Sci U S A **101**(35): 12783-8.

Mohaghegh, P., J. K. Karow, R. M. Brosh Jr, Jr., V. A. Bohr and I. D. Hickson (2001). "The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases." Nucleic Acids Res **29**(13): 2843-9.

Morris, R. J., A. Perrakis and V. S. Lamzin (2003). "ARP/wARP and automatic interpretation of protein electron density maps " Methods Enzymol **374**: 229-244.

Nair, D. T., R. E. Johnson, L. Prakash, S. Prakash and A. K. Aggarwal (2005). "Rev1 employs a novel mechanism of DNA synthesis using a protein template." Science **309**(5744): 2219-22.

Nair, D. T., R. E. Johnson, L. Prakash, S. Prakash and A. K. Aggarwal (2006). "Hoogsteen base pair formation promotes synthesis opposite the 1,N6-ethenodeoxyadenosine lesion by human DNA polymerase iota." Nat Struct Mol Biol **13**(7): 619-25.

Nakayama, K., N. Irino and H. Nakayama (1985). "The recQ gene of Escherichia coli K12: molecular cloning and isolation of insertion mutants." Mol Gen Genet **200**(2): 266-71.

Nishino, T., K. Komori, Y. Ishino and K. Morikawa (2005). "Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms." Structure **13**(8): 1183-92.

Nishino, T., K. Komori, D. Tsuchiya, Y. Ishino and K. Morikawa (2005). "Crystal structure and functional implications of Pyrococcus furiosus hef helicase domain involved in branched DNA processing." Structure **13**(1): 143-53.

Olsen, G. J. and C. R. Woese (1997). "Archaeal genomics: an overview." Cell **89**(7): 991-4.

Oshige, M., N. Aoyagi, P. V. Harris, K. C. Burtis and K. Sakaguchi (1999). "A new DNA polymerase species from Drosophila melanogaster: a probable mus308 gene product." Mutat Res **433**(3): 183-92.

Patel, S. S. and I. Donmez (2006). "Mechanisms of helicases." J Biol Chem **281**(27): 18265-8.

Pause, A. and N. Sonenberg (1992). "Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A." Embo J **11**(7): 2643-54.

Picha, K. M., P. Ahnert and S. S. Patel (2000). "DNA binding in the central channel of bacteriophage T7 helicase-primase is a multistep process. Nucleotide hydrolysis is not required." Biochemistry **39**(21): 6401-9.

Powell, H. (1999). "The Rossmann Fourier autoindexing algorithm in MOSFLM." Acta Crystallogr D **55**(10): 1690-1695.

Richmond, E. and C. L. Peterson (1996). "Functional analysis of the DNA-stimulated ATPase domain of yeast SWI2/SNF2." Nucleic Acids Res **24**(19): 3685-92.

Rittinger, K., P. A. Walker, J. F. Eccleston, S. J. Smerdon and S. J. Gamblin (1997). "Structure at 1.65 A of RhoA and its GTPase-activating protein in complex with a transition-state analogue." Nature **389**(6652): 758-62.

Sampath, A., T. Xu, A. Chao, D. Luo, J. Lescar and S. G. Vasudevan (2006). "Structure-based mutational analysis of the NS3 helicase from dengue virus." J Virol **80**(13): 6686-90.

Scheffzek, K., M. R. Ahmadian, W. Kabsch, L. Wiesmuller, A. Lautwein, F. Schmitz and A. Wittinghofer (1997). "The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants." Science **277**(5324): 333-8.

Seki, M., F. Marini and R. D. Wood (2003). "POLQ (Pol theta), a DNA polymerase and DNA-dependent ATPase in human cells." Nucleic Acids Res **31**(21): 6117-26.

Seki, M., C. Masutani, L. W. Yang, A. Schuffert, S. Iwai, I. Bahar and R. D. Wood (2004). "High-efficiency bypass of DNA damage by human DNA polymerase Q." Embo J **23**(22): 4484-94.

Selby, C. P. and A. Sancar (1993). "Molecular mechanism of transcription-repair coupling." Science **260**(5104): 53-8.

Sengoku, T., O. Nureki, A. Nakamura, S. Kobayashi and S. Yokoyama (2006). "Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa." Cell **125**(2): 287-300.

Sharma, S., J. A. Sommers, S. Choudhary, J. K. Faulkner, S. Cui, L. Andreoli, L. Muzzolini, A. Vindigni and R. M. Brosh, Jr. (2005). "Biochemical analysis of the DNA unwinding and strand annealing activities catalyzed by human RECQ1." J Biol Chem **280**(30): 28072-84.

Sheldrick, G. M. and T. R. Schneider (1997). SHELXL: High-Resolution Refinement. Methods in Enzymology. **277:** 319–343.

Shima, N., S. A. Hartford, T. Duffy, L. A. Wilson, K. J. Schimenti and J. C. Schimenti (2003). "Phenotype-based identification of mouse chromosome instability mutants." Genetics **163**(3): 1031-40.

Silvian, L. F., E. A. Toth, P. Pham, M. F. Goodman and T. Ellenberger (2001). "Crystal structure of a DinB family error-prone DNA polymerase from Sulfolobus solfataricus." Nat Struct Biol **8**(11): 984-9.

Singleton, M. R., M. S. Dillingham, M. Gaudier, S. C. Kowalczykowski and D. B. Wigley (2004). "Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks." Nature **432**(7014): 187-93.

Singleton, M. R., M. R. Sawaya, T. Ellenberger and D. B. Wigley (2000). "Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides." Cell **101**(6): 589-600.

Singleton, M. R., S. Scaife and D. B. Wigley (2001). "Structural analysis of DNA replication fork reversal by RecG." Cell **107**(1): 79-89.

Singleton, M. R. and D. B. Wigley (2002). "Modularity and specialization in superfamily 1 and 2 helicases." J Bacteriol **184**(7): 1819-26.

Sladek, F. M., M. M. Munn, W. D. Rupp and P. Howard-Flanders (1989). "In vitro repair of psoralen-DNA cross-links by RecA, UvrABC, and the 5'-exonuclease of DNA polymerase I." J Biol Chem **264**(12): 6755-65.

Staley, J. P. and C. Guthrie (1999). "An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p." Mol Cell **3**(1): 55-64.

Stano, N. M., Y. J. Jeong, I. Donmez, P. Tummalapalli, M. K. Levin and S. S. Patel (2005). "DNA synthesis provides the driving force to accelerate DNA unwinding by a helicase." Nature **435**(7040): 370-3.

Story, R. M. and T. A. Steitz (1992). "Structure of the recA protein-ADP complex." Nature **355**(6358): 374-6.

Subramanya, H. S., L. E. Bird, J. A. Brannigan and D. B. Wigley (1996). "Crystal structure of a DExx box DNA helicase." Nature **384**(6607): 379-83.

Swagemakers, S. M., J. Essers, J. de Wit, J. H. Hoeijmakers and R. Kanaar (1998). "The human RAD54 recombinational DNA repair protein is a double-stranded DNA-dependent ATPase." J Biol Chem **273**(43): 28292-7.

Tackett, A. J., Y. Chen, C. E. Cameron and K. D. Raney (2005). "Multiple full-length NS3 molecules are required for optimal unwinding of oligonucleotide DNA in vitro." J Biol Chem **280**(11): 10797-806.

Takata, K., T. Shimizu, S. Iwai and R. D. Wood (2006). "Human DNA polymerase N (POLN) is a low fidelity enzyme capable of error-free bypass of 5S-thymine glycol." J Biol Chem **281**(33): 23445-55.

Tanner, N. K. (2003). "The newly identified Q motif of DEAD box helicases is involved in adenine recognition." Cell Cycle **2**(1): 18-9.

Tanner, N. K., O. Cordin, J. Banroques, M. Doere and P. Linder (2003). "The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis." Mol Cell **11**(1): 127-38.

Terwilliger, T. C. (2002). "Automated structure solution, density modification and model building." Acta Crystallogr D **58**(11): 1937-1940.

Turck, D. (1992). Weiterentwicklung eines Programms fuer Molekuelgraphik und Elektronendichtemanipulation und seine Anwendung auf verschiedene Protein-Strukturaufklaerungen. Munich, Technical University. **PhD thesis**.

Tuteja, N. and R. Tuteja (2004). "Unraveling DNA helicases. Motif, structure, mechanism and function." Eur J Biochem **271**(10): 1849-63.

Uemori, T., Y. Ishino, H. Doi and I. Kato (1995). "The hyperthermophilic archaeon Pyrodictium occultum has two alpha-like DNA polymerases." J Bacteriol **177**(8): 2164-77.

Uljon, S. N., R. E. Johnson, T. A. Edwards, S. Prakash, L. Prakash and A. K. Aggarwal (2004). "Crystal structure of the catalytic core of human DNA polymerase kappa." Structure **12**(8): 1395-404.

Umezu, K., K. Nakayama and H. Nakayama (1990). "Escherichia coli RecQ protein is a DNA helicase." Proc Natl Acad Sci U S A **87**(14): 5363-7.

Vandewiele, D., A. R. Fernandez de Henestrosa, A. R. Timms, B. A. Bridges and R. Woodgate (2002). "Sequence analysis and phenotypes of five temperature sensitive mutator alleles of dnaE, encoding modified alpha-catalytic subunits of Escherichia coli DNA polymerase III holoenzyme." Mutat Res **499**(1): 85-95.

Velankar, S. S., P. Soultanas, M. S. Dillingham, H. S. Subramanya and D. B. Wigley (1999). "Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism." Cell **97**(1): 75-84.

Venema, J. and D. Tollervey (1999). "Ribosome synthesis in Saccharomyces cerevisiae." Annu Rev Genet **33**: 261-311.

Wechsler, J. A. and J. D. Gross (1971). "Escherichia coli mutants temperature-sensitive for DNA synthesis." Mol Gen Genet **113**(3): 273-84.

Whitby, M. C. and R. G. Lloyd (1998). "Targeting Holliday junctions by the RecG branch migration protein of Escherichia coli." J Biol Chem **273**(31): 19729-39.

Wong, I. and T. M. Lohman (1992). "Allosteric effects of nucleotide cofactors on Escherichia coli Rep helicase-DNA binding." Science **256**(5055): 350-5.

Wu, J., A. K. Bera, R. J. Kuhn and J. L. Smith (2005). "Structure of the Flavivirus helicase: implications for catalytic activity, protein interactions, and proteolytic processing." J Virol **79**(16): 10268-77.

Wu, L. and I. D. Hickson (2003). "The Bloom's syndrome helicase suppresses crossing over during homologous recombination." Nature **426**(6968): 870-4.

Xie, P. (2006). "Model for helicase translocating along single-stranded DNA and unwinding double-stranded DNA." Biochim Biophys Acta **1764**(11): 1719-29.

Yarranton, G. T., R. H. Das and M. L. Gefter (1979). "Enzyme-catalyzed DNA unwinding. A DNA-dependent ATPase from E. coli." J Biol Chem **254**(23): 11997-2001.

Yoakum, G. H. and R. S. Cole (1978). "Cross-linking and relaxation of supercoiled DNA by psoralen and light." Biochim Biophys Acta **521**(2): 529-46.

Yoshimura, M., M. Kohzaki, J. Nakamura, K. Asagoshi, E. Sonoda, E. Hou, R. Prasad, S. H. Wilson, K. Tano, A. Yasui, L. Lan, M. Seki, R. D. Wood, H. Arakawa, J. M. Buerstedde, H. Hochegger, T. Okada, M. Hiraoka and S. Takeda (2006). "Vertebrate POLQ and POLbeta cooperate in base excision repair of oxidative DNA damage." Mol Cell **24**(1): 115-25.

Zheng, N., E. Fraenkel, C. O. Pabo and N. P. Pavletich (1999). "Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP." Genes Dev **13**(6): 666-74.

Zhou, B. L., J. D. Pata and T. A. Steitz (2001). "Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain." Mol Cell **8**(2): 427-37.

7 Supplementary material

7.1 Abbreviations

Å	Ångström (=10^{-10} m)
AA	amino acid or residue
ATP	adenosine triphosphate
AMP-PNP	5´- adenylyl-imido-triphosphate
bp	base pair(s)
BSA	bovine serum albumin
ca.	circa
CCD	charge coupled device
CSL	Cep1 synthetic lethality
DALI	distance matrix alignment
DEPC	Diethylpyrocarbonate
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
ds	double-stranded
DSB	double strand break
E.coli	*Escherichia coli*
e.g.	*exempli gratia* (for example)
HPLC	high performance liquid chromatography
HR	homologous recombination
kb	kilobase pair(s);
LB	Luria-Bertani
M	molar
MAD	multiple-wavelength anomalous dispersion
Mfd	mutation frequency decline
MPD	
MR	molecular replacement
MTR	mRNA transport regulator
MW	Molecular weight
NCBI	National Center for Biotechnology Information
n.d.	not determined
NER	nucleotide excision repair
NHEJ	non-homologous end-joining
NMR	nuclear magnetic resonance spectroscopy
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol

PEI	polyethyleneimine
pI	isoelectric point
PMSF	phenyl methyl sulfonyl fluorid
PNK	polynucleotide kinase
PVDF	polyvinyliden fluorid
RMSD	root mean square deviation
RNA	ribonucleic acid
RP-HPLC	reversed phase HPLC
RRP	ribosomal RNA processing factor
SAD	single-wavelength anomalous diffraction
SDS	sodium dodecyl sulphate
SLS	swiss light source
ssDNA	single stranded DNA
ss	single-stranded
TAE	tris-acetate-EDTA
TB	tris-borate
TF	transcription factor
TLC	thin layer chromatography
TLS	translesion synthesis
v	volume
w	weight
wt	wild-type

7.2 Amino acids and nucleotides

One letter code	Three letter code	Amino acid
A	Ala	alanine
B	Asx	asparagine or aspartic acid
C	Cys	cystein
D	Asp	aspartic acid
E	Glu	glutamic acid
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
L	Leu	leucine
K	Lys	lysine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
X	Xaa	unknown or other
Y	Tyr	tyrosine
Z	Glx	glutamine or glutamic acid

One-letter code	Nucleobase
A	adenine
C	cytosine
G	guanine
T	thymine

8 Curriculum vitae

Personal Details

Name	Katharina Büttner
Date of Birth	June 10, 1977
Place of Birth	Wolfratshausen, Germany
Nationality	German
Martial status	single

Education

1983-1987	Grundschule Icking
1987-1989	Volks- und Hauptschule Wolfratshausen
1989-1994	Realschule Geretsried
1994-1998	Theodolinden Gymnasium München

Academic Training

1998 – 2003	Ludwigs Maximilians University (LMU) of Munich Studies in Biology (Diplom)
2002 – 2003	LMU Munich, Gene Center/ Institute of Biochemistry Diploma Thesis: "Struktur- und Funktionsaufklärung der ABC ATPase HP68" Supervisor: Prof. Dr. Karl-Peter Hopfner
since December 2003	LMU Munich, Gene Center / Institute of Biochemistry PhD Thesis in Structural Biology Supervisor: Prof. Dr. Karl-Peter Hopfner

9 Acknowledgements

With this, I would like to thank all the people who have contributed to this work.

I am very grateful to my supervisor Prof. Dr. Karl-Peter Hopfner for giving me the opportunity to work in his laboratory, for his guidance and constructive advice over all the years. It has been a privilege to work with you.

Special thanks go to current, former and non-members of the Hopfner lab – Aaron Alt, Nora Assenmacher, Derk Bemeleit, Dr. Sheng Cui, Harald Dürr, Sebastian Fenn, Song Gao, Julia Griese, Sophia Hartung, Annette Karcher, Axel Kirchhofer, Alfred Lammens, Dr. Katja Lammens, Diana Pippig, Alexandra Schele, Christian Schiller and Dr. Gregor Witte – for their help in so many ways. You are fireflies in warm summer nights, and I will carry the pictures of you in my heart wherever I go. Furthermore, I would like to thank Sebastian Nehring for his help and effort during his Bachelor thesis.

With all my heart, I want to thank my family. This work is dedicated to you. Your love, support and advice is unpayable, and I am indepted to you until the end of my days.

Bernhard, I thank you for being a part of my life – and for being the rope that keeps me from falling.

Many thanks to all of my wonderful friends out there. Your sun is shining through the nights of my scientific parallel world.

I know that these few words, written on a blank sheet of paper, are not enough to express my gratefulness, and my humility to your help, patience, love and friendship. And I apologize for all the times I failed to be more thankful.