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

Analyzing Protein - Nucleic Acid Complexes using Hybrid Methods

I. The DNA Damage Checkpoint Protein DisA

II. Structural Biochemistry of RNA Turnover by the Exosome

Sophia Hartung

aus
Würzburg

München, 2008

Erklärung

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

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe erarbeitet.

München, am 01.9.2008

....
(Sophia Hartung)

Dissertation eingereicht am 01.09.2008
1. Gutachter Herr Prof. Dr. Karl-Peter Hopfner
2. Gutachter Herr Prof. Dr. Roland Beckmann
Mündliche Prüfung am 14. Oktober 2008

The presented thesis was prepared in the time from January 2005 to July 2008 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:

Hartung, S. and Hopfner K. P. (2007).
The exosome, plugged. *EMBO Rep* **8**(5): 456-7.

Witte, G.*, **Hartung, S.*,** Büttner, K. and Hopfner K. P. (2008).
Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol Cell* **30**(2): 167-78.

* These authors contributed equally to this work.

Table of Contents

1 Analysis of Large Protein Complexes and Their Ligands....1
  1.1 Preface....1
  1.2 Methods for Structural Analysis....1
    1.2.1 X-ray Crystallography....2
      1.2.1.1 Preface....2
      1.2.1.2 Crystallizing Proteins....3
      1.2.1.3 Structure Determination by X-ray Diffraction....4
      1.2.1.4 The Phase Problem....5
      1.2.1.5 Refinement....6
    1.2.2 SAXS....7
      1.2.2.1 Preface....7
      1.2.2.2 Structure Determination by Small Angle X-ray Scattering....8
      1.2.2.3 The Guinier Approximation and Porod's Law....10
      1.2.2.4 The Pair Distribution Function....11
      1.2.2.5 Ab initio Modeling....12
      1.2.2.6 Computation of Scattering Patterns from Crystal Structures....13
  1.3 Biochemical Activity Assays....14

2 Material and Methods....16
  2.1 Material....16
  2.2 Software....16
  2.3 Methods....17
    2.3.1 Cloning, Expression and Purification Methods....17
      2.3.1.1 Cloning....17
      2.3.1.2 Protein Expression....18
      2.3.1.3 Purification of Exosome Proteins and Complexes....19
      2.3.1.4 Purification of Exosome Complexes with Bound RNA....19
    2.3.2 Biochemical Assays....20
      2.3.2.1 Diadenylate Cyclase Assays....20
      2.3.2.2 RNAse Assays....21
        2.3.2.2.1 Radioactive labeling of oligonucleotides....21
        2.3.2.2.2 Assay Conditions....21
        2.3.2.2.3 Urea Gel Electrophoresis....21
      2.3.2.3 EMSA (electrophoretic mobility shift assay)....22
      2.3.2.4 Fluorescent Labeling of the Exosome for Single-Molecule Experiments....22
    2.3.3 Crystallization and Structure Determination....23
      2.3.3.1 Exosome in Complex With Small RNA Molecules....23
      2.3.3.2 Csl4 S1-ZnR Domain....23
      2.3.3.3 Exosome core with copurified RNA....24
    2.3.4 SAXS Experiments and Data Processing....24
      2.3.4.1 DisA....24
      2.3.4.2 Exosome....25

3 The DNA Integrity Scanning Protein A (DisA)....26
  3.1 The DisA Protein and its Influence on Sporulation....26
  3.2 The Crystal Structure of T. maritima DisA....27
  3.3 Cyclic Purine Nucleotides As Second Messengers....29
    3.3.1 cAMP and cGMP....29
    3.3.2 C-di-GMP and Associated Enzymes....30
      3.3.2.1 Diguanylate Cyclase Activity and GGDEF Domains....31
      3.3.2.2 Phosphodiesterase Activity and EAL Domains....32

        3.3.2.3 PilZ as First Proposed c-di-GMP Binding Domain 33
  3.4 The Role of Holliday Junctions and Fork Structures in DNA Repair 33
    3.4.1 Double Strand Breaks 34
    3.4.2 Stalled Replication Forks 35
    3.4.3 DNA Damage Checkpoints 35
  3.5 Aim Of The Project 35

4 Results – DisA 37
  4.1 Identification of the Correct Quaternary Structure of DisA 37
  4.2 Identification of the Enzymatic Activity of DisA 41
    4.2.1 DisA is a Specific Diadenylate Cyclase 41
    4.2.2 The Influence of Mg2+ and Active Site Mutants on the Activity 44
  4.3 The Influence of Different DNA Molecules on DisA Activity 44
  4.4 The Influence of Azide 46
  4.5 DisA and DNA Binding 47

5 Discussion – DisA 49
  5.1 SAXS as Complementary Method to X-ray Crystallography 49
  5.2 The DisA Octamer 49
  5.3 Reliability of SAXS Structures 50
  5.4 The Moving Foci of DisA 51
  5.5 c-di-AMP and Diadenylate Cyclase Activity 52
  5.6 The Effect of Holliday Junctions on the Activity 54

6 The Exosome and RNA Metabolism 57
  6.1 The Exosome and Quality Control 57
  6.2 Exosome-like Complexes in Bacteria, Archaea and Eukaryotes 58
    6.2.1 Bacterial PNPase and RNase PH 58
    6.2.2 Comparison of Exosome-like Complexes 59
  6.3 Functions of the Exosome 60
    6.3.1 The Exosome in Cytoplasm and Nucleus 60
      6.3.1.1 The Exosome in the Cytoplasm 61
      6.3.1.2 The Exosome in the Nucleus 62
    6.3.2 RNA Degradation 62
      6.3.2.1 RNA Degradation by the Exosome in Archaea 62
      6.3.2.2 RNA Degradation by the Exosome in Yeastb 63
    6.3.3 RNA Polymerization Activity of the Archaeal Exosome 64
  6.4 Cofactors of the Exosome Complex 65
    6.4.1 Mtr4 and the TRAMP Complex in the Nucleus 66
    6.4.2 The Ski Complex 67
    6.4.3 Sequence Specific Cofactors 68
  6.5 Aim of the Project 69

7 Results – Exosome 70
  7.1 Exosome Structures in Solution 70
  7.2 Crystal Structures of the Archaeal Exosome 72
    7.2.1 Csl4-Exosome Wild Type with RNA Bound to the Active Site 72
    7.2.2 Csl4-Exosome Y70ARrp42 with RNA Bound to the Active Site 75
    7.2.3 Crystal Structure of the Archaeal Csl4 S1 and Zn-ribbon Domain 75
  7.3 The Archaeal Exosome Bound to a Large RNA Molecule 76
    7.3.1 Preparation 76
    7.3.2 Structural Analysis in Solution 77
    7.3.3 Crystal Structure of the Complex 78

7.4 Enzymatic Activities of Archaeal Exosome Complexes....80
  7.4.1 Variants of Exosome Complexes Used for Activity Assays....80
  7.4.2 RNA Degradation and Polymerization Assays....81
    7.4.2.1 RNase Activity of the Archaeal Exosome....81
    7.4.2.2 Polymerization Activity of the Archaeal Exosome....82
  7.4.3 Quantification of RNase Assays....84
    7.4.3.1 Theoretical Models for RNA Degradation....84
    7.4.3.2 Experimental Data....88
    7.4.3.3 Data Analysis and Rate Constants....90
7.5 RNA Binding to Exosome Complexes....93

8 Discussion – Exosome....94
8.1 Low Resolution Structures of Exosome Complexes in Solution....94
8.2 The Atomic Structure of RNA Bound to the Exosome Active Site....95
8.3 Atomic Structure of the Isolated Csl4 S1- and Zn-ribbon-domains....97
8.4 The Exosome with a Substrate from E. coli Cells....98
8.5 Enzymatic Activities of the Exosome....99
8.6 The Cleavage-and-Binding Model for RNA Degradation....100
8.7 Evolution and the Exosome....104

9 Summary....107

10 Appendix....108
10.1 Summary of SAXS measurements....108
10.2 Summary of X-ray Crystallographic Experiments....108
10.3 Supplementary Data Concerning DisA....109
  10.3.1 Sequence Alignment....109
  10.3.2 Mass spectrometry results, which identified c-di-AMP....109
10.4 MATLAB Scripts....110
  10.4.1 Cleavage-only model....110
  10.4.2 Cleavage-and-binding model....111
10.5 Experimental Data from RNase Assays and Fits....112
10.6 Denaturing PA Gel and Agarose Gel Showing the RNA Ligand....114

11 References....115

12 Curriculum Vitae....121

13 Acknowledgments....122

1 Analysis of Large Protein Complexes and Their Ligands

1.1 Preface

During the last years an important milestone in scientific progress was the sequencing of several eukaryotic genomes. The resulting availability of the primary structure of many proteins opens a variety of new possibilities for understanding protein functions. Using computational methods it is possible to predict the secondary or even the tertiary structure of proteins. Homology searches allow for the comparison of proteins that evolved during evolution. Insights into the function of one protein can often be transferred to its homologs. In addition, this sequence information makes it possible to recombinantly produce the protein of interest and analyze it *in vitro*.

All proteins need to fold into specific three-dimensional conformations to be able to perform their functions. For most proteins it is even not enough to be properly folded, they are not functional in their isolated state and need at least one interaction partner to perform their task in the living cell.

Most cellular functions like DNA replication, transcription and mRNA translation require the coordinated action of a large number of proteins that are assembled in an array of multi-protein complexes. In these complexes the correct composition and structure is essential for functionality. Additionally, most biological processes are connected and regulated by dynamic signaling networks of interacting proteins that transfer signals, ligands or impulses to a downstream effector.

All these facts clearly show that the examination of single isolated proteins will not be sufficient to understand most of the cellular processes. It is essential to expand experiments to the analysis of protein complexes, their composition, ligands, binding partners and activities.

1.2 Methods for Structural Analysis

Three dimensional structural information is extremely important for the understanding of protein complexes. However, even with the most advanced light microscopes a fast and convenient determination of protein structures is still not possible. More complicated and time-consuming methods have to be used, especially for solving a protein structure at low resolutions.

The different methods structural biologists use to determine structures generally involve measurements on vast numbers of identical molecules at the same time. At present, the mainly used methods are nuclear magnetic resonance (NMR), electron microscopy / electron cryomicroscopy (EM/cryo-EM) and X-ray crystallography.

NMR is performed in aqueous solution, which allows to monitor the binding of ligands to a protein or determine the structure of different conformational states of a protein. The biological samples can be analyzed *in vitro* close to physiological conditions. Nowadays NMR spectrometer with very high magnetic fields are available, but even with multidimensional spectra the method is limited to proteins smaller than 70 kDa because of overlapping signals. To analyze large protein complexes, NMR is only suitable in very special cases.

In contrast, EM studies work perfectly especially on large complexes. For EM experiments the proteins have to be immobilized, which results in an environment that is less physiological than in NMR experiments. Another disadvantage of EM is the limited resolution. Although the resolution of the latest structures reaches down to approximately 5-10 Å, it is impossible to fit side chains of amino acids into the obtained electron densities. Additionally, flexible parts of the molecules can not be visualized, because the different orientations are lost during averaging.

In this thesis two different structural methods were used, X-ray crystallography to obtain structures at atomic resolution and Small Angle X-ray Scattering (SAXS) to analyze structures in solution, a method that recently became increasingly important in the biological field.

1.2.1 X-ray Crystallography

1.2.1.1 Preface

Structure analysis of proteins using protein crystallography is used since the late 1950s and still remains the most widely used method for visualizing atomic structures of proteins and nucleic acids. Compared to other methods it possibly provides the most detailed picture of a biological molecule.

The number of macromolecular structures deposited in the Protein Data Bank now exceeds 51 000, with more than 85% determined using crystallographic methods. Thousands of studies describing such structures have been published in scientific literature and many Nobel prizes in chemistry or medicine have been awarded to protein crystallographers. The progress in structure determination has accelerated during the last years due to the introduction of powerful new algorithms and computer programs for diffraction data collection, structure solution, refinement and presentation. The availability of highly energetic X-rays at synchrotrons strongly improved data quality and tunable beamlines nowadays allow for multi-wavelength experiments for phasing.

To understand the enzymatic mechanisms of proteins in the cell, one single structure is often not sufficient. The determination of more structures, for example with different binding partners or ligands, can give insights into the mode of function of an enzyme.

With help of structural information at high resolutions a working model can be proposed and biochemical activity assays with the wild-type protein in comparison to structure derived mutants can be used to verify this model.

One disadvantage of X-ray crystallography is the fact that the sample has to be in a crystalline state. Proteins only form crystals when their shape is very homogeneous and rigid. Flexible loops can hinder the crystallization process, so they have to be removed and cannot be visualized. Some proteins are even so flexible that it is not at all possible to crystallize them. This of course restrains the method drastically. As enzymes can be interpreted as molecular machines they normally have to be flexible to perform their tasks.

Therefore it has become clear that it needs the application of hybrid methods to answer biological questions as completely as possible. In the present thesis X-ray crystallography was combined with SAXS to gain information about the atomic composition of a complex and its behavior in solution.

1.2.1.2 Crystallizing Proteins

The process of crystallizing a protein or protein complex of interest is in most cases the crucial step that makes structure determination difficult. However the growth of high quality crystals is inevitable for the generation of good diffraction patterns. Crystals are generally solid and consist of molecules that are packed in a regularly ordered, repeated pattern extending in all three spatial dimensions. In contrast to small molecules like salt, proteins are not very rigid and generally have many degrees of freedom which reduces the conditions for crystallization drastically. Therefore hundreds of different crystallization conditions have to be screened. In general, they all consist of three components: First, all proteins are very sensitive to the surrounding pH, why the condition contains a buffer to fix the pH of the solution. Second, some precipitant is added that lowers the solubility of the proteins. Third, different additives can change the condition slightly and thereby increase the probability of crystallization.

The most common approach is to gradually lower the solubility of the protein, which in this case means a slow increase in precipitant and protein concentration. However, if this is done too quickly, the protein will precipitate from solution and is useless for structure determination. Crystal growth in solution consists of two steps: nucleation of a crystal followed by its growth. Normally in the initial screens only small crystals can be found that cannot be used for diffraction studies. Therefore the conditions resulting in small crystals have to be refined by changing the condition only slightly to improve crystal quality and growth.

Different techniques can be used to achieve slow increase of protein and precipitant concentration. Mainly used are the sitting and hanging drop vapor diffusion method. Thereby, a drop of protein solution is suspended over a reservoir containing buffer and precipitant. The drop slowly equilibrates with the reservoir solution by diffusion, leaving the drop with optimal crystal growth conditions.

1.2.1.3 Structure Determination by X-ray Diffraction

After successful production of a crystal, it is mounted in a nylon loop and flash frozen with liquid nitrogen so that it can be placed in the X-ray beam and rotated. Freezing is extremely important for biological samples as it reduces the radiation damage caused by the highly energetic X-rays, as well as the noise in the Bragg peaks due to thermal motion (Debye-Waller effect). However, especially when the water content in a crystal is too high, crystal packing can be damaged during freezing. Therefore crystals are generally pre-soaked in a cryoprotectant solution prior to freezing. The best suited cryoprotectant is determined by screening different candidates like for example glycerol, MPD and 1,4-butanediol. The frozen crystal is exposed to an X-ray beam (nowadays mainly from synchrotrons) and rotated. The resulting diffraction patterns are recorded by a detector.

The observed diffraction results from the electrons in the outer shells of all atoms in the periodically arranged biomolecule. The process of diffraction is actually a combination of two separate and simultaneous operations, scattering and interference. The scattering depends only on the interaction between the X-rays and the protein crystal. When the X-ray photons collide with the atoms, the oscillating electric component of the photons induces oscillations mainly in the electrons. The oscillating electrons and nuclei then emit secondary "scattered" X-rays of the same energy as the incident photons. The scattered waves then interfere with one another because of the periodic nature of the crystalline sample and produce the diffraction pattern. This diffraction pattern is directly related to the electron distribution in the crystals and can be explained by Bragg's law.

With a detector the position as well as the intensity of every reflection is detected.

The position of each detected reflection corresponds to the overall scattering from one particular set of Bragg planes, which are labeled with reciprocal space coordinates (h,k,l), also known as Miller indices. Therefore the geometry (space group and cell dimensions) of the unit cell can be determined with the knowledge of the positions of the reflections.

The intensity of every measured reflection is proportional to the square of the structure factor amplitude $|F_{hkl}|$. The structure factor describes the way in which an incident beam is scattered by the atoms of a crystal unit cell:

$$F_{hkl} = \sum_{j=1}^{all\ atoms} f_j \cdot \mathrm{e}^{2\pi i(hx_j + ky_j + lz_j)} \quad (1)$$

Where the sum is over all atoms in the unit cell and f_j is the atomic form factor for atom i that describes the different scattering powers of the elements. During scattering a difference in phase occurs, as the atoms are spatially distributed in the unit cell. This phase shift information is taken in account by the exponential term. The scattering power of an atom is additionally influenced by the temperature-, B- or Debye-Waller factor, which describes the attenuation of scattering caused by thermal motion or quenched disorder.

The order of the crystals is the reason for a periodicity in the electron density. Therefore it is possible to apply an inverse Fourier transformation to the structure factor and obtain the electron density ρ:

$$\rho_{xyz} = \frac{1}{V} \sum_{hkl} F_{hkl} \mathrm{e}^{-2\pi i(hx+ky+lz)} \quad (2)$$

Where V is the volume of the unit cell and the sum is over all combinations of h, k and l. The coordinates x, y and z represent the three-dimensional coordinates of a point within the unit cell. With this equation the reciprocal space with coordinates h, k and l is transformed into the real space with coordinates x, y and z.

The problem in solving this equation is the structure factor F_{hkl}. From the reflections only the amplitudes $|F_{hkl}|$ and not the corresponding phases are known. The determination of the phases is a very central problem in the structure determination process and is therefore known as the "phase problem".

1.2.1.4 The Phase Problem

The phases are evident in the following reformulation of equation (2):

$$\rho_{xyz} = \frac{1}{V} \sum_{hkl} |F_{hkl}| \mathrm{e}^{-2\pi i(hx+ky+lz-\alpha_{hkl})} \quad (3)$$

Where α_{hkl} is the corresponding phase to $|F_{hkl}|$. As it is not possible to directly calculate the phases, indirect methods were developed over the last years to determine them. No direct relationship exists between the phases and the amplitudes, only via the molecular structure or electron density. Some prior knowledge about the electron density or structure is therefore the basis for all phasing methods.

Nowadays the mainly used methods are single-wavelength anomalous dispersion (SAD), multi-wavelength anomalous dispersion (MAD) and molecular replacement. More detailed information can be found in textbooks (e.g. Drenth, 1999; McPherson, 2001; Blow, 2002).

In this thesis the phase problem was solved with molecular replacement and this method will be discussed in more detail:

Molecular replacement can only be used, when the structure is similar to an already known atomic structure, e.g. an apoenzyme structure can be used for structure determination of the ligand bound form. For molecular replacement first a Patterson map has to be calculated. The Patterson map is an inter-atomic vector map and can be calculated without phase information. The Patterson function is given as equation 4.

$$P_{uvw} = \frac{1}{V} \sum_{hkl} |F_{hkl}|^2 e^{-2\pi(hu+kv+lw)} \quad (4)$$

The peaks in the Patterson function are the interatomic distances weighted by the product of the number of electrons in the atoms concerned. The Patterson map includes all the information about the structure that can be extracted from the data without the phases. By comparison of the calculated Patterson map of the known structure with the one of the new data the phases can be estimated. The known model is rotated and translated, until its Patterson map fits the new one best. The programs use maximum likelihood based algorithms to evaluate the results. The highest correlation is obtained when the two Patterson maps (from the known and unknown structure) are in a similar orientation. In the translation function, the orientated model can now be correctly positioned by translating it to the correct coordinates within the asymmetric unit.

This is followed by the estimation of the new phases by calculating the phase angles implied in the model structure that has been correctly orientated and positioned in the cell. The calculated phases α_{hkl} are now attached to the measured structure amplitudes $|F_{hkl}|$ to receive an estimate of the structure factor and of the electron density.

Usage of these estimated phases results in a model bias coming from the model structure. It is therefore essential to calculate a 2 $|F_o|$ - $|F_c|$ electron density map and built a new atomic model into the density. This is followed by iterative refinement processes, trying to find the best agreement between the observed structure factors and the calculated ones.

1.2.1.5 Refinement

To monitor the quality of the model, the R-factor (for residual or reliability factor) is calculated:

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|} \quad (5)$$

With $|F_{obs}|$ being the experimentally determined, observed structure factor amplitudes and $|F_{calc}|$ the ones calculated from the model. The higher the R-factor, the more the calculated amplitudes differ from the observed ones. Therefore the final goal is always to refine the model until the R-factor is as low as possible. A similar quality criterion is R_{free}, which is calculated from a subset (~5-10%) of reflections that were not included in the structure refinement.

For manual rebuilding of the initial model a difference density map ($|F_o|-|F_c|$) is used, which shows the difference between the true and the currently modeled structure. With this map, parts that exist in the structure, but are not yet included in the model, as well as wrong parts of the model can be identified. In addition to manually improving the model, refinement programs are applied that refine the position and B-factors of every atom. These computational refinement programs imply restraints for e.g. bond lengths, bond angles, torsion angles and chiral volumes. After some rounds of manual and computational refinement, better phases and a decrease of the R-factor are achieved.

1.2.2 SAXS

1.2.2.1 Preface

Only rigid parts of a protein structure can be determined at atomic resolution via X-ray crystallography, but also structural information about the flexible parts of a macromolecule are of great interest. Therefore, a second method besides crystallography should be carried out in solution to gain information about flexible regions of the sample. In this thesis SAXS was the method of choice to address these questions. SAXS is a technique that is carried out in solution and it offers the potential for obtaining structural information relatively easy, it does not require crystals and is therefore perfect for analyzing systems possessing substantial flexibility. SAXS as a method for biological macromolecules lagged many decades behind crystallography because the application of SAXS to macromolecules is a surprisingly complex problem. Even with good X-ray quality from new generation high energy synchrotrons and suitable computational power from advanced computer clusters, SAXS is still limited by the relatively small number of effective observations in a given solution scattering curve.

At the first glance it seems that SAXS provides even less information than EM. Yet, there are several advantages: SAXS experiments can be performed in a very short timescale and relatively small amounts of sample are needed. Moreover, as the samples are measured in solution many different and preferably physiologic buffer conditions can be screened. And as already mentioned, in contrast to EM SAXS can provide information about flexible parts of the sample. Especially by combining SAXS with crystallography and computational methods, a powerful tool is formed. The scattering observations from SAXS data combined with crystal structures can provide more accurate and complete models of protein and RNA structures, conformations, interactions and assemblies in solution.

An example for the combination of SAXS with protein crystallography is shown in figure 1 (from Putnam et al., 2008).

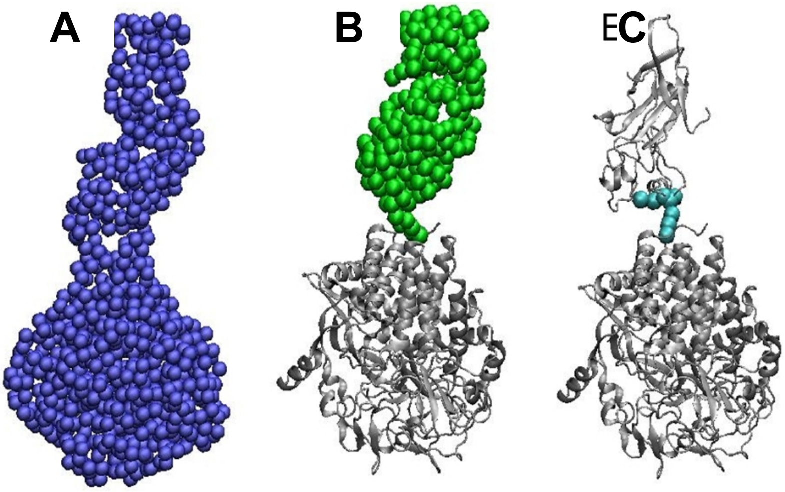


Figure 1: SAXS model of the complexed cellulase (from Putnam et al., 2008). **A**: *ab initio* model representing residues as beads calculated with the program GASBORp (blue). **B**: Reconstitution of one missing module represented by an ensemble of dummy residues (green). The secondary structural elements of the known atomic structure is shown in gray. **C**: Rigid body modeling applied on the known atomic structures (gray) in combination with *ab initio* modeling of the linker region (cyan) using the program BUNCH.

The full-length cellulase protein is composed of two domains and could not be crystallized due to a flexible linker between the two domains. Still is was possible to determine the atomic structure of the two domains separately. The SAXS structure of the full-length protein was calculated and the two crystal structures were modeled into the SAXS density. Using *ab initio* calculations, the connecting flexible linker could be positioned between the two domains. To further optimize the model, the best-fit conformation could be searched for by applying molecular dynamics on the linker region.

1.2.2.2 Structure Determination by Small Angle X-ray Scattering

In a SAXS experiment – similar to X-ray diffraction – the X-ray wave is scattered mainly by the electrons surrounding the atoms of the protein. As SAXS is performed in solution, the proteins are spatially averaged and the wave does not encounter any regularly repeated elements like it does in a crystal. Therefore no interference can occur and no diffraction pattern can be observed. But still the scattering can be recorded as a function of the scattering angle and structural information can be extracted.

SAXS is a contrast method were the scattering signal is derived from the difference in the average electron density of solvent and the electron density due to the molecule in the solution. As higher density differences give higher scattering intensities, nucleic acids give much stronger signals compared to proteins.

To describe the scattering from assemblies of atoms it is convenient to introduce the scattering length density distribution ρ(r), which is equal to the total scattering length of the atoms per unit volume. SAXS experiments on biomolecules in solution require separate measurements of the scattering from sample and solvent, as the contribution of the solvent scattering to the total scattering is relatively high. The scattering from the biomolecule is thus often referred to as excess scattering over the solvent scattering. Assuming the solvent has a constant scattering density ρ_s, the scattering amplitude from a single particle relative to that of an equivalent volume of solvent is given by the Fourier transform of the difference in scattering density $\Delta\sigma(r) = \sigma(r) - \sigma_s$:

$$A(s)=\int_V \Delta\rho(r)\,\mathrm{e}^{isr}\,dr \quad (6)$$

with A being the scattering amplitude and s the momentum transfer $s = 4\pi\lambda^{-1}\sin(\theta)$, where 2θ is the total scattering angle relative to the forward direction. The integration is performed over the particle volume.

In a scattering experiment the intensity I(s) (not the amplitude) is detected, which is proportional to the number of scattered photons. As the molecules are randomly distributed the intensity of the ensemble is a function proportional to scattering of a single particle averaged over all orientations. For the analysis of SAXS data from biological macromolecules two important restrictions are introduced, which greatly simplify the problem: first the system is postulated to be monodisperse and second no correlation between the molecules should exist.

Using equation 6 the spherically averaged single particle intensity can be calculated and by integrating in spherical coordinates one gets

$$I(s)=4\pi\int_0^{D_{max}} r^2\gamma(r)\frac{\sin sr}{sr}\,dr \quad (7)$$

where *γ(r)* is the spherically averaged autocorrelation function of the excess scattering density, which is equal to zero for distances exceeding the maximum particle diameter D_{max}. In practice the function $p(r) = r^2 \gamma(r)$ is called pair or distance distribution function and corresponds to the distribution of distances between elements inside the molecule, weighted by the excess density distribution. This pair distribution function is commonly used to analyze scattering data from biomolecules and is computed by the inverse transformation

$$p(r)=\frac{r^2}{2\pi^2}\int_0^\infty s^2 I(s)\frac{\sin sr}{sr}dr \quad (8)$$

1.2.2.3 The Guinier Approximation and Porod's Law

Besides the pair distribution function an approximation is important that is only valid for scattering at very small values of momentum transfer (s → 0). It was derived by Guinier (Guinier, 1939) and has long been the most important tool in the analysis of scattering from isotropic systems, the so called Guinier equation is:

$$I(s)\simeq I_0 e^{-\frac{1}{3}R_g^2 s^2} \quad (9)$$

where I_0 is the intensity at $s = 0$ and R_g is the radius of gyration of the molecule. It is clear that the radius of gyration can be calculated directly from the scattering data when using only very small angle data. A Guinier plot, where ln(I(s)) is plotted versus s^2 should give a linear function with I_0 as intercept and a slope that can be used to calculate R_g. Only very homogeneous samples with no attractive or repulsive forces between molecules will give a linear Guinier plot. Therefore it is also a method do test the quality of the sample, but one has to keep in mind that it is still an approximation and only valid in the range of $s \cdot R_g < 1.3$.

The radius of gyration characterizes the size of a particle and is defined as the square root of the average squared distance of each scatterer from the particle center. The advantage of the radius of gyration is the fact that it can be calculated by numerical integration from any structure, for example a crystal structure, and compared to the value obtained from the measurement.

Besides R_g, the forward scattering intensity I_0 can be determined. I_0 is related to the molecular mass of the sample and independent from the shape. It is only dependent on the squared contrast of the particle, the concentration of the sample and the intensity of the transmitted beam. With known sample concentration and with a reference sample like BSA (bovine serum albumin) for proteins, the molecular mass can be calculated.

Analysis of high s-values gives information regarding molecular shape. For a folded macromolecule the intensity of the scattering falls off by Porod's law (Kratky and Porod, 1949):

$$I(s) \propto s^{-4} \quad (10)$$

This correlation is only true for a uniform density of the scatterer, which breaks down at high s values when atomic resolution information begins to contribute more and more to the scattering profile. Therefore Porod's law is - like the Guinier approximation - only valid for a certain position of the scattering curve. When the Porod plot ($s^4 \cdot I(s)$ versus s^4) at high angles is not linear, the samples shows internal inhomogeneities.

1.2.2.4 The Pair Distribution Function

The pair distribution function (equation 8) is the Fourier transform of the scattering profile *I(s)* and in principle both contain the same information. It can be explained like a two-dimensional Patterson function (see 1.2.1.4) that shows the frequency with which vectors of a length r connect two volume elements within the molecule. It is the real space representation and is therefore much more intuitive that the raw data. As shown in figure 2 the shape of the pair distribution function provides information about the shape of the molecule.

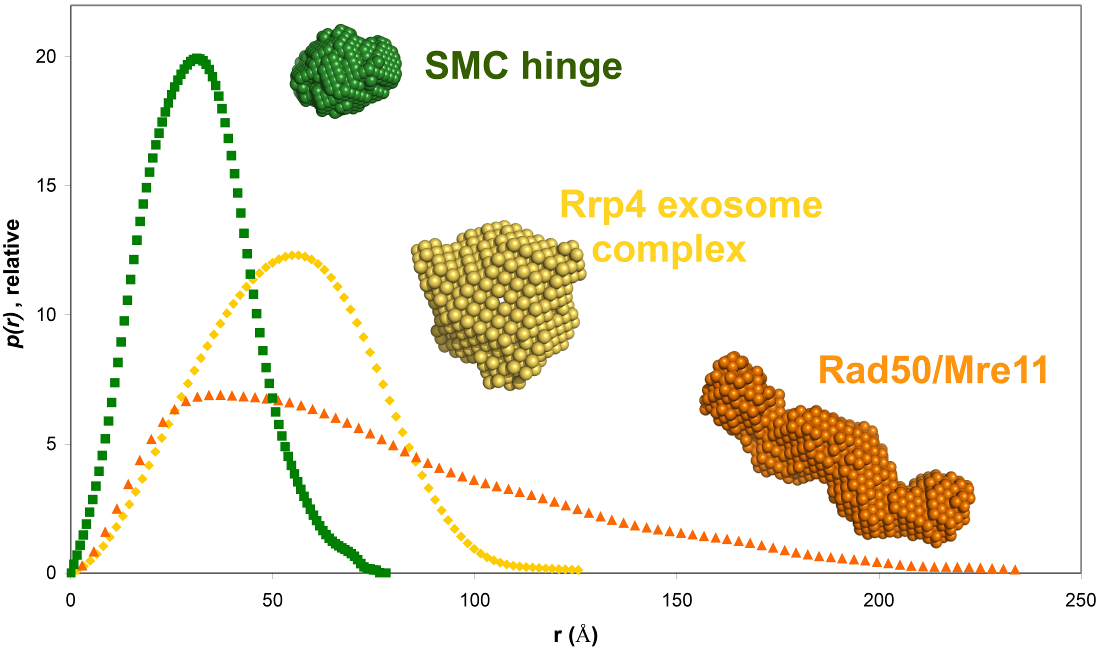


Figure 2: p(r) distribution and bead model of three proteins with different overall shape. The following proteins are shown: in dark green the small and globular hinge domain of the SMC protein from *Pyrococcus furiosus*, in yellow the barrel-like structure of the *Archaeoglobus fulgidus* Rrp4 exosome complex and in orange the elongated Rad50/Mre11 complex from *Thermotoga maritima*.

SAXS structures and the corresponding *p(r)* distribution of three different structure are shown:

the Rad50/Mre11 complex from *Thermotoga maritima* (Bemeleit, 2008), the hinge domain of *Pyrococcus furiosus* SMC protein (Griese, unpublished) and the Rrp4 exosome complex from the *Archaeoglobus fulgidus*. A globular protein like the SMC hinge domain gives a fairly symmetric shape with one quite narrow peak. An elongated protein like the Rad50/Mre11 complex gives a skewed *p(r)* distribution and a barrel-like protein like the exosome gives one peak that is broader than it would be for a sphere with the same R_g.

Reliable computation of *p(r)* is essential for further analysis in terms of three dimensional models. Therefore one has to be very careful when extrapolating I(s) in both directions using the Guinier approximation and Porod's law not to produce artifacts.

1.2.2.5 Ab initio Modeling

As described above some overall parameters can be computed directly from the scattering data without the calculation of any structural model: the radius of gyration (R_g), the molecular mass and the maximum particle diameter (D_{max}). It is clear that the reconstruction of a three-dimensional structure from a one-dimensional scattering curve is a more complicated problem.

In 1998 Chacon et al. showed that the electron density map can be approximated in terms of an assembly of beads or dummy atoms. This bead model can be used to fit the scattering data. To achieve reasonable models it helps to add physical constraints, like imposing a uniform density on the interior of the molecule. Fitting the data leads to a multidimensional minimization problem that can be numerically solved. The program DAMMIN (Svergun, 1999) uses a simulated annealing procedure which imposes a compactness criterion. Improvement for protein solution structure calculations was achieved by the program GASBOR (Petoukhov and Svergun, 2003) where the loose composition of beads (dummy atoms) is changed to a string of beads in the size of an average amino acid (dummy residues). GASBOR is the method for structure calculation used in this thesis. It compares thousands of configurations of a chain of dummy residues (as many as amino acids in the protein of interest) to the measured scattering curve. Like in DAMMIN, simulated annealing is used with additional penalties for non-protein-like density. GASBOR is always run at least 10 times and the results are aligned and averaged for a final shape. Through comparison of the different models in the context of compatibility one can evaluate the reliability of the final model. In a perfect monodisperse system all runs converge on a similar structure.

It is surprising that each existing three dimensional structure of a protein should be described by one unique scattering curve, but Sokolova et al. (2003) could show that all deposited structures in the protein data bank have distinguishable calculated SAXS profiles. However, current SAXS *ab initio* shape predictions are more accurately low resolution envelopes (~15Å), whose calculated scattering curves are consistent with the experimental scattering curve. Very basic protein properties are used to significantly reduce the number of acceptable three dimensional shapes. The comparison of *ab initio* SAXS structures of proteins with their crystal structure demonstrates the surprisingly high accuracy of the programs.

However, the confidence level in shape reconstruction has to be improved. A more quantitative measure of the solution uniqueness is needed, especially when no structural information about the sample is available. But as soon as structural information is available, GASBOR is extremely useful to gain additional information about a system. Moreover, mixed populations have to be handled with care, as the *ab initio* prediction will then give averaged conformations and can be very misleading (Heller, 2005). Similar to the results from other single techniques, SAXS results have to be considered carefully for their reliability.

1.2.2.6 Computation of Scattering Patterns from Crystal Structures

For the comparison of SAXS data with atomic models from crystallography one can calculate *ab initio* solution structure models and superimpose them with the crystal structure. This may take some time, as the calculation of *ab initio* models needs a lot of computational power. However, the data of the two different experiments can be compared on the level of SAXS scattering curves: it is much faster to calculate the theoretical scattering curve of a known crystal structure and compare this with the solution scattering data. First, the *p(r)* distribution is calculated by evaluating all interatomic distances in the structure. The commonly used program CRYSOL (Svergun et al., 1995) uses spherical harmonics envelopes that cover the entire model and then transforms the *p(r)* distribution to an intensity profile.

The main problem in evaluating the solution scattering from atomic coordinates is the correct calculation of the influence of solvent scattering. Thereby the hydration shell that surrounds macromolecules in solution has to be considered. The observed scattering of a protein is mainly given by the difference between the scattering of the protein with its ordered solvation layer and the excluded volume that takes into account the scattering of displaced solvent. The scattering of the excluded volume can be determined by defining the shape of the protein, filling it with electron density equivalent to bulk solvent and calculate the scattering (Fraser et al., 1978).

For computational determination of the solvation layer of a protein, the hydration shell is modeled by explicitly placing water molecules on the surface (Fujisawa et al., 1994) or by surrounding the particle by a continuous envelope representing the solvation shell of 3Å with a density that can differ from both bulk density and the solute (Svergun et al., 1995).

1.3 Biochemical Activity Assays

Determination of protein structures is a major step to understand protein functions. However, structural data only gives snapshots of all the motion and reactions that take place in the cell. These pictures indeed help to make progress in understanding a particular problem. However, for a complete understanding it is inevitable to identify and analyze the activity of a protein by biochemical activity assays. For some very common enzymatic activities, a functional assay is already established. Still the assay conditions have to be adapted to the present system. In this thesis, a protein with an enzymatic activity that was not yet described was analyzed. In this case an activity assay had to be developed.

When the substrate of a protein is a mono- or polynucleotide, it is often very useful to work with radioactively labeled molecules. The isotope ^{32}P, a β-emitter with a half-life of 14.3 days, is routinely used in life-science laboratories, either directly incorporated in nucleoside mono-, di- or triphosphates or to produce DNA and RNA probes with a radiolabeled 5' phosphate. Radiolabeled substrates have two dominant advantages over fluorescent labels: first the sensitivity is extremely high – only very low concentrations can be detected via phospho-imaging. Second, the incorporated radiolabel does not change the chemical properties of the substrate. Fluorophores are mostly very hydrophobic and therefore can influence the binding behavior of the nucleotide.

To verify the correct functionality of an assay it is essential to perform negative controls. Analysis of the atomic structure of the active site of an enzyme allows to conclude which amino acid side chains are essential for catalyzing the reaction or for substrate binding. The contacts can be via for example hydrogen bonds or ionic interactions. By introducing a point mutation into the protein sequence, the chemical properties of the side chain can be changed and the enzyme may not be able to perform the reaction any more. For example, structural analysis of AAA-ATPases showed that the exchange of a glutamate to a glutamine in the active site completely abolishes activity in most cases (Lammens et al., 2004). Only if the wild type protein and the mutant version – purified in exactly the same way – show differing behavior in the activity assay one can be sure that the activity of the protein of interest is monitored.

In the present thesis two different protein complexes were analyzed structurally and biochemically: the DNA integrity scanning protein A (DisA) from *Bacillus subtilis*, a DNA damage checkpoint protein with a novel enzymatic activity and the archaeal exosome, a macromolecular complex involved in RNA degradation and surveillance.

2 Material and Methods

2.1 Material

Unless specified otherwise, all common chemicals were ordered by Merck (Darmstadt), Roth (Karlsruhe) or Sigma (Deisenhofen) in the highest available purity. The enzymes were obtained from Invitrogen (Karlsruhe), MBI Fermentas (St. Leon-Rot) or New England Biolabs (Frankfurt). All chromatographic material and columns was obtained from GE Healthcare (Munich). Radioactive material (α-, γ-^{32}P-ATP and α-^{32}P-GTP) was purchased from GE Healthcare (Munich) and Hartmann Analytik (Braunschweig). Crystallization screens, crystallization grade reagents and crystallization tools were obtained from Hampton Research (Laguna Niguel, USA), Qiagen (Hilden) or Jena Bioscience (Jena). The DNA oligonucleotides for cloning were ordered from MWG (Martinsried) and RNA oligonucleotides for crystallization and RNAse assays were synthesized from Biomers (Ulm). All used oligonucleotides were HPLC purified. Thin layer chromatography plates for diadenylate cyclase assays were PEI-Cellulose F plates from Merck (Darmstadt).

Bacterial strains used in this theses are XL-1 blue (recA1 endA1 gyrA96 thi-1 hsdR17 supE44relA1 *lac*[F´*proAB lacIqZΔdM15Tn10*(Tetr)]) from Stratagene (La Jolla, USA) and Rosetta (DE3) (F- *ompT hsdSB* (rB$^-$ mB$^-$) *gal dcm lacY1* (DE3) pRARE (CmR)) from Novagen (Schwalbach/Ts). The cells were grown in LB media containing 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).

2.2 Software

For primary analysis of SAXS data different programs from the ATSAS 2.1 software package were used (Konarev et al., 2006). All *ab initio* modeling was performed using GASBOR (Svergun et al., 2001) and multiple model alignment, averaging and superposition of crystal structures with SAXS structures was performed with SUPCOMB and DAMAVER (Volkov and Svergun, 2003).

Data processing of crystallographic data was performed with XDS (Kabsch, 1993), for many other applications suitable programs from the CCP4 program suite were used (CCP4, 1994). Manual building in electron density maps was done with COOT (Emsley and Cowtan, 2004). Refinement of models was performed with CNS (Brunger et al., 1998) and PHENIX (Afonine et al., 2005).

All figures of protein or nucleic acid structures were created with PyMOL from DeLano

Scientific (Palo Alto, USA).

For analysis and quantification of diadenylate cyclase, RNAse assays and EMSAs the ImageQuant Software from GE Healthcare (Munich) was used.

Simulation and curve fitting of the RNase assays was performed with MATLAB from MathWorks (Munich).

2.3 Methods

2.3.1 Cloning, Expression and Purification Methods

2.3.1.1 Cloning

Standard methods for molecular biology were performed according to standard protocols. This includes polymerase chain reaction, restriction enzyme digestion, ligation of DNA fragments, preparation of competent cells, transformation, amplification of plasmid DNA and analysis by agarose gel electrophoresis. Plasmid DNA preparation and purification of DNA fragments was performed with the NucleoSpin®-Plasmid Quick Pure Kit and NucleoSpin® Extract II (Macherey-Nagel, Düren), respectively.

Point mutations were introduced by PCR-based site-directed mutagenesis (Ho et al., 1989). Two complementary oligonucleotides encoding the desired mutation are used to generate two DNA fragments with overlapping ends in the first amplification reaction (see table 1 for a list of all 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). In the case of bicistronic vectors the whole cassette with both genes was amplified. Further employed standard methods for working with and analysis of proteins like for example SDS-PAGE, determination of protein concentration and protein concentrating were used as described in Sambrook and Russell 2001.

Table 1: Oligonucleotides used for cloning

Oligonucleotide	Sequence (5' → 3' direction)
exosome forward and reverse primer	
afRrp41 for NdeI	AAAAAA **CATATG** TCGGAATTCAATGAAAAACCAGAA
afRrp41 rev Hind III	AAAAAA **AAGCTT** TCAGGCATCTTCACCACCCTCTG
afRrp42 for Nco I	AAAAAAA **CCATGG** GCCCTGAAGACATCCTTGTGGACATT
afRrp42 rev Not I	AAAA **GCGGCCGC** TTAAATTTCCTTAAATTTCTCCCTCAG
afRrp4 for Nco I	AAAAAAA **CCATGG** GCAGGAAGATAGTACTGCCAGGAGAT

Table 1: continued

Oligonucleotide	Sequence (5' → 3' direction)
afRrp4_cHis rev Not I	AAAAAAA **GCGGCCGC** TTGAATTCCGACATCTGCCTTCCT
afCsl4 for Nde I	AAAAAA **CATATG** AGATTCGTAATGCCGGGAGAT
afCsl4 for S1 Nde I	AAAAAA **CATATG** ATGGGAGATGTTGTCCTCGGCA GA
afCsl4 rev Hind III	AAAAAA **AAGCTT** CTACCACTCTCCCTTGCCGTAAT
primer for mutagenesis	
afRrp42_Y70A_for	CCGGGCGAGCCC **GCT** CCGGACACCC
afRrp42_Y70A_rev	GGGTGTCCGG **AGC** GGGCTCGCCCGG
afRrp41_C43A_for	CTGACGGGTCT **GCG** TACCTTGAAATG
afRrp41_C43A_rev	CATTTCAAGGTA **CGC** AGACCCGTCAG
afRrp41_C140S_for	GCAGGACAGCC **AGT** CTGAATGCTGC
afRrp41_C140S_rev	GCAGCATTCAG **ACT** GGCTGTCCTGC
afRrp42_C248A_for	GATGTGAGCATAAAC **GCC** GCGAGGAAGCTGAG
afRrp42_C248A_rev	CTCAGCTTCCTCGC **GGC** GTTTATGCTCACATC
afRrp41_Cys_STOP_Hind_rev	TTTTT **AAGCTT** TCA **GCA** GGCATCTTCACC
afRrp41_K51E_for	CCTTGAAATGGGGAAGAAC **GAA** GTAATTGCAGCGGTTTTC
afRrp41_K51E_rev	GAAAACCGCTGCAATTAC **TTC** GTTCTTCCCCATTTCAAGG
afRrp41_K37C_for	GCAAGCGTTTTA **TGT** AGGGCTGACGGG
afRrp41_K37C_rev	CCCGTCAGCCCT **ACA** TAAAACGCTTGC
afRrp42_D143C_for	ATCCACGCTCTT **TGT** GATGACGGAAAC
afRrp42_D143C_rev	GTTTCCGTCATC **ACA** AAGAGCGTGGAT

2.3.1.2 Protein Expression

A list of all used expression plasmids is given in table 2.

Table 2: Expression plasmids

#	insert	vector	restriction sites	tag	comments
1	afRrp41-afRrp42	pET-21	Nde I, Not I	-	bicistronic
2	afRrp41-afRrp42	pET-29	Nde I, Not I	C-6xHis	bicistronic
3	afRrp4	pET-29	Nco I, Not I	C-6xHis	-
4	afCsl4	pET-21	Nde I, Hind III	-	-
5	afCsl4	pET-28	Nde I, Hind III	C-6xHis	-
6	afCsl4_S1+ZnR	pET-28	Nde I, Hind III	C-6xHis	Csl4 missing the N-terminal domain
7	afRrp41 (R65E)-afRrp42	pET-21	Nde I, Not I	-	bicistronic – neck mutation
8	afRrp41 (D180A)-afRrp42	pET-21	Nde I, Not I	-	bicistronic – active site mutation
9	afRrp41 (R65E+D180A)-afRrp42	pET-21	Nde I, Not I	-	bicistronic – active site and neck mutation
10	afRrp41-afRrp42 (Y70A)	pET-21	Nde I, Not I	-	bicistronic – tyrosine mutation
11	afRrp42 (C248A)	pET-28	Nco I, Not I	-	Rrp42 for fluorescent labeling
12	afRrp41 (C43A+C140S)	pET-21	Nde I, Hind III	N-6xHis / C-Cys	Rrp41 for fluorescent labeling
13	afRrp41 (K51E)-afRrp42	pET-29	Nde I, Not I	C-6xHis	bicistronic – interface mutation
14	afRrp41 (K37C)	pET-21	Nde I, Hind III	-	cysteine mutations for crosslink experiment
15	afRrp42 (D143C)	pET-28	Nco I, Not I	C-6xHis	

For overexpression of recombinant proteins, competent *E. coli* Rosetta (DE3) cells (Novagen, Schwalbach/Ts) were always freshly transformed with plasmid DNA carrying the gene of interest. In case of coexpression of two subunits, plasmids with resistance for two different antibiotics were chosen. Cells were grown at 37°C in LB medium in the presence of the appropriate antibiotics to an optical density at 600 nm of 0.8 - 1.0. Protein expression was induced with 0.17 mM IPTG (Roth, Karlsruhe) and temperature was cooled down to 18°C. After shaking for approximately 15 h at 18°C cells were harvested by centrifugation at 4°C. Cell pellets were flash frozen in liquid nitrogen and stored at -80°C.

DisA proteins were cloned and purified by Gregor Witte as described in Witte et al., 2008.

2.3.1.3 Purification of Exosome Proteins and Complexes

Clones of wild-type exosome constructs were obtained from Katharina Büttner (Büttner, 2005). For purification, cells were resuspended in lysis buffer (for all buffers, see table 3) and disrupted by sonication. Cell debris was removed by centrifugation. During a heat step of 80°C for 15 minutes most *E. coli* protein impurities were heat denatured and removed by centrifugation. His-tagged constructs were first purified using 3ml Ni^{2+}-NTA material (Qiagen, Hilden) using a batch method. The resin was subsequently washed with wash buffer I and for nucleic acid removal with high salt wash buffer II. Protein was then eluted with elution buffer. Fractions were analyzed quantitatively for protein content using Bradford protein assay (Bio-Rad, Munich). Protein containing fractions were pooled, diluted 1:4 with dilution buffer, and further purified by anion exchange chromatography (HiTrap Q, GE healthcare, Munich), only in the case of the Rrp41-Rrp42-Rrp4 complex with cation exchange chromatography (HiTrap SP, GE healthcare, Munich) using the Äkta System (GE Healthcare, Munich). Prior to the experiments, the ion exchange columns were equilibrated with buffer A and protein was eluted with a gradient of 20 column volumes from buffer A to buffer B. Peak fractions were pooled, concentrated with centrifugal devices (Amicon® Ultra, Millipore, Billerica, USA) and loaded onto a Superdex S200 26/60 size exclusion chromatography column equilibrated with gelfiltration buffer). Peak fractions were concentrated to different concentrations according to the following experiments, flash frozen in liquid nitrogen and stored at -80°C. Cysteine mutants for labeling with fluorescent dyes were purified in the constant presence of 2mM β-mercaptoethanol or 5mM DTT to avoid misfolding.

2.3.1.4 Purification of Exosome Complexes with Bound RNA

For structural studies the Rrp41-Rrp42 and the Rrp41-Rrp42-Rrp4 complex were purified in complex with endogenously bound *Escherichia coli* RNA. To achieve a stable and homogeneous protein-RNA complex, some changes in the purification strategy were performed: the Ni-material was not washed with high salt when the complex was bound. In all buffers the salt concentration was never higher than 250mM. After the Ni-column the complex was loaded on an anion exchange column to remove as much unbound nucleic acids as possible. More than one size exclusion chromatography column was run, to assure the total removal of free RNA. Only when only one distinct peak was eluted, the fractions were pooled, concentrated and flash frozen.

Table 3: Buffers used for protein purification and biochemical assays

Buffer	Composition
protein purification	
lysis buffer	50mM Tris pH 7.4 (25°C), 250mM NaCl
wash buffer I	50mM Tris pH 7.4 (25°C), 250mM NaCl, 10mM imidazole
wash buffer II	50mM Tris pH 7.4 (25°C), 1.5M NaCl, 10mM imidazole
elution buffer	50mM Tris pH 7.4 (25°C), 250mM NaCl, 250mM imidazole
dilution buffer	50mM Tris pH 7.4 (25°C)
buffer A	50mM Tris pH 7.4 (25°C), 100mM NaCl
buffer B	50mM Tris pH 7.4 (25°C), 1M NaCl
gelfiltration buffer	20mM Tris pH 7.4 (25°C), 200mM NaCl
biochemical assays	
diadenylate cyclase buffer	40mM Tris pH 7.5 (25°C), 100mM NaCl, 10mM $MgCl_2$
TLC running buffer	1M NH_4SO_4, 1.5M KH_2PO_4, pH 3.6
RNAse buffer	20mM Tris pH 7.6 (25°C), 60mM KCl, 10% glycerol, 2mM DTT, 0.1% PEG 8000
stopping buffer	0.75 g/l bromphenol blue, 0.75 g/l xylene cyanol, 25% (v/v) glycerol, 50% formamide
20x TBE buffer	1.8M Tris, 1.8M boric acid, 40mM EDTA
labeling buffer	50mM Tris (pH 7.5), 150mM NaCl and 25% glycerol

2.3.2 Biochemical Assays

2.3.2.1 Diadenylate Cyclase Assays

For diadenylate cyclase assays 3.5 µM DisA (monomer concentration) was incubated with or without different amount of varying DNA molecules in 40 mM Tris (pH 7.5), 100 mM NaCl and 10 mM $MgCl_2$. The reaction was started with the addition of 100 µM ATP containing 1:600 radioactively labeled α-^{32}P-ATP (3000Ci/mmol, GE Healthcare) and incubated at 50°C for *Tma*DisA and at 30°C for *Bsu*DisA. As negative control the same assay was performed with GTP and α-^{32}P-GTP instead of ATP. At different time points 0.5 µl of the reaction mix was applied on a PEI-Cellulose F thin-layer chromatography (TLC) plate (Merck, Darmstadt) and left to dry. The application of the protein on to the TLC plate and the drying stops the reaction and makes it possible to follow the cyclase activity through time.

For analysis the TCL plates were developed in 1 M NH_4SO_4 and 1.5 M KH_2PO_4, pH 3.6, dried and exposed on a Storage Phosphor Screen (GE Healthcare, Munich) and scanned with a StormScanner (GE Healthcare, Munich). The intensity of the various radioactive species was calculated by quantifying the intensities of the relevant spots using the ImageQuant software.

2.3.2.2 RNAse Assays

2.3.2.2.1 Radioactive labeling of oligonucleotides

To visualize the different oligonucleotides used in RNase assays, they were radioactively labeled. The radioactive labeling of RNA was carried out by mixing 100 pmol of the relevant oligonucleotide with 30 µCi γ-^{32}P-ATP, and 10 U T4 Polynucleotide Kinase (Fermentas, St. Leon-Rot) and the supplied buffer in a 20 µl reaction for 1 h at 37°C following the provided instructions. The reaction contained additionally 0.8 U/µl RNasin (Promega, Mannheim). Free ATP was removed from the oligonucleotides using MicroSpin G-25 columns (GE Healthcare, Munich).

2.3.2.2.2 Assay Conditions

To analyze the RNase and polymerization activity of exosome complexes from A*rchaeoglobus fulgidus*, the proteins were incubated in 20 mM Tris (pH 7.6), 60 mM KCl, 10 % glycerol, 2 mM DTT, 0.1 % PEG 8000. In addition different amounts of varying divalent ions like Mg^{2+} or Mn^{2+} were used. The reaction was started by adding radioactively labeled RNA molecules of different lengths in changing concentrations and incubated at 50°C. At different time points 2µl of the sample were mixed with 2 µl stopping buffer (0.75 g/l bromphenol blue, 0.75 g/l xylene cyanol, 25 % (v/v) glycerol, 50 % formamide) and analyzed by gel electrophoresis.

2.3.2.2.3 Urea Gel Electrophoresis

Gels for analysis of oligonucleotide products was performed in a Sequi-Gen GT sequencing cell (BioRad, Munich) for 21 cm x 40 cm x 0.4 mm gels. Sinalization of the glass plates was performed to avoid sticking of the gel to the walls. 40 ml gel solution containing 8 M urea, 20 % polyacrylamide and TBE buffer (table 3) was prepared and the polymerizing reaction of the gel was started by addition of 18 µM TEMED and 180 µl 10% APS (ammonium peroxodisulfate). Gels were poured horizontally and run in TBE buffer. Gels were pre-run at 40 Watt, until the gel temperature reached 45 – 50°C. After thorough washing of the gel wells, 3 µl of sample were loaded and the gel was run for 2 h at 40 Watt.

For analysis of RNA bands the gel was removed from the cell and exposed on a Storage

Phosphor Screen (GE Healthcare, Munich) for 1 h and scanned with a StormScanner (GE Healthcare, Munich). The intensity of the various radioactive species was calculated by quantifying the intensities of the relevant bands using the ImageQuant software.

2.3.2.3 EMSA (electrophoretic mobility shift assay)

To test binding of exosome complexes and single exosome subunits to RNA, binding assays were performed with different radioactively labeled oligoribonucleotides as substrate (see 2.3.2.2.1). For each reaction, different concentrations of protein were incubated with 50 fmol RNA substrate in the same buffer that was used for the activity assays in a 20 µl reaction for 10 minutes at room temperature. Reactions contained additionally 0.8 U/µl RNasin (Promega, Mannheim). The samples were mixed 5:1 with loading buffer (1.5 g/l bromophenol blue, 1.5 g/l xylene cyanol, 50% (v/v) glycerol), and RNA:protein complexes were resolved on a native 6% polyacrylamide gel. Gels were run in 0.5x TB-Buffer for 2 hours at 80 V. The gels were exposed on a Storage Phosphor Screen (GE Healthcare, Munich) for 1 h and scanned with a Storm Scanner (GE Healthcare, Munich). The intensity of the various radioactive species was calculated by quantifying the intensities of the relevant bands using the ImageQuant software.

2.3.2.4 Fluorescent Labeling of the Exosome for Single-Molecule Experiments

In collaboration with the group of Taekjip Ha at the University of Illinois, USA single-molecule FRET (fluorescence resonance energy transfer) experiments were performed. For these experiments fluorescently labeled exosome complexes had to be produced. Labeling was done with a Cy3 mono-functional maleimide which is particularly suitable for the selective labeling of molecules containing free sulfhydryl groups, such as cysteine residues. Therefore a single cysteine mutant of the exosome had to be produces to ensure specific labeling. All existing cysteines in the Rrp41 protein (C43 and C140) were mutated to alanines and an additional cysteine was introduced at the C-terminus. The protein was purified together with Rrp42 in presence of reducing agents.

Labeling of the complex with Cy3 was performed using the Cy3 Maleimide mono-Reactive Dye pack from GE Healthcare (Munich) in labeling buffer (50 mM Tris (pH 7.5), 150 mM NaCl and 25% glycerol) according to the supplied protocol. To optimize the labeling efficiency the reaction was carried out under oxygen free atmosphere in a glove box. After incubation over night the labeled protein had to be separated from remaining free dye. Therefore the complex was applied to a HiTrap Desalting column (GE Healthcare, Munich), followed by an anion exchange column. As the Cy3 dye is extremely hydrophobic it does not bind to the column and can so be efficiently removed from the protein complex. The labeling efficiency was determined by measuring an absorption spectrum and using the molar extinction coefficients of Dye and protein: Cy3: ε_{552nm} = 150 000 M^{-1} cm^{-1} and Rrp41-Rrp42 hexamer: ε_{280nm} = 61 200 M^{-1} cm^{-1}, the calculation has to be corrected for the absorbance of the dye at 280 nm (approximately 8% of the absorbance at 552 nm).

2.3.3 Crystallization and Structure Determination

2.3.3.1 Exosome in Complex With Small RNA Molecules

120 µM Csl4-exosome wt and Y70A^{Rrp42} (=27 g/l) in gel filtration buffer was incubated with 400 µM RNA (3.3fold excess) for 10 minutes on ice. Different RNA molecules were used: 10mer polyA, 8mer polyU, 6mer CCUCCU and 6mer CCACCA. The protein:RNA complex was crystallized by sitting drop vapour diffusion technique by mixing 1 µl protein and 1 µl of reservoir solution (0.1 M NaAcetate, pH 4.6, 30% MPD, 100 mM NaCl) at 20°C.

In addition the Csl4-exosome was crystallized without RNA and the crystals were soaked with the four different RNA molecules. For the soaking experiment Csl4 exosome alone was crystallized in the same condition and 1 µl 0.5 mM RNA in the crystallization solution was pipetted to the drop. After 2 h crystals were washed for 10 min in reservoir solution.

Prior to data collection, crystals were directly mounted in nylon loops and flash frozen in liquid nitrogen. Datasets was recorded at the ID-14-2 beamline (ESRF, Grenoble, France) to 2.6 Å. Data from 180 images (0.5° rotation) were integrated and scaled with XDS (Kabsch 1993). A model of the apo-Csl4-exosome complex (Büttner et al., 2005) was used as a search model for molecular replacement using PHASER (CCP4, 1994). Refinement to 2.6 Å was performed with CNS (Brunger et al., 1998) and PHENIX (Afonine et al., 2005). In the additional electron density a 6mer RNA molecular was positioned using COOT (Emsley and Cowtan, 2004). Refinement of the complete complex at 2.6 Å followed iterative cycles of manual model completion with COOT and refinement with CNS.

2.3.3.2 Csl4 S1-ZnR Domain

The truncated version of Csl4 (lacking the N-terminal domain) was concentrated to 5.5 mg/ml and screened for crystallization conditions. The protein was crystallized by sitting drop vapour diffusion technique by mixing 1 µl protein and 1 µl of reservoir solution. Crystals grew in 100 mM HEPES, pH 7.5, 20% PEG 4000 and 10% isopropanol. The condition was refined to

100 mM HEPES pH 7.8; 18% PEG 4000, 10% isopropanol and 10% glycerol and crystals were directly flash frozen in liquid nitrogen. A dataset was recorded at the PX beamline (SLS, Villigen, CH) to 1.9 Å. Data from 360 images (0.5° rotation) were integrated and scaled with XDS (Kabsch 1993). Refinement to 1.9 Å was performed with CNS (Brunger et al., 1998) and PHENIX (Afonine et al., 2005).

2.3.3.3 Exosome core with copurified RNA

Approximately 10 g/l exosome complex with copurified RNA in gel filtration buffer was crystallized by sitting drop vapour diffusion technique by mixing 1 µl protein and 1 µl of reservoir solution (0.1 M HEPES, pH 7.5, 5% MPD, 10% PEG 6000) at 20°C. Crystals grew after some days and were mounted in nylon loops and flash frozen in liquid nitrogen. A dataset was recoded to 3 Å at the PX beamline (SLS, Villigen, Switzerland). Data from 360 images (1° rotation) were integrated and scaled with XDS (Kabsch 1993). A model of the exosome core (Büttner et al., 2005) was used as a search model for molecular replacement using PHASER (CCP4, 1994). Refinement to 3.0 Å was performed with CNS (Brunger et al., 1998) and PHENIX (Afonine et al., 2005).

2.3.4 SAXS Experiments and Data Processing

Results from all SAXS measurements are summarized in table A1 in the appendix.

2.3.4.1 DisA

T. maritima DisA protein for SAXS experiments was supplied by Gregor Witte in gel filtration buffer. Samples were measured at the SIBYLS beamline (Advanced Light Source, Berkeley, CA, USA) with a concentration of 2, 4, 6 and 8 mg/ml. PRIMUS (Konarev et al., 2003) was used to primary analyze the data and exclude the possibility of protein aggregation. In the case of concentration dependent aggregation or attractive forces between molecules the normalized scattering curves would diverge at small s-values. To exclude the possibility of radiation damage three different measurements with the same sample were performed: one short (6 seconds) one long (60 seconds) and a final short (6 seconds) experiment. In the case of radiation damage during the long experiment, the first and the third scattering curve would differ. Good agreement between the two is a clear indicator of a protein structure that is insensitive to X-rays. The radius of gyration was calculated using the Guinier plot in the linear region (constraint: $s \cdot R_g < 1.3$) and the calculation of the pair distribution function was done with GNOM within PRIMUS. Thereby the correct maximum particle diameter D_{max} was iteratively determined by evaluating the resulting R_g value, the I(s)-fit and the shape of the p(r)-distribution. Theoretical SAXS curves of the crystal structure were calculated using CRYSOL (Svergun et al., 1995). *Ab initio* modeling of the DisA solution structure was done with GASBORp (Svergun et al., 2001). Thereby different particle symmetries were assumed and more than 10 identically calculated models were aligned and averaged using DAMAVER and SUPCOMB (Volkov and Svergun, 2003).

2.3.4.2 Exosome

The different exosome complexes were measured in at least three concentrations in gel filtration buffer at the SIBYLS beamline (Advanced Light Source, Berkeley, CA, USA) and the X-33 beamline (EMBL/DESY, Hamburg). Primary data analysis was performed like with the DisA protein (see 2.3.4.1). Pair distribution functions were calculated with GNOM (Svergun, 1992) and *ab initio* models were calculated with GASBORp and GASBORi (Svergun et al., 2001), were GASBORp uses the pair distribution function for *ab initio* modeling and GASBORi the non-transformed scattering intensity. For apo-exosome complexes a P3 symmetry was assumed. Alignment of at least 15 *ab initio* models and averaging was performed with DAMAVER and SUPCOMB.

3 The DNA Integrity Scanning Protein A (DisA)

3.1 The DisA Protein and its Influence on Sporulation

During cell division all organisms have to promote correct replication of chromosomes and their distribution to the two daughter cells. Both processes are very complex and require as intact chromosomes as possible. To avoid loss or change of genetic information, cells developed mechanisms to inhibit the progression of cell multiplication when the chromosomes are damaged. These so-called checkpoints (Hartwell and Weinert, 1989) seem to be universal to all cell types. In 2006 the DNA integrity scanning protein A (DisA) of the gram-positive soil bacterium *Bacillus subtilis* has been identified as a new and noteworthy checkpoint protein by Bejerano-Sagie and coworkers.

Bejerano-Sagie et al. identified the *disA* gene as the last gene in an operon that consists of six different genes involved in protein degradation, competence and DNA repair. They could show that the deletion of the *disA* gene has no effect on cell division, sporulation or competence of DNA damage response in vegetative cells. Interestingly, they could show that in the presence of DNA damages, sporulation is inhibited by DisA. On the basis of this finding they examined the expression of early sporulation genes via their transcription factor Spo0A and concluded that DisA is at least partly required to inhibit Spo0A activity in case of occurring damaged DNA .

The authors propose that DisA scans the genome of *B. subtilis* for damage and reports possible abnormalities to the cell-cycle machinery before starting the sporulation process. Using a time-resolved fluorescence microscopy approach with fluorescently labeled DisA in living *B. subtilis* cells, they were able to visualize the movement of DisA foci in the living cell (figure 3).

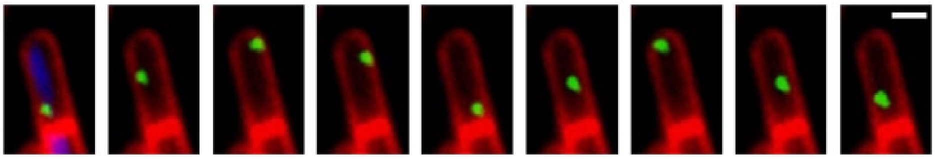


Figure 3: Visualization of moving DisA foci in B. subtilis cells (Bejerano-Sagie et al. 2006). The dynamic localization of DisA-GFP foci is demonstrated by time-lapse microscopy from individual cells before polar division. Cells are stained with FM4-64 (red) and DAPI (blue) and photographed at ~ 300 msec intervals. The scale bar corresponds to 1 µm.

When introducing DNA lesions in the cell they observed that the majority of DisA-GFP (green fluorescent protein) loci were no longer moving but seem to be stalled at various positions within the cell. By introducing a double strand break at a known position they could even presume that the DisA-GFP locus is stopping at the approximate position of the lesion.

As a consequence of their data Bejerano-Sagie et al. proposed the following model: the DisA expression increases when sporulation condition are sensed. At a certain protein level, DisA forms globular structures that move rapidly within the cell and scan for DNA lesions. Upon detection of a lesion the DisA complex looses its mobility and stalls – thereby triggering a cellular response that leads to a temporary block to initiation of sporulation. After repair of the lesion the arrest is abolished and the sporulation processes can be initiated or continued.

The results presented by Bejerano-Sagie et al. give remarkable new insights into the processes in which DisA seems to be involved and in the downstream effects of this checkpoint protein. As the authors focused in their work mainly on *in vivo* data, further detailed biochemical and biophysical analysis of the DisA-protein can provide answers to open questions, such as the size of a DisA focus and its composition. The exact mechanisms of DisA DNA lesion recognition is not at all understood and the question remains, if DisA directly recognizes DNA abnormalities or if one or more interacting proteins are additionally involved. Furthermore, the downstream events triggered by DisA, which can induce sporulation arrest by an unknown signaling pathway, remain to be investigated.

3.2 The Crystal Structure of T. maritima DisA

As already mentioned above, a very powerful method to understand protein complexes, their function and their duty in the living cell is the determination of the atomic structure of the complex. To shed more light on the properties of the DisA protein, the crystal structure of the *Thermotoga maritima* homolog of DisA was solved (Witte et al. 2008), which shows high sequence homology to the *Bacillus* variant of DisA (more than 55 %).

After purification of both *B. subtilis* (Bsu) and *T. maritima* (Tma) DisA proteins, analytical size exclusion chromatography and analytical ultracentrifugation experiments showed that DisA forms an asymmetric-shaped octameric complex in solution under the present conditions.

The obtained crystal structure of *Tma*DisA is shown in figure 4. DisA forms a dumbbell-shaped structure with dimensions of 80 x 55 x 30 Å and consists of three structural domains. Two different globular domains at the N- and C-terminus of the protein are connected by a third domain that contains an elongated bundle of three α-helices. No reasonable structural homologue of this middle domain could be found by database searches, but relatives of the two globular domains could be found:

The C-terminal domain is with only 62 residues relatively small and possesses a typical helix-hairpin-helix (HhH) fold. This motif is known to be associated with DNA phosphate backbone binding in numerous DNA-interacting proteins. Therefore the C-terminal part of DisA is very likely the DNA-binding site.

On the opposite side of the protein, at the N-terminus, a larger globular domain with 145 amino acids is located. Comparison of this domain with structurally similar proteins gave only one structural neighbor in the DALI database (Holm and Sander, 1996). This uncharacterized prokaryotic domain of unknown function has not been further characterized, yet.

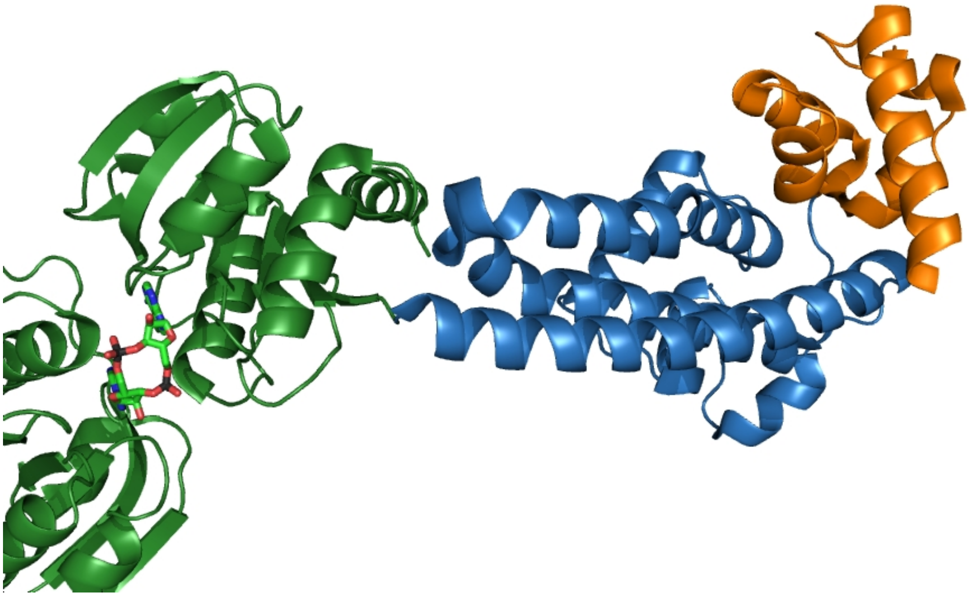


Figure 4: Crystal structure of the DisA monomer from *Thermotoga maritima*. Shown are the three domains of DisA in ribbon representation and the bound ligand as sticks. The smaller C-terminal DNA binding domain is shown in orange, the nucleotide binding domain in green and the connecting helical domain in blue. The N-terminal domain of the second interacting DisA molecule is only partly shown. The ligand is bound by the two opposing DisA N-termini.

Especially worth mentioning is a finding in the structure of DisA that at first could not be explained:

In the electron density a non-protein ligand could be seen, which was clearly coordinated by the N-terminus of DisA (figure 4). This ligand was located at the interface between two opposing DisA molecules and on a symmetry axis. The electron density at this position resembled a cyclic purine nucleoside, but none of the commonly known small biological molecules could be fitted into the density. During crystallization no additive that could explain this density was added so this molecule must have been picked up endogenously or even been synthesized by DisA itself during recombinant expression in the *Escherichia coli* cells. The shape and dimension of the density as well as the chemical nature of the corresponding protein-binding pocket strongly argues for the presence of a bis-(3'-5')-cyclic dimeric adenosine monophosphate (c-di-AMP). By the use of MALDI-TOF mass spectrometry of the whole protein complex, this assumption could be verified.

However, the occurrence of the so far unknown c-di-AMP molecule that is well-coordinated by DisA residues is puzzling. Neither the existence nor the origin and function of this molecule have been described so far. In addition to this, it still has to be proved that c-di-AMP is the natural ligand of DisA, how it was produced and that it was not by accident copurified as an artifact from recombinant expression in *E. coli*.

3.3 Cyclic Purine Nucleotides As Second Messengers

Second messengers are elements of the signal transduction cascade. In the second messenger system a small and thus fast diffusing signaling molecule is rapidly generated and/or released, which then activates effector proteins within the cell to exert a cellular response.

Secondary messenger systems can be activated by diverse means. In the case of small cyclic molecules the increase of the second messenger in the cell is regulated via the activation of cyclases, which synthesize cyclic nucleotides. These small molecules may then go on to exert their effect by binding to and thereby (in)activating effector molecules such as protein kinases, ion channels and a variety of other proteins, continuing and down-streaming the signaling cascade.

3.3.1 cAMP and cGMP

The discovery of the second messenger adenosine 3',5' cyclic monophosphate (cAMP) and the molecular analysis of its role in hormone action (Sutherland, 1972) opened a new era in biological research. Only shortly later guanosine 3',5' cyclic monophosphate (cGMP) was found to be an important player in eukaryotic signal transduction (Fedorov, 1976). cAMP is synthesized by an adenylyl cyclase (AC) from ATP, whereas cGMP is synthesized by the corresponding guanylyl cyclase (GC) from GTP (for chemical structures see figure 5).

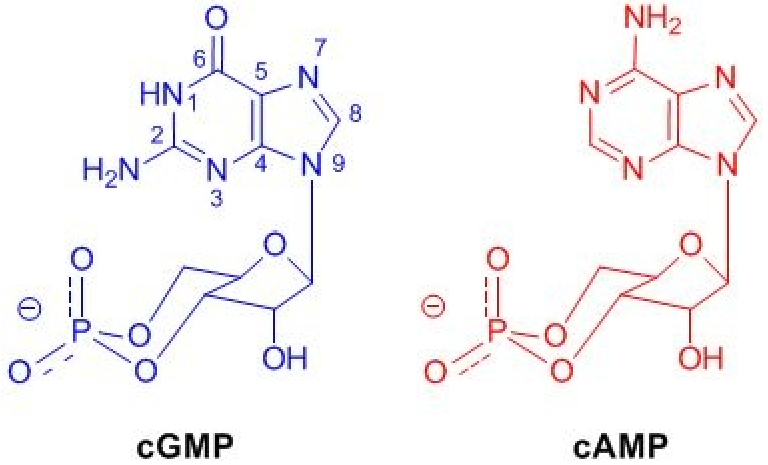


Figure 5: cGMP and cAMP: Cyclic adenosine monophosphate and guanosine monophosphate are formed through introduction of a chemical bond between the 5'-phosphate and the oxygen of the 3'-OH-group.

Both are degraded to nucleoside monophosphates by cyclic nucleotide phosphodiesterases (PDEs). The levels of cAMP and cGMP in the cell are tightly balanced by the opposing actions of purine nucleotide cyclases and PDEs, both of which can occur in multiple isoforms, often with distinct subcellular locations. This arrangement facilitates fine tuning of cellular responses. In their roles as eukaryotic second messengers, cAMP and cGMP can bind to and activate a number of intracellular receptors; these include cyclic nucleotide-dependent protein kinases, cyclic nucleotide-gated ion channels, PDEs, a Rap1 guanine nucleotide-exchange factor (Epac) (de Rooij et al., 1998) and GAF domains. The protein kinases are key regulators of these signaling pathways and their activation triggers a cascade of downstream events that leads to a cellular response. In many prokaryotes, the biological activity of cAMP is mediated through its role in transcriptional activation (Baker and Kelly, 2004; Lucas et al. 2000; Kaupp and Seifert, 2002).

Whereas cAMP signaling is common to both prokayotes and eukaryotes, cGMP does not seem to be present in bacterial cells. However, there is increasing evidence that the cyclic dimer of cGMP, c-di-GMP plays a critical role in bacterial signaling (D'Argenio and Miller, 2004; Jenal, 2004).

3.3.2 C-di-GMP and Associated Enzymes

Although the cyclic nucleotide bis-(3'-5')-cyclic-di-guanosine monophosphate (c-di-GMP, figure 6) was already discovered in the 1980s, its global role in bacterial signaling has become apparent only recently in the view of the growing bacterial genome sequence information available. C-di-

GMP affects various cellular functions principally related to the transition between a motile, single-cell state and an adhesive, surface-attached, multicellular state called “biofilm” (Hall-Stoodley and Stoodley, 2005). C-di-GMP is synthesized out of two GTP molecules by diguanylate cyclases and hydrolyzed again by c-di-GMP specific phosphodiesterases, both of which occur in soluble and membrane-bound variants. Homologs of these proteins are only found in eubacteria, which implies that c-di-GMP signaling is specific to this domain of life. The domain responsible for cyclase activity is named GGDEF domain and the diesterase activity is associated with the EAL domain. Recently, the predicted enzymatic activities of these domains that are named for their most obvious conserved amino acids, was experimentally verified (Paul et al. 2004, Chan et al. 2004, Chang et al. 2001).

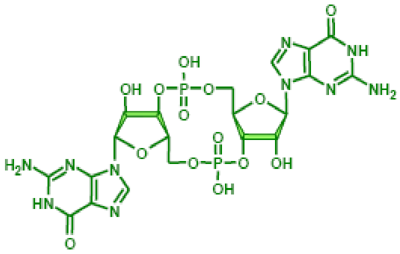


Figure 6: C-di-GMP signaling molecule: One c-di-GMP molecule is synthesized by a diguanylate cyclase from two GTP molecules through release of two pyrophosphates.

3.3.2.1 Diguanylate Cyclase Activity and GGDEF Domains

Proteins with GGDEF domain are very specific for GTP as a substrate and the only detectable product is c-di-GMP. Like the well-characterized adenylyl and guanylyl cyclases (Baker et al, 2004), diguanylate cyclases (DGCs) seem to function as dimers of trimers, whereby oligomerization is an intrinsic property of the GGDEF domains (Ryjenkov et al., 2005). The activity of the GGDEF domains is stimulated by phosphorylation-induced dimerization, which is thought to be mediated by an N-terminal receiver domain. One of the best characterized proteins with GGDEF domain is PleD - named like this, because mutations in the *pleD* gene cause pleiotropic effects (Sommer and Newton, 1989; Aldridge et al., 2003; Paul et al., 2007). PleD is a member of the response regulator family required for pole development in *Caulobacter crescentus* (Aldridge et al., 2003). Biochemical analysis of purified PleD, which consists of two N-terminal receiver domains arranged in tandem and a C-terminal GGDEF output domain, showed that this protein is a diguanylate cyclase. An important progress in understanding GGDEF domains was the atomic structure of PleD in complex with c-di-GMP solved by Chan et al. in 2004 (figure 7).

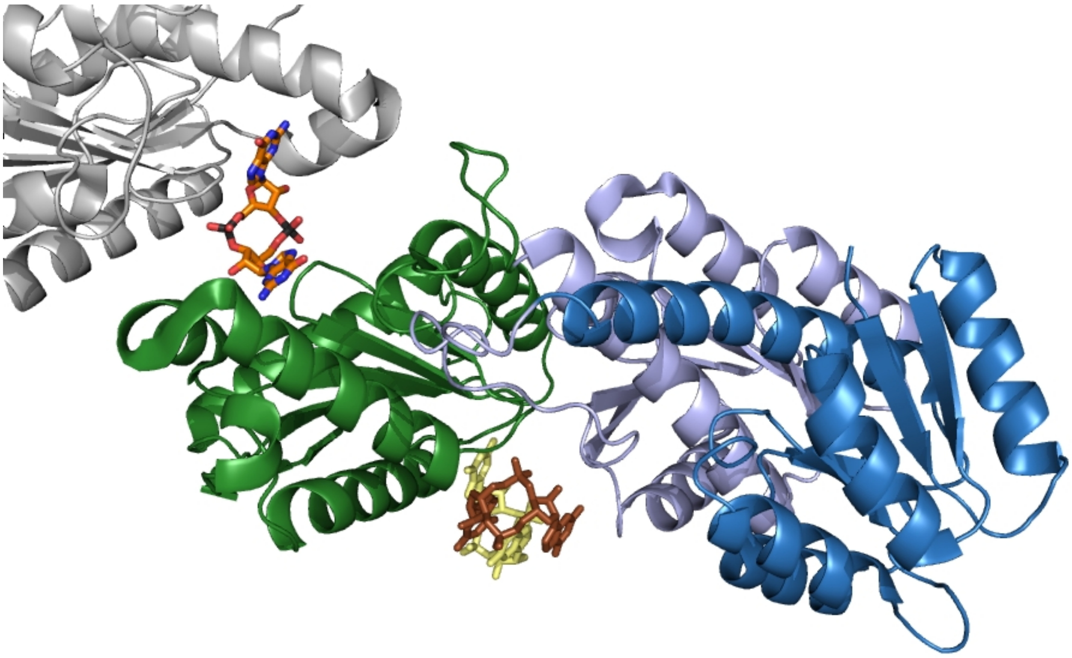


Figure 7: Crystal structure of PleD from *Caulobacter vibiroides* with bound c-di-GMP: The two receiver domains are shown in blue and the catalytic GGDEF domain in green. The c-di-GMP molecules are shown as sticks, the one in the active site is shown in orange and the two c-di-GMP molecules bound to the regulatory site are shown in brown and light yellow.

The diguanylate cyclase domain shows a tertiary structure closely similar to that of the catalytic core of adenylyl cyclases and to the “palm” domain of DNA polymerases. These domains are functionally related and are catalyzing 3'-5' phosphodiester bond formation. Nucleotide/protein interactions in the active side of the GGDEF domain clearly explain the preference of these cyclases for GTP in contrast to ATP. The insights gained from the atomic structure allow to propose a binding mode for GTP and a model for the catalytic mechanism. The PleD crystal structure revealed an additional binding site for c-di-GMP nested in the interface between the GGDEF domain and the central receiver-like domain (figure 7). High-affinity binding of c-di-GMP to this site is responsible for the observed strong non-competitive product inhibition of the cyclase activity and leads to the conclusion that allosteric feedback inhibition in diguanylate cyclases is a general regulatory principle of c-di-GMP signaling.

3.3.2.2 Phosphodiesterase Activity and EAL Domains

Shortly after identifying GGDEF domains as the responsible proteins for diguanylate cyclase activity, the phosphodiesterase activity was unambiguously assigned to EAL domains (Christen at al., 2005). For several EAL proteins it could be shown that c-di-GMP is rapidly hydrolyzed to pGpG, whereas the further hydrolysis to two GMP molecules is much slower and does not seem to be a physiologically relevant activity of this domain. Observations made with isolated EAL domains showed very effective hydrolysis of c-di-GMP, therefore these proteins were proposed to be active as monomers.

Besides, catalytically non-active EAL domains could be found (Schmidt et al., 2005), which led to the conclusion that they might have adopted alternative regulatory functions. Furthermore, the significance and catalytic function of proteins containing both GGDEF and EAL domains is still unclear. All biochemically analyzed enzymes possessed either cyclase or diesterase activity. The existence of bifunctional enzymes is quite likely, because there are bacterial genomes that encode only one GGDEF+EAL protein, but have no separate genes for either a cyclase or a phosphodiesterase.

3.3.2.3 PilZ as First Proposed c-di-GMP Binding Domain

Recently the 118 amino acid PilZ domain was identified as a candidate for c-di-GMP binding (Amikam and Galperin, 2006). It had already been discussed as a candidate, because PilZ domain proteins are part of biosynthesis pathways that are regulated by c-di-GMP. The knockout of the name-giving protein PilZ, which is a stand alone protein in the type IV pili biosynthesis pathway without assigned function, shows the same phenotype as cells lacking the FimX protein, a multi domain GGDEF-EAL protein (Huang et al., 2003). This observation indicates that PilZ is part of a c-di-GMP signaling pathway or even the protein, which is located directly downstream of c-di-GMP.

3.4 The Role of Holliday Junctions and Fork Structures in DNA Repair

The integrity of the genetic information is essential for the sustainment of life. However, it cannot be prevented that damage from various threads (e.g. radiation) occurs to the genetic blueprint. To overcome this problem and to keep the genomic integrity, organisms developed sophisticated repair mechanisms to limit corruption to their DNA. Failure of these systems can be fatal or have severe consequences such as cancers and other heritable or sporadic disorders. Different signals and recognition mechanisms can lead to the activation of the respective repair pathway.

3.4.1 Double Strand Breaks

Different kinds of damage can change the DNA structure. Probably the most severe kind of damages are double strand breaks, in which both strands of the DNA are broken, because this can lead to genome rearrangements. Cells developed two different methods of repairing double strand breaks: non-homologous end joining and recombinational repair (also known as template-assisted repair) (Siede et al., 2006). Recombinational repair requires the presence of a (nearly) identical sequence to be used as a template for repair of the break. The enzymatic machinery responsible for this repair process is closely related to the machinery responsible for homologous recombination during meiosis. Like during homologous recombination, a special DNA tertiary structure called Holliday junction occurs. A Holliday junction is a mobile junction between four strands of DNA (figure 8).

The structure is named after Robin Holliday, who proposed it in 1964 (Holliday, 1964). Because of the fact that these junctions are formed between highly homologous sequences, they can slide up and down the DNA. In bacteria, this sliding (or branch migration) is facilitated by the RuvABC complex (Kowalczykowski, 2000). An important progress in understanding the recognition and binding process was the solution of the atomic structure of RuvA bound to a Holliday junction (figure 8, Ariyoshi et al., 2000). The arrangement of four HhH-motifs in the tetrameric structure of the RuvA protein is obviously essential for the binding to this highly symmetric DNA element.

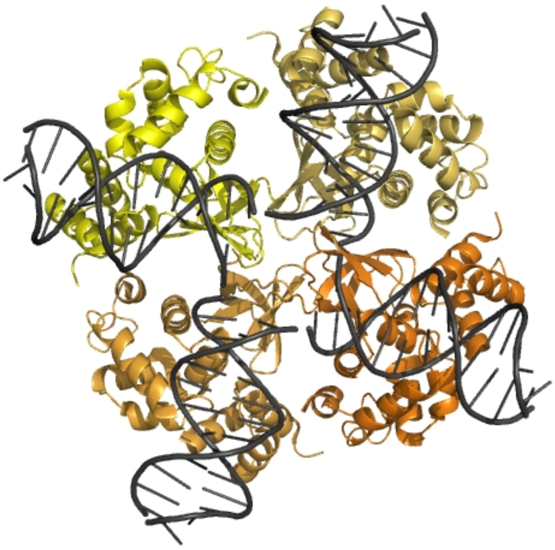


Figure 8: Crystal structure of RuvA from *Escherichia coli* bound to a Holliday junction: The four copies of RuvA are shown in ribbon representation in different colours. The bound Holliday junction is shown in gray. In the crystal a four-fold symmetry axis is located in the middle of the RuvA tetramer and the Holliday junction.

3.4.2 Stalled Replication Forks

During DNA replication three DNA double strands meet together forming the so called replication fork. Normally this replication fork is moving along the DNA sequence together with all involved proteins (Kornberg and Baker, 2005). The replication process is highly processive and accurate, but any obstacles can be a potential thread in this dynamic fork progression. If for example the replisome runs into a site with DNA lesions caused by UV-light, for example thymidine-dimers, this can lead to a complete stop of movement of the replication fork (stalling). The stalled replication fork can then be recognized by the corresponding repair factor and a damage response system is activated.

3.4.3 DNA Damage Checkpoints

The DNA repair mechanisms need some time to recover the correct situation in the cell. Therefore it has to be assured that the cell does not continue cell division as long as the repair enzymes are not yet finished with their work. Therefore, so called cell-cycle checkpoints are activated after the detection of DNA damage. DNA damage checkpoints occur at the G1/S and G2/M boundaries of the cell cycle. Some checkpoint proteins respond to DNA double-strand breaks, disruptions in chromatin structure or to stalled replication forks (Siede et al., 2003). Normally kinases are involved, which phosphorylate downstream targets in a signal transduction cascade, eventually leading to cell cycle arrest.

3.5 Aim Of The Project

The main objective of this project was the understanding of the biochemical function of DisA from *T. maritima* and its *B. subtilis* homolog. Surprisingly, in the crystal structure of *T. maritima* DisA, electron density for a c-di-AMP molecule was identified which could also be proved by mass spectrometry. The existence of c-di-AMP as biological molecule was not described in the literature, yet. As the structure of DisA in complex with c-di-AMP is not a proof for a physiological role of the newly identified molecule, further biochemical studies are needed. It has to be clarified, whether DisA picked up the c-di-AMP molecule from the *E. coli* cells during expression or if it synthesized it *de novo* from ATP. Furthermore, the possibility that DisA actually prefers binding to the thoroughly investigated second messenger c-di-GMP, or even can synthesize c-di-GMP from GTP, has to be ruled out. Using *in vitro* biochemical assays the cellular function of DisA will be analyzed.

Additionally, the correct octameric organization of *T. maritima* DisA is still not clear. Even though size exclusion chromatography and analytical ultracentrifugation could show that DisA is a homo-octamer, the crystal packing of DisA in the *Tma*DisA crystals allow two different octameric assemblies. To get answers to this problem a Small Angle X-ray Scattering (SAXS) approach will be used to identify the correct DisA octamer.

4 Results – DisA

4.1 Identification of the Correct Quaternary Structure of DisA

To understand the function of the DNA integrity scanning protein A (DisA) better, DisA from the thermophile bacterium *Thermotoga maritima* (*Tma*DisA) was crystallized by Gregor Witte and the atomic structure was solved (PDB entry 3C1Y). As shown in figure 9, *Tma*DisA crystallized in a long fiber-like chain of interconnected tetramers. The interface within the tetramer is especially mediated by the helical domain in the middle of the protein (shown in blue) and is with approximately 4400Å^2 quite large and very likely a stable physiological interface. From size exclusion chromatography and analytical ultracentrifugation studies, it was already clear that DisA forms a complex in solution that is definitely larger than a tetramer. The results strongly support the existence of an octameric complex in solution.

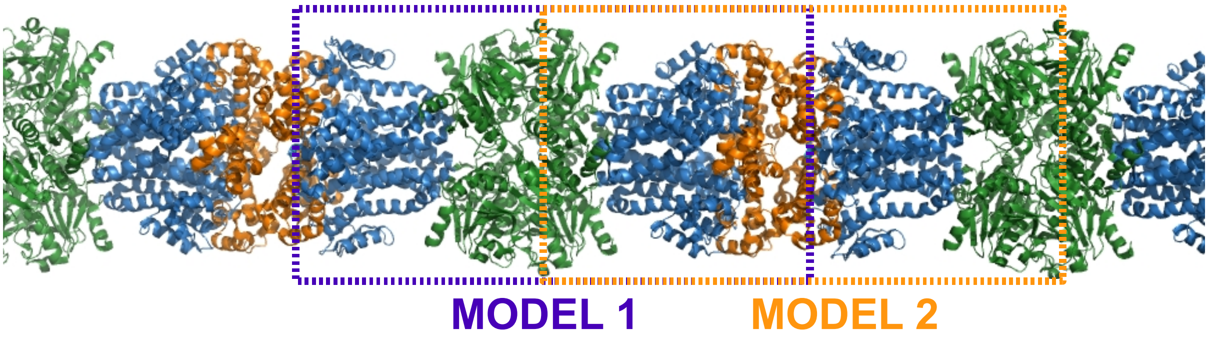


Figure 9: Organization of DisA tetramers in the crystal. The two possible octameric DisA complexes are indicated. In model 1 (blue) two tetramers interact via the N-terminal nucleotide binding domains and in model 2 (orange) the interaction is mediated via the C-terminal DNA binding domains.

Figure 9 depicts the two different possible octameric assemblies, which can be found in the crystal structure: in model 1 the two tetramers interact via the N-terminal nucleotide binding domain (green) and in model 2 the interaction is mediated via the postulated DNA binding domain (orange) at the C-terminus. In model 1 the bound c-di-AMP ligand would be buried in the interface, whereas in model 2 it would be accessible to the solvent.

To identify which of the two octameric forms is existent in solution, SAXS studies were performed. SAXS curves were measured at different protein concentrations and no concentration dependent aggregation or the existence of attractive forces between the complexes could be observed. In addition, the sensitivity of the protein to radiation was analyzed by replicate exposures of the sample to X-rays and the complex seemed to be structurally very stable even after long exposure. The measured scattering curve of DisA is shown in figure 10A.

To determine the radius of gyration the ln(s) vs. s^2 - Guinier plot (Guinier, 1939) was used by analyzing the region of small s-values (see figure 10B). The R_g was determined from the slope of the linear regression curve extrapolated to $s^2 = 0$ to be 52.8 Å for all protein concentrations (considering $s \cdot R_g < 1.3$ to fulfill the constraints for the Guinier approximation).

For transformation of the scattering data into the pair-distribution function, the maximum particle diameter D_{max} was determined to be 176 Å, which fits very well to the longest distance of slightly less than 175 Å in the crystal structure. The calculated p(r)-distribution is shown in figure 10C and shows the typical shape of an elongated, cylindrical molecule.

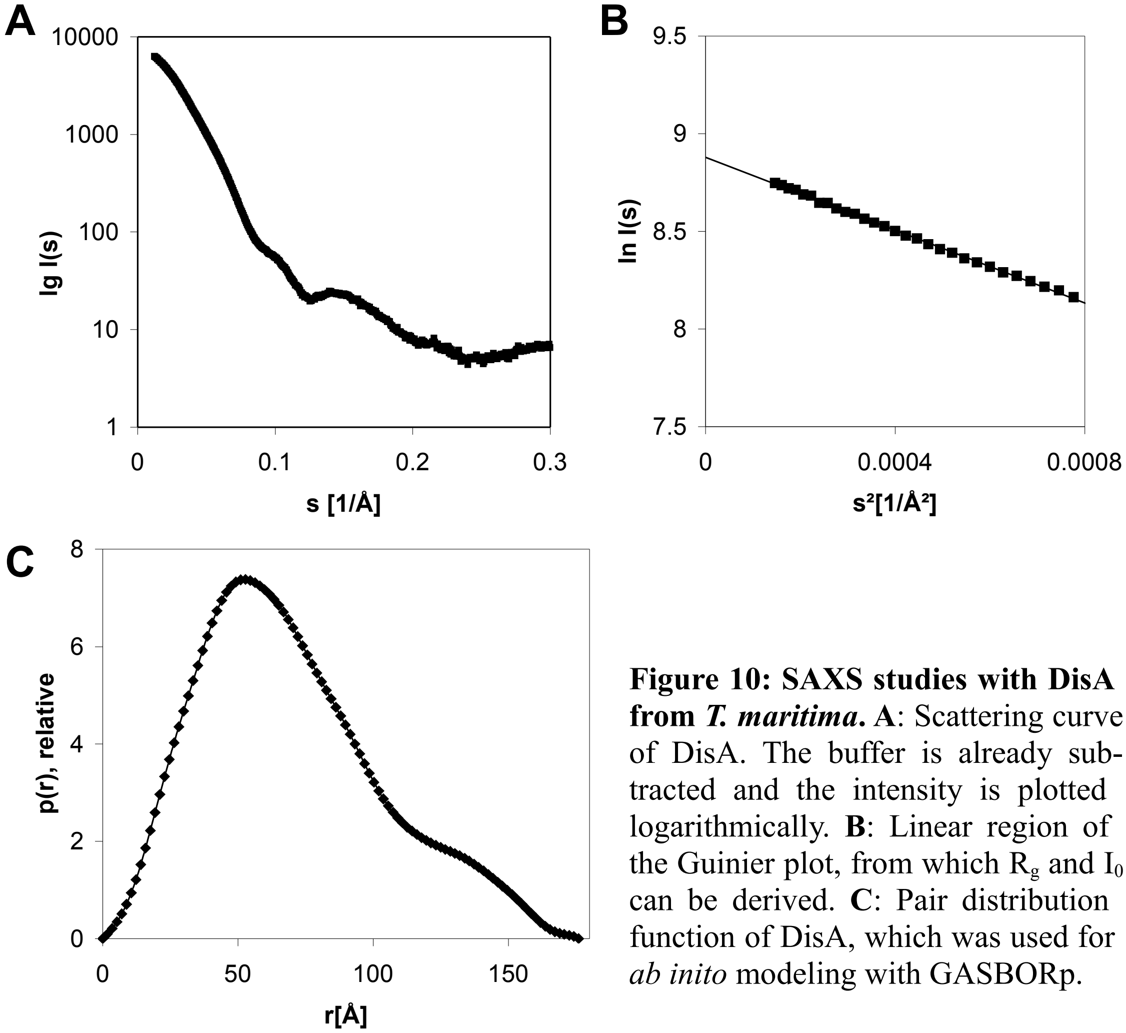


Figure 10: SAXS studies with DisA from *T. maritima*. **A**: Scattering curve of DisA. The buffer is already subtracted and the intensity is plotted logarithmically. **B**: Linear region of the Guinier plot, from which R_g and I_0 can be derived. **C**: Pair distribution function of DisA, which was used for *ab inito* modeling with GASBORp.

The solution structure was determined using the p(r) distribution and the dummy residue approach (Svergun et al., 2001) with different symmetry constraints for modeling. As the DisA structure shows a 4-fold symmetry and a 2-fold symmetry with perpendicular symmetry axes, four different solution structures were calculated: implicating no symmetry (P1), one forced 2-fold symmetry axis (P2), one forced 4-fold symmetry axis (P4) and forced 2- and 4-fold symmetry axis (P42). For all four structures the *ab initio* structure calculation was repeated 10 times using identical input parameters and the structures were then aligned and averaged. The obtained structures are shown in figure 11 as bead models.

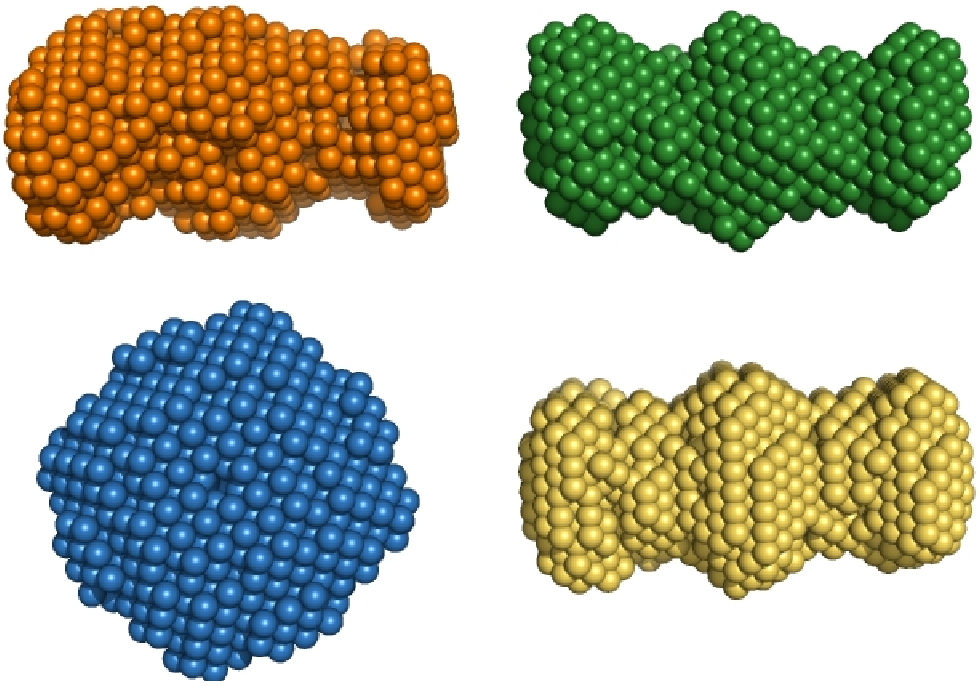


Figure 11: DisA solution structures calculated with different implicated symmetries. The depicted bead models of SAXS solution structures were calculated with GASBORp using the pair distribution function of DisA shown in figure 9C. 10 structures were aligned and averaged. For *ab initio* modeling different symmetry constraints were used: no symmetry (orange), 2-fold (green) and 4-fold (blue) and both 2- and 4-fold symmetry (yellow).

The final bead model calculated with forced single 4-fold symmetry did not give a reasonable result, because the program has located the symmetry axis in a wrong orientation. All other three structures show a similar overall shape.

To determine which of the two possible crystallographic octamers is present in solution, the crystal structures were overlaid with the *ab initio* solution structure using the SUPCOMB program (Kozin and Svergun, 2000). The superposition of the bead model obtained with the 2-fold symmetry with both octamer models is shown in figure 12. Model 1 with facing N-termini of DisA fits very well to the solution structure, whereas for model 2 some structural elements clearly stick out of the envelope and are missing at other positions.

In addition to comparison at structural level, the two models were compared with the SAXS data at the level of the scattering curves.

The theoretical scattering curves of both possible octameric models were calculated using CRYSOL and overlaid with the measured scattering curve of the DisA complex. Figure 13 shows the comparison between calculated and measured scattering data. Whereas the calculated scattering curve for model 1 only shows small deviations from the measured data, model 2 clearly does not fit the experimental scattering curve. Thus, model 1 is existent in solution and the crystal structure of DisA resembles its solution structure very much.

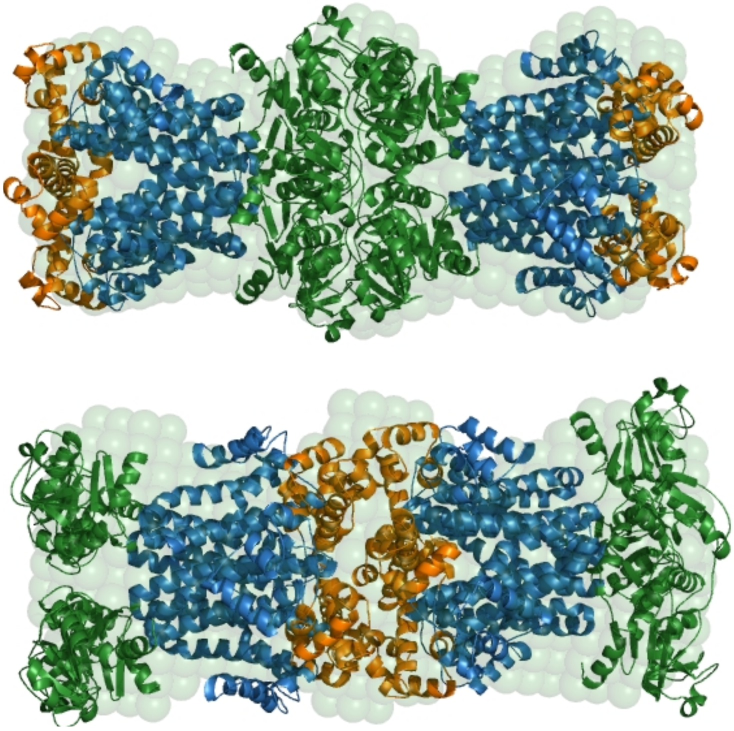


Figure 12: Superposition of the two possible octamer models with the SAXS structure. Clearly model 1 (top) fits better to the solution structure than model 2.

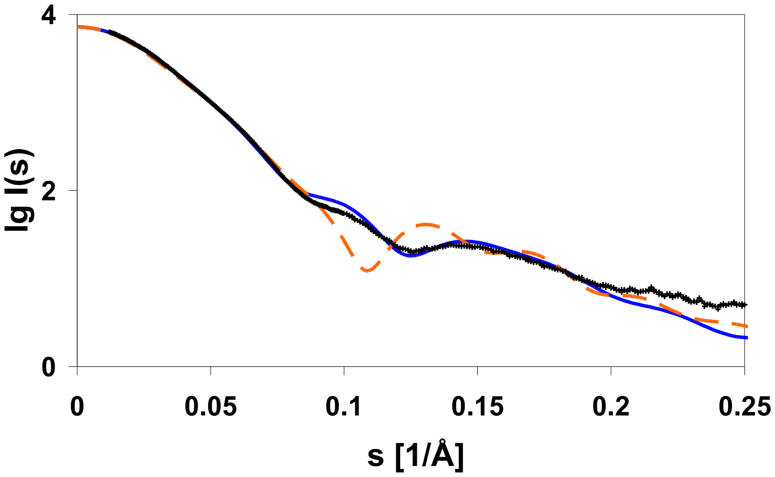


Figure 13: Experimental scattering curve in comparison to theoretical scattering curves calculated from the two possible DisA assemblies: The calculated scattering curve for model 1 (solid blue line) agrees very well with the experimental curve (black crosses), while the calculated curve for model 2 (dashed orange line) substantially disagrees.

4.2 Identification of the Enzymatic Activity of DisA

4.2.1 DisA is a Specific Diadenylate Cyclase

During model building of the DisA structure, additional electron density for a ligand bound to the N-terminal domain of DisA could be observed. Using mass spectrometry the ligand was proved to be a bis-(3'-5')-cyclic dimeric adenosine monophosphate (c-di-AMP). As no such molecule was added during purification or crystallization it must have been co-purified. However, this ligand has not been described as a biological molecule in literature, yet. Thus, it remained to be analyzed whether DisA bound endogenous c-di-AMP during expression in *E. coli* cells (assuming that c-di-AMP exists in *E. coli*) or synthesized it from substrates available in the cells.

Even though c-di-AMP is unknown so far, a similar molecule, c-di-GMP, is already well characterized as widespread second messenger in prokaryotes (Camilli and Bassler, 2006; Jenal and Malone, 2006; Römling and Amikam, 2006) and a biochemical assay for its synthetase was already established (Christen et al., 2005). To test whether DisA may be the synthetase for c-di-AMP a diadenylate cyclase assay was developed by adaptation from the diguanylate cyclase activity assay described by Christen et al. (2005). DisA was incubated with radioactively labeled α^{32}P-ATP as a substrate and the formation of c-di-AMP was monitored over time. Separation of the substrate ATP and the reaction product c-di-AMP was achieved with thin layer chromatography in a suitable running buffer as described in Christen et al. (2005).

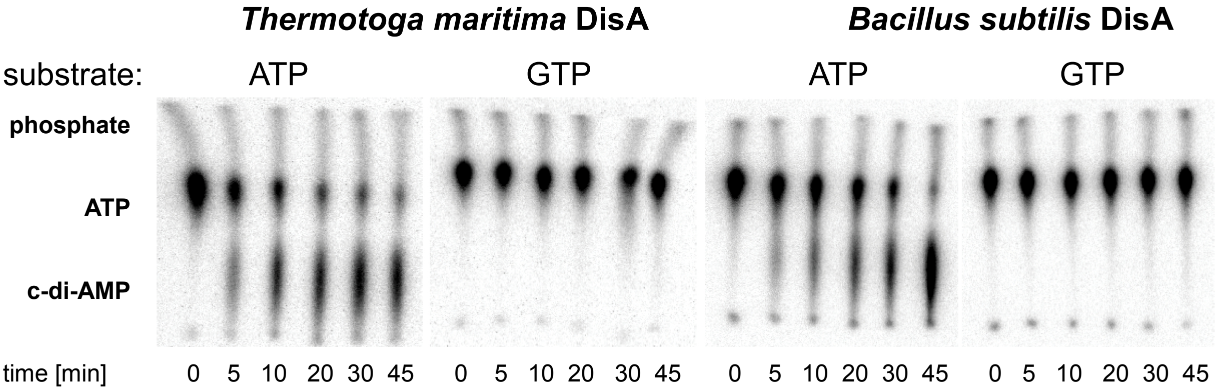


Figure 14: Dinucleotide cyclase activity of DisA is specific for ATP. Thin-layer chromatography of assay products for *Tma*DisA and *Bsu*DisA carried out at 50°C and 30°C respectively. With both DisA proteins the amount of c-di-AMP is increasing while ATP concentration is decreasing. Activity can only be observed with 100 µM ATP and not with equal amount of GTP.

The experiment was able to prove the existence of a reaction product behaving similar to c-di-GMP (Christen et al., 2005) on a TLC plate. However, the activity seems to be low as shown in figure 14. Due to the fact that all earlier published *in vivo* experiments with DisA (Bejerano-Sagie et al., 2006) were carried out with the homologue from *Bacillus subtilis*, both proteins *Bsu*DisA and *Tma*DisA were tested for activity (figure 14).

The assays with *Thermotoga maritima* DisA had to be performed at 50°C as the source organism is a thermophile and the *Tma*DisA activity increased strongly with temperature (figure 15A-C)). Assays with *Bsu*DisA were carried out at 30°C. It could be shown that both proteins are equally active. For reasons of comparison and due to the fact that ATP gets more easily hydrolyzed at higher temperatures, all further studies were carried out with *Bsu*DisA protein at 30°C.

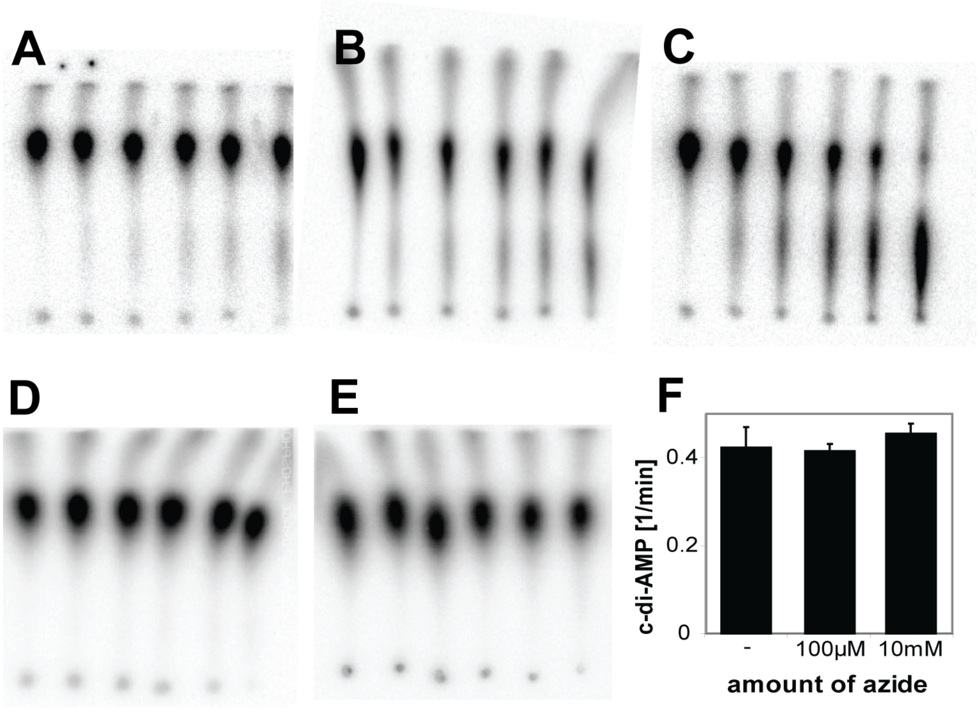


Figure 15 A-C: The cyclase activity of *Tma*DisA is temperature dependent. The cyclase assay was carried out at 30°C, 40°C and 50°C and an increase in activity can be observed for higher temperatures. **D:** Mg^{2+} ions are necessary for enzymatic activity of DisA. **E:** Using cordycepin triphosphate, which is missing the 3'-OH at the ribose, no reaction product can be observed. **F:** Sodium azide does not affect the enzymatic activity of DisA. Time points were taken as shown in figure 14.

To proof that the *in vitro* synthesized product of the reaction is in fact a c-di-AMP molecule, two independent reactions were performed in parallel: one with radioactively labeled ATP and a second one with non radioactive ATP and both were incubated for 1 h. Using the radioactively labeled sample, the position of the spot was identified and the equivalent part on the second TLC-plate was scratched off the plate and incubated in water for 30min at 50°C. This sample was analyzed by negative mode MALDI-TOF mass spectrometry using sinapinic acid as a matrix. In comparison to the reference (matrix alone) the spectrum of the sample showed a prominent peak at 657 Da which corresponds to the mass of single-charged c-di-AMP (see appendix 9.3.2).

While c-di-AMP was not discovered as relevant biological molecule before, the closely related c-di-GMP is well characterized. To ensure that DisA did not use ATP as a substrate, because its cellular level is much higher than that of GTP, the assay was carried out under the same conditions only exchanging ATP with GTP as a substrate. No comparable diguanylate cyclase activity could be detected, excluding the possibility that DisA may be a diguanylate cyclase (figure 14).

In a bacterial cell more variants of nucleotide triphosphates exist that could possibly be a substrate for DisA. To ensure that DisA uses only ATP as substrate, competition assays were carried out in the presence of radioactive ATP and equal amounts of competing unlabeled ATP (control), GTP, ITP, CTP or UTP. Only non radioactive ATP was able to compete with radioactively labeled ATP. This result indicates that neither one of this triphosphates can compete in binding at the active site of DisA nor in being a substrate for transformation into a cyclic dinucleotide (figure 16).

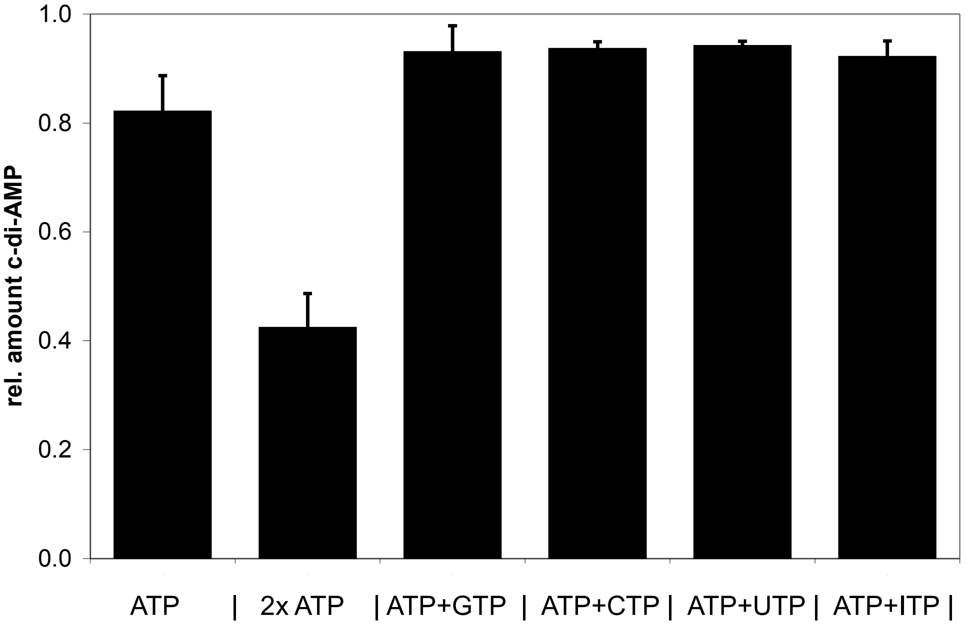


Figure 16: Competition assay to verify the specificity of DisA. DisA was incubated with 100µM α-^{32}P-ATP and 0 or 100 µM of the indicated cold NTPs. The relative amount of synthesized c-di-AMP is shown (mean of three independent experiments ± standard deviation). Only ATP is able to compete with α-^{32}P-ATP, indicating that DisA is a specific diadenylate cyclase.

4.2.2 The Influence of Mg^{2+} and Active Site Mutants on the Activity

Comparison of the diadenylate cyclase active site with other enzymes that use ATP as substrate indicated that DisA may need Mg^{2+} for activity. Testing the activity of DisA without any divalent ions in the reaction buffer verified this assumption, as no c-di-AMP could be detected after 45 min (see figure 15D).

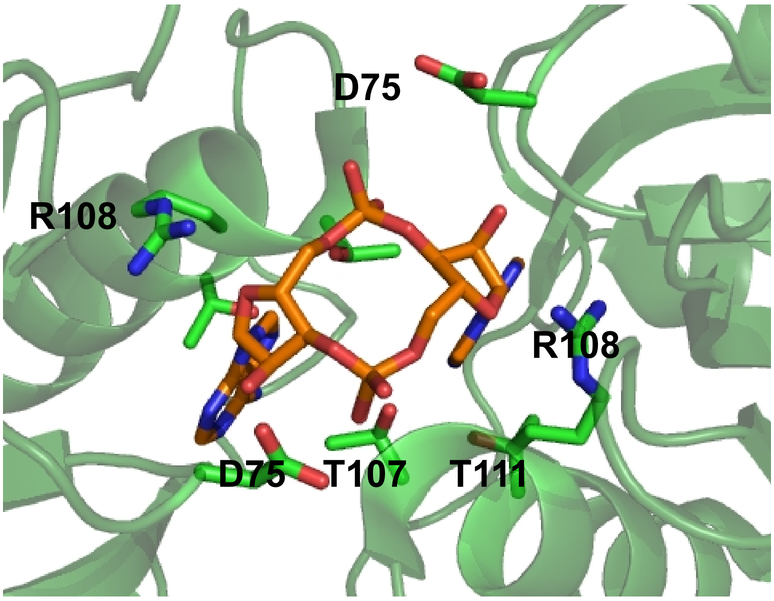


Figure 17: Detailed view of the active site of *Tma*DisA. Some amino acids that seem to be important for substrate binding or catalysis are shown in sticks.

A closer look to the structure of DisA at the active site (figure 17) suggested some amino acids which may be important for either binding ATP or catalyzing the cyclase reaction. To test their importance in c-di-AMP formation mutants were supplied by Gregor Witte and tested in the cyclase assay. The following mutants resulted in total abolishment of diadenylate cyclase activity: D75A, D75N, R108A+H109A+R110A and T107V+T111V. In addition, the cyclase activity was tested with an ATP analogue as substrate that is missing the 3'-OH at the ribose (cordycepin triphosphate). A proposed nucleophilic attack from the α-phosphate of one ATP molecule on the 3'-OH of the second ATP molecule should then be impossible and indeed no c-di-AMP formation could be observed with this substrate under the used conditions (figure 15E).

4.3 The Influence of Different DNA Molecules on DisA Activity

DisA was postulated to be a DNA damage checkpoint protein (Bejerano-Sagie et al., 2006) that recognizes directly or indirectly DNA lesions and induces an arrest in cell cycle progression. The fact that biochemical analysis of DisA revealed specific adenylate cyclase activity raises the question, if the activity is influenced by (damaged) DNA. Therefore, the activity of DisA was tested in the presence of different synthetic DNA substrates that are equivalent to DNAs occurring in DNA damages and DNA repair (Figure 18).

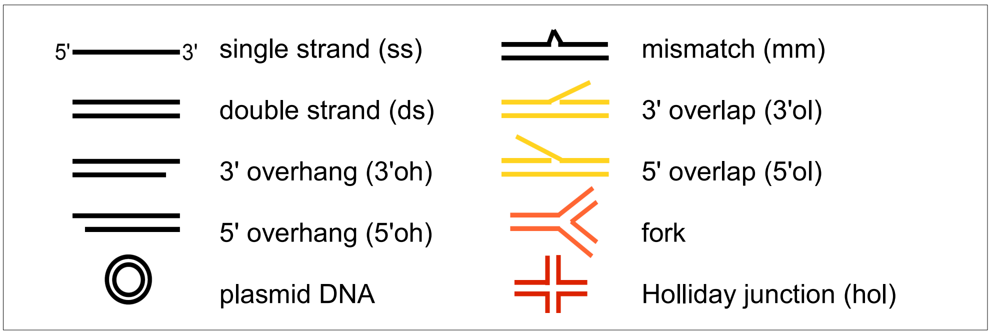


Figure 18: DNA molecules used to test for their influence on DisA activity. DNA structures that showed an effect on DisA activity are colored.

DisA was incubated with increasing amounts of all DNA molecules shown in figure 18 and diadenylate cyclase activity was measured as described before. Single-stranded and double-stranded DNA of different lengths, as well as plasmid DNA and dsDNA with 3'- and 5'-overhangs had no detectable effect on DisA activity. Thus, DNA ends and therefore double-strand breaks are unlikely to be the primary recognition site for DisA – which in consequence leads to the idea that DisA possibly does not bind to the damage itself but to DNA structures which are emerging from DNA-damages.

During the recognition and repair of DNA damage different DNA secondary structures can arise like e.g. fork structures, Holliday junctions or DNA with single-stranded flap structures.

The effect of these DNA molecules on DisA activity is shown in figure 19. Whereas 3'- as well as 5'-single-strand flap structures only moderately affected DisA activity by slightly decreasing its ability to synthesize c-di-AMP, branched nucleic acids that resemble three-way and four-way DNA junctions strongly modulated the activity. In comparison to branched nucleic acid, the presence of ssDNA and dsDNA had no effect on the activity of DisA within error range. Addition of 3'- or 5'-single stranded flap structures leads to a reduction of diadenylate cyclase activity of around 50 % for a concentration of two DNA molecules per DisA octamer. In presence of three- and four-way DNA junctions the effect is even stronger. Two fork-like DNA molecules or two Holliday junctions can reduce the activity to approximately 20 % compared to the activity in absence of DNA.

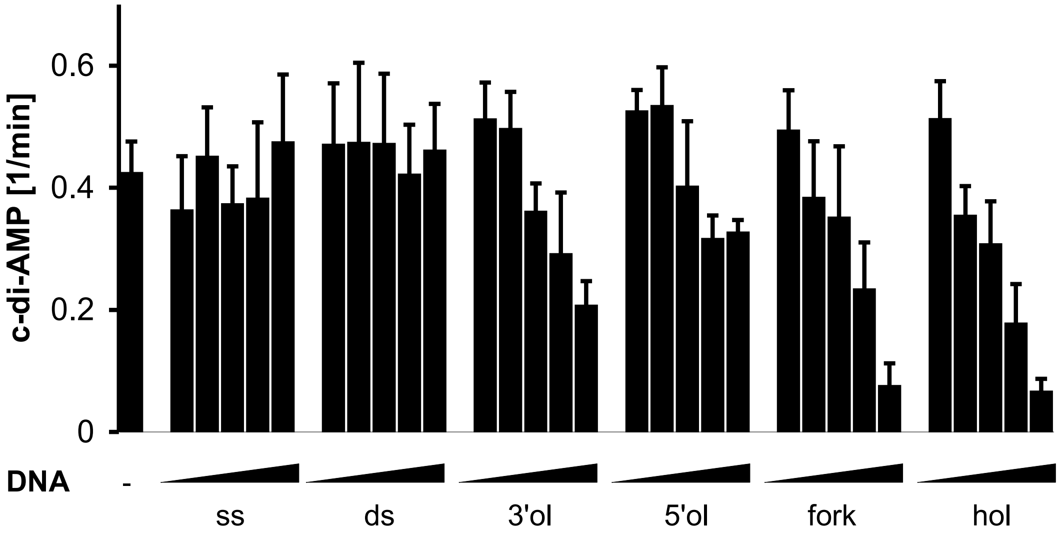


Figure 19: Effect of different DNA molecules on DisA activity. Diadenylate cyclase activity of 3.5 µM DisA is unaffected by increasing amounts of ssDNA or dsDNA(- = no DNA; triangles = increasing concentrations of 0.65, 1.3, 2.0, 2.4, 3.0 and 6.5 µM DNA). 3' flaps and 5' flaps (3'ol and 5'ol) moderately reduce activity. Three- and four-way junctions substantially inhibit the activity of DisA. Data represent the mean of six independent experiments ± standard deviation.

Summarized, this data suggests that the catalytic activity of DisA is regulated by binding of the complex to branched nucleic acids, well-known structures occurring during DNA repair pathways.

4.4 The Influence of Azide

Using fluorescence microscopy Berjerano-Sagie et al.(2006) analyzed the movement of DisA-GFP foci *in vivo*. In the presence of the ATPase inhibitor sodium azide they detected an immediate stalling of DisA. On the basis of this observation they concluded that mobility of DisA is energy dependent. As the diadenylate cyclase assays clearly shows that DisA uses ATP as a substrate, sodium azide might inhibit this activity similar to the inhibition of ATPases. To test this hypothesis, DisA activity was analyzed in presence of up to 10 mM sodium azide in the assay buffer and no differences in enzymatic activity could be detected. Therefore, the effect of sodium azide on the movement of DisA does not seem to be due to direct influence on the diadenylate cyclase activity (see figure 15F).

4.5 DisA and DNA Binding

The obvious effect of DNA on the activity of DisA can only be explained by binding of the complex to DNA, followed by a conformational change which then influences the activity of the protein. The C-terminal domain of DisA was identified as potential DNA binding domain by structural homology searches (Holm and Sander, 1997). It is highly homologous to helix-hairpin-helix (HhH) domains that are common non-specific DNA binding domains. In electrophoretic mobility shift experiments DisA has a more than two times higher affinity to Holliday junction DNA compared to normal dsDNA.

Of special interest in this context is the HhH domain of the Holliday junction binding protein RuvA, as the structure of this protein was solved in complex with a Holliday junction (PDB entry 1C7Y, Ariyoshi et al., 2000). To get a principal idea about the possible mode of DNA binding of DisA, the HhH domain of RuvA and the bound DNA was superimposed with the C-terminal domain of DisA using SSM superpose implemented in COOT (figure 20).

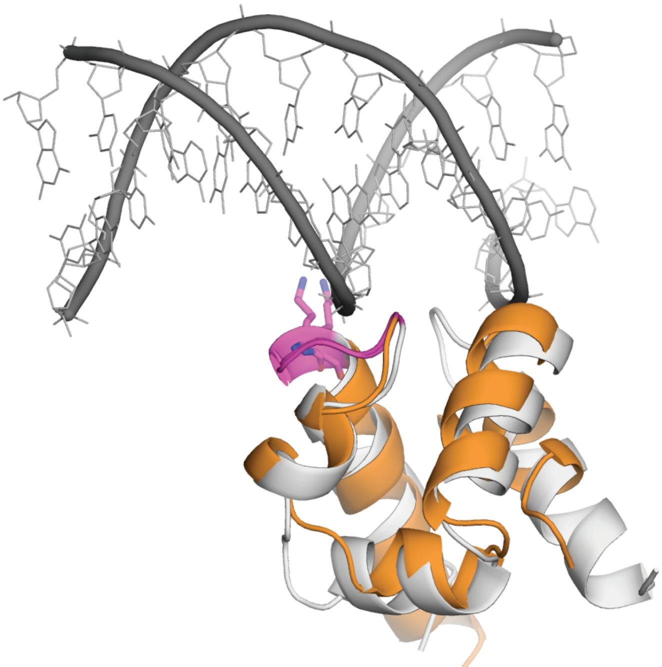


Figure 20: Superposition of the HhH domain from *Tma*DisA (orange) with the HhH domain of RuvA bound to DNA (grey, PDB code 1C7Y). The prominent DNA binding loop (magenta, including side chain sticks for a conserved lysine) is highly conserved between HhH domains from DisA and RuvA, indicating a conserved mode of DNA backbone interaction.

Comparison of the sequence of the DisA HhH domain with other HhH domains gave a very conserved patch of amino acids within a DNA binding loop (figure 20, magenta). To prove that the HhH domain of DisA is in fact the DNA binding motif, one of the conserved residues that shows backbone interactions with the DNA, glycine 334, was mutated to a glutamate and this protein expressed and purified by Gregor Witte. In electrophoretic mobility shift assays the G334E mutant shows a drastically reduced binding to the DNA substrates, proving the involvement of this domain in DNA binding. To analyze whether the HhH-motif is involved in regulation of the diadenylate cyclase activity, the G334E mutant was also used for cyclase assays.

The mutant protein was tested for influence of Holliday junctions on the activity (figure 21). Compared to wild-type protein the inhibitory effect of the Holliday junction is drastically diminished in case of the G334E mutant, which indicates a decrease in specific binding to branched nucleic acids. This result strengthens the proposed role of the HhH domain in being the binding site for Holliday junctions and also the fact that binding of DNA to these motifs leads to structural changes which are somehow translated to the N-terminal domain to regulate the activity.

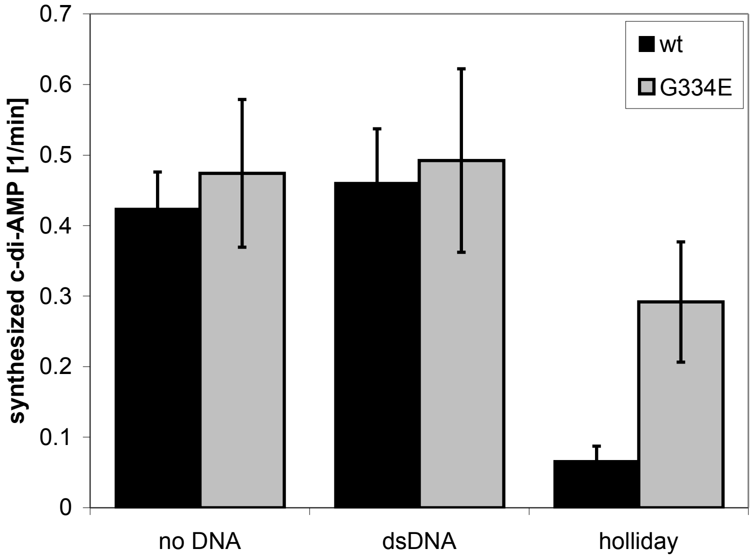


Figure 21: Influence of glycine 334 on Holliday junction binding. The inhibitory effect of the Holliday junction on DisA activity is much lower for the G334E mutant compared to the wild-type. This indicates that glycine 334 is important for the affinity of DisA to DNA and that DNA-binding to the HhH is the trigger for changes in activity of the N-terminal domain.

5 Discussion – DisA

5.1 SAXS as Complementary Method to X-ray Crystallography

The identification of the correct octameric assembly of the DisA complex perfectly shows the power of SAXS when structural information is already available. When using crystallography as a method in structural biology it is always difficult to proof that the determined atomic structure is not only a crystallographic artifact. It has to be shown that the proteins or complexes have the same fold and assemble to the same overall structure in solution as they do in the crystal. Furthermore the conditions in which some proteins crystallize are far from physiological, for example some crystals grow under extreme pH-values or very high salt concentrations. With SAXS one can determine a structure under conditions that are as physiologic as possible. Summarized it can be said that SAXS is very important for protein crystallographers to proof the relevance of their structures for biological studies.

In addition – as it was shown for the *Tma*DisA protein – it is often not easy to distinguish between real, physiological interfaces within a protein crystal and crystallographic contacts that do not exist in solution. The calculation of theoretical scattering curves of atomic models is possible in only some minutes and a SAXS experiment does not require much time. This makes SAXS an easy method to determine or verify an oligomeric assembly that can be seen in a protein crystal.

5.2 The DisA Octamer

From the results of different experiments, the DisA protein was already proposed to form an octameric complex in solution. From the way the protein crystallized, no distinct octamer could be identified, but two possible models were found. The two tetramers could either be arranged in a head-to-head of in a foot-to-foot orientation. Looking at the two different interfaces already gives a first hint to the correct complex. In the first model the interaction is mediated via the N-terminal nucleotide binding domain. This interface is substantially larger than the interface in the second model, where two C-terminal HhH domains interact with each other. In addition, the amino acids mediating the interaction in the first model are highly sequence conserved and hydrophobic in nature. This already gives promising hints to a physiologic interface in contrast to the interface between the HhH domains, which has more hydrophilic residues and resembles more a crystal contact.

Still it had to be experimentally verified that one of the proposed octameric models matches the solution structure of DisA. The results of the SAXS studies proof without any doubt that model 1 is present in solution and that the four nucleotide binding sites are located in the center of the complex.

5.3 Reliability of SAXS Structures

When determining a solution structure with SAXS a three-dimensional model has to be created from a two dimensional scattering curve, which represents averaged data from all possible orientations of the molecule. Although it is possible to calculate an *ab initio* structure using only the scattering data, it is very helpful to have at least some additional structural information. In the case of DisA even before solving the crystal structure, it was already known that the protein forms an octamer in solution. Using this information it was possible to assume different symmetries of the oligomer, use this information as an additional input parameter and improve the model.

However, like in most theoretical fitting algorithms one has to be very careful not to use false constraints and reasonable input parameters. A significant disadvantage of SAXS is the fact that no real parameter for quality control of the result exists. In X-ray crystallography the R-factor, (especially R_{free}) enables the scientist to estimate the reliability of the achieved result. In SAXS every scattering curve will result in some three dimensional bead model and the biologist then has to decide to the best of his knowledge whether this structure is reliable or not. For example, a protein complex that has a special symmetry in a crystal is often less symmetric in solution. Therefore it is very important to proof that an implicated symmetry constraint is correct.

As shown for the DisA complex, it is important to try different symmetries in comparison to the no-symmetry approach. The four SAXS solution structures from the DisA protein show that the 2-fold and 4-fold symmetry implicated in the calculations was correct, as the variation between the structures is negligible.

The shape of the structure obtained with one 4-fold symmetry axis is obviously different to the other three models. In this case the program located the symmetry axis in the wrong orientation and then the iterative minimization method resulted in the wrong structure. However, comparison of all obtained structures with the symmetry-free structure clearly proofs which result is correct.

In summary, SAXS is extremely powerful when structural information is already available and provides supplemental information of the solution structure of the protein. It can also be used as an *ab inito* approach, but then one has to be very careful with the interpretation and reliability of the obtained structure to be certain not to look at artifacts. In principle, one scattering curve can be explained by two completely different structures. However, comparative database searches with theoretical scattering curves of proteins available in the PDB did not reveal overlaps of two different structures with the same theoretical scattering curve (Sokolova et al., 2003).

5.4 The Moving Foci of DisA

From their *in vivo* observations using *B. subtilis* cells, Bejerano-Sagie et al. (2006) concluded that the cells increase expression of DisA upon sensing sporulation conditions, the protein assembles to a globular structure and moves rapidly along the chromosome searching for DNA lesions. These data are all based on observations made using GFP-labeled DisA. However, it is still not understood how this single, rapidly moving focus is composed and how is moves. The structural analysis in this work shows that DisA forms a stable octamer in solution. For DNA polymerases that are only loosely localized close to the replication fork 20-40 proteins where already detectable as fluorescent foci (Lemon and Grossman, 1998). In the case of DisA, the fluorescent GFP-fusion molecules are very tightly arranged. This raises the possibility that only a few DisA complexes might form part if not all of the detectable *in vivo* focus as the local concentration of GFP-fluorophors is very high due to the compact assembly.

The mobility of DisA in wild type cells had a mean velocity of 0.22µm/s and was immobilized in cells that contained a double strand break or were treated with azide. Using fluorescence correlation spectroscopy the diffusion constants for protein complexes similar or larger in size to octameric DisA could be measured to be in the range of 1-3 $\mu m^2/s$ (Wang et al., 2006). As these numbers are quite similar, DisA movement may at least be in part driven by brownian diffusion. This assumption would also explain the finding that chromosomal DNA is not needed for DisA mobility (Bejerano-Sagie et al., 2006).

The stalling of the foci in the presence of DNA double strand breaks can be explained by the fact that DisA has higher affinity to Holliday junctions than to "normal" double stranded DNA. This suggests that the immobilization does not occur at the double strand itself, but at branched nucleic acids like Holliday junctions of stalled replication forks. These types of branched nucleic acids are not present in a normally replicating cell but are formed during DNA double-strand break induced recombination, chromosome missegregation or as a consequence of replication forks stalling at a lesions. These data lead to the conclusion that DisA might be a more general sensor for unsegregatable chromosomes and might fulfill its role as a checkpoint protein via coupling the recognition of unfinished repair with the activity of c-di-AMP synthesis.

5.5 c-di-AMP and Diadenylate Cyclase Activity

C-di-AMP is related to, but distinct from c-di-GMP. The molecule has not been described as a biological molecule that is existing in or synthesized by any pro- or eukaryotic organism, yet. In contrast to this the related c-di-GMP is widely studied and was identified as a bacterial signaling molecule. It serves as the focal point for several extracellular sensory inputs and plays a role in regulating complex cellular processes that are also regulated by quorum sensing. Similar to the important second messenger cAMP the level of c-di-GMP is regulated by the activity of enzymes that synthesize it from GTP and others that degrade it again to GMP. The cyclase activity is associated with GGDEF domains, while the phosphodiesterase activity is associated with EAL domains (the names stand for the conserved amino acid sequence motif in the active site). These domains as well as c-di-GMP as a second messenger are only found in eubacteria and are absent in archaea and eukaryotes (Galperin, 2005).

Although no structural or sequence homology between GGDEF domains and DisA can be found, it is quite likely that c-di-AMP works as a second messenger in a similar way to c-di-GMP. Although c-di-AMP synthesis may have evolved independently of c-di-GMP synthesis, the general overall architecture of the DisA octamer, with central pairs of DAC (diadenylate cyclase) domains and peripheral HhH domains resembles to some extent the allosterically regulated diguanylate cyclase PleD (figure 7, Chan et al., 2004).

A structure similarity search for the DAC domain of DisA in the protein data bank using the DALI search tool (Holm and Sander, 1996) revealed only one structure that shows the same overall fold (figure 22). The identified protein, a structure contributed by a structural genomics consortium, belongs to the uncharacterized cluster of orthologous groups COG1624 and has outside the DAC domain no sequence homology to DisA. The superposition of the DAC domain of DisA with this "domain of unknown function" (DUF) from *Bacillus cereus* (PDB code 2FB5) shows that most of the side chains that are involved in c-di-AMP binding in DisA are conserved and also present in the DUF (see appendix 9.3.1 for a sequence alignment).

Therefore, it is quite likely that this protein of unknown function possesses a DAC domain and functions as diadenylate cyclase. Although DisA sequences (including the DAC and HhH domains) are limited to a subset of eubacterial genomes, the DAC alone is widespread among eubacterial genomes and can be identified also in archaea. In contrast, the GGDEF domains correlated with c-di-GMP synthesis are restricted to eubacteria.

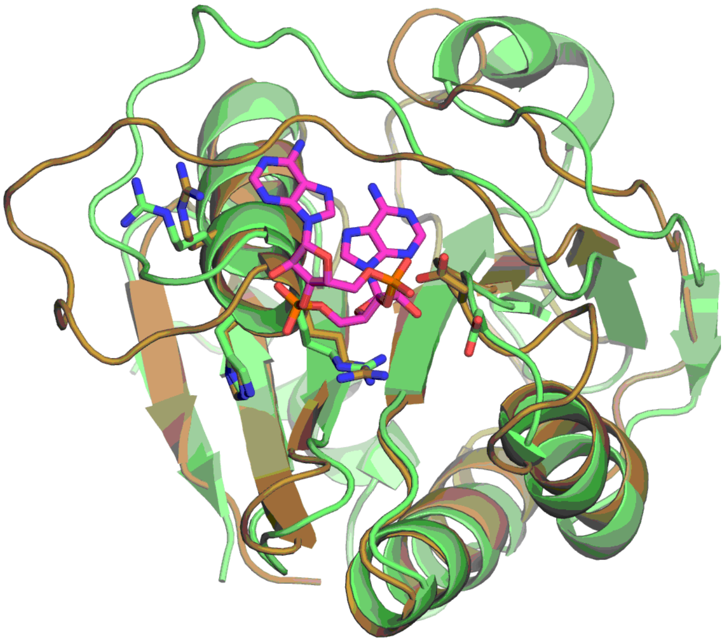


Figure 22: Superposition of the N-terminal di-adenylate cyclase domain of DisA with the C-terminal domain of a protein with unknown function. The N-terminal domain of DisA is shown in green, the c-di-AMP molecule in magenta and the structurally related protein found by DALI search is shown in brown (PDB code 2fb5). Side chains involved in binding and possibly in synthesis of c-di-AMP are shown as sticks. These residues are conserved between both proteins, suggesting that the homologous protein also possesses di-adenylate cyclase activity.

Combining all results published by Bejerano-Sagie et al. 2006 and the results of this thesis allows the proposition of a model for the function of DisA and c-di-AMP in the cell:

C-di-AMP might play a general role in the regulation of some cellular processes that are associated with DNA repair and damage signaling. Using DisA-null cells Bejerano-Sagie and coworkers could show that DNA damages induce premature timing of sporulation. However, present results do not explain whether c-di-AMP has a direct or indirect effect on sporulation. The *in vitro* data suggests that a higher level of c-di-AMP could be present in healthy cells leading to normal cell division or sporulation. The presence of damaged DNA and the resulting abnormal branched nucleic acids would lead to a decrease in c-di-AMP level, signaling the cells the presence of an abnormal situation. This will possibly induce a signaling cascade that results in a stop of sporulation. Only after successful repair of the lesions the enzymatic activity of DisA is back at the normal level – because of the lack of binding targets – and so signals the cell to resume dividing.

Sequence similarity searches for homologous proteins of DisA revealed that they are also found in non-sporulating eubacteria. As mentioned above, the DAC domains, which synthesize c-di-AMP are even found archaea, allowing the conclusion that c-di-AMP is possibly not only involved in sporulation signaling, but has a more general role in the regulation of the cell division process via recognition of unsegregatable chromosomes.

The relatively low affinity of DisA to DNA and the low c-di-AMP synthesis rate may indicate that other factors like interacting proteins, special DNA sequences or DNA structural specificities are important *in vivo*.

5.6 The Effect of Holliday Junctions on the Activity

When looking at the structure of DisA octamers from the top, the HhH domains are arranged in a way that offers a plausible mechanism for the effect of DAC-regulation by binding to Holliday junctions (figure 23A). The four HhH domains are arranged at both sides of the DisA octamer in a geometry that strikingly resembles the four HhH domains in the Holliday junction binding protein RuvA.

Based on the structural similarity between HhH motifs of DisA and RuvA, a similar geometry of HhH domains bound to DNA is very likely. In the case of RuvA each arm of the Holliday junction is bound by one HhH domain (Ariyoshi et al., 2000; Hargreaves et al., 1998; Rafferty et al., 1996; Roe et al., 1998). Superposition of the DNA bound RuvA with the HhH domain of DisA reveals a common fold and shows that some potential DNA-binding residues are present in both HhH domains (figure 20). Indeed it could be shown that the mutation of residue G334 to a glutamate strongly reduces the capability of the Holliday junction to inhibit diadenylate cyclase activity of DisA. These data support the model of a DNA-binding mode of DisA that is related to RuvA. However, in the available DisA structure the HhH-motifs are not all capable of binding DNA without conformational changes, as they seem to be in a hinged position. Supposing similar geometries using RuvA as a template, the HhH motifs of DisA require slight rotation and tilt to reach the orientation to be well positioned to bind branched nucleic acids. This movement very likely involves also parts of the helical DisA middle domain which then might affect activity by changes in the N-terminal DAC-domains (figure 23C).

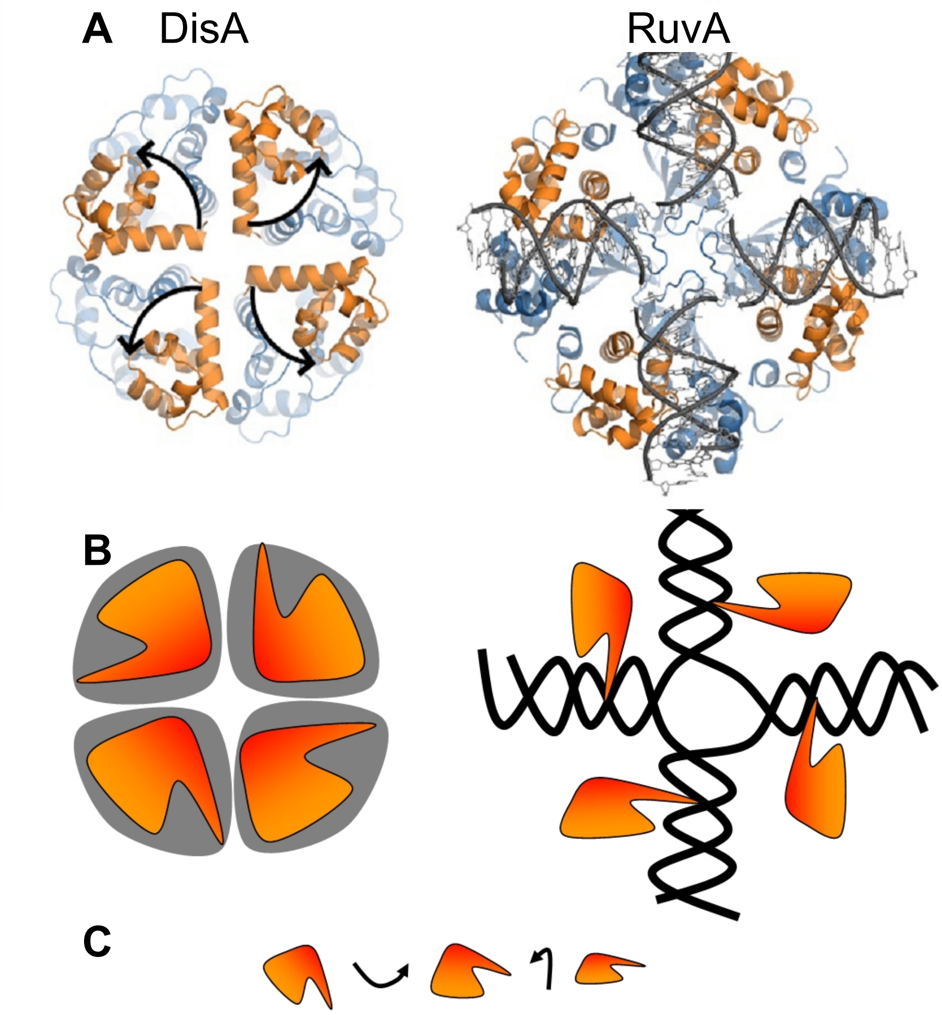


Figure 23: HhH domains of DisA and RuvA in comparison. A: Top view of DisA (left) and RuvA bound to a holliday junction (PDB entry 1C7Y, Ariyoshi et al., 2000), HhH domains are shown in orange. **B:** Schematic view of the orientation of the HhH domains. In order to be orientated in the same way as the HhH domains of RuvA, the HhH domains of DisA would have to turn and tilt (**C**).

Considering all data, a mechanistic model for the role of DisA as a checkpoint protein can be proposed (figure 24): In a normal, healthy cell with no chromosomal damage, DisA synthesizes c-di-AMP which serves as a signal for normal cell division progress. In the presence of branched nucleic acids like recombination intermediates or stalled replication forks DisA binds to these targets, thereby losing the ability to synthesize c-di-AMP. This decrease in c-di-AMP level is a signal for the presence of unsegregatable chromosomes and leads (possibly through further downstream signaling) to a halt in the cell cycle progression. Still some aspects of this model – like the mode of DNA binding – are speculative and need to be tested.

The determination of the atomic structure of DisA bound to a Holliday junction would help to understand, how the DNA can affect the active site although it is positioned at the opposing end of the molecule. In addition, *in vivo* studies with c-di-AMP to proof and further analyze its role as second messenger could help to support the model that was proposed in this thesis.

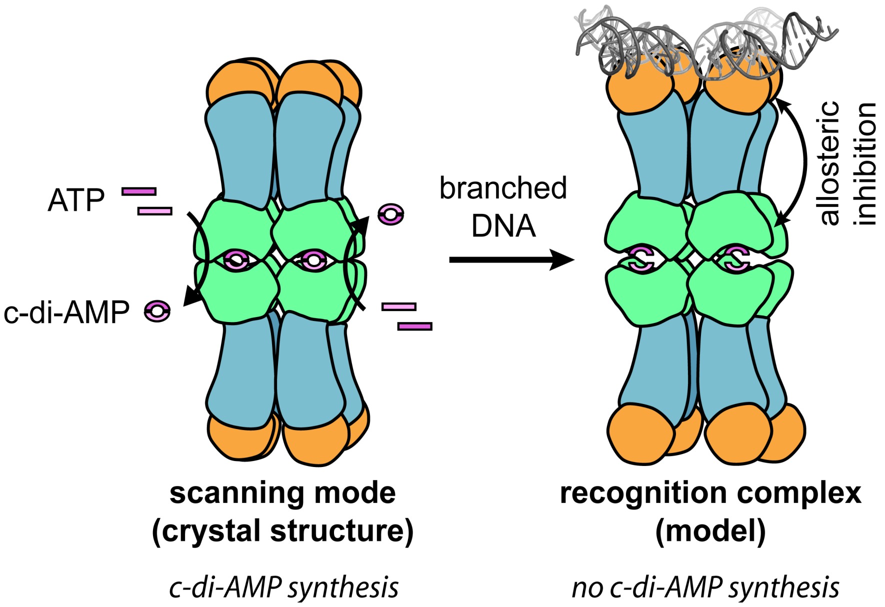

ATP
c-di-AMP
branched DNA
allosteric inhibition
scanning mode (crystal structure)
c-di-AMP synthesis
recognition complex (model)
no c-di-AMP synthesis

Figure 24: Mechanistic model for the role of DisA. In the absence of chromosomal damage DisA synthesizes c-di-AMP. Recognition of branched nucleic acids leads to inhibition of c-di-AMP synthesis, signaling the presence of unsegregatable chromosomes.

6 The Exosome and RNA Metabolism

6.1 The Exosome and Quality Control

The variety and amount of RNA molecules that are produced in the cells of all organisms is remarkable. There are mRNAs from thousands of genes as well as many RNA molecules that are not translated into proteins and function directly as RNAs. These non-coding RNAs include tRNAs (transfer RNAs involved in translation), rRNAs (ribosomal RNAs), snRNAs (spliceosomal small nuclear RNAs), snoRNAs (small nucleolar RNAs involved in processing of rRNAs) and a diversity of other RNAs that influence processes ranging from chromosome replication to mRNA translation (Warner, 1999). In addition, RNA molecules can be combined with proteins and form a ribonucleoprotein (RNP). Similar to proteins, many RNAs are only functional when folded into a complex structure. Not only in the case of mRNAs, non-functional mutated RNA can arise by synthesis from mutant genes, transcriptional errors, premature termination, post-transcriptional editing events and mistakes in RNA processing. These errors may result in RNA misfolding and/or failure to assemble with proteins to form functional RNPs.

When the cell recognizes defective or non-functional RNAs, rapid degradation of these RNAs is induced. Additionally, posttranscriptional degradation is an important method to regulate cellular mRNA levels. Efficient surveillance machineries are responsible to distinguish between defective and non-defective RNAs and to monitor the process that converts the primary RNA transcript to a mature RNP.

Two main pathways involved in RNA degradation have been identified in eukaryotes: the RNA is degraded from the 5'-end either by the 5' exonuclease Rat1 (Xrn2 in humans) or by Xrn1 (Kastenmayer and Green, 2000). These enzymes are known to be involved in the degradation of decapped mRNAs. This thesis concentrates on the second pathway, which is responsible for degradation from the 3'-end by the so called exosome.

In archaea and eukaryotes the exosome is a key component of the RNA surveillance machinery (Mitchell et al., 1997). The exosome is a multiprotein complex with 3' → 5' exonuclease activity. It degrades many RNAs that are targeted by surveillance activities in the nucleus as well as in the cytoplasm. In addition, it is responsible for the precise trimming of 3'-ends of several nuclear RNA precursors (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).

The yeast exosome was identified in 1997 as a 3' → 5' exonuclease that is responsible for the 3′ end maturation of 7S rRNA to mature 5.8S rRNA (Mitchell et al., 1996). Two years later the human homologs were identified as components of the known PM-Scl particle, a multi-subunit complex recognized by autoimmune sera of patients suffering from polymyositis-scleroderma overlap syndrome (Allmang et al., 1999).

6.2 Exosome-like Complexes in Bacteria, Archaea and Eukaryotes

Exosome complexes are found in archaea and eukaryotes and are homologous to the bacterial polynucleotide phosphorylase (PNPase). The PNPase, which was also found in mitochondria and chloroplasts, is therefore often referred to as an exosome-like complex. From known crystal structures it is assumed that all exosomes and the PNPases are structurally highly conserved.

The eukaryotic core exosome consists of six polypeptides with sequence similarity to the phosphate-dependent 3'→5' exoribonuclease RNase PH and three protein K homology (KH) and/or S1-domain-containing RNA-binding proteins. In eukaryotes, additional subunits include the RNase R homolog Rrp44/Dis3 (identified in the yeast exosome) and the RNase D homolog Rrp6 (only present in the nuclear isoform). The archaeal exosome exhibits a simplified subunit composition and is assembled from three copies each of two RNase PH-like proteins (Rrp41 and Rrp42), Rrp4 (**r**ibosomal **R**NA **p**rocessing proteins) and/or Csl4 (**C**ep1 **s**ynthetic **l**ethal **4**) (Evguenieva-Hackenberg et al., 2003).

6.2.1 Bacterial PNPase and RNase PH

Remarkably, the Polynucleotide phosphorylase (PNPase) can catalyze both the processive 3' → 5' degradation of RNA and the reverse reaction, meaning the addition of nucleotides to the 3'-end (Grunberg-Manago, 1999). PNPase was the first identified enzyme that can catalyze the formation of polynucleotides from ribonucleotides without the need of a template. When a mixture of nucleotide diphosphates is available, the PNPase produces a random polymer. In *E. coli* it was thought to be mostly involved in degradation, but more and more often its activity in the polymerization of heteropolymeric tails has been reported (Mohanty and Kushner, 2000). In spinach chloroplasts, cyanobacteria and gram-positive bacteria PNPase is even suggested to be the major polyadenylating enzyme (Bollenbach et al., 2004). In *E. coli* the PNPase is partly associated with the endoribonuclease RNase E, an RNA helicase, enolase, and possibly other proteins in a high molecular weight complex called the degradosome (Regnier and Arraiano, 2000). Figure 25 shows the atomic structure of PNPase in comparison with exosomes.

RNase PH is the other member of phosphate dependent 3'→5' exoribonucleases in bacteria. It is a relatively small single domain protein that is known to be involved in trimming tRNA precursors at their 3' ends (Li and Deutscher, 1996). Homologs of RNase PH exist in all three kingdoms of life and form the core of the exosome complexes in archaea and eukaryotes. In bacteria, six RNase PH polypeptides form a hexameric ring that is similar to that of PNPase and the exosome (Ishii et al., 2003).

6.2.2 Comparison of Exosome-like Complexes

All exosome-like complexes contain two subunits that are related to the RNase PH enzyme. These subunits are therefore termed RNase PH-like (PH) domains. The other two subunits are homologous to the KH and S1 domains, which are characteristic for RNA binding proteins.

The crystal structures of the bacterial PNPase, the archaeal exosome and the human exosome were determined by X-ray crystallography and are shown in figure 25 (Symmons et al., 2000; Büttner et al., 2005; Liu et al., 2006).

All structures are ring-shaped with the PH domains surrounding a central channel that can accommodate a single-stranded RNA molecule (see figure 25). In bacteria the complex consists of a ring with six copies of one polypeptide chain. In archaea three different proteins form a hetero-trimeric complex. The eukaryotic exosome, which is active in the cytoplasm and nucleus, consists of 10 – 11 different proteins and is the most complex version of all. However, alignment of the proteins points to a high structural similarity between the archaeal and eukaryotic exosomes and the PNPases: The core structure is in all cases a similar ring shaped complex consisting of six different PH-domains and three subunits with an RNA binding surface like KH or S1 domains. In addition to this highly conserved complex, a novel subunit was found in yeast: a hydrolytic 3' → 5' exonuclease (Rrp44 or Dis3), which belongs to the family of RNase R and RNase II in bacteria.

In the archaeal exosome three active sites that catalyze the degradation and polymerization of RNA are located at the inner side of the processing chamber. In contrast, the nine-subunit core of the yeast and the human exosome have no phosphate dependent RNase activity (Dziembowski et al, 2007; Liu et al, 2006). Instead, hydrolytic processive and distributive exonuclease activities could be observed for the ectopic components Rrp44 and Rrp6 (Mitchell et al., 1997; Briggs et al., 1998).

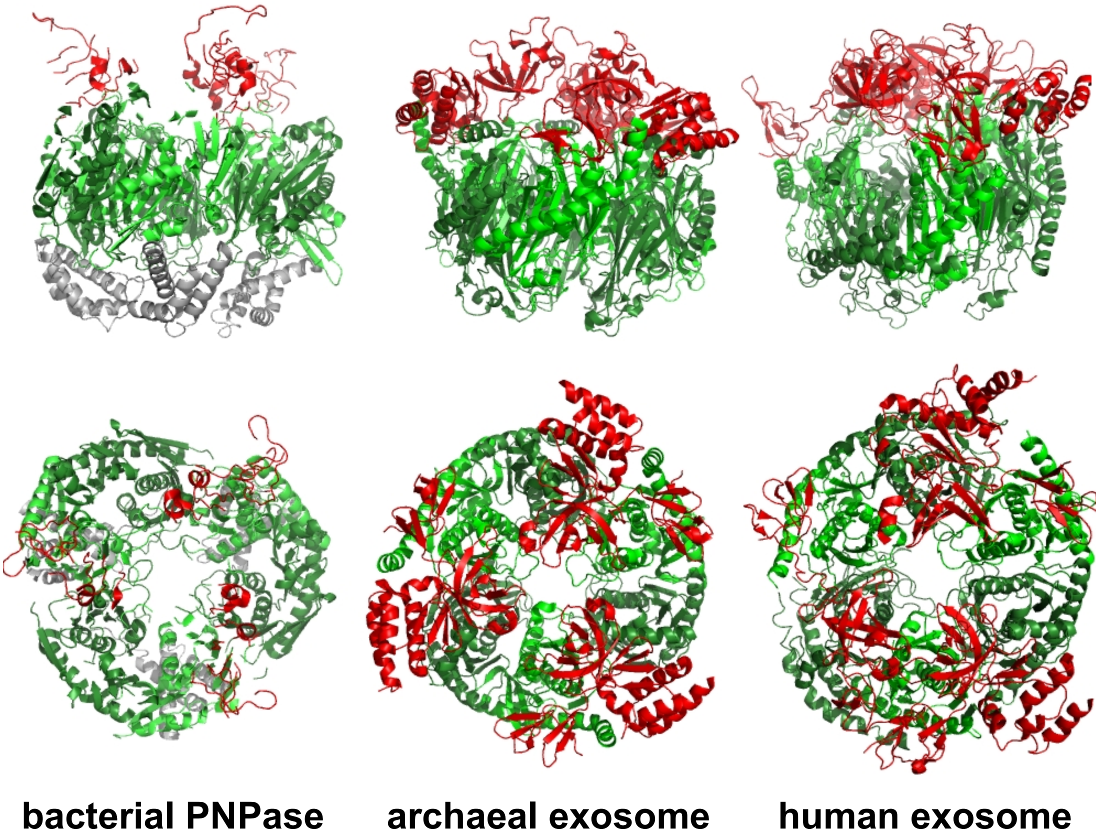


Figure 25: Crystal structures of bacterial PNPase, archaeal exosome and human nine-subunit exosome. The exosome clearly is evolutionary conserved. PH like domains are shown in green (Rrp41 homolog: light green, Rrp42 homolog: dark green) and the KH and S1 domain containing cap proteins are shown in red. S1 and KH domains of PNPase are incompletely resolved and may be mobile. (PDB entries: PNPase: 1E3P; archaeal exosome with Rrp4 as cap protein: 2BA0; human exosome: 2NN6)

6.3 Functions of the Exosome

6.3.1 The Exosome in Cytoplasm and Nucleus

Mitchell et al. (1997) identified nuclear and cytoplasmic forms of the exosome in eukaryotes, especially the yeast exosome is best studied and characterized. It could be shown that the nuclear and cytoplasmic exosome complexes share ten common components, but differ in the presence of the GTPase Ski7 in the cytoplasmic complex and the RNase Rrp6 and the putative nucleic acid binding protein protein Rrp47 (also known as Lrp1) in the nuclear complex. Table 4 summarizes the most important components of the exosome and exosome-like complexes (Allmang et al., 1999; Burkard and Butler, 2000; van Hoof et al., 2000; Araki et al., 2001; Peng et al., 2003; Mitchell et al., 2003).

Table 4: Comparison of exosome complexes and description of their proteins

bacterial homolog	archaeal exosome	yeast exosome	description of subunits
PNPase	Rrp41	Rrp41 Rrp46 Mtr3	Part of the hexameric ring structure and homologous to archaeal Rrp41 – in archaea a phosphate dependent exonuclease, in eukaroytes catalytically inactive
	Rrp42	Rrp42 Rrp43 Rrp45	Part of the hexameric ring structure and homologous to archaeal Rrp42 – in archaea and eukaryotes catalytically inactive
	Rrp4	Rrp4 Rrp40	Cap proteins that have S1and KH RNA binding domains
	Csl4	Csl4	Cap protein that has an S1 domain and a zinc-ribbon domain
RNase R	-	Rrp44 (Dis3)	Member of the RNase II familiy
-	-	Rrp6	Similar to *E.coli* RNase T/D familiy – present only in the nuclear exosome
-	-	Rrp47	Possible RNA binding protein – present only in the nuclear exosome
-	-	Ski7	Putative GTPase – present only in the cytoplasmic exosome

Proteins for which *in vitro* exonuclease activity could be detected are highlighted in gray.

6.3.1.1 The Exosome in the Cytoplasm

For the cytoplasmic exosome mRNAs are the only known substrates. RNA levels are not only determined by their rate of synthesis, but also by the rate of degradation. Therefore, mRNA turnover rates are an integral component of the control of gene expression. With the help of the Ski2 protein the exosome is responsible for the 3' → 5' degradation of mRNAs.

In addition, the exosome degrades mRNAs with structural defects in two different pathways: in the non-stop decay pathway the exosome eliminates transcripts lacking a termination codon (Frischmeyer et al., 2002). In the nonsense-mediated decay pathway it degrades mRNAs with nonsense mutations that were recognized by the exon-junction complex (Lejeune et al., 2003). The human exosome is additionally involved in the rapid decay of mRNAs with AU-rich sequence elements in the 3'-untranslated region (Chen et al., 2001). Via AU-rich elements the expression of proteins like growth factors and proto-oncogenes is kept transient.

Finally the exosome is involved in cleavage of 5' fragments of mRNAs. These fragments can occur in the no-go decay pathway, which targets mRNAs on which translation has stalled (Doma and Parker, 2006).

6.3.1.2 The Exosome in the Nucleus

In the nucleus the exosome is involved both in the maturation of RNAs by trimming the RNA to a precisely defined end point and in the rapid and complete degradation of defective RNAs. More precisely it is responsible for the maturation of the 5.8S rRNA, snRNAs and snoRNAs and the degradation of aberrant nuclear precursors of mRNAs, tRNAs and rRNAs. This dual role of the exosome requires specific signals that are associated with the different substrates. Thereby the complex must be able to structurally distinguish between all types of RNA. This mechanisms remain unclear, but it is very likely that the tertiary structure of the substrates as well as specific associated proteins play an important role.

6.3.2 RNA Degradation

For all exosome and exosome-like complexes a exoribonuclease activity was observed. However the position of active sites and the mode of degradation vary from archaea to yeast/human.

6.3.2.1 RNA Degradation by the Exosome in Archaea

The archaeal exosome as well as the bacterial PNPase possess phosphate dependent exonuclease activity. The active site is located at the inner side of the RNase PH ring, hence the RNA is degraded inside a processing chamber, similar to protein degradation by the proteasome. Although both Rrp41 and Rrp42 from the archaeal exosome adopt the RNase PH fold which is characteristic of phosphorolytic 3' → 5' exoribonucleases, only Rrp41 has catalytic activity (Lorentzen et al., 2005). All three active sites within the complex are only accessible via the central channel and the cap protein Rrp4 strongly influences the binding affinity to RNA (Oddone et al., 2007). From these results a model for RNA recognition and degradation was proposed and is shown in figure 26 (Büttner et al., 2005; Lorentzen et al., 2007). The model assumes that the 3' end of the RNA substrate is recruited by the S1-pore of the Rrp4/Csl4 cap and then threaded through the neck of the central channel to the processing chamber, where it binds to one of the active sites and is processively degraded.

6.3.2.2 RNA Degradation by the Exosome in Yeastb

In eukaryotes the equivalent active sites are not functional and no catalytic activity could be detected from the RNase PH ring (Liu et al., 2006). However, two additional factors that are only present in the eukaryotic exosome, namely Rrp6 and Rrp44 (also known as Dis3) showed hydrolytic exonuclease activity *in vitro* (Burkard and Butler, 2000; Dziembowski et al., 2007). Only Rrp44 is part of the core exosome and present in the nucleus and cytoplasm. From native mass spectrometry analysis (Hernandez et al., 2006) the position of the Rrp44 protein relative to the other 9-subunits could be detected: Rrp44 interacts with three adjoining proteins from the RNase PH ring, namely Rrp45, Rrp41 and Rrp42. In addition a negative-stain electron microscopy model of the yeast exosome at 20 Å resolution showed how Rrp44 associates with an 8-subunit exosome (missing the Csl4 protein) (Wang et al., 2007). From interpretation of the electron density map, the position of Rrp44 was located at the bottom of the RNase PH ring at the postulated exit-site of the RNA. Lorentzen et al. (2008) solved the crystal structure of yeast Rrp44 and showed that the presence of the other 9-subunits modulates the hydrolytic exonuclease activity of Rrp44. The 10-subunit exosome complex was not able to unwind RNA and can only fully degrade ssRNA. However, Rrp44 alone can easily degrade duplex RNA molecules with a 3' overhang.

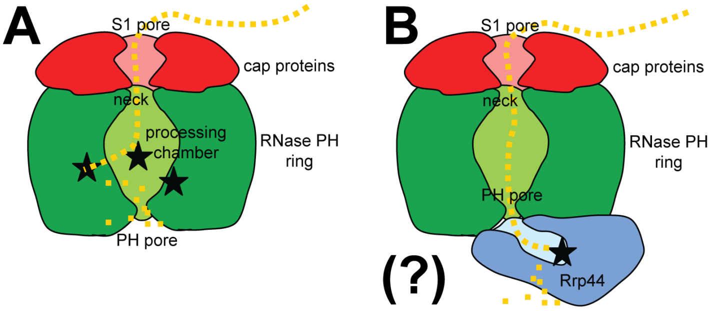


Figure 26: Model for RNA recognition and degradation by the archaeal and yeast exosome. A: Archaeal exosome. The RNase PH ring with the processing chamber is shown in green, the RNA binding cap proteins are shown in red. **B:** Yeast exosome. The RNAse PH ring without active sites is shown in green, the RNA binding cap proteins are shown in red and the hydrolytic exoribonuclease Rrp44 in blue. The proposed path of the RNA is shown in yellow. Active sites are indicated with a star. The model for the eukaryotic exosome is still speculative.

From all this recent data the following model can be proposed (figure 26): the overall architecture of the 9-subunit core of the exosome seems to be essential for correct functioning of the complex, as no catalytic activity could be detected *in vitro*. The RNA is proposed to bind to the exosome cap proteins in a similar way compared to the archaeal exosome. As no active site is existent in the chamber of the RNase PH ring, the RNA possibly leaves the exosome at the PH pore, where it may be directly guided to the active site of the attached Rrp44 protein. Therefore, the 9-subunit complex would function as regulator and could influence substrate specificity for its active subunit Rrp44.

In addition to the Rrp44 protein, the exosome can associate with a second exonuclease in the nucleus, the Rrp6 protein. Rrp6 participates both in RNA processing and quality control. The structure of yeast Rrp6 has a conserved RNase D core with a flanking helicase and RNase D C-terminal domain. The N-terminal domain is proposed to mediate the interaction with the exosome core. Rrp6 has distributive exonucelase activity and was shown to be involved in degradation of unstructured polyadenylated RNAs (Carneiro et al., 2007). This activity is not affected on assembly with the 10-subunit exosome (Liu et al., 2006). Degradation of structured RNA substrates by Rrp6 was shown to be promoted by the cofactor protein Rrp47 (Stead et al., 2007). Rrp47 is a nucleic acid binding protein and seems to facilitate binding to structured RNA elements. *Rrp6* is not an essential gene, but its deletion results in many RNA processing defects (Wyers et al., 2005).

This observation suggests that although partly overlapping in function, Rrp44 and Rrp6 have different substrate specificities *in vivo*.

6.3.3 RNA Polymerization Activity of the Archaeal Exosome

Polyadenylation is an important post-transcriptional modification of prokaryotic, eukaryotic and organellar RNA. In bacteria and organelles, such as mitochondria and chloroplasts, polyadenylation is transient and occurs mainly on RNAs as part of their decay pathway (Dreyfus and Regnier, 2002; Kushner, 2004). Also in human mitochondria non-abundant, polyadenylated RNA fragments were identified (Slomovic et al., 2005). In general, the mRNA decay process consists of endonucleolytic cleavage, addition of degradation-stimulating poly(A)-tails or poly(A)-rich sequences to the cleavage product and subsequent exonucleolytic degradation. In contrast to this form of degradation-stimulating polyadenylation, stable poly(A)-tails are added to mature 3'-end of most nuclear encoded mRNAs and are important for proper translation initiation, mRNA stability and in some cases nuclear export (Edmonts, 2002). This coexistence of stabilizing and degradation-stimulating poly(A)-tails needs a thorough regulation.

The main polyadenylating enzyme in *E. coli* is the poly(A) polymerase I. However, it was shown that for all heteropolymeric poly(A)-rich tails that were found in cyanobateria, gram-positive bacteria, *Streptomyces* and *Bacillus*, as well as in spinach chloroplasts, PNPase was the polymerizing enzyme. Therefore the nucleotide composition of a post-transcriptionally added extension can indicate the identity of the responsible enzyme (Mohanty and Kushner, 2000; Bollenbach et al., 2004).

The archaeal exosome is very similar to PNPase, so it is not surprising that heteropolymeric RNA tails could be detected in hyperthermophilic and methanogenic archaea (Portnoy et al., 2005). Not all archaeal organisms contain an exosome complex and it could be shown that heteropolymeric polyadenylation occurs only in organisms containing an exosome (Portnoy and Schuster, 2006).

Unlike RNA polymerases, PNPase and the archaeal exosome do not require a template and cannot copy one. When they are supplied with a mixture of ribonucleotide diphosphates (NDPs), the ensuing polymerization reaction forms a tail with random sequence. As described above, the enzymes can not only catalyze 5' → 3' polymerization, but also processive 3' → 5' phosphate dependent RNA degradation (Büttner et al., 2005; Grunberg-Manago, 1999). As a phosphorylase they use phosphate to cleave the phosphodiester bond. In contrast to hydrolysis, the phosphorolysis reaction is very close to equilibrium and therefore most reversible. *In vitr*o the direction of activity – either degradation or polymerization – can be controlled by the relative concentrations of NDPs and phosphate.

The molecular mechanism of RNA polyadenylation and degradation has been most intensively studied in *E. coli*. It was shown that PNPase is mainly active in degradation and only to a substantial degree it is active in polymerizing heteropolymeric tails (Mohanty and Kushner, 2000; Deutscher and Li, 2001). In contrast, in spinach chloroplasts, cyanobacteria and gram-positive bacteria the exosome is thought to be the major polyadenylating enzyme (Rott et al., 2001; Slomovic et al., 2006).

6.4 Cofactors of the Exosome Complex

The exosome is involved in the maturation and degradation of many different RNAs and it soon became obvious that it needs specific cofactors which help to distinguish between all these RNA species. In the last years a diverse array of activators and cofactors have been identified that activate the exosome on defined classes of transcripts. The most important activators in the context of a yeast cell are shown in figure 27.

6.4.1 Mtr4 and the TRAMP Complex in the Nucleus

To access the active site of the exosome, RNA substrates have to pass through a narrow pore that makes it impossible for double-stranded or structured RNA to be degraded. The DExH-box RNA helicase Mtr4 (also known as Dob1) has nucleic acid dependent ATPase activity, can unwind duplex RNA in the 3' → 5' direction and binds to single stranded RNA, especially to short poly(A) substrates (Bernstein et al., 2008). Mtr4 was shown to be required for most nuclear activities of the exosome (LaCava et al., 2005; Vanacova et al., 2005). Mtr4 is believed to assist the exosome as it moves through structured regions of its RNA substrates and thereby enables the RNA to reach the active site behind the exosome pore.

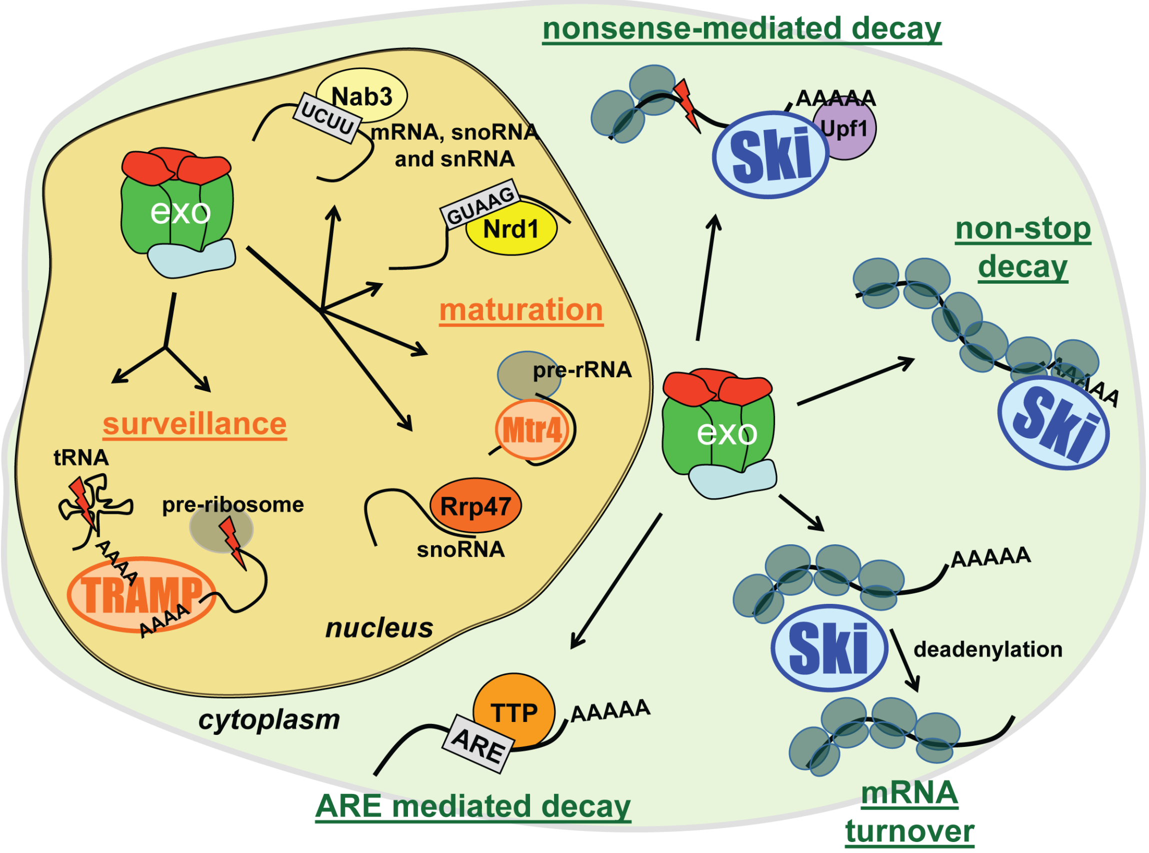

nonsense-mediated decay
Nab3
UCUU
mRNA, snoRNA
and snRNA
AAAAA
Upf1
Ski
exo
GUAAG
Nrd1
non-stop
decay
maturation
pre-rRNA
Mtr4
AAAAA
Ski
surveillance
tRNA
pre-ribosome
exo
Rrp47
snoRNA
AAAA
TRAMP
AAAA
AAAAA
nucleus
Ski
deadenylation
cytoplasm
TTP
AAAAA
ARE
ARE mediated decay
mRNA
turnover

Figure 27: Activation of the exosome by different cofactors in a yeast cell. Many different substrates and activating cofactors have been identified to activate the exosome in the nucleus and the cytoplasm of a yeast cell. The various pathways are described in chapter 6.4 (adapted from Houseley et al., 2006).

The action of Mtr4 on some RNAs occurs in the context of the TRAMP complexes, which consists of a Poly(A) polymerase (Trf4 or Trf5), a zinc-knuckle protein (Air1 or Air2) and the Mtr4 protein (LaCava et al., 2005; Wyers et al., 2005; Vanacova et al., 2005; Kadaba et al., 2004 and 2006). The TRAMP complex is thought to bind RNA through the zinc-knuckle, putative

RNA-binding domains that are present in Air1 and Air2. Mtr4 might then actively recruit the exosome to the RNA, as its depletion leads to RNA hyper-adenylation *in vivo*, which indicates that polyadenylation has been uncoupled from degradation (Houseley and Tollervey, 2006). The Trf4/5 protein adds short poly(A) tails to the substrate, thereby forming a favorable substrate for the exosome, and Mtr4 helps to dissolve RNA secondary structures.

Mtr4 processes a diverse array of pre-RNAs together with the exosome: maturation of the 5.8S ribosomal RNA and processing of snRNAs such as the U4 small nuclear RNA (Allmang et al., 1999). As part of the TRAMP complex it is additionally involved in the degradation of defective tRNAs and precursor ribosomal RNAs before they are exported to the cytoplasm.

Two distinct forms of TRAMP complexes, TRAMP4 and TRAMP5 have been identified in yeast. They show clear substrate preferences *in vivo*, but their mode of selectivity is still unclear.

Both the exosome and the TRAMP4 complex interact with an additional RNA binding protein complex called the nuclear pre-mRNA down-regulation complex 1 (Ndr1) (Vasiljeva and Buratowski, 2006). This complex consists of the RNA helicase Sen1 and the proteins Ndr1 and Nab3 (nuclear polyadenylated RNA-binding) that recognize specific sequence motifs on RNAs (Steinmetz and Brow, 1998; Carroll et al., 2004). This complex is required for transcription termination of snRNA and snoRNA genes (Steinmetz et al., 2001). *In vitro* the Nrd-complex can directly stimulate exosome degradation of substrates with Ndr1- and Nab3-binding motifs. *In vivo* it probably helps to bring the exosome to specific RNA substrates.

6.4.2 The Ski Complex

In the cytoplasm the exosome needs the specific cofactor Ski7, a GTPase with homology to translation factors for most of its activities (Araki et al., 2001). In addition, the Ski2, Ski3 and Ski8 proteins from a complex known as the Ski complex are involved in most of the cytoplasmic activities of the exosome:

The *ski* (superkiller) genes were identified via mutations that cause overexpression of a killer toxin encoded by the endogenous double-stranded RNA (Toh et al., 1978). It could be shown that not only the exosome, but the proteins Ski2, Ski3 and Ski8 are required for the 3'-mRNA degradation (Anderson and Parker, 1998). Ski2p is a putative RNA helicase with homology to Mtr4 (Widner and Wickner, 1993), Ski3p is a tetratricopeptide repeat (TPR) protein (Rhee et al., 1989) and Ski8p contains five WD-40 (beta-transducin) repeats (Matsumoto et al., 1993). The three Ski proteins form a stable complex, which is localized in the cytoplasm (Brown et al., 2000). Mutations in the three genes inhibit 3' → 5' mRNA decay, but do not affect other functions of the exosome (Anderson and Parker, 1998). Therefore the Ski-complex was proposed to be the cofactor of the exosome in the degradation of mRNAs in yeast.

Ski7, another member of the Ski proteins, is permanently associated with the exosome in the cytoplasm. The *ski7* gene was initially identified as one of the *ski* gene family (Benard et al., 1999). It was shown that Ski7 is also required for 3' → 5' mRNA degradation and acts in the same pathway as the Ski-complex, because deletion of the *ski7* gene caused impaired 3'-mRNA decay similar to *ski2*, *ski3* or *ski8* deletions (van Hoof et al., 2002). When introducing a mutation in one of the genes of the Ski-complex, the complex did not assemble any more. In contrast, the Ski complex was still intact in the *ski7* mutant, suggesting that Ski7 is not a member of the Ski complex and is not required for the formation of the complex (Brown et al., 2000). Therefore it seems that the role of Ski7 in mRNA degradation differs from the rest of the Ski complex.

The Ski complex together with the exosome and Ski7 function in the nonsense-mediated decay and the non-stop decay mRNA surveillance pathways (Mitchell and Tollervey, 2003; Lejeune et al., 2003; van Hoof et al., 2002).

6.4.3 Sequence Specific Cofactors

Some cofactors were identified that recruit the exosome to RNA substrates with specific sequences. They include the Nrd1 protein (see 6.4.1), which enhanced the RNA degradation activity of the exosome *in vitro* and *in vivo* (Arigo et al., 2006; Vasiljeva and Buratowski; 2006). The precursors of many RNA species contain Ndr1-binding sites, which probably function as targets for exosome-mediated degradation (Steinmetz et al., 2001). It is likely that during normal processing of RNAs with Ndr1 binding site, this binding site is removed and hence only correctly processed RNAs are protected from exosome-mediated degradation.

In human cells, ARE (AU-rich elements) mediated degradation is an important mRNA turnover pathway that involves the recruitment of the exosome (Mukherjee et al., 2002). AREs are found in the 3'-untranslated region of many mRNAs that code for proto-oncogenes, nuclear transcription factors and cytokines. They represent the most common determinant of RNA stability in mammalian cells and are known to target mRNAs for rapid degradation (Barreau et al., 2005). Mukherjee et al. (2002) suggested that certain subunits of the human exosome specifically bind to AREs causing an ARE-dependent degradation of mRNAs. Other studies have shown that several ARE-binding proteins, for instance TTP (Tristetraprolin) and KSRP (KH splicing regulatory protein), are physically associated *in vitro* with the exosome and are required for preferential degradation of ARE-containing mRNAs by the exosome (Chen et al., 2001; Gherzi et al., 2004). Tran et al. (2004) proposed the involvement of the RNA helicase RHAU. RHAU displaces mRNA stabilizing proteins from the ARE and then recruits the exosome to the RNA to facilitate mRNA decay. ARE-mediated decay can also occur in yeast, where the activity requires a TTP homolog – the role of the exosome has not yet been reported.

It is still not much known about the way all these cofactors are able to activate the exosome. However, a physical recruitment of the exosome to specific RNA substrates seems to play a key role in stimulating RNA degradation both in the nucleus and the cytoplasm.

6.5 Aim of the Project

Exosomes occur in archaeal and eukaryotic cells as main 3' → 5' exonucleases that are involved in the maturation and surveillance of many different types of RNA. Recent structural studies revealed the atomic structure of the archaeal and the human exosome complex (Büttner et al, 2005; Liu et al., 2006). The archaeal exosome possesses phosphate dependent exonuclease activity and its active site is located inside of the 9-subunit core complex. In eukaryotes, this active sites are not functional any more, but an additional subunit, Rrp44, is bound to the 9-subunit core and provides hydrolytic exonuclease activity. However, although the position and structure of the active sites changed from archaea to eukaryotes, the overall structure of the core exosome is highly conserved and extremely similar. This points to the fact that the architecture of the exosome 9-subunit core has a regulatory function, which seems to be unchanged from archaea to humans.

Based on the structural insights from Büttner (2007), the aim of this thesis was a more detailed investigation of the archaeal exosome. Structural as well as biochemical methods will be used to reveal the mode of function and the enzymatic activities of the exosome complex.

Using Small Angle X-ray Scattering the known crystal structures will be compared with the conformations of the Rrp4 exosome, Csl4 exosome as well as the 6-subunit complex in solution.

The atomic structure of the exosome complex with a bound RNA substrate will be determined and the RNA-exosome structure together with biochemical data will be used to propose a model for the mode of enzymatic degradation.

7 Results – Exosome

7.1 Exosome Structures in Solution

The crystal structures of the two different 9-subunit archaeal exosome complexes were solved (Büttner et al., 2005) and are shown in chapter 6, figure 25. The three dimensional organization of a protein in a crystal can differ from its shape in an aqueous solution. Therefore SAXS studies were performed to determine a low resolution structure of the exosome complexes in solution and under physiologic conditions. The complexes were purified as described in chapter 2 and three different concentrations (5, 10 and 15 mg/ml) were measured at the SIBYLS beamline (ALS, Berkeley, USA). No aggregation and intermolecular forces between the complexes could be observed. With numerous consecutive measurements of the same sample it could be shown that the scattering of exosome samples was not influenced by long exposure to high energy X-ray radiation.

The obtained scattering curves have a shape that is typical for globular proteins and are shown in figure 28.

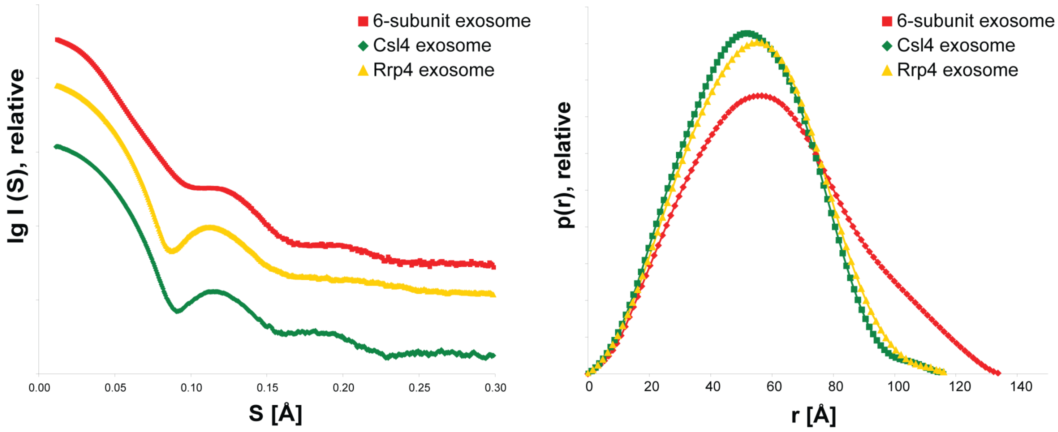


Figure 28: SAXS curves and pair distribution functions of three archaeal exosome complexes. The curves of the 6-subunit exosome are shown in red, for the 9-subunit Csl4-exosome in green and for the 9-subunit Rrp4 exosome in yellow.

Using the Guinier approximation for small s-values, the radii of gyration were determined. The three complexes have the following R_g: 6-subunit complex: 46.2 Å; 9-subunit Csl4 exosome: 41.4 Å and 9-subunit Rrp4 exosome: 39.6 Å. The pair distribution functions were calculated with GNOM and are shown in figure 28. Using GASBORp for the 9-subunit complexes and GASBORi for the 6-subunit complexes, 15 structures were modeled with the *ab initio* dummy residue approach for all three complexes. For the modeling process a three-fold symmetry was assumed. The obtained structures were aligned and averaged using DAMAVER. The resulting bead models were superimposed with the equivalent crystal structures and are shown in figure 29.

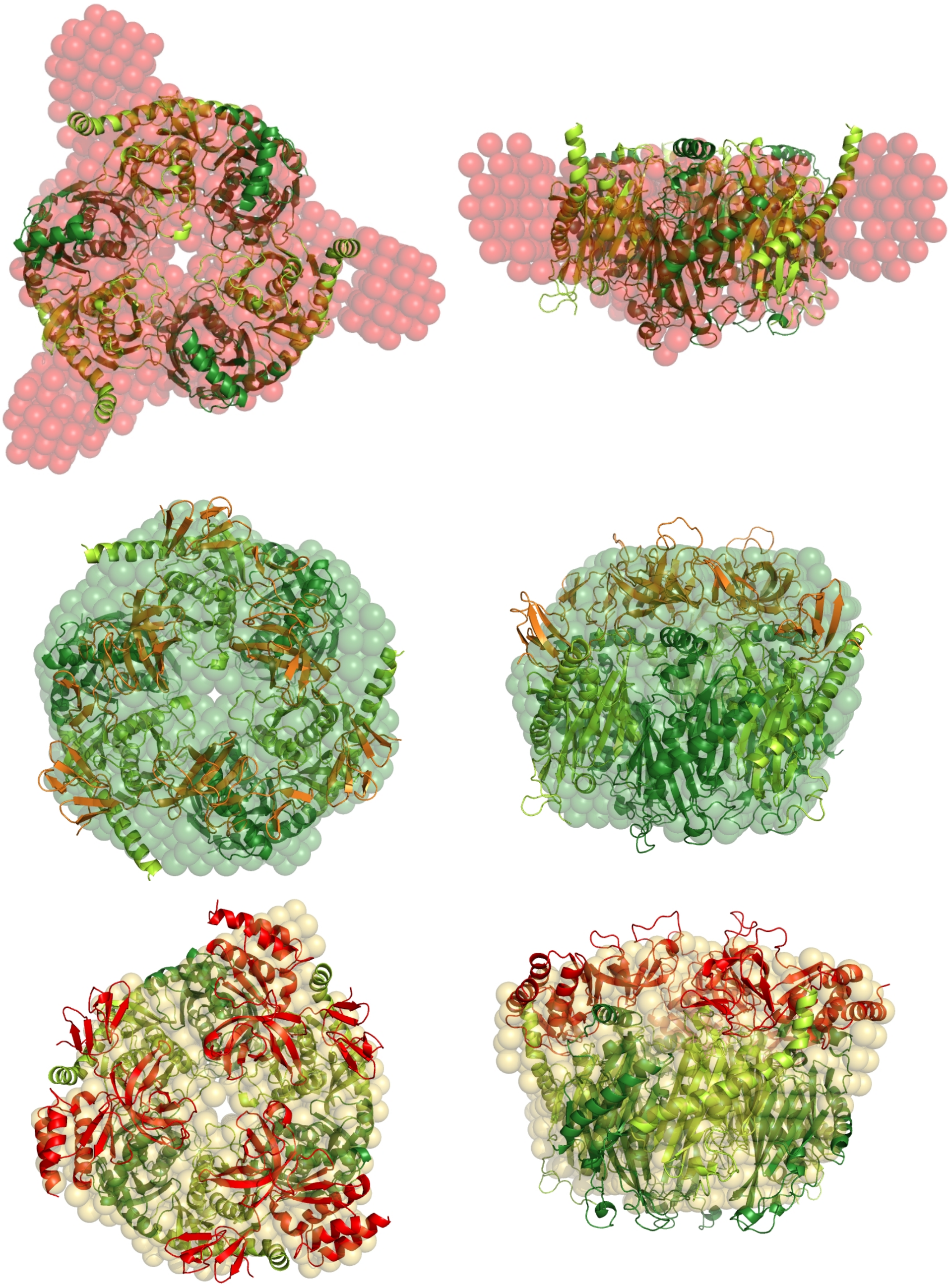


Figure 29: Low resolution structures of exosome complexes in solution superimposed with the equivalent crystal structure. The 6-subunit exosome is shown in red, the 9-subunit Csl4 exosome in green and the 9-subunit Rrp4 exosome in yellow. For all three complexes the structures are shown from the top and from the side. Superposition was done with SUPCOMB.

7.2 Crystal Structures of the Archaeal Exosome

7.2.1 Csl4-Exosome Wild Type with RNA Bound to the Active Site

The archaeal exosome subunits Rrp41, Rrp42 and Csl4 were coexpressed in *E. coli* cells and purified as described in chapter 2. The correct formation of a 9-subunit complex was verified using size-exclusion chromatography. The protein was concentrated to 20 mg/ml, which corresponds to approximately 90 µM 9-subunit complex. Synthesized RNA molecules with different length (from 6 bases to 12 bases) were added to a final concentration of 0.4 mM RNA. Crystallization screens with 0.3 µl + 0.3 µl drops and 50 µl reservoir solution resulted in the growth of different crystal forms in different conditions. Several conditions were refined and the obtained crystals were tested for diffraction. Best diffracting crystals grew in 100mM Na-acetate, pH 4.6; 100mM NaCl and 30% MPD and are shown in figure 30A. In parallel, the protein complex was crystallized without RNA in the same condition and crystals were soaked with RNAs of different lengths over night. All obtained crystals were directly flash frozen in liquid nitrogen and data sets were recorded at the PX beamline (SLS, Villigen, CH) and the ID-14-2 (ESRF, Grenoble, F). Datasets diffracted to 2.4 Å – 3 Å and indexing with XDS determined a unit cell of a = 138.27 Å, b = 138.27 Å, c = 262.27 Å, $\alpha = 90°$, $\beta = 90°$, $\gamma = 90°$ and the space group $P4_322$. The structures were solved with molecular replacement using the atomic structure of the Csl4-exosome (pdb-code 2BA1) as search model. For further details of data collection see appendix table A2.

Difference electron density maps were calculated and additional electron density for some RNA bases in the active site became visible (figure 30 C and D). The crystals produced with co-crystallization or soaking with RNA substrates smaller than 10 bases all resulted in a comparable density map. They possessed the same space group, diffracted to similar resolutions and differed only in the first decimal place of the unit cell constants. The best electron density was used for model building and refinement. The obtained difference density map is shown in figure 30. Clearly, an RNA molecule is bound to all three active sites. A model for the RNA was manually built into the density using COOT (figure 30) and refinement was done with CNS and PHENIX. Density for four bases was clearly visible in the active sites, for the following base only parts of the density could be detected. This observation was independent from the length of used RNA: even in crystals containing an 8-mer RNA molecule electron density for not more that six bases could be identified. Crystals that were soaked with a 10- or 12-mer RNA dissolved after some minutes. This points to a binding region of the RNA that interferes with crystal packing.

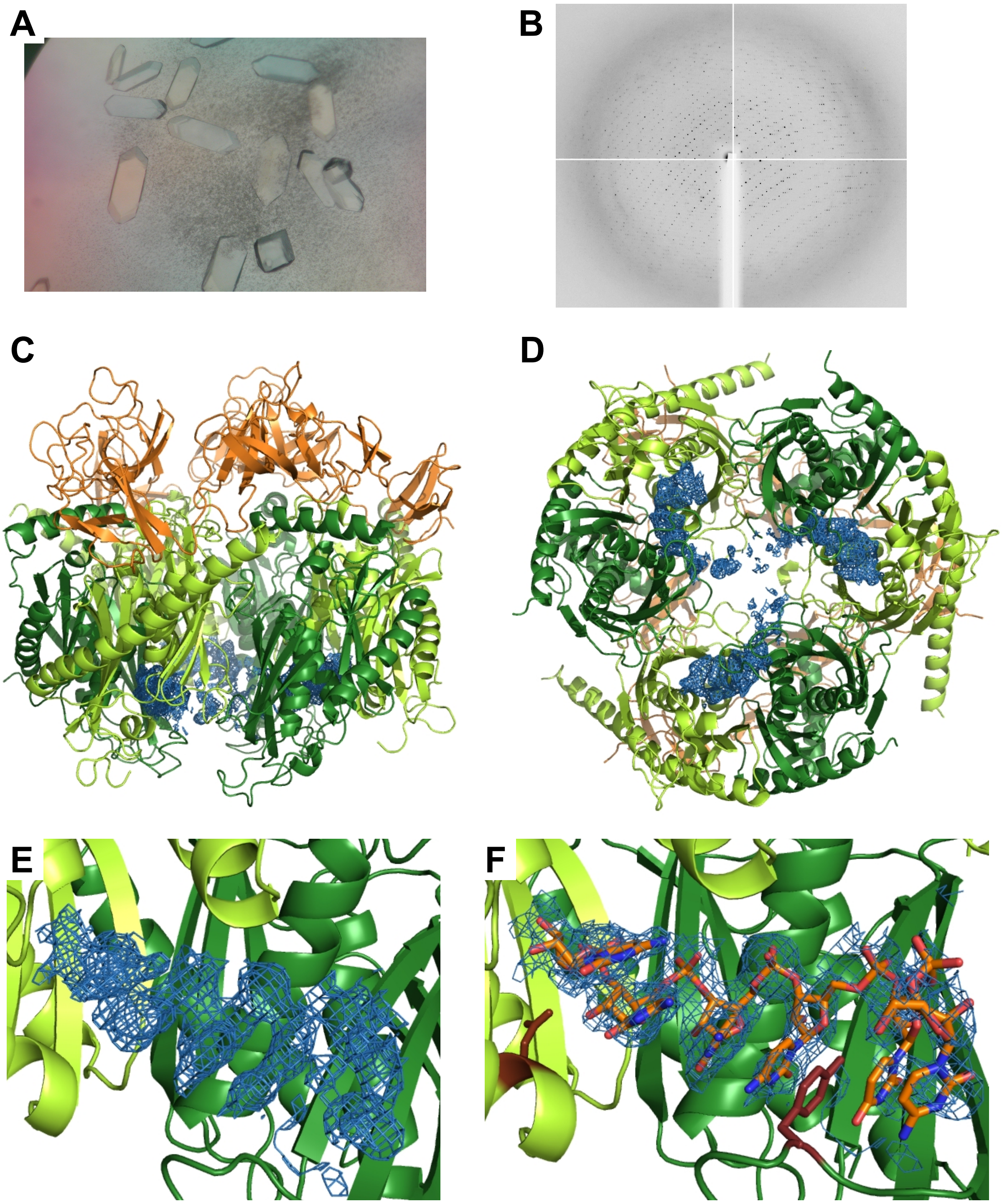


Figure 30: Solving the atomic structure of RNA (CCCCUC) bound to the active site of the archaeal exosome. A: Co-crystals of the exosome-RNA complex. **B:** Diffraction pattern of the crystals shown in A. **C-F:** Csl4-exosome complex with additional electron density (blue, $|F_o| - |F_c|$ map, 1σ contour) in the active site in side-view (C), in bottom view (D) and a closeup of one active site (E). In panel F a model for a 6-mer RNA is built into the obtained density, residues Y70 and D180 are shown in sticks (brown). Color code of exosome subunits: Rrp41 – light green; Rrp42 – dark green and Csl4 – orange.

In agreement to former results (Büttner et al., 2005), which showed that the proposed active site of Rrp42 is non-functional, RNA could only be detected in the active sites of the Rrp41 protein.

The RNA is situated in a binding cleft (figure 31) that involves both subunits of one hetero-dimer of Rrp41 and Rrp42. This finding explains, why the Rrp41 subunit alone is not capable of degradation of RNA, but needs the presence of Rrp42 (Lorentzen et al., 2005). The aspartate 180 that was shown to be essential for exosome activity is located close to the 3' phospho-diester bond to catalyze the degradation reaction.

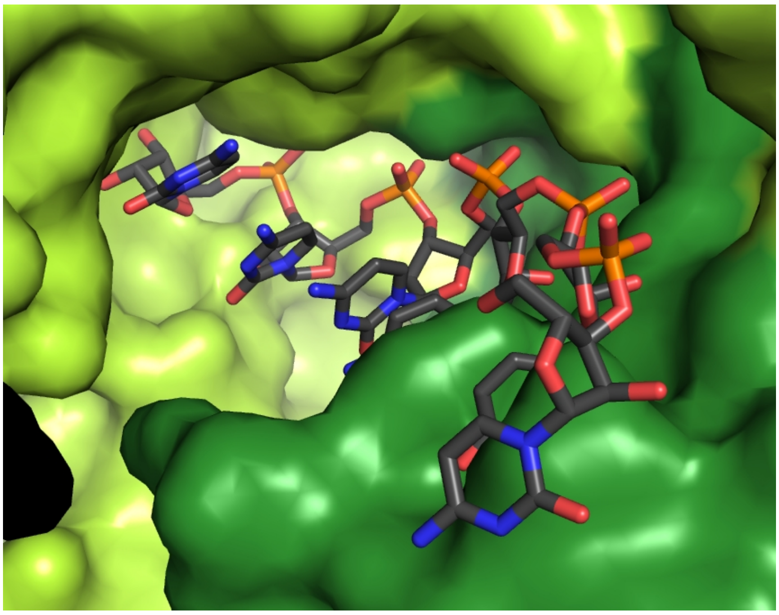


Figure 31: RNA binding cleft of the archaeal exosome with a 6-mer RNA molecule. The Rrp41 subunits is shown in light green and the Rrp42 subunit in dark green. The interactions between the RNA and protein are mainly via the backbone and the bases point away from the protein surface.

The RNA is recognized mainly via electrostatic interactions between the phosphate backbone and a set of arginine residues. In addition, π-stacking between the tyrosine 70 of Rrp42 and the fourth base of the RNA (counting from the active site) stabilized the position of the RNA. All these interactions are sequence independent, which explains why the exosome degrades its substrates in a sequence inspecific manner.

Interactions between the 2'-OH of the ribose and the exosome can explain the specificity of the exosome for RNA in contrast to DNA, where this OH-group is missing. An important contact is detectable between arginine 114 and the OH-group of the fourth base.

7.2.2 Csl4-Exosome Y70A^{Rrp42} with RNA Bound to the Active Site

From the analysis of the atomic structure of the exosome-RNA complex, the tyrosine residue 70 of the Rrp42 protein was proposed to play an important role in RNA binding. Therefore the tyrosine was mutated to an alanine. The exosome (Y70A^{Rrp42}) was expressed and purified in the same way as the wild-type exosome and the protein was co-crystallized with the same RNA molecule (CCCCUC). The structure was solved as it was done with the wild-type complex. In this mutant version, RNA could also be detected in all three active sites. In the wild-type protein four bases were clearly visible and two more bases only partly. In the Y70A^{Rrp42} mutant of the exosome, only for the first four bases electron density could be observed, and the intensity of this density was lower in all three active sites. This points to a decreased affinity of the RNA to the mutant exosome and a role of tyrosine 70 in RNA binding. A closeup view of the superposition of both structures in the region of tyrosine 70 is shown in figure 32.

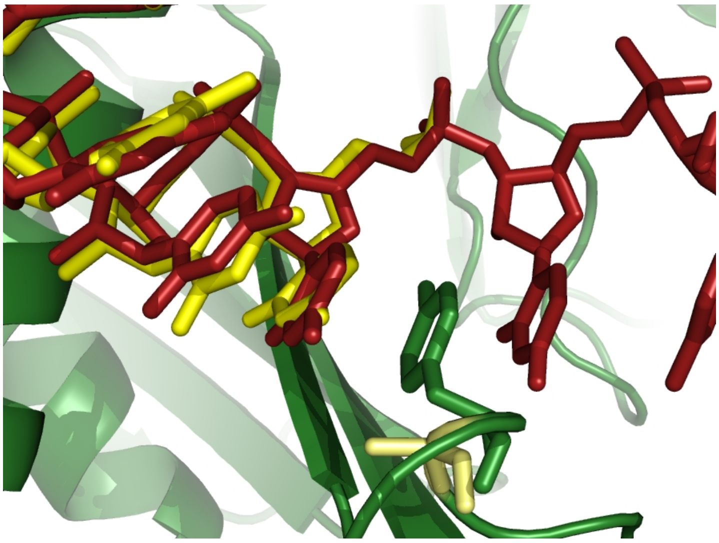


Figure 32: Closeup view of Rrp42 tyrosine 70. The RNA from the wild-type structure is shown in red sticks, the tyrosine in green. The RNA from the mutant structure and the alanine that replaces the tyrosine is shown in yellow.

7.2.3 Crystal Structure of the Archaeal Csl4 S1 and Zn-ribbon Domain

To learn more about the role of the exosome cap proteins, the Csl4 protein was purified on its own and crystallization screens were setup. The full-length protein did not result in any crystal growth, so the possibly flexible N-terminal domain (residue 1 – 49) was removed and this truncated version of Csl4 was purified. The protein was concentrated to 5.5 mg/ml and screened for crystallization conditions. Crystals grew in 100 mM HEPES, pH 7.5, 20% PEG 4000 and 10% isopropanol. The condition was refined to 100 mM HEPES pH 7.8; 18% PEG 4000, 10% isopropanol and 10% glycerol and crystals were directly flash frozen in liquid nitrogen. A dataset was recorded at the PX beamline (SLS, Villigen, CH). The crystal diffracted to 1.8 Å and indexing with XDS determined a unit cell with a = 42.02 Å, b = 75.62 Å, c = 76.61 Å, $\alpha = \beta = \gamma = 90°$ and the space group $P2_12_12_1$. The structure was solved with molecular replacement using the S1 and Zn-ribbon domain of the exosome-bound Csl4 protein as search model. All parameters of structure solution and refinement are summarized in table A2 in the appendix. To compare the obtained structure with the exosome-bound Csl4, both structures were superimposed (figure 33). Clearly the two RNA binding domains (S1 and Zn-ribbon) do not change their conformation when they are isolated and not bound to the exosome. This may point to a possible role of the Csl4 protein, when isolated from the exosome complex.

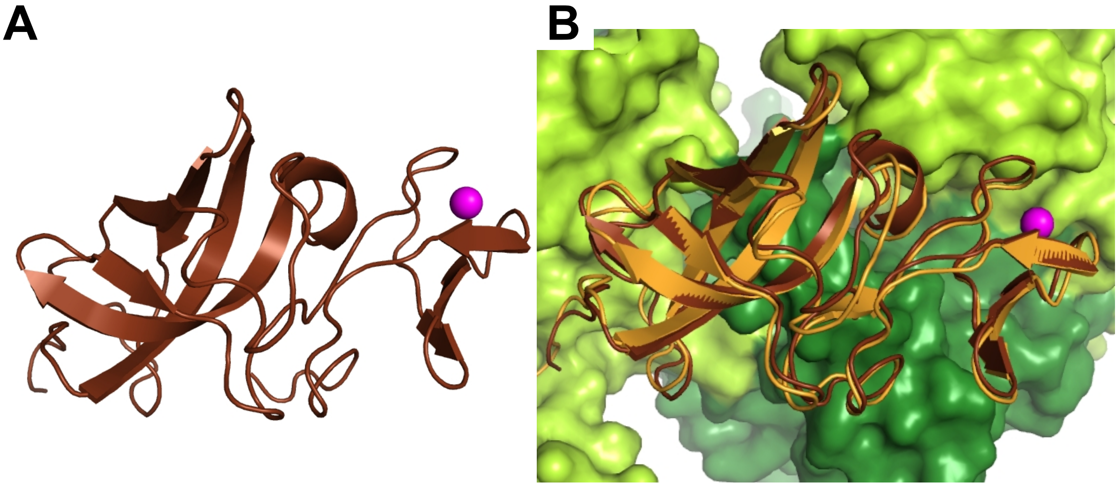


Figure 33: Crystal structure of the archaeal Csl4 protein lacking the N-terminal domain (A). The zinc atom is shown in magenta. **B:** Superposition of the isolated S1- and Zn-ribbon domain with the archaeal exosome shows the same overall fold. The surface representation of the exosome 6-subunit core is shown in light (Rrp41) and dark (Rrp42) green, the exosome bound Csl4 is shown in orange and the isolated S1- and Zn-ribbon domains in brown.

7.3 The Archaeal Exosome Bound to a Large RNA Molecule

7.3.1 Preparation

The binding of RNA to the active site of the archaeal exosome complex was analyzed with structural methods by determination of the atomic structure described above.

More insight into the role of the cap proteins in RNA binding and recognition was gained using biochemical assays. Structural information about the cap proteins in complex with RNA would increase the knowledge and could help to better understand the duty of the conserved cap protein as part of the exosome complex. All trials to crystallize the exosome with *in vitro* synthesized

RNA bound to the cap proteins failed. Either crystals did not grow at all or the RNA was only found in the active site. The truncated cap proteins (S1- and Zn-ribbon-domain of Csl4 or S1- and KH-domain of Rrp4) were also used for crystallization setups with short RNA molecules and no crystals could be obtained.

A totally different approach was therefore used to achieve structural results of the exosome complex with RNA bound to the cap proteins. During the purification of recombinantly produced exosomes, high amounts of nucleic acids were detected to be bound to the complex. Therefore the complex was washed with a buffer containing a high salt concentration during Ni-affinity purification. With this washing step, pure protein without the contamination of nucleic acids could be obtained.

For the following analyses the exosome complex was purified without this high-salt washing step. The resulting protein-nucleic acid complex was stable up to 600 mM salt on the anion exchange column and eluted as a single peak from the size exclusion chromatography column. For analysis the complex was applied on a 2 % agarose gel in loading buffer containing 8m urea. The nucleic acid could be stained with ethidium bromide and traveled in the gel as one single band (appendix 9.6). To determine the length of this RNA, the RNA was separated from the protein using a buffer containing 8M urea and the NucleoSpin RNA clean-up kit (Macherey-Nagel, Düren). The isolated RNA was dephosphorylated with alkaline phosphatase and radioactively labeled by transformation of the γ-phosphate from γ-^{32}P-ATP with T4 polynucleotide kinase. The labeled RNA was analyzed on a denaturing PA gel and the size was estimated to be between 50 and 70 bases using a calibration curve from RNAs with known length.

7.3.2 Structural Analysis in Solution

The obtained nucleic acid-exosome complex was used for SAXS studies. Therefore the complex was concentrated to an absorption at 280 nm of A_{280} = 55 and measured in a 1:0, 1:1 and 1:2 dilution to evaluate the concentration dependency of scattering. The complex did not show concentration dependent aggregation and was not affected by long exposure to high-energy X-rays. The obtained scattering curve and the pair distribution function are shown in figure 34. The radius of gyration was determined to be 46.8 Å using the Guinier approximation for small s-values. Thus the R_g is substantially larger that the one for the Rrp4 exosome complex without RNA, which was determined to be 39.6 Å.

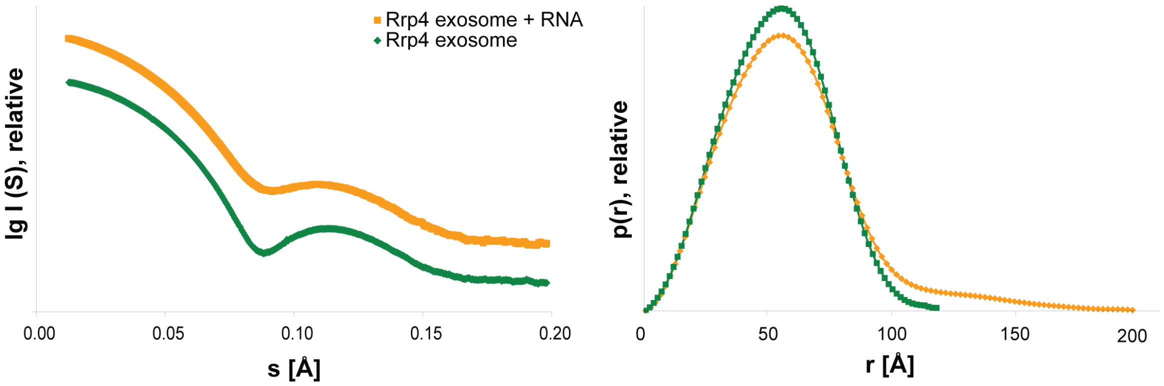


Figure 34: SAXS curves and pair distribution function of the Rrp4 exosome (green) compared to the RNA bound complex (orange). The R_g for the RNA bound complex was determined to be 46.8 Å.

10 different *ab initio* models were calculated with GASBORp from the pair distribution function shown in figure 45, aligned and averaged. The resulting bead model was superimposed with the crystal structure of the Rrp4 exosome and is shown in figure 35.

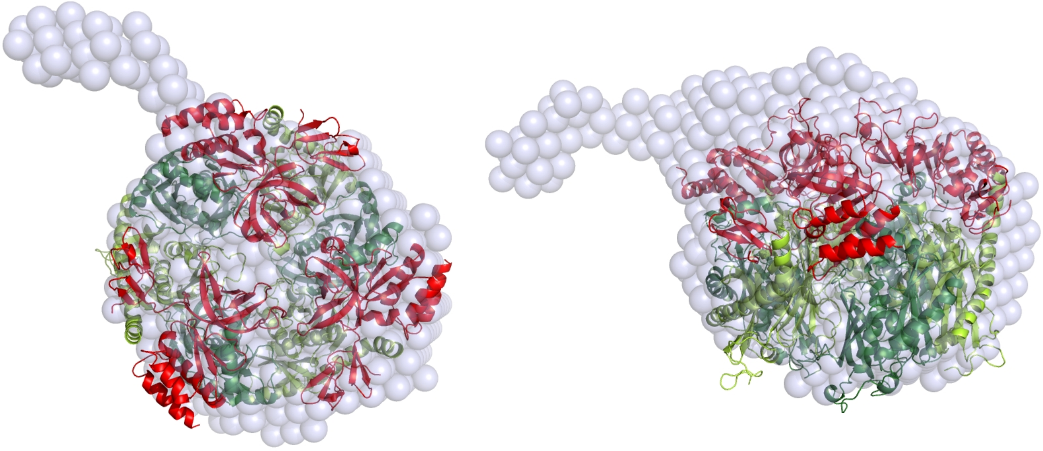


Figure 35: Superposition of the Rrp4-exosome crystal structure with the SAXS structure of the Rrp4-exosome-RNA complex. The bead model (light blue) was calculated from the pair distribution function shown in figure 45 and 10 models were aligned and averaged. The complex is shown in side and top view. (Rrp41 – light green; Rrp42 – dark green; Rrp4 – red).

7.3.3 Crystal Structure of the Complex

In the same way as the sample for SAXS measurements, a second batch was prepared for crystal setups. Cubic crystals grew in 100 mM Tris pH 7.5; 100 mM NaCl; 50 mM $MgCl_2$ and 1% PEG 4000. Crystals were soaked in reservoir solution with 15 % 1,4-butanediol as cryoprotectant and flash frozen in liquid nitrogen. The crystals diffracted to 3 Å and a data set was recorded at the PX beamline (SLS, Villigen, CH).

Indexing with XDS determined a unit cell of a = b = 112.77 Å, c = 179.25 Å, α = β = 90°, γ = 120° and the space group R3. Molecular replacement with PHASER using the 9-subunit Rrp4 exosome as search model did not lead to a reasonable result. However, with the 6-subunit complex as search model, phases could be obtained. This electron density map had high similarity to the map from the 6-subunit exosome, but showed additional density on top of the core ring, at the position where normally the cap proteins are located (see figure 36). Although this electron density was clearly visible, it was not good enough to manually built an atomic model into it.

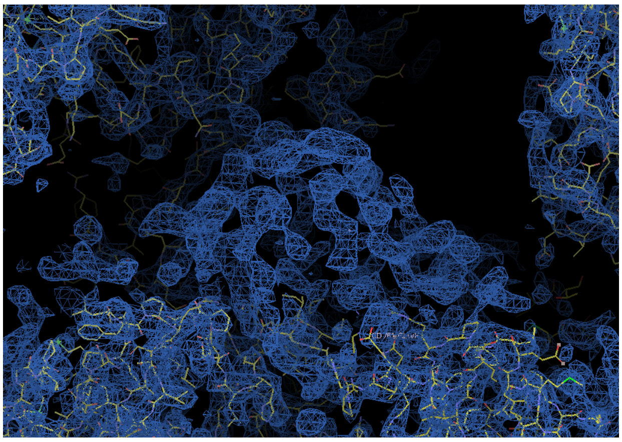


Figure 36: 2 $|F_o|$ - $|F_c|$ electron density map after molecular replacement with the 6-subunit exosome as search model. Clearly a protein is bound on top of the core exosome.

After molecular replacement the phase information is usually not very good for the parts of the molecule that are distant from the search model. Therefore a second approach was used to improve the additional electron density. After searching with the 6-subunit core, single subunits of the Rrp4 protein (N-terminal domain, S1-domain and KH-domain) were used for an additional search round. Only the S1-domain was found by PHASER and was located at a slightly changed position compared to the apo complex. In the regions surrounding the S1-domain there was still additional density visible. Especially the helices of the KH-domain could be identified and were manually modeled into the density using COOT. Rigid body modeling followed by refinement with CNS resulted in the model shown in figure 37. Superposition of the structure with the already known Rrp4 exosome apo structure shows a tilting of the cap protein towards the pore of the processing chamber (see figure 37).

Still there is more additional density visible for the N-terminal domain, the structure has to be further refined and more parts of the model have to be fitted into the density.

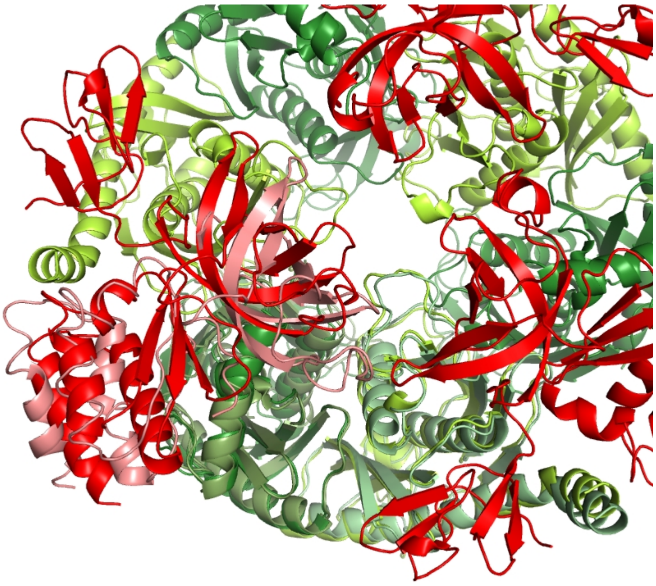


Figure 37: Superposition of the crystal structure of the apo Rrp4 exosome (pdb-code 2BA0) with the preliminary atomic model from the Rrp4 exosome purified with endogenous *E. coli* RNA. The Rrp4 protein from the apo structure is shown in red and the S1 and KH domains from the new structure are shown in salmon.

7.4 Enzymatic Activities of Archaeal Exosome Complexes

The archaeal exosome possesses three active sites within the processing chamber (Büttner et al., 2005). The complex is known to catalyze two different reactions: either the phosphate dependent 3' → 5' degradation of single stranded RNA molecules or the reverse reaction, the addition of single bases to the 3'-end of a single stranded RNA molecule using a nucleotide-di-phosphate. In this thesis both activities were studied in detail by the use of denatured polyacrylamide (PA) gelelectrophoresis.

7.4.1 Variants of Exosome Complexes Used for Activity Assays

For the analysis of the enzymatic activities, different variants of the archaeal exosome were used. First, the three wild-type exosome complexes were studied: 6-subunit exosome, Csl4 9-subunit exosome and Rrp4 9-subunit exosome. Second, mutant versions were tested: the tyrosine mutant Y70A (see 7.1.2) and the neck-mutation R65E (Büttner et al., 2005). Additionally, an interface mutation was introduced (K51E^{Rrp41}) that leads to formation of only the Rrp41/Rrp42 dimer and destroys the hexameric ring. The complex was purified analogous to the wild-type protein and the existence of a dimer was tested by size exclusion chromatography. The different variants are shown schematically in figure 38.

As negative control, the active site mutation D180A^{Rrp41} was used.

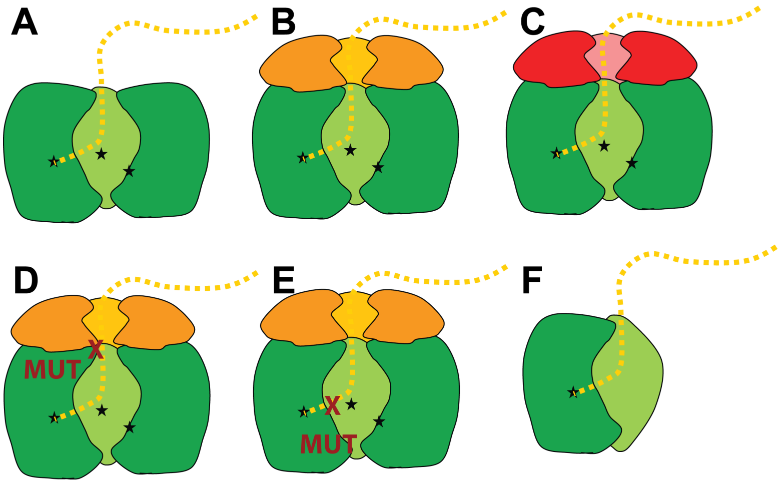


Figure 38: Exosome variants that were used for enzymatic activity assays. A-C: Wild-type exosome complexes: 6-subunit exosome (A); Csl4-exosome (B); Rrp4-exosome (C). **D+E:** Putative RNA binding mutants. Neck mutant (R65E^{Rrp41}) (D) and tyrosine mutant (Y70A^{Rrp42}) (E). **F:** Interface mutant (K51E^{Rrp41}) resulting in a single Rrp41/Rrp42 dimer.

7.4.2 RNA Degradation and Polymerization Assays

7.4.2.1 RNase Activity of the Archaeal Exosome

To gain insights into the mode of RNA degradation, exosome complexes were incubated with RNA substrates and the reaction products were monitored with time. As radioactively labeled RNA with α-^{32}P-phosphate group at the 5'-end was used, not the cleaved nucleotide diphosphate but the remaining RNA molecule could be detected. The exosome complexes originate from the thermophile organism *Archaeoglobus fulgidus*, thus the assays were performed at 50°C.

Figure 39 shows the denaturing gel of an RNase assay with a 30mer poly(A) RNA as substrate. The three different exosome complexes 6-subunit exosome, Csl4 exosome and Rrp4 exosome were tested for activity in the presence of 10 mM $MgCl_2$ and 10 mM phosphate; only 10 mM $MgCl_2$ or none of both.

The Csl4 exosome is degrading the polyA-RNA faster than the Rrp4 exosome does. The 6-subunit exosome is also capable of degrading the substrate, but with even lower activity. The final product of degradation is a three-mer of RNA (not visible on this gel, see figure 40). Clearly the exosome is also able to degrade RNA in a phosphate independent way with much lower activity and it needs a bivalent cation for catalysis. It could be shown that the exosome works with Mn^{2+} in a similar way as Mg^{2+}. The active site mutant D180G as well as the double mutant D180G-R65A did not show any RNase activity.

The same experiment was done with single stranded DNA and no activity could be observed with all complexes.

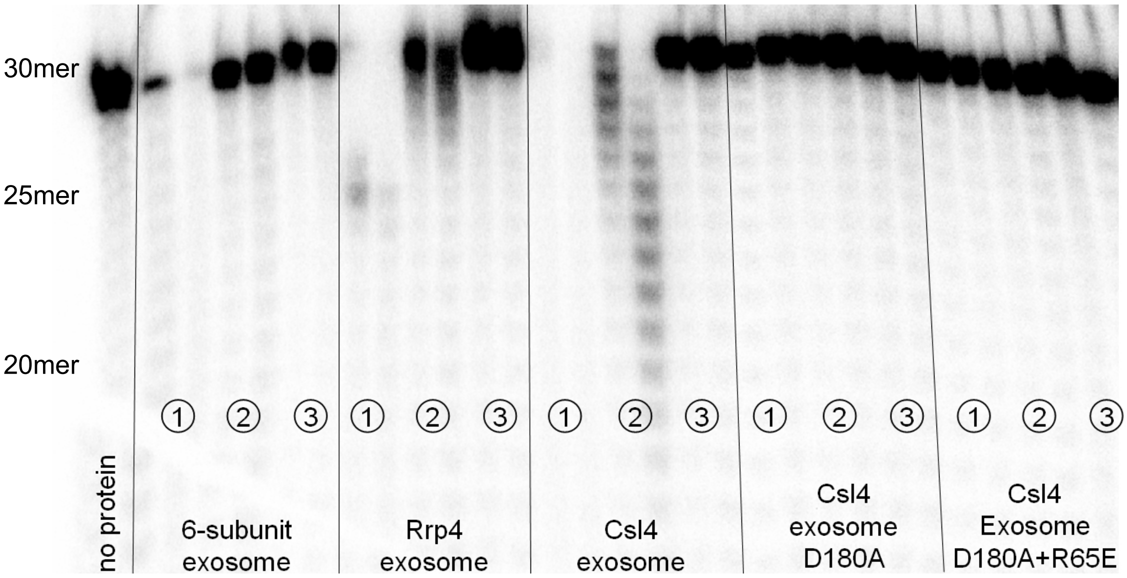


Figure 39: Denaturing PA gel for the analysis of RNase activity of different exosome complexes. Radioactively labeled RNA was visualized. Every complex was incubated with 10 mM $MgCl_2$ and 10 mM phosphate (1); only with 10 mM $MgCl_2$ (2) and with none of both (3). Samples were taken after 2 min and 6 min. For clearness only the upper part of the gel is shown. The final product of degradation is a three-mer of RNA (not visible on this gel, see figure 40)

7.4.2.2 Polymerization Activity of the Archaeal Exosome

In addition to the phosphate dependent RNase activity, the archaeal exosome is capable of catalyzing the reverse reaction. Thereby it uses a nucleotide-di-phosphate and adds it to the 3'-end of a single stranded RNA molecule. To analyze this enzymatic activity, the exosome was incubated in the same conditions, as for monitoring RNA degradation – just by exchanging phosphate with ADP. A denaturing polyacrylamide gel showing both the degradation and polymerization activity is shown in figure 40.

The length of the different RNA degradation products is known and the distance they cover in the gel can be measured. With this data it is possible to create a calibration curve and estimate the length of polyadenylated RNA molecules. From the gel shown in figure 40 the length of polyadenylated RNA was estimated to be approximately 130 bases.

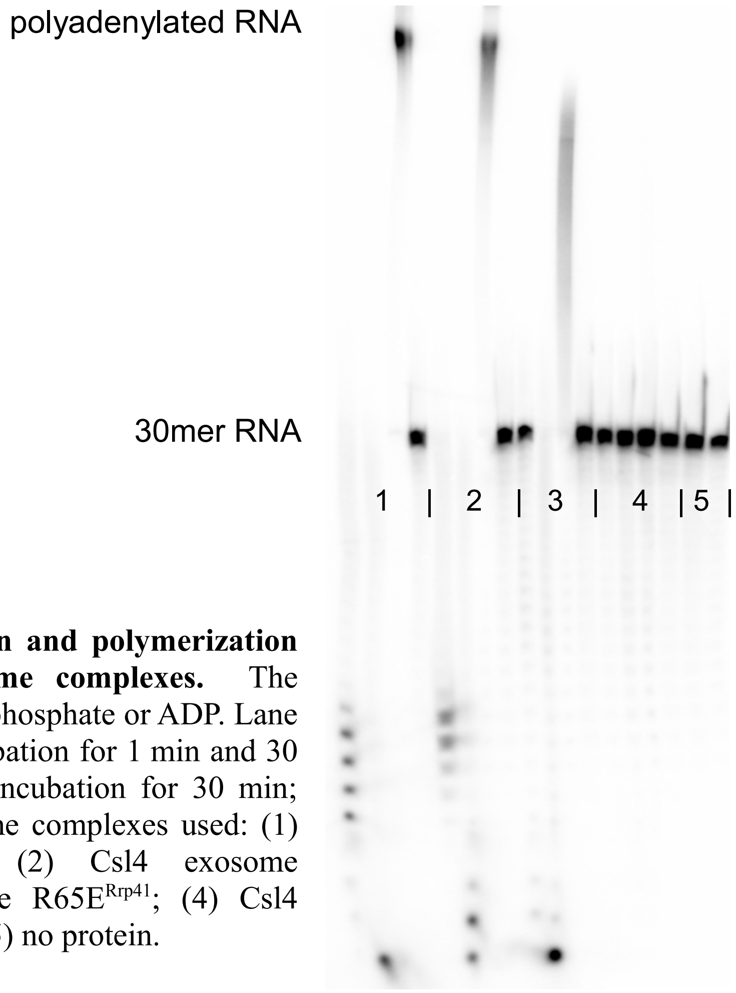


Figure 40: RNA degradation and polymerization activity of archaeal exosome complexes. The proteins were incubated with phosphate or ADP. Lane 1+2: 10mM phosphate – incubation for 1 min and 30 min; lane 3: 10mM ADP – incubation for 30 min; lane 4: none of both. Exosome complexes used: (1) Csl4 exosome wild-type; (2) Csl4 exosome Y70A^{Rrp42}; (3) Csl4 exosome R65E^{Rrp41}; (4) Csl4 exosome D180G+R65E^{Rrp41}; (5) no protein.

To test for substrate specificity of the archaeal exosome, the same experiment was repeated with 10 mM of ADP, CDP, GDP and UDP. The experiment was performed with all three exosome wild-type complexes and the result is shown in figure 41. Clearly the exosome uses all different nucleotide-di-phosphates, but with different preferences. Additionally, a difference in substrate preference can be seen for the different exosome complexes.

It could be shown that the archaeal exosome is able to degrade RNA from the 3'-end in the presence of phosphate and add nucleotides to the 3'-end in the presence of NDPs. In the living cell normally a mixture of phosphate and NDPs is available for the exosome. Therefore the activity of the archaeal exosome was tested in the presence of both phosphate and ADP (see figure 41B). The polyadenylation reaction seems to be energetically favored in comparison to the degradation reactions: In the reaction mix containing the same concentration of ADP and phosphate, the reaction product is an RNA molecule, which is larger than the initial 30mer.

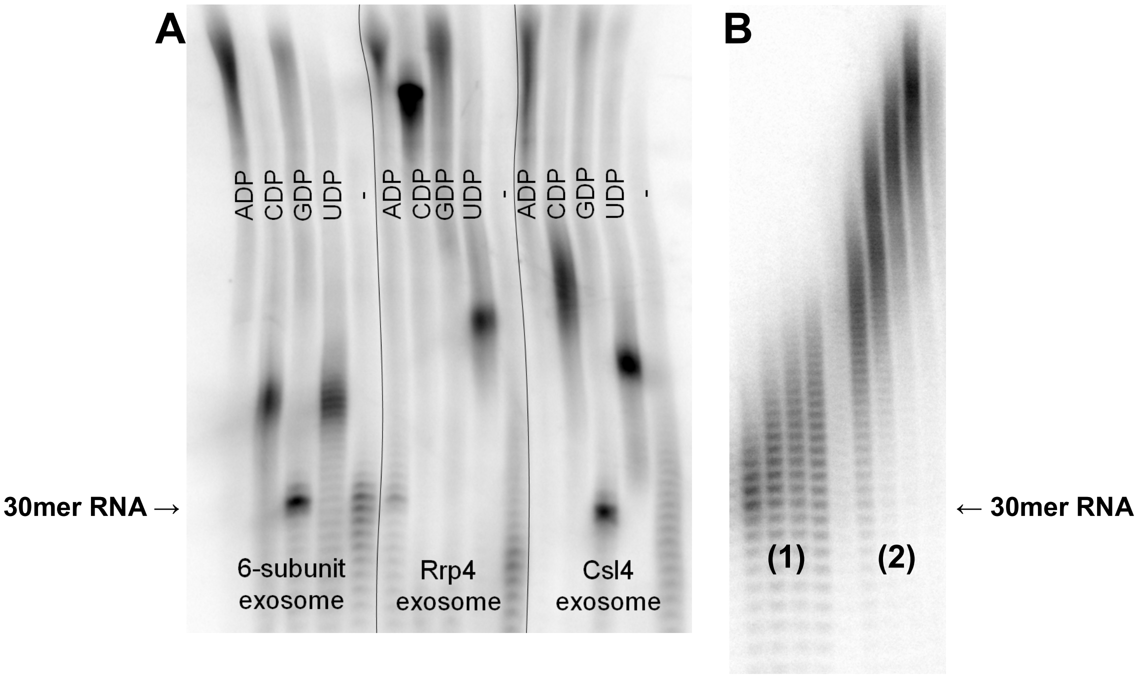


Figure 41: Polymerization activity of the archaeal exosome. A: All three exosome wild-type complexes can use the four different nucleotide-di-phosphates as substrate to add bases to the 3'-end of a polyA-RNA. 250 nM exosomes were incubated with 10 mM NDPs for 1 h. **B:** Polymerization and degradation activity of the exosome are competing with each other. (1) The Csl4 exosome was incubated with 20 mM phosphate and 10 mM ADP. (2) The Csl4 exosome was incubated with 10 mM phosphate and 10 mM ADP. For both reactions four time points are shown: 45 s, 1.5 min, 2.25 min and 3 min. The polymerization reaction seems to be energetically favored.

7.4.3 Quantification of RNase Assays

To gain more insights into the mode of RNA degradation of the archaeal exosome, RNase assays were performed as described above. For a better resolution up to 35 time points were taken with time intervals that allow monitoring of the reaction in more detail (see figure 40). In the next chapter two possible models are introduced to describe the degradation reaction and are then compared to the experimental data.

7.4.3.1 Theoretical Models for RNA Degradation

At first glance, an obvious model for the RNase activity of the archaeal exosome would be a simple degradation reaction that is only described by the rate constant for cleavage k_c. The reaction could then be described by the following set of equations, considering the initial substrate is an RNA with 30 bases:

$$\frac{d[RNA_{30}]}{dt} = -k_{c(30)} \cdot [RNA_{30}]$$
$$\frac{d[RNA_{29}]}{dt} = k_{c(30)} \cdot [RNA_{30}] - k_{c(29)} \cdot [RNA_{29}]$$
$$\vdots$$
$$\frac{d[RNA_4]}{dt} = k_{c(5)} \cdot [RNA_5] - k_{c(4)} \cdot [RNA_4]$$
$$\frac{d[RNA_3]}{dt} = k_{c(4)} \cdot [RNA_4]$$

with [RNA$_n$]: concentration of the n-mer RNA and k$_{c(n)}$: cleavage rate for degradation of the n-mer RNA.

Using these equations, the degradation process can be simulated with a different cleavage rate for each RNA length. This model will be termed cleavage-only model and an example for a simulation using this model is shown in figure 42. The simulation was calculated using MATLAB and the scripts are shown in the appendix 9.4.

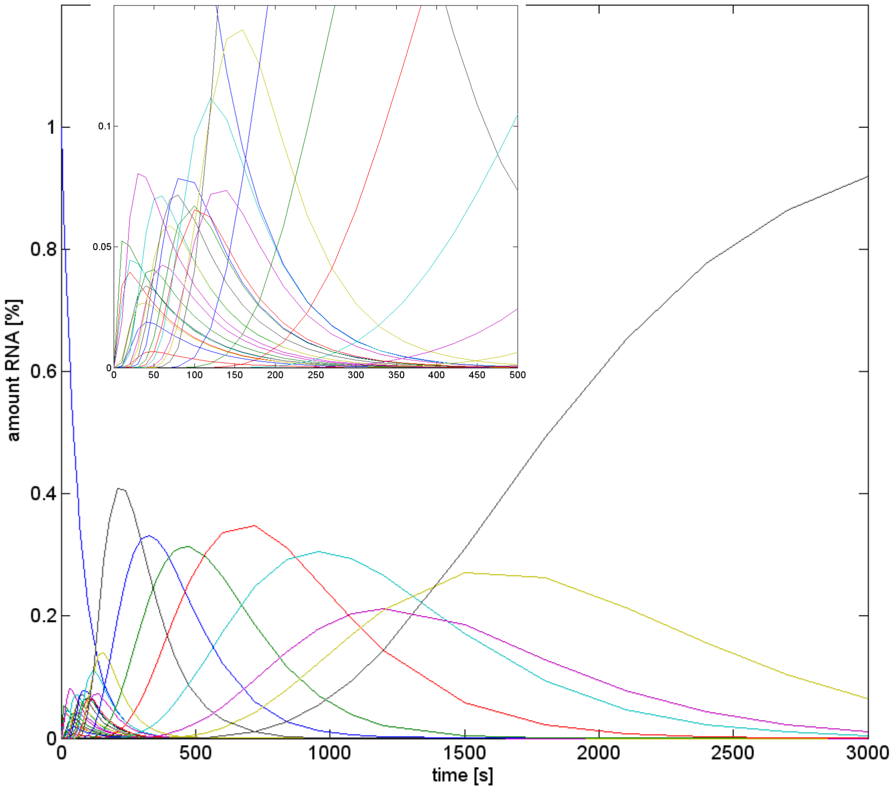


Figure 42: Simulation for RNA degradation described by the cleavage-only model. Different values for the cleavage rate were assumed for all 27 steps of degradation. The inset shows an enlargement of the curves for small times. (Color code: 30/23/16/9mer; 29/22/15/8mer; 28/21/14/7mer; 27/20/13/6mer; 26/19/12/5mer; 25/18/11/4mer; 24/17/10/3mer)

Since the cleavage-only model is extremely simple, a second and more comprehensive model was proposed: This model does not only account for cleavage reactions, but also for association and dissociation of the RNA to and from the exosome complex. In addition, the reverse reaction of the degradation, the polyadenylation, can be included into the model: this reaction needs ADP, which is formed during the degradation reaction. Therefore polyadenylation may become important towards the end of the experiment. The change of RNA concentration with time for the initial RNA molecule, the 30mer, can be described with the second model, termed cleavage-and-binding model, by the following equations:

$$\frac{d[RNA_{30,free}]}{dt} = -k_{a(30)} \cdot [RNA_{30,free}] + k_{d(30)} \cdot [RNA_{30,bound}]$$

$$\frac{d[RNA_{30,bound}]}{dt} = k_{a(30)} \cdot [RNA_{30,free}] - k_{d(30)} \cdot [RNA_{30,bound}] - k_{c(30)} \cdot [RNA_{30,bound}] + k_{p(29)} \cdot [RNA_{29,bound}]$$

with $[RNA_{n,free}]$ and $[RNA_{n,bound}]$: concentration of free and bound n-mer RNA.

The four rate constants used for these equations are (where n is again the length of the described RNA):

$k_{c(n)}$ = cleavage rate – describing the degradation of an RNA_n molecule to the product RNA_{n-1}

$k_{a(n)}$ = association rate – describing the association of an RNA_n molecule to the exosome

$k_{d(n)}$ = dissociation rate – describing the dissociation of the RNA_n molecule from the exosome

$k_{p(n)}$ = polymerization rate – describing the reaction of addition of a base to the substrate RNA_n

It has to be considered that the denaturing gels do not allow for discrimination of free and bound RNA. The intensity of the observed band corresponds to the sum $[RNA_{n,free}] + [RNA_{n,bound}]$ of free and bound RNA.

For 29 mer to 4 mer RNA, the amount of bound RNA does not only decrease due to degradation, but it additionally increases due to being the degraded product of RNA+1. Furthermore, the RNA can decrease due to the polymerizing reaction. Therefore the equations have to be extended as follows:

$$\frac{d[RNA_{29,free}]}{dt} = -k_{a(29)} \cdot [RNA_{29,free}] + k_{d(29)} \cdot [RNA_{29,bound}]$$

$$\frac{d[RNA_{29,bound}]}{dt} = k_{a(29)} \cdot [RNA_{29,free}] - k_{d(29)} \cdot [RNA_{29,bound}] - k_{c(29)} \cdot [RNA_{29,bound}] + k_{p(28)} \cdot [RNA_{28,bound}] + k_{c(30)} \cdot [RNA_{30,bound}] - k_{p(29)} \cdot [RNA_{29,bound}]$$

Since the 3mer is not degraded further, the corresponding reactions are described by:

$$\frac{d[RNA_{3,free}]}{dt} = -k_{a(3)} \cdot [RNA_{3,free}] + k_{d(3)} \cdot [RNA_{3,bound}]$$

$$\frac{d[RNA_{3,bound}]}{dt} = k_{a(3)} \cdot [RNA_{3,free}] - k_{d(3)} \cdot [RNA_{3,bound}] + k_{c(30)} \cdot [RNA_{30,bound}] - k_{p(3)} \cdot [RNA_{3,bound}]$$

The cleavage-and-binding model contains four different types of rate constants, namely cleavage, association, dissociation and polymerization rate constants. In principle, each of these rate constants can depend on RNA length, resulting in 4 x 28 fit parameters. However, in order to ensure a fair comparison between the cleavage-only model and the cleavage-and-binding model, the latter should not contain more fit parameters than the former. Moreover, from a biochemical point of view it is very likely that some of these rate constants are independent of RNA length. An example for a simulation of the degradation reaction using the cleavage-and-binding model with only 23 model parameters is shown in figure 43 (for the MATLAB scripts see appendix 9.4).

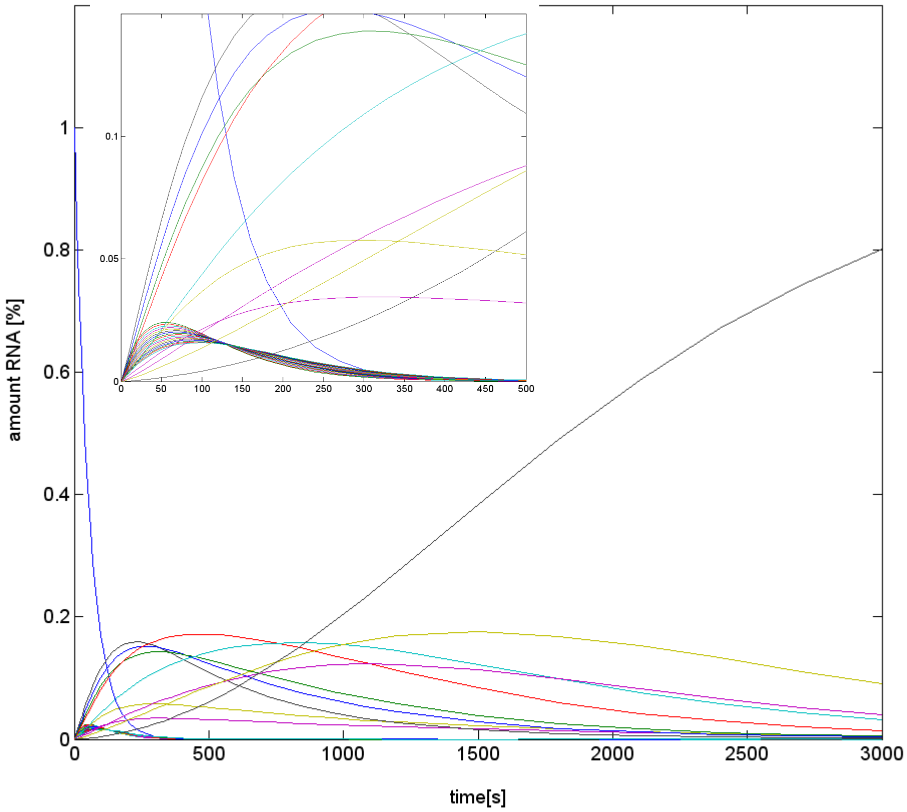
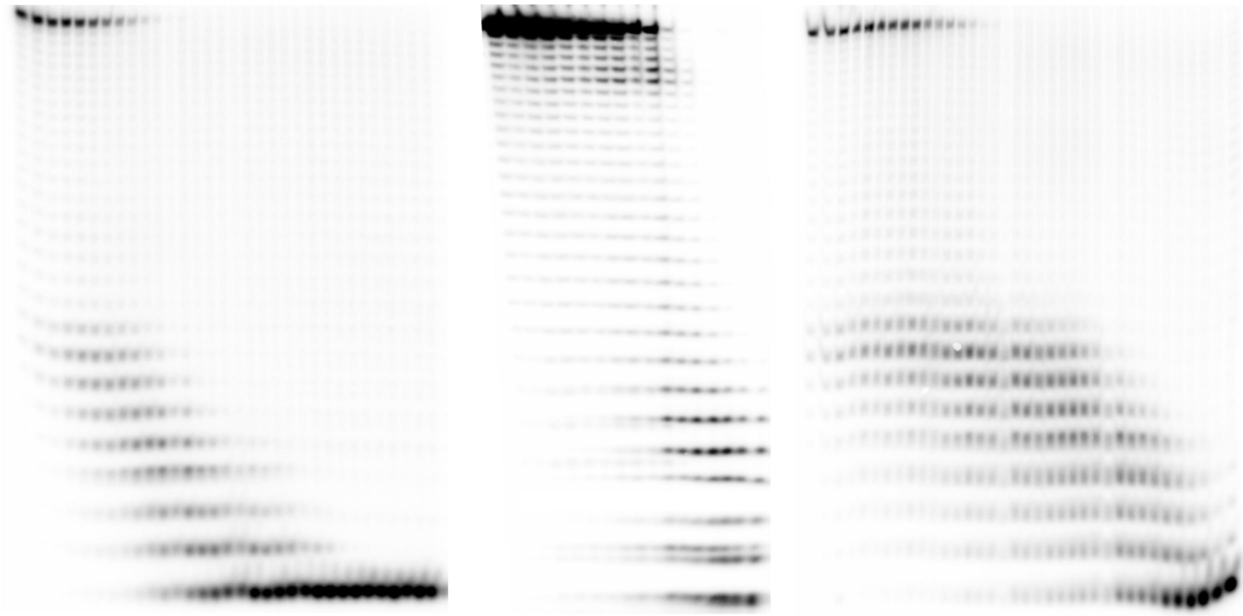


Figure 43: Simulation for RNA degradation described by the cleavage-and-binding model. The cleavage rate was kept constant and different binding and dissociation rates were assumed for the last 10 steps of degradation. The polymerization rate was assumed to be zero. The inset shows an enlargement of the curves for small time points. (Color code: 30/23/16/9mer; 29/22/15/8mer; 28/21/14/7mer; 27/20/13/6mer; 26/19/12/5mer; 25/18/11/4mer; 24/17/10/3mer)

By fitting both models to the experimental data, it is possible to check whether one of the models describes the RNA degradation by the exosome (at least to first order).

7.4.3.2 Experimental Data

A Csl4 exosome **B** Rrp4 exosome **C** 6-subunit exosome



D Csl4 exosome R65E **E** Csl4 exosome Y70A **F** 2-subunit exosome

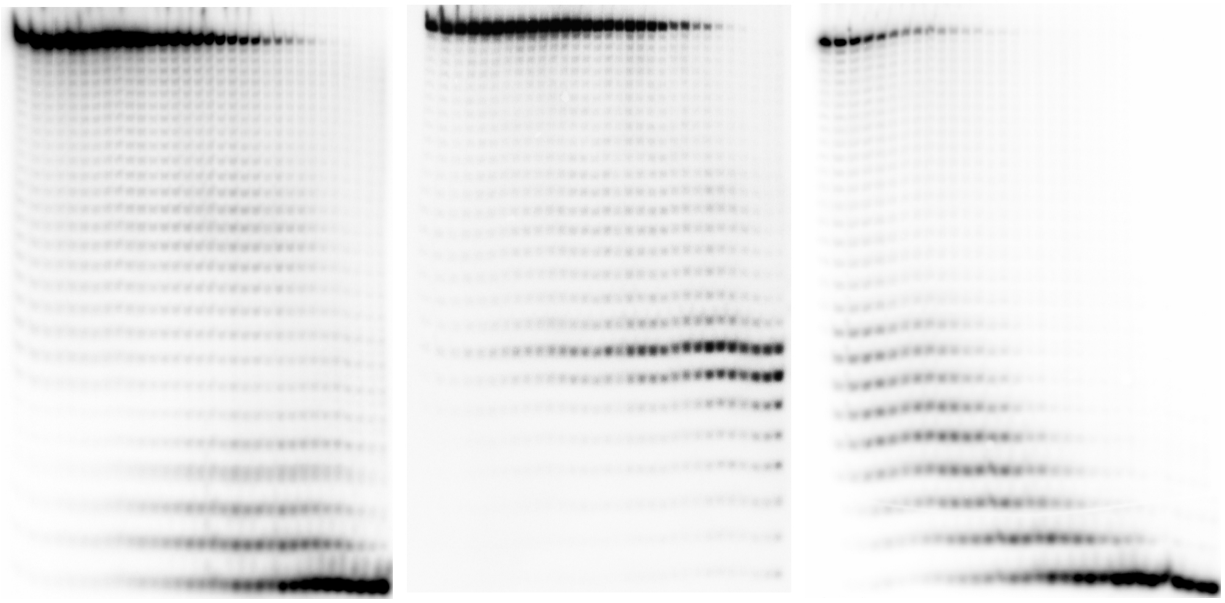


Figure 44: RNase assays of the different exosome variants with a 30mer poly(A) RNA as substrate. The exosome variants used in this assays are shown in figure 34. Radioactively labeled degradation products were visualized in a denaturing polyacrylamide gel.

RNA degradation of a 30mer poly(A) substrate by the different exosome variants described in figure 38 was monitored by denatured PA gel electrophoresis as described before. 30nM (Csl4-exosome, Rrp4-exosome and interface mutant) or 60nM (6-subunit exosome, R65E^{Rrp41} and Y70A^{Rrp42} Csl4-exosome) of the complexes were mixed with 120nM radioactively labeled RNA. The first eight samples were taken every 10 seconds. Then the time interval was subsequently increased. The last sample was taken after 30 min. The resulting gels are shown in figure 44.

Every single band on the gel was quantified using the ImageQuant software. The background was subtracted. From the intensities of the bands the percentage of the RNA of a certain length relative to the amount of total RNA at a certain time was calculated. Figure 45 shows an example for the concentration distribution of all RNA species over time for the Csl4-exosome, determined from the gel shown in figure 44A.

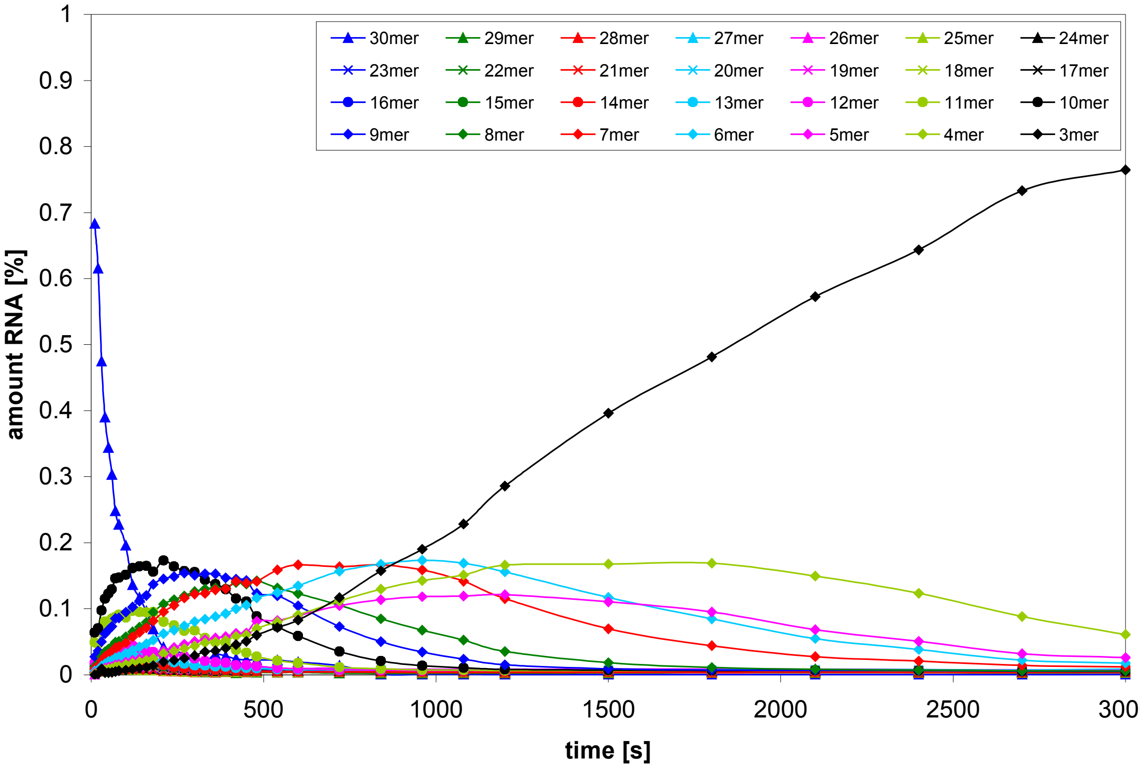


Figure 45: Degradation of a 30mer polyA-RNA by the Csl4-exosome. The amount of the different RNAs that emerge during the reaction are shown as percentage of total RNA and are plotted against time. RNAs of a length from 29mer to 15mer can hardly be detected. The final product of the reaction is a 3mer.

The other five gels were analyzed in the same way. The plots are shown in the appendix 9.5.

7.4.3.3 Data Analysis and Rate Constants

In order to analyze the experimental data and to fit the models described in 7.3.3.2, several MATLAB scripts were written (see appendix 9.4). The MATLAB function *ode15s* was used for the simulation of the differential equations. A least squares fit was performed by minimizing the quadratic differences between simulated curves and experimental data with the MATLAB function fminsearch. *fminsearch* finds the minimum of a scalar function of several parameters, starting at an initial estimate. The correct initial estimated had to be found iteratively.

First, the cleavage-only-model was considered. Even when all 27 cleavage constants were presumed to be different, a good fit could not be achieved (see figure 46).

To proof that the second model, the cleavage-and-binding model, describes the experimental data better, 24 different constants were used to fit the model to the data. The following constraints were imposed, to reduce the theoretically possible set constants:

For simplicity the polymerization rate was assumed to be $k_p = 0$, which is reasonable for high phosphate concentrations (10 mM phosphate in this assay compared to 3.6 µM ADP, when all RNA molecules would be totally degraded).

In addition the cleavage rate was assumed to be independent of the length of RNA and thus does not change during the reaction. Finally association and dissociation rates were assumed to stay constant for the first 15 reactions, meaning that they are identical for the 30mer – 15mer. The resulting fit for the Csl4-exosome is shown in figure 46.

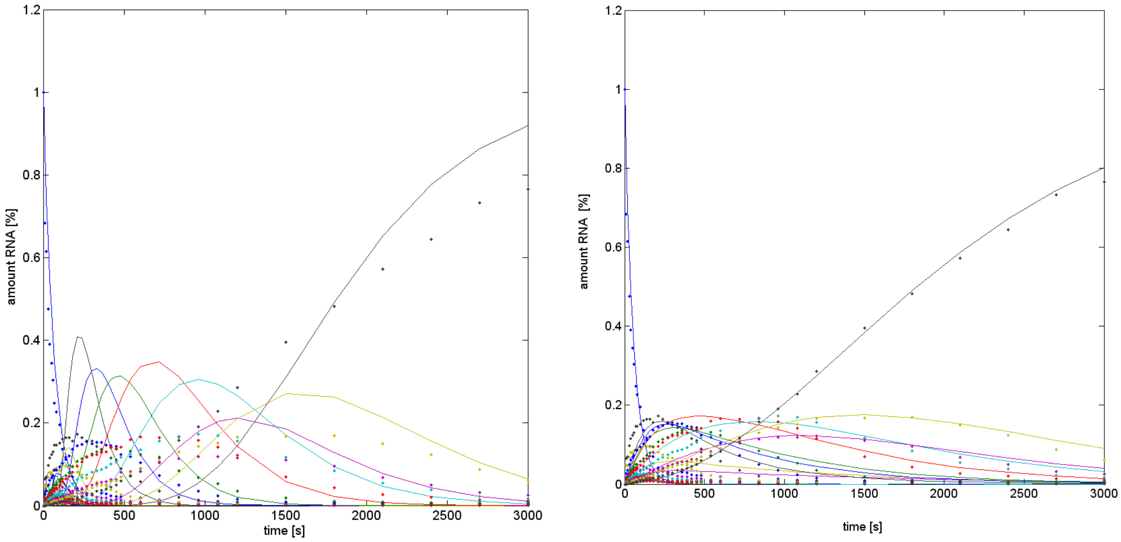


Figure 46: Fit of the data from RNA degradation by the Csl4-exosome using the cleavage-only model (left) and the cleavage-and-binding model (right). Clearly the cleavage-and-binding model fits the experimental data better than the cleavage-only model. The color code for the different RNAs is the same as shown in figure 41.

In view of the extremely poor result of the cleavage-only-model, we will restrict our analysis in the following to the cleavage-and-binding model. The corresponding fits for the other exosome variants are shown in the appendix 9.5 and the determined rate constants are given in table 5. To estimate the quality of the fits, the coefficient of determination R^2 was calculated. R^2 describes the proportion of variability in a data set that is accounted for by the model. Accordingly, for $R^2 = 1$ the model perfectly fits the data, whereas an R^2 of zero indicates that the model totally fails to explain the data. The definition of the coefficient of determination is

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}}$$

where SS_{tot} is the total sum of squares and SS_{err} is the residual sum of squares and

$$SS_{tot} = \sum_{i=1}^{n} [[RNA_{30}]^{\exp}(t_i) - \langle [RNA_{30}]^{\exp} \rangle]^2 + \sum_{i=1}^{n} [[RNA_{29}]^{\exp}(t_i) - \langle [RNA_{29}]^{\exp} \rangle]^2 + \cdots$$

$$\cdots + \sum_{i=1}^{n} [[RNA_3]^{\exp}(t_i) - \langle [RNA_3]^{\exp} \rangle]^2$$

$$SS_{err} = \sum_{i=1}^{n} [[RNA_{30}]^{\exp}(t_i) - [RNA_{30}]^{fit}(t_i)]^2 + \sum_{i=1}^{n} [[RNA_{29}]^{\exp}(t_i) - [RNA_{29}]^{fit}(t_i)]^2 + \cdots$$

$$\cdots + \sum_{i=1}^{n} [[RNA_3]^{\exp}(t_i) - [RNA_3]^{fit}(t_i)]^2$$

Here $[RNA_N]^{\exp}(t_i)$ and $[RNA_N]^{fit}(t_i)$ are the measured and fitted concentrations of the N-mer at time t_i and $\langle [RNA_N]^{\exp} \rangle = \sum_{i=1}^{n} [RNA_N]^{\exp}(t_i)/n$ is the time average of the concentration of the N-mer. The calculated R^2-values for every fit are given in table 5.

In all cases the dissociation rate increased with decreasing RNA size. An example for the dissociation rates is shown in figure 47 for the Csl4 exosome complex wild-type and $Y70A^{Rrp42}$ mutant. The dissociation rate is very low for long RNAs and increases with decreasing RNA size. As expected the RNA binding mutant $Y70A^{Rrp42}$ has overall higher dissociation rates than the wild-type.

Table 5: Rate constants derived from fitting the cleavage-and-binding model to the experimental data.

complex	k_c		k_a and k_d for different RNA lengths											R^2
			30-15	*14+13*	*12*	*11*	*10*	*9*	*8*	*7*	*6*	*5*	*4*	
6-subunit exosome	10	**bind**	0.022	0.022	0.0005	0.0001	0.053	0.054	0.05	0.05	0.05	0.031	0.31	0.97
		diss	0.27	0.27	0	0.14	6.6	8.3	10.0	14.7	12.6	13.6	30	
Csl4 exosome	12.4	**bind**	0.019	0.019	0.0007	0.001	0.0032	0.003	0.0039	0.005	0.004	0.005	0.01	0.94
		diss	0.85	0.85	0.53	1	4.77	5.84	7.97	15.3	17.35	16.69	57.35	
Rrp4 exosome	0.30	**bind**	0.007	0.007	0.0001	0.0001	0.0001	0.0007	0.0007	0.0007	0.0001	0.0001	0.0001	0.96
		diss	0.022	0.022	0.002	0.0035	0.007	0.028	0.066	0.109	0.052	0.059	0.234	
2-subunit exosome	2.3	**bind**	0.03	0.03	0.002	0.001	0.145	0.144	0.127	0.10	0.1	0.101	0.093	0.96
		diss	1.49	1.49	2.15	4.2	5.9	4.9	7.1	10.5	14.1	14.7	31	
Csl4 exo R65E^{Rrp41}	15.6	**bind**	0.003	0.003	0.0001	0.0001	0.083	0.92	0.068	0.075	0.185	0.053	0.049	0.95
		diss	0.96	0.96	0.81	0.88	0.39	1	6.01	17.67	87.83	43.78	147.89	
Csl4 exo Y70A^{Rrp42}	15.9	**bind**	0.013	0.0005	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.002	0.91
		diss	1.46	1.2	3.88	20.34	26.31	30.41	30	37.12	34.98	41.2	78.7	

The rates are all given in 1/s. R^2 was calculated as described. k_c = cleavage rate; k_a = association rate and k_d = dissociation rate.

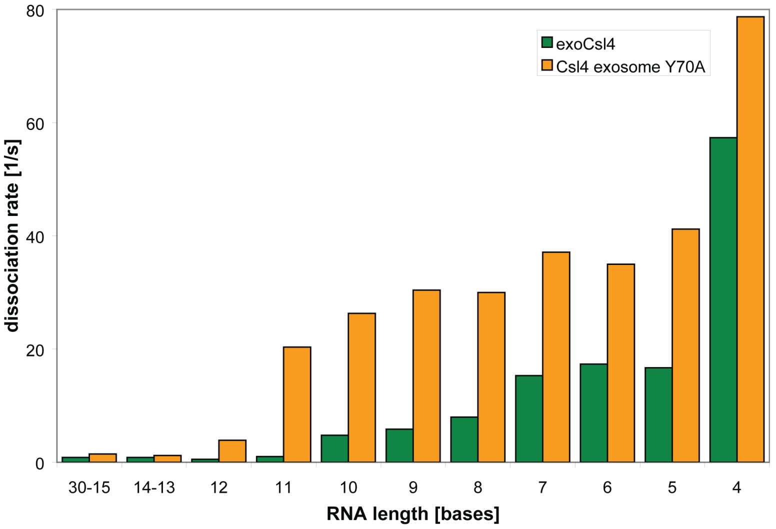


Figure 47: Dissociation rate determined from fitting the cleavage-and-binding model to the experimental data from denaturing PA gels. The change of rates with RNA length is shown for the wild-type Csl4 exosome (green) and the Y70A^{Rrp42} mutant (orange).

7.5 RNA Binding to Exosome Complexes

To learn more about the role of the cap proteins for RNA binding, recognition and subsequent degradation or polymerization, electrophoretic mobility shift assays (EMSA) were performed. Exosome proteins were incubated with the radioactively labeled RNA molecule for 10 min and the protein-RNA complex was separated from the free RNA in a native PA gel. Not only the 9-subunit complexes were tested, but also the cap proteins alone. 30mer RNAs with three different sequences were used: a polyA RNA, a polyU RNA and polyCCU RNA. The resulting gels are shown in figure 48. The different RNAs show different affinities to the 9-subunit complexes. Mutation of the arginine 65, which reaches into the neck of the exosome strongly decreases affinity, but does not abolish it. The cap proteins alone also bind RNA, but with weaker affinity.

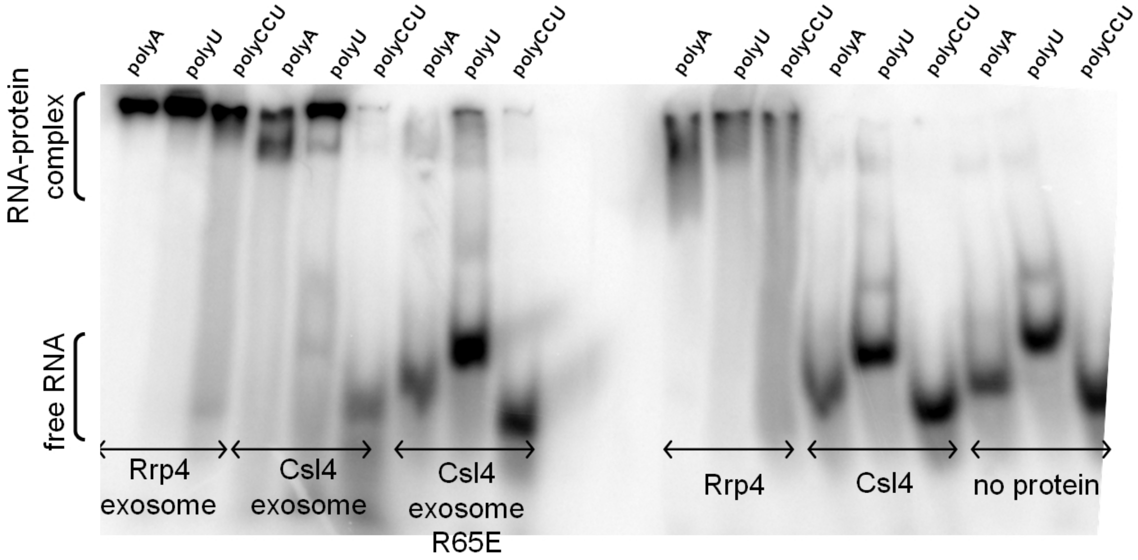


Figure 48: RNAs with different sequences have different affinities to exosome proteins. 30mer RNA molecules with three different sequences were incubated with the two 9-subunit exosome complexes and with the cap proteins alone. Both the arginine 65 in the neck of the complex and the cap proteins seem to influence substrate binding.

8 Discussion – Exosome

With the combination of biochemical assays, structural biology and theoretical modeling new insights into the enzymatic activity of the exosome and especially its mode of degradation could be obtained. With SAXS the known crystal structures could be verified, with crystallography the binding of RNA to the active site could be resolved to atomic resolution and the results form biochemical assays allow for the proposal of a model for degradation. All obtained results are discussed in the next section in detail.

8.1 Low Resolution Structures of Exosome Complexes in Solution

The crystal structures of the two different archaeal 9-subunit exosome complexes (*Archaeoglobus*) were solved by Büttner et al. (2005) and of a 6-subunit archaeal exosome (*Sulfolobus*) by Lorentzen et al. (2005). In addition, the crystal structure of the 9-subunit complex of the human exosome was solved in 2006 by Liu et al. All structures show a very similar overall fold, but they were obtained by X-ray crystallography and it cannot be excluded that the structures differ from the shape of the complexes in solution. Ramos et al. (2006) did SAXS studies with the *Pyrococcus* archaeal exosome to gain insights into the conformation of the complex in solution. When measuring the scattering of the Rrp4 protein alone they predicted an elongated conformation of the protein in solution. *Ab initio* modeling of the whole 9-subunit complex resulted in a rigid central structure with three extended, protruding and flexible arms (see figure 49). Therefore they proposed that in solution a high flexibility is present in the cap proteins of all exosome complexes.

All known crystal structures of exosome complexes with cap proteins show very well ordered cap proteins and nothing hints to high flexibility in these parts. To test the *Archaeoglobus* exosome complexes for their conformation in solution and especially for potential flexibility in the cap proteins, SAXS studies with both 9-subunit exosomes and with the 6-subunit exosome were performed. The results for the 9-subunit complexes show very high similarity to the crystal structures and it does not seem that bigger parts of the proteins possess intrinsic flexibility (figure 49).

As the sequence homology between the archaeal exosomes is very high, it is quite likely that under our conditions also the *Pyrococcus* exosome would have a similar structure in solution as the *Archaeoglobus* homolog.

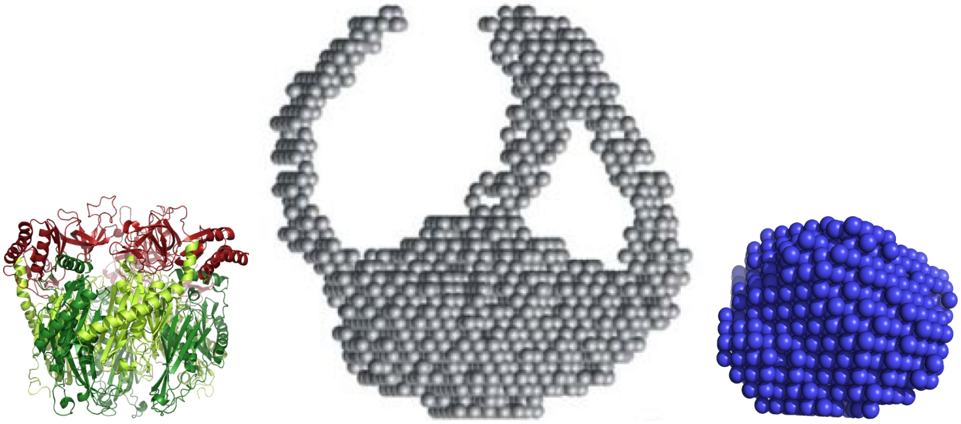


Figure 49: Comparison of the SAXS structure of the *pyrococcus* Rrp4 exosome (grey) from Ramos et al, (2006) with the crystal structure and SAXS structure of the *archaeoglobus* Rrp4 exosome (blue). The structures are scaled to compare the size. Clearly the *archaeoglobus* Rrp4 exosome has the same overall fold in solution compared to the crystal structure and the two SAXS structures are very different.

The 6-subunit exosome has a slightly different conformation in solution compared to the crystal structure. We do not have the crystal structure of the *Archaeoglobus* 6-subunit exosome. But comparison of the two core rings obtained from the different 9-subunit exosomes shows one difference: the end of the C-terminal helix of the Rrp41 protein is in one complex stretched and in the other one bent. Exactly at the position where we could identify this structural difference, we can see additional density in the SAXS structure. The SAXS model leads to the conclusion that the C-terminus of the Rrp41 protein is unstructured as long as no cap protein is bound and folds only upon binding of either Rrp4 or Csl4.

Summarized we propose that under physiological conditions and in the living cell the cap proteins bind to the exosome in a way that is not significantly different to the crystal structures.

8.2 The Atomic Structure of RNA Bound to the Exosome Active Site

The archaeal exosome was crystallized with a short RNA molecule bound to all three active sites. As no phosphate was added to the crystallization solutions, a stable RNA complex could be formed even with the wild-type protein. Different lengths of RNA (6 – 12 bases) were used for crystallization setups and soaking experiments. Addition of a 6mer or an 8mer to the exosome resulted in both cases in clear electron density for four bases in all three active sites and for some weak density for two more bases. This could be seen for cocrystallization as well as for soaking experiments.

When longer RNA was used (10 bases and more) no crystals could be obtained and soaking of these RNAs into crystals of the native complex resulted in dissolving of the crystals. This indicates, that binding of the longer RNA molecules to the protein complex somehow destroyed crystal packing. As the cap proteins are involved in crystals contacts, it is possible that these larger RNA molecules reached out of the processing chamber and bound to the cap proteins whereupon they interfered with crystal packing.

Obviously the first four bases of the RNA substrate bind very tightly to the active site of the exosome. The RNA is located in a horizontal binding cleft with the active site at one end. The four bases that are positioned in this cleft are clearly visible. For the next bases only parts of the electron density are visible, therefore the following five to six bases seem to be flexible. From the 10th base on the RNA seems to have contacts to the upper parts of the exosome and supposedly to the RNA binding elements of the cap proteins.

A set of arginine residues are located in the binding groove and charged interactions ensure tight binding of the first four bases. A closer look at the active site with bound RNA shows that residues from both the Rrp41 and the Rrp42 protein are involved in binding of the RNA molecule. This explains why the Rrp41 protein alone is not capable of degrading RNA and why the Rrp42 protein is needed for catalytic activity. Moreover the structure explains the sequence unspecificity of the exosome, as all interactions between the RNA and the protein are mainly via the phosphates from the RNA backbone and not via characteristic features of single bases. Interactions with the 2'OH-group of the ribose, which is missing in DNA, can explain how the exosome distinguishes between DNA and RNA and why only RNA is a substrate.

The only obvious interaction between the complex and an RNA base is a sequence unspecific π-stacking of tyrosine 70 from the Rrp42 protein with the fourth base. Mutation of this tyrosine residue and structural analysis revealed a lower intensity of electron density for the first four bases of the RNA and no density at all for the 5th and the 6th base. This supports the assumption that this residue is important for RNA binding to the complex and correct orientation of the substrate towards the active site. Because of these structural observations concerning the tyrosine, biochemical activity assays with the mutant Y70A^{Rrp42} were carried out and are described below.

Still the question remains, if the RNA reaches the active site through the top of the proteins via involvement of the cap proteins, or through the bottom of the complex, which is much closer to the active site. RNA-exosome complexes with longer RNA molecules were crystallized by Lorentzen et al. (2007) and Navarro et al. (2008). Lorentzen used the *Sulfolobus* Rrp4 exosome and an RNA molecule with secondary structured parts at the 5'-end, which is designed to be only partially trimmed rather than totally degraded. Navarro used the *Pyrococcus* 6-subunit exosome and a 10mer poly(A) RNA. Both groups could see the four bases in the active site, very similar to the structure obtained in this thesis. In addition they found electron density for one RNA base at the narrowest constriction of the central channel, in the neck of the processing chamber. At this position a conserved arginine residue points towards the pore and binds the RNA base via a charged interaction. These results support the theory that the path of the RNA is via the cap proteins and the neck to the processing chamber and the active sites.

8.3 Atomic Structure of the Isolated Csl4 S1- and Zn-ribbon-domains

Two different cap proteins of the archaeal exosome complex are known: the Rrp4 protein that possesses high homology to the eukaryotic Rrp4 and Rrp40 proteins and the Csl4 proteins, which has only one homolog of the same name in eukaryotes. The cap proteins are very similar in the composition of one small N-terminal domain and two RNA binding domains. Both possess a S1-domain, Rrp4 an additional KH-domain and Csl4 an additional Zn-ribbon domain. Still there is one striking difference when looking at the positions of the genes for these proteins in the genome: the *rrp4* gene is located directly next to the genes for the core ring of the exosome *rrp41* and *rrp42*. The *csl4* gene in the contrary is located in a different operon. This means that the mRNA level and thus also the expression level of the Rrp41, Rrp42 and Rrp4 proteins is expectedly relatively similar in the cell. The amount of Csl4 protein however is regulated by a different promoter and can therefore differ from the amount of exosome core. This can be an indicator for a role of the Csl4 protein independent from the exosome – especially when more Csl4 than core is available in the cell.

To analyze if Csl4 is actually able to fold properly when it is not bound to the exosome, the structure of the S1 and Zn-ribbon domain was solved. The atomic structure of the isolated S1 and Zn-ribbon domain of the Csl4 protein shows the same fold as the exosome bound version. Not only is the fold of the two domains identical, but also their orientation relative to each other is not changed.

This finding could suggest a possible role for Csl4: under certain conditions the *csl4* gene may be stronger transcribed than the rest of the exosome and Csl4 exists in the cell in an isolated form. Then it may bind RNAs that are substrates for the exosome. Possibly it can thereby recruit the RNA to the exosome complex. The Csl4-RNA complex might have higher affinity to the exosome core than Csl4 alone. Thus the two Csl4 proteins could be exchanged and the RNA is thereby positioned in a way that it can reach the active site and is further processed or degraded.

8.4 The Exosome with a Substrate from E. coli Cells

It is very difficult to achieve homogeneous complexes of RNA bound to the cap proteins of the exosome. This is due to the many different RNA binding sites within the caps as well as to the fact that the exact sequences of substrates for the exosome are not known. So it is very difficult to synthesize a physiological substrate for usage in structural studies.

Therefore a totally different approach was used to gain information about the role of the cap proteins in substrate recognition and binding. For this experiment the Rrp4 exosome was chosen, as it showed a much higher affinity to RNA compared to the Csl4 exosome in the assays. The complex was overexpressed in *E. coli* cells and purified together with the RNA substrate, which bound to it within the bacterial cell. After thorough purification of the complex it was used for SAXS studies as well as crystallization setups.

The structure of this complex in solution shows a clear difference to the Rrp4 exosome without RNA. Without RNA a clear hole could be observed in the middle of the exosome ring, where the RNA is supposed to enter the processing chamber. This hole is not visible anymore in the presence of RNA. Furthermore additional density could be observed on top of the exosome and a tail is sticking out. Looking at the difference density that is only present in the RNA bound structure, is looks like the RNA molecule is protruding from the processing chamber through the S1-pore, clings to the Rrp4 protein until the rim of the exosome and has a flexible end.

To resolve this structure to an atomic level, the same complex was crystallized. Using the whole Rrp4 exosome as search model, no phases could be obtained with molecular replacement. Only with the 6-subunit exosome and the isolated S1-domain a reasonable electron density map could be calculated. Density for the two missing domains was visible in the difference map and both domains could be modeled. Still this structure is not yet refined to its final state. It is already obvious that the Rrp4 protein changed its conformation to a considerable degree. From the current density map it cannot be said whether RNA is visible in the structure. It can be speculated that the conformational change in the Rrp4 proteins is induced by a bound RNA molecule. The RNA may not be bound homogeneously or may be too flexible to be visible in a crystal structure. Moreover it is still not known whether more than one RNA species are present and the identification of this bound molecule is inevitable. More refinement and model building procedures of the structure could help to improve the structure and the it could still be possible that electron density for bound RNA appears during these steps.

8.5 Enzymatic Activities of the Exosome

The bacterial PNPase (Polynucleotide Phosphorylase) and the archaeal exosome are homologous concerning their structure and their catalytic activity. Both complexes have a similar overall fold and are able to degrade single stranded RNA from the 3'-end, but cannot cope with structured substrates without the help of cofactors (Büttner et al., 2005; Lorentzen et al., 2005). In addition, both complexes are able to catalyze the reverse reaction. For the archaeal exosome it could be shown that it polyadenylates RNAs at the 3'-end *in vitro* (Büttner et al., 2005). Some archaeal organisms do not possess an exosome and in these organisms no polyadenylation could be observed (Portnoy and Schuster, 2006), therefore the exosome is believed to be the main polyadenylating enzyme in archaea.

Still the mode of degradation remained unclear. Recent structural studies of RNA-exosome complexes showed one RNA molecule bound to the active site and the neck (Lorentzen et al., 2007; Navarro et al, 2008). Due to these results it is now quite certain that the RNA binds to the cap proteins on top of the exosome and is then recruited into the processing chamber and to the active site.

To gain insights into the different roles of the cap proteins and the mode of RNA degradation and polymerization of the archaeal exosome, biochemical activity assays were performed.

For all exosome complexes used in this thesis the RNase assays revealed a 3mer as the final end product of degradation. This fits very nicely to the structural results from the Y70A^{Rrp42} mutant: this tyrosine strongly stabilizes the RNA in the active site by stacking to the fourth base. As soon as the 4mer is degraded to a 3mer, this interaction is lost and the affinity of the RNA to the exosome will be extremely low. The 3mer will dissociate from the active site and will most probably not associate again.

Additionally, it could be shown that the speed of degradation depends on the composition of the cap. For poly(A) RNA the Csl4 exosome showed the highest degradation activity and the lowest enzymatic activity was observed for the 6-subunit core exosome without any cap protein bound. This hints to a sequence specificity and regulative role of the cap proteins.

In literature only phosphate dependent RNase activity was described for the exosome so far. In this thesis it could be shown that the complex is able to hydrolytically degrade RNA with much lower activity as long as magnesium is present in the reaction buffer. Similarly to the phosphate dependent degradation, the Csl4 exosome is more active than the Rrp4 exosome and nearly no hydrolytic activity could be observed for the 6-subunit exosome. This also points to a regulatory role and possibly sequence specificity in RNA binding of the cap proteins.

For analysis of the polymerization activity of the exosome, the complex was not only incubated with ADP, but also with other nucleotide-di-phosphates. The preference for all tested nucleotides is different, but all of them were added to the 3'-end of the substrate RNA. Besides ADP, the exosome easily incorporates the other purine base, GDP. In contrast, incubation with the pyrimidine bases CDP and UDP results in much shorter polymerized tails. A difference can be observed between the 6-subunit/Csl4 exosome and the Rrp4 exosome: They all use the different nucleotides in a similar manner, only the Rrp4 exosome added the pyrimidine bases much more effectively to the template than the other two complexes. This could hint to a second role of the cap proteins: not only could they influence the substrate selectivity of the exosome, but they possibly somehow also influence the equilibrium between degradation and polymerization. This could only be explained by a possible structural and allosteric effect of the cap proteins on the whole complex.

As in all living cells never either ADP or phosphate exist, but always a mixture of both with changing ratios, the activity of the exosome was tested in the presence of a mix of ADP and phosphate. The results show that the exosome reaches an equilibrium between degradation and polymerization after some minutes. With equal concentrations of phosphate and ADP, the resulting RNA is longer than the substrate RNA and only with a 10-fold excess of phosphate the average length of RNA stays the same. That means that the polymerization reaction is energetically favored over the degradation reaction under the present conditions. This experiment was done with the Csl4 exosome and a different result may be achieved with the Rrp4 exosome.

8.6 The Cleavage-and-Binding Model for RNA Degradation

In this thesis denaturing sequencing gels were used to visualize the degradation of RNA by the archaeal exosome. In the assays the phosphate concentration was chosen to be very high, to ensure that mainly the phosphate dependent degradation is present. The experiment allowed for the time resolved monitoring of the reaction, the detection of every single RNA species that is formed during the reaction and thereby a closer analysis of the mode of degradation. Two models were proposed to characterize the reaction: The first model, the cleavage-only model, describes only the degradation of the RNA base by base. The second model, the cleavage-and-binding model describes in addition the association and dissociation of every single RNA that occurs during the reaction. Furthermore, the possibility that – especially towards the end of the reaction when more and more ADP is formed – the polyadenylation reaction coexists besides the degradation was included into the model.

The two models were fitted to the experimental data using the same number of free fitting parameters and clearly the cleavage-only model describes the data very poorly compared to the cleavage-and-binding model. Therefore the consideration of association and dissociation of the RNA to and from the exosome is essential when describing the degradation reaction.

During the first cycles of fitting it became clear that the following simplifications can be made:

The polymerization rate, which theoretically could influence the whole reaction especially towards the end, can be assumed as zero. This is reasonable as the phosphate concentration in this experiment was very high and even at the point when all RNA is degraded, it still was more than 100-times higher than the ADP concentration.

The active site was assumed to be independent from the length of RNA bound to the exosome and therefore the cleavage rate was set constant for the whole reaction. From the fitting results is became clear that the association and dissociation rates are both always constant for RNA molecules larger than 15 bases. Therefore for the final fitting procedure these rates were set constant for long RNAs.

Using these constraints the fitting could be done with around 24 different constants. The quality of a fit cannot be judged visually. Therefore a parameter to estimate the quality of the fits, the R^2-value was introduced ($R^2 = 1$ is a perfect fit). The R^2 values, were for all curves higher than 0.9 indicating good agreement between the fitted curves and experimental data.

In total 24 different constants were obtained for each complex (see table 5). The results are a first order approximation and the model describes a simplification of this very complex system. Still a lot of information can be extracted from the results.

To analyze and discuss this information and especially for better comparison, the constants were summarized and normalized as follows: All obtained constants from one experiment were divided by the cleavage rate constant, which results in a normalization of all data to an identical k_c. In addition, the association and dissociation rate constants for long RNAs (30 – 15 bases) were compared to an average of the association and dissociation rate constants for short RNAs (10 – 4 bases). That is to say, not the constants for every single RNA that occurs during the experiment will be discussed, but we compare only the constants for RNAs longer than 15 bases with the constants for RNAs shorter than 10 bases. The calculated rate constants are shown in table 6.

Table 6: Summarized rate constants obtained from fits of the cleavage-and-binding model to the experimental data.

complex	k_c	association		dissociation	
		k_a long RNAs	k_a short RNAs	k_d long RNAs	k_d short RNAs
6-subunit exosome	1	$2\cdot10^{-3}$	$5\cdot10^{-3}$	0.03	1.10
Csl4 exosome	1	$1.5\cdot10^{-3}$	$3\cdot10^{-4}$	0.07	0.90
Rrp4 exosome	1	$2.3\cdot10^{-2}$	$1\cdot10^{-3}$	0.07	0.175
2-subunit exosome	1	$1.5\cdot10^{-2}$	0.05	0.66	4.20
Csl4 exosome R65E^{Rrp41}	1	$2\cdot10^{-4}$	0.015	0.06	1.67
Csl4 exosome Y70A^{Rrp42}	1	$8\cdot10^{-4}$	$2\cdot10^{-4}$	0.09	2.10

Rate constants are normalized to identical cleavage rates. Rates are given in 1/s. k_c = cleavage rate; k_a = association rate and k_d = dissociation rate.

The following observations could be made:

For long RNAs the cleavage rate constant is more than one order of magnitude larger than the dissociation rate constant. This explains the processivity of the exosome: as soon as an RNA molecule is bound to the exosome it will be degraded so quickly that it has no time to dissociate again. For the small RNA molecules the dissociation rate constant is in the same order of magnitude as the cleavage rate constant. This explains why small RNA molecules are degraded so much slower compared to the large ones: constant dissociation and association of RNA slows down the degradation reaction.

Looking at the association rate constants they seem to be extremely small. This explains why the exosome needs some time until it starts degrading the RNA. An explanation could be the architecture of the exosome: the RNA has to not only bind to the cap proteins of the complex, but the 3'-end has to find its way through the S1 pore and the neck of the processing chamber to finally reach the active site. It is of course quite unlikely that the RNA finds this way in a normal diffusion driven way and it is not surprising that binding of the 3'-end to the active site happens only with a low time rate.

Additionally to the wild-type exosome complexes, two proposed RNA binding mutants were tested: the neck mutant R65E^{Rrp41} and the tyrosine mutant Y70A^{Rrp42}. As expected, the dissociation rate constants of these mutants are larger and naturally the association rate constants are smaller compared to the wild-type proteins.

To learn more about the relevance of the hexameric architecture of the exosome core, an interface mutant was created that forms only one dimer of Rrp41 and Rrp42 and cannot form the trimer of dimers. This mutant cannot form a processing chamber. The fact that this version of the exosome is still active, proofs that the hexameric ring is not essential for catalytic activity. This interface mutant was used with the same concentration of active sites compared to the normal hexamer. Therefore three active sites were accessible for the RNA and as association is diffusion driven, it makes sense that the association rate is larger than for the “closed” complex. Of course the dissociation rate for RNAs of all sizes is larger compared to the wild-type complexes, as the RNA is not fixed within the processing chamber but can freely dissociate from the active site. Additionally interesting is a comparison of the absolute cleavage rates. For most of the complexes the rate is in the magnitude of 15/s, but for the interface mutant it is only 2/s. It may be possible that the formation of the hexameric ring and the existence of the processing chamber can influences the cleavage rate by stabilizing the active site and changing its environment.

Most different compared to the other wild-type complexes are the constants derived for the Rrp4 exosome. The relative binding rate is surprisingly high. When comparing the absolute cleavage rates of all complexes with each other the rate for the Rrp4 complex is 15 times smaller than expected. Possibly the Rrp4 protein as a cap changes the architecture of the whole exosome and thereby also of the active site. This means the binding of the cap proteins could have an allosteric effect on the active site and thus Rrp4 can decrease the cleavage rate compared to Csl4. This is of course very speculative and has to be tested in further detail.

In summary the following conclusion can be drawn from the results:

The data shows that RNA binds to the exosome relatively slow. Still it remains unclear, if the 3'-end can be threaded through the extremely narrow neck of the processing chamber in such a timescale. Probably the whole complex is less rigid when no RNA is bound. The interface between Rrp41 and Rrp42, which forms the active site, is with 3000 $Å^2$ larger than the interface connecting three of these dimers to a hexameric ring (2500 $Å^2$). The smaller interface might loosen a little bit and could thereby increase the diameter of the neck. This could enable a faster association of the RNA to the active site. As soon as RNA is bound to the complex it will be stabilized by ionic interactions between the RNA and mainly arginine residues (e.g. $R65^{Rrp41}$) thus ensuring processivity.

In addition, the importance for RNA binding of the arginine in the neck of the exosome becomes much clearer: RNA molecules that are long enough to reach from the active site to the neck, have much higher affinity to the complex. As soon as this interaction is lost, the affinity drops strongly and the RNA dissociates from the complex easily. Just from qualitative analysis of the denaturing gels in figure 40 the difference between the wild-type and the $R65E^{Rrp41}$ mutants are obvious. The accumulation of RNAs from 10 to 12 bases cannot be detected with the mutant, but is only visible when in presence of the arginine residue.

For short RNAs that cannot reach the arginine residue, the tyrosine 70 from Rrp42 is extremely important for the affinity: the Y70A^{Rrp42} mutant shows extremely low affinity for RNAs smaller 10 bases.

The architecture of the exosome is very conserved and it is quite likely that the observations about RNA association and dissociation made in this thesis might be also relevant for the eukaryotic exosome.

8.7 Evolution and the Exosome

The role of the exosome in all organisms and especially the differences in enzymatic activity between the exosomes from all kingdoms of life from a phosphate dependent to a hydrolytic RNase are of widespread interest. This thesis as well as new results from other groups allow some speculations and the proposition of a model about the events that changed the exosome during evolution and the importance of its architecture.

The changes from the bacterial PNPase to the archaeal exosome are mainly structurally. The enzymatic activities are quite similar as both enzymes catalyze the phosphate dependent 3' → 5' degradation of single stranded RNAs as well as the reverse reaction, the addition of nucleotides to the 3'-end (Slomovic et al., 2008). Structurally some changes occurred: The bacterial enzyme is a homo-trimer that has an RNA binding surface on top and – like the archaeal homolog – the active site is located within the processing chamber. The archaeal exosome however consists of three polypeptide chains each of which exists as three copies within the complex. Although the overall structure is still relatively similar, the RNA binding surface changed. Not only one protein can possibly bind to the core ring, but two different proteins do and it cannot be excluded that a mixed complex is also existent. The additional protein, the Csl4 protein, is located in a different operon in the archaeal genome compared to all other three proteins (Rrp41/Rrp42/Rrp4). This can lead to different expression levels of the proteins and can thereby have an additional regulatory role. RNA binding experiments carried out in this thesis showed a high influence of the arginine residue in the neck of the core ring to the RNA binding affinity. In addition, it could be shown that RNAs with varying sequences have different affinities to the exosome. The same pattern that was observed for the 9-subunit complexes was found for the equivalent cap proteins alone, just with lower affinity. This supports the idea that the two cap proteins bind to different RNA substrates and that a change in their expression level can shift the substrate specificity of the exosome.

The change from one version of RNA binding surface on top of the exosome in the bacterial PNPase to two versions in the archaeal exosome is continued: in eukaryotes a third RNA binding protein that binds on top of the exosome and is homolog to Rrp4 was identified. In comparison to archaea, where exosome complexes with only Rrp4 or only Csl4 can exist, in eukaryotes only mixed complexes with all three different cap proteins were found. Moreover the whole exosome is much more complex, as no protein is present in more than one copy, but six different proteins form the core ring. Summarized this means a change of four possible proteins in the archaeal complex to a set of at least nine different proteins in eukaryotes.

The most important change happened from archaea to eukaryotes concerning the enzymatic activity of the exosome. The following events could have led to these discrepancies:

The reaction of phosphate dependent cleavage of the phosphodiester bond is different to the hydrolytic cleavage regarding their thermodynamic equilibrium. In the phosphate dependent reaction the reverse reaction is much more probable whereas in the hydrolytic reaction the reverse reaction happens hardly ever. For a biological organism the phosphate dependent cleavage is thus a reaction that has to be thoroughly regulated. Obviously the development of a hydrolytic RNase as an additional part of the exosome complex facilitated the regulation of RNA degradation.

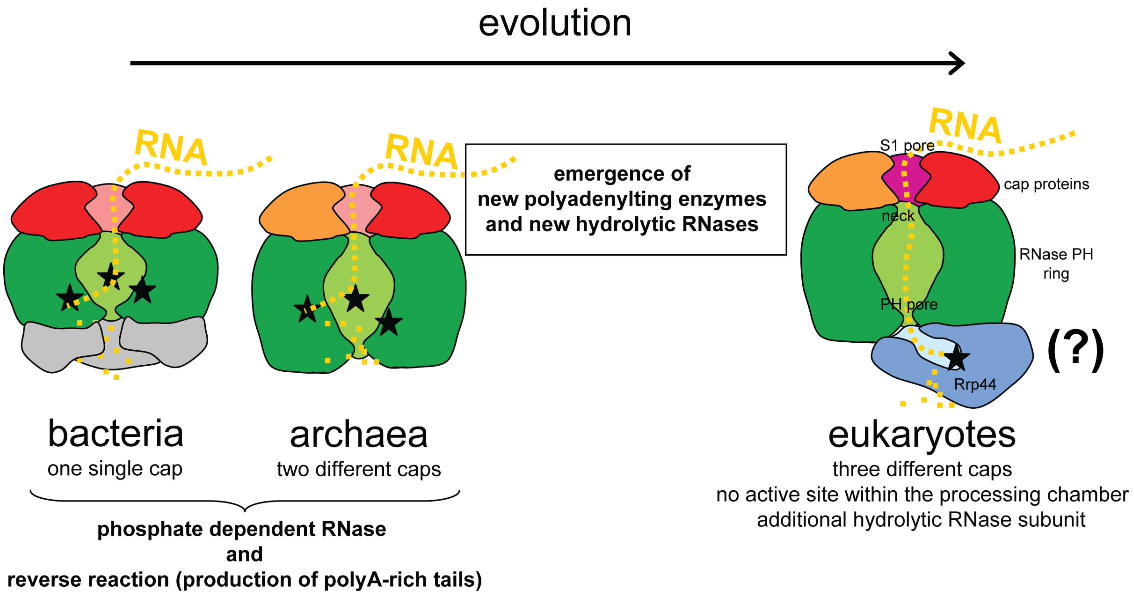

evolution
RNA
RNA
RNA
emergence of
new polyadenylting enzymes
and new hydrolytic RNases
S1 pore
cap proteins
neck
RNase PH ring
PH pore
Rrp44
(?)
bacteria
one single cap
archaea
two different caps
eukaryotes
three different caps
no active site within the processing chamber
additional hydrolytic RNase subunit
phosphate dependent RNase
and
reverse reaction (production of polyA-rich tails)

Figure 50: Changes in exosomes during evolution. Black stars indicate the active sites that changed from phosphate dependent RNase activity within the processing chamber (green) to hydrolytic RNase activity in the additional subunit Rrp44 (blue). Different varieties of cap proteins (red, orange and pink) emerged during evolution from an internal RNA binding surface in the bacterial PNPase and two different cap proteins in archaea to a set of three different caps in eukaryotes. The path of the RNA in the eukaryotic exosome is still speculative.

It seems that is was easier to enable the phosphate dependent active site and introduce a new hydrolytic active site instead. Thus the active site within the processing chamber was disabled and the complex was expanded to a 10-subunit core. This idea is supported by the fact that phosphate dependent RNases exist in eukaryotes only in mitochondria and chloroplasts. In bacteria and archaea the exosome had the additional task of synthesizing poly(A)-rich tails. It could be shown that the enzyme uses all nucleotide-di-phosphates that are available and can add them to the 3'-end of RNA. It is quite likely that the development from enzymes that produce hetero-polymeric tails to specific poly(A) polymerases happened in parallel to the change from phosphorolytic to hydrolytic RNases.

The availability of atomic structures of both the archaeal and the human exosome helps to understand, why all nine core subunits of the exosome are essential for survival in all organisms although none of them possesses any enzymatic activity in eukaryotes. As shown in figure 25 the overall structure of the two exosome complexes is extremely similar. As they do not have the same enzymatic activity, the architecture seems to play a very important role for its functions in RNA surveillance. Through the emergence of new polyadenylating enzymes and new hydrolytic RNases, the enzymatic activity of the exosome was not needed anymore. However not only the active site alone is an important feature of the exosome, but the cap proteins play a role in substrate binding and recognition. In addition, the architecture of the processing chamber disables the complex to degrade structured RNA molecules. The degradation of these substrates is instead regulated by additional cofactors like helicases. Recent results about the Rrp44 protein (Lorentzen et al., 2008) support this theory: the yeast Rrp44 protein alone degraded all RNA molecules, even when secondary structures were present. In complex with the 9-subunit exosome however, only single stranded RNA was degraded. Therefore it is very likely that the mechanisms of the exosome in RNA binding, substrate specificity and regulation did not change from archaea to eukaryotes. Only the position of the active site and the mode of degradation changed – and a different group of enzymes took over the task of addition of poly(A)-tails to RNAs.

9 Summary

The comprehensive analysis of complex biological processes cannot be achieved using only one single biochemical technique. In this thesis two different protein complexes were analyzed, both structurally and biochemically. The results obtained demonstrate the benefits of using hybrid methods to address challenging biological questions.

The DisA protein (DNA integrity scanning protein A) is known for its involvement in DNA damage checkpoint initiation in *Bacillus subtilis* by controlling sporulation. From the crystal structure of DisA the oligomeric organization of the complex could not be unambiguously determined. Structural studies of the complex under physiological conditions revealed an octameric organization with four nucleotide binding sites in the center of the complex and two DNA binding surfaces at the opposing ends. Unexpectedly, DisA was found to possess diadenylate cyclase activity and synthesizes a novel cyclic nucleotide, c-di-AMP. This molecule is reminiscent of, but distinct from c-di-GMP, an emerging prokaryotic regulator of complex cellular processes. Diadenylate cyclase activity was shown to be unaffected by linear DNA or DNA ends but strongly suppressed by branched nucleic acids such as Holliday junctions.

These results indicate that DisA synthesizes c-di-AMP as a signal for DNA structures that interfere with chromosome segregation. Homologies between the nucleic acid binding domain of DisA and proteins with unknown functions in archaea and prokaryotes imply a more general role for c-di-AMP as novel second messenger.

The exosome, the second protein studied, is a multisubunit 3' → 5' exoribonuclease complex that participates in degradation and processing of cellular RNA. Archaeal exosomes consist of three active and three inactive RNase PH subunits that form a hexameric ring and a processing chamber (Rrp41 and Rrp42). Two different RNA binding proteins (Rrp4 and Csl4) can bind as a trimeric cap to this core structure, form an RNA binding surface for the substrate and thus recruit and deliver the RNA to the active sites. The combination of structural studies with biochemical activity assays allowed the proposal of a model for RNA degradation. The data revealed that the process of binding and dissociation of RNA to and from the exosome strongly influences activity and that the architecture of the processing chamber is crucial for the processivity of the enzyme. Additionally, the results indicate that by binding different RNA substrates the two cap proteins Rrp4 and Csl4 influence the overall activity and the specificity of the complex.

Biochemical assays, structural biology and theoretical modeling, together with the latest results from other research groups, support a model that describes the evolution of the exosome and exosome-like complexes: Whereas the enzymatic activity changed from a phosphate dependent to a hydrolytic RNase, the structure of the complex is very conserved from bacteria to human. Therefore, the observed influence of the architecture on the biochemical properties of the exosome, like specificity and processivity, are likely to be also conserved.

10 Appendix

10.1 Summary of SAXS measurements

Table A1: Summarized results of SAXS measurements

protein	DisA	6-subunit exo	Rrp4 exo	Csl4 exo	Rrp4 exo + RNA
beamline	SIBYLS, ALS	SIBYLS, ALS	SIBYLS, ALS	SIBYLS, ALS	X33, DESY
protein conc.	2, 6, 8 mg/ml	5, 10, 15 mg/ml	5, 10, 15 mg/ml	5, 10, 15 mg/ml	A_{280} = 13.75, 27.5, 55
R_g [Å]	52.8	46.2	39.6	41.4	46.8
D_{max} [Å]	176	134	119	117	200
modeling program	GASBORp	GASBORi	GASBORp	GASBORp	GASBORp
number of averaged models	10	15	15	15	10

10.2 Summary of X-ray Crystallographic Experiments

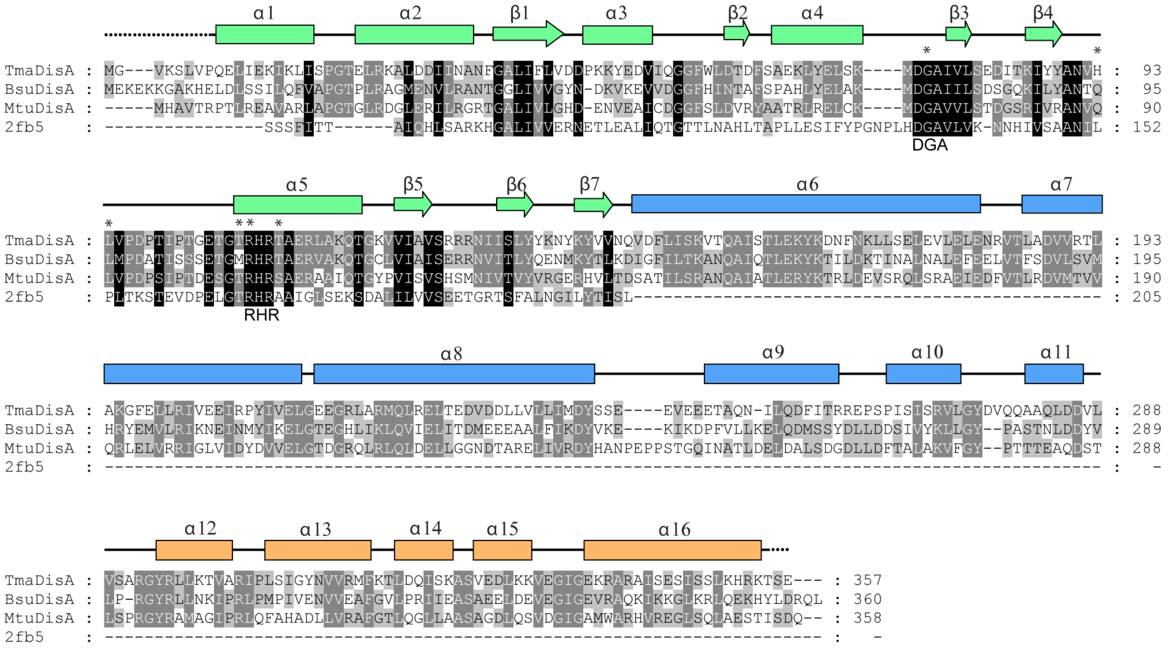
Table A2: Data collection, structure solution and refinement statistics

	Csl4-exosome + RNA	Csl4-exosome Y70A^{Rrp42} + RNA	Csl4 S1-ZnR domain	Rrp4-exo + RNA
DATA COLLECTION				
space group	$P4_322$	$P4_322$	$P2_12_12_1$	R3
cell dimensions	a = b= 138.2 Å; c = 262.0 Å $\alpha = \beta = \gamma = 90°$	a = b = 137.7 Å; c = 261.8 Å $\alpha = \beta = \gamma = 90°$	a = 42.02 Å; b = 75.62 Å c = 76.61 Å $\alpha = \beta = \gamma = 90°$	a = b =112.2 Å; c = 178.6 Å $\alpha = \beta = 90°; \gamma = 120°$
beamline	ID-14-2, ESRF	PX I, SLS	PX I, SLS	PX I, SLS
wavelength	0.933 Å	1.006 Å	1.000 Å	0.972 Å
resolution	2.4 Å	3 Å	1.9 Å	3 Å
R_{sym}	5.6 % (41.3 %)	6.7 % (31.5 %)	6.7 % (41.9 %)	5.7% (36.8%)
$I/\sigma I$	18.9 (3.82)	23.7 (6.10)	14.15 (4.90)	20.7 (3.59)
completeness	94.5 % (86.4%)	98.9 % (96.9 %)	91.9 % (83.4%)	92.7 % (78.4 %)
# of unique reflect.	94,417	50,904	18,281	15,872
redundancy	4.90	7.22	5.05	5.06
REFINEMENT				
resolution (Å)	85 – 2.4 Å	20 – 3 Å	40 – 1.9 Å	not yet refined to a final structure
# of reflect.	99,465	51,207	19,188	
R_{work}/R_{free}	19.4 % / 24.8 %	18.8 % / 26.6 %	22.0 % / 26.3 %	
ligand	6mer RNA (CCCCUC)	6mer RNA (CCCCUC)	-	
water molecules	901	18	96	
rmsd				
bond lengths	0.006 Å	0.007 Å	0.006 Å	
bond angles	1.019°	1.036°	1.038°	

Values in parentheses are for highest-resolution shell. R_{free} is calculated for a randomly chosen 5% of reflections.

10.3 Supplementary Data Concerning DisA

10.3.1 Sequence Alignment



```
TmaDisA : MG---VKSLVPQELIEKIKLISPGTELRKALDDIINANFGALIFLVDDPKKYEDVIQGGFWLDTDFSAEKLYELSK----MDGAIVLSEDITKIYYANVH :  93
BsuDisA : MEKEKKGAKHELDLSSILQFVAPGTPLRAGMENVLRANTGGLIVVGYN-DKVKEVVDGGFHINTAFSPAHLYELAK----MDGAIILSDSGQKILYANTQ :  95
MtuDisA : -----MHAVTRPTLREAVARLAPGTGLRDGLERILRGRTGALIVLGHD-ENVEAICDGGFSLDVRYAATRLRELCK----MDGAVVLSTDGSRIVRANVQ :  90
2fb5    : ----------------SSSFITT------AIQHLSARKHGALIVVERNETLEALIQTGTTLNAHLTAPLLESIFYPGNPLHDGAVLVK-NNHIVSAANIL : 152

TmaDisA : LVPDPTIPTGETGTRHRTAERLAKQTGKVVIAVSRRRNIISLYYKNYKYVVNQVDFLISKVTQAISTLEKYKDNFNKLLSELEVLELENRVTLADVVRTL : 193
BsuDisA : LMPDATISSSETGMRHRTAERVAKQTGCLVIAISERRNVITLYQENMKYTLKDIGFILTKANQAIQTLEKYKTILDKTINALNALEFEELVTFSDVLSVM : 195
MtuDisA : LVPDPSIPTDESGTRHRSAERAAIQTGYPVISVSHSMNIVTVYVRGERHVLTDSATILSRANQAIATLERYKTRLDEVSRQLSRAEIEDFVTLRDVMTVV : 190
2fb5    : PLTKSTEVDPELGTRHRAAIGLSEKSDALILVVSEETGRTSFALNGILYTISL----------------------------------------------- : 205

TmaDisA : AKGFELLRIVEEIRPYIVELGEEGRLARMQLRELTEDVDDLLVLLIMDYSSE----EVEEETAQN-ILQDFITRREPSPISISRVLGYDVQQAAQLDDVL : 288
BsuDisA : HRYEMVLRIKNEINMYIKELGTEGHLIKLQVIELITDMEEEAALFIKDYVKE----KIKDPFVLLKELQDMSSYDLLDDSIVYKLLGY--PASTNLDDYV : 289
MtuDisA : QRLELVRRIGLVIDYDVVELGTDGRQLRLQLDELLGGNDTARELIVRDYHANPEPPSTGQINATLDELDALSDGDLLDFTALAKVFGY--PTTTEAQDST : 288
2fb5    : ---------------------------------------------------------------------------------------------------- :   -

TmaDisA : VSARGYRLLKTVARIPLSIGYNVVRMFKTLDQISKASVEDLKKVEGIGEKRARAISESISSLKHRKTSE--- : 357
BsuDisA : LP-RGYRLLNKIPRLPMPIVENVVEAFGVLPRIIEASAEELDEVEGIGEVRAQKIKKGLKRLQEKHYLDRQL : 360
MtuDisA : LSPRGYRAMAGIPRLQFAHADLLVRAFGTLQGLLAASAGDLQSVDGIGAMWARHVREGLSQLAESTISDQ-- : 358
2fb5    : ------------------------------------------------------------------------ :   -
```

Figure A1: Sequence alignment of DisA proteins (COG1623) of *Thermotoga maritima* (TmaDisA), *Bacillus subtilis* (BsuDisA) and *Mycobacterium tuberculosis* (MtuDisA) along with a structurally identified homolog from *Bacillus cereus* (hypothetical membrane spanning protein COG1624; PDB code 2FB5, residues 76 – 205). Conserved residues are shaded and the secondary structure of TmaDisA is shown on top of the alignment. Notable functional motifs are annotated (DGA and RHR). Residues that bind the c-di-AMP are indicated (*).

10.3.2 Mass spectrometry results, which identified c-di-AMP

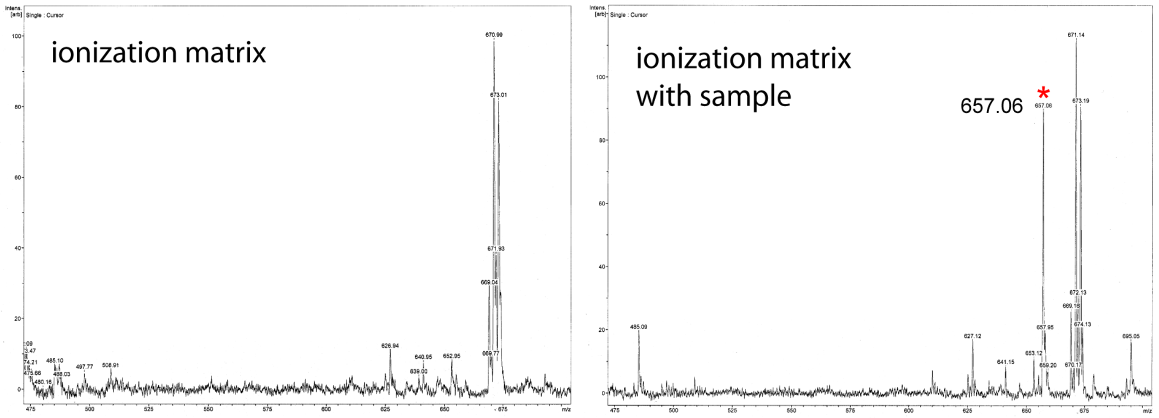


Figure A2: Isolated product from TLC plates produced by DisA analyzed by negative mode mass spectrometry on a Bruker Autoflex II MALDI-TOFMS instrument. The left panel is the m/z spectrum of the sinapinic acid ionization matrix (control), the right panel shows the sample ionized in sinapinic acid. The prominent peak (red *) corresponds to the expected mass of singly charged c-di-AMP (657 Da).

10.4 MATLAB Scripts

In order to fit the models to the experimental data, several scripts are needed. The scripts exsosquares_m1.m (see section 9.4.1) and exosquares_m2.m (see section 9.4.2) simulate the differential equations of the cleavage-only model and the cleavage-and-binding model, respectively, and calculate the quadratic differences between simulated curves and measured data. To minimize the quadratic differences, a second script (exominsquares.m, see below) is needed. This script additionally reads in the experimental data.

exominsquares.m

```
dd=load('-ascii','experimental_data.txt'); % load experimental data


%no experimental data for t=0 -> insert t=0, y_t=1
%Indices of matrix dd have to agree with experimental data
t = [0;dd(2:27,1)];
y_t =[1,zeros(1,27);dd(2:27,2:29)];

pstart = [0.022, 0, 0.04, 0.03,  0.4, 15];  %start parameters for finding the minimum

% minimize quadratic differences
perg=fminsearch(@(par) exosquares_mx(par,t,y_t),pstart) %x=1,2 depending on model
```

10.4.1 Cleavage-only model

exsosquares_m1.m

```
function squares = exosquares_m1(params, t, y_t)

global ks

ks=[...]; % write here fit parameters as params(1), params(2), ...

x0 = [1;zeros(27,1)]; %initial condition for t=0

[T, X] = ode15s(@reakglei, t, x0); %solve differential equation

figure(1) %plotting the experimental data (y_t) and the fits (X)
plot(T,X);
hold on;
plot(T,y_t,'*','MarkerSize',4);
hold off;

gew=zeros(length(T),28);  %matrix for weighted fit
gew(1,:)=ones(1,28).*T(1)./T(length(T));
for i=2:length(T)
gew(i,:)=ones(1,28).*(T(i)-T(i-1))./T(length(T));
end;

quadrier=(X-y_t).*(X-y_t).*gew;

squares=sum(sum(quadrier)); %quadratic differences between experimental data and fit

function xdot = reakglei(tin, x) %cleavage-only model

global ks

xdot=zeros(28,1);

xdot(1) =-ks(1)*x(1); %30mer
for i=2:27
xdot(i)=-ks(i)*x(i)+ks(i-1)*x(i-1); %(31-i)mer
end;
xdot(28)=ks(27)*x(27);%3mer
```

10.4.2 Cleavage-and-binding model

exosquares_m2.m

function squares = exosquares_m2(params, t, y_t)

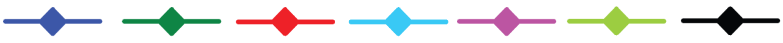
```
global kb kd ks kk
kb=[ ... ]; % write here fit parameters as params(1), params(2), ...
kd=[ ... ];
ks=ones(1,27).*...;
kk=zeros(1,27);

x0 = [1;0;zeros(54,1)]; %initial condition for t=0

[T, X] = ode15s(@reakglei, t, x0); %solve differential equation

merkonz=zeros(length(T),28); %adding the concentrations of bound and free RNA
freikonz=zeros(length(T),28);
gebkonz=zeros(length(T),28);
for i=1:28
merkonz(:,i)=X(:,2*i-1)+X(:,2*i);
freikonz(:,i)=X(:,2*i-1);
gebkonz(:,i)=X(:,2*i);
end;

save exosome.dat merkonz -ASCII %saving the solution

figure(1) %plotting the experimental data (y_t) and the fits (merkonz)
plot(T,merkonz);
hold on;
plot(T,y_t,'*','MarkerSize',4);
hold off;

gew=zeros(length(T),28); %matrix for weighted fit
gew(1,:)=ones(1,28).*T(1)./T(length(T));
for i=2:length(T)
gew(i,:)=ones(1,28).*(T(i)-T(i-1))./T(length(T));
end;

quadrier=(merkonz-y_t).*(merkonz-y_t).*gew; % needed for calculation of quadratic
differences between experimental data and fit

%different options:
%minimizing only the quadratic differences for 30mer
%squares=sum(sum(quadrier(:,1)))

%minimizing the quadratic differences for 30mer (i=17) and for 14mer - 3mer bis 3 (i=28)
%squares=sum(sum(quadrier(:,21))) + sum(sum(quadrier(:,18:22)));

%minimizing the quadratic differences for 30mer - 13mer
%squares=sum(sum(quadrier(:,1:16)));

%minimizing the quadratic differences for all curves
squares=sum(sum(quadrier));

%calcualation of coefficient of determination
Sserr=sum(sum((merkonz-y_t).*(merkonz-y_t))); %calculation of R² (quality of the fit)
SStot=sum(sum((y_t-ones(size(t))*mean(y_t)).*(y_t-ones(size(t))*mean(y_t))));
1-SSerr/SStot %coefficient of determination, will be writen to Matlab console

function xdot = reakglei(tin, x) %cleavage and binding model

global kb kd ks kk

xdot=zeros(56,1);
xdot(1)  =-kb(1)*x(1)+kd(1)*x(2); %free 30mer
xdot(2)  = kb(1)*x(1)-kd(1)*x(2)-ks(1)*x(2)+kk(1)*x(4); %bound 30mer

for i=2:27
xdot(2*i-1)=-kb(i)*x(2*i-1)+kd(i)*x(2*i); % free(31-i)mer
xdot(2*i)=kb(i)*x(2*i-1)+ks(i-1)*x(2*i-2)-(kk(i-1)+ks(i)+kd(i))*x(2*i)+kk(i)*x(2*i+2);
%bound(31-i)mer
end;

xdot(55)  =-kb(28)*x(55)+kd(28)*x(56);  %free 3mer
xdot(56)  = kb(28)*x(55)+ks(27)*x(54)-(kk(27)+kd(28))*x(56); %bound 3mer
```

Script for drawing the simulated curves with higher resolution

```
function dummy()

global kb kd ks kk
kb=[...]; %rate constants as determined by fitting
kd=[...];
ks=ones(1,27).*...;
kk=zeros(1,27);

x0 = [1;0;zeros(54,1)];
dd=load('-ascii','experimental_data.txt');

tmess=[0;dd(2:27,1)]; %times of experimental data
y_t=[1,zeros(1,27);dd(2:27,2:29)]; %experimental data in y_t
t=linspace (0, 900, 3000)'; %3000 points in between t=0s and t=900s
[T, X] = ode15s(@reakglei, t, x0); %solve differential equation

merkonz=zeros(length(T),28); %adding the concentrations of bound and free RNA
freikonz=zeros(length(T),28);
gebkonz=zeros(length(T),28);
for i=1:28
merkonz(:,i)=X(:,2*i-1)+X(:,2*i);
freikonz(:,i)=X(:,2*i-1);
gebkonz(:,i)=X(:,2*i);
end;

table = [t, merkonz]; %saving the solution
save exosome.dat table -ASCII

figure(1) %plotting the data
plot(T,merkonz);
hold on;
plot(tmess,y_t,'*','MarkerSize',4);
hold off;

i=28; % plotting only the curve for the i-mer
figure(2)
plot(T,merkonz(:,i));
hold on;
plot(tmess,y_t(:,i),'*','MarkerSize',4);
hold off;

function xdot = reakglei(tin, x) %cleavage and binding model

global kb kd ks kk
xdot=zeros(56,1);
xdot(1)   =-kb(1)*x(1)+kd(1)*x(2);   %free 30mer
xdot(2)   = kb(1)*x(1)-kd(1)*x(2)-ks(1)*x(2)+kk(1)*x(4); % bound 30mer

for i=2:27
xdot(2*i-1)=-kb(i)*x(2*i-1)+kd(i)*x(2*i); %free (31-i)mer
xdot(2*i)=kb(i)*x(2*i-1)+ks(i-1)*x(2*i-2)- ...
(kk(i-1)+ks(i)+kd(i))*x(2*i)+kk(i)*x(2*i+2); %bound (31-i)mer
end;

xdot(55)  =-kb(28)*x(55)+kd(28)*x(56);  %free 3mer
xdot(56)  = kb(28)*x(55)+ks(27)*x(54)-(kk(27)+kd(28))*x(56);%bound 3mer
```

10.5 Experimental Data from RNase Assays and Fits

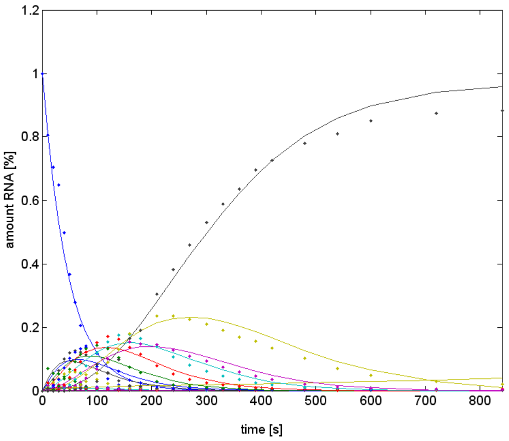
The experimental data and the corresponding fits are shown for all six complexes. The color code is for all plots as follows:



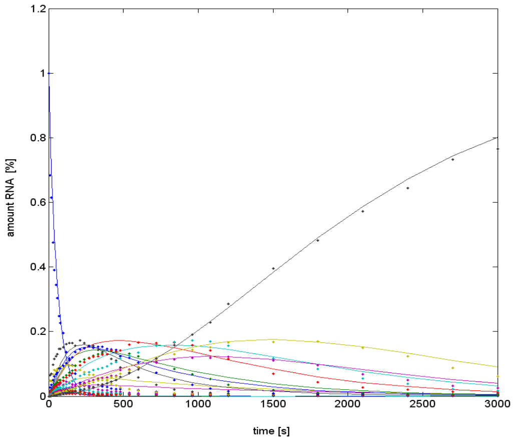
30mer	29mer	28mer	27mer	26mer	25mer	24mer
23mer	22mer	21mer	20mer	19mer	18mer	17mer
16mer	15mer	14mer	13mer	12mer	11mer	10mer
9mer	8mer	7mer	6mer	5mer	4mer	3mer

The data was fitted using the cleavage-and-binding model described in chapter 7.3.3. The obtained rate constants are given in table 5.

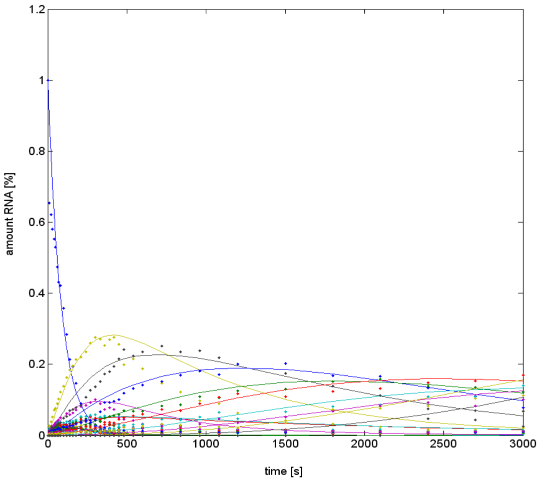
6-subunit exosome



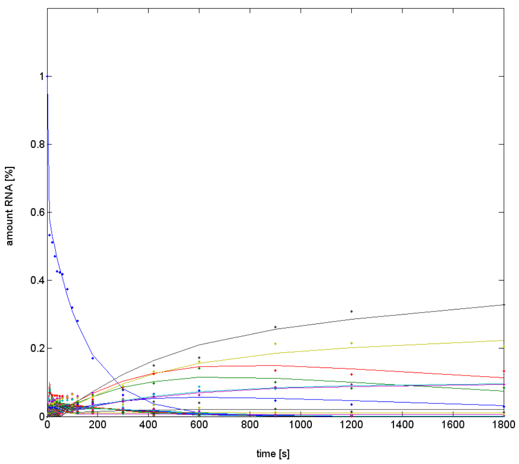
Csl4 exosome



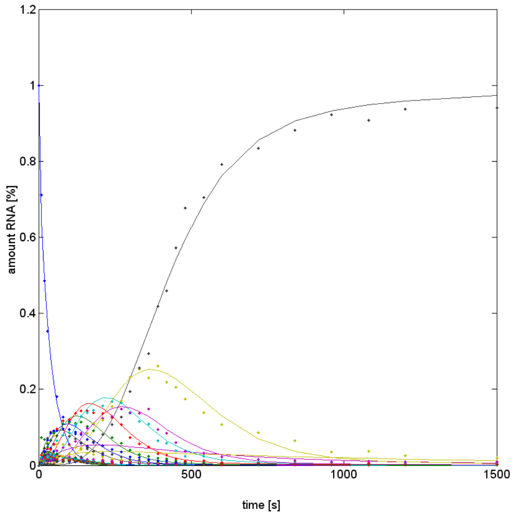
Csl4 exosome Y70A^{Rrp42}



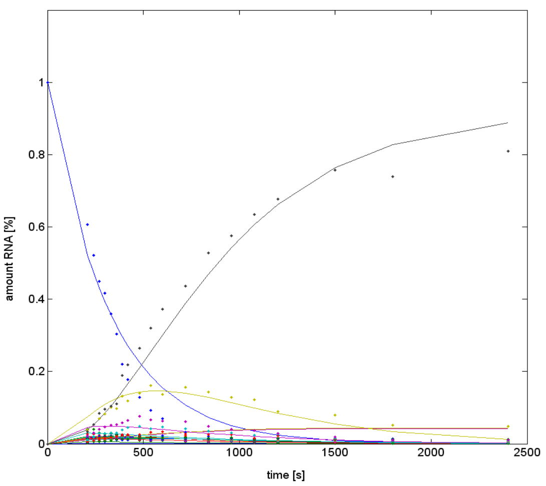
Rrp4 exosome



2-subunit exosome

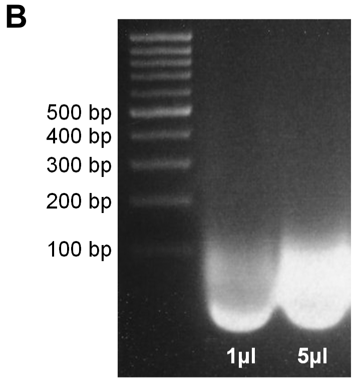


Csl4 exosome R65E^{Rrp41}



10.6 Denaturing PA Gel and Agarose Gel Showing the RNA Ligand

A
50 – 70 bases


B
500 bp
400 bp
300 bp
200 bp
100 bp
1µl
5µl

A: Denaturing PA gel of the RNA substrate, that was purified with the exosome from *E. coli* cells. On the right the bands of an RNase assay are shown that were used for the calibrationcurve to estimate the size of the RNA at the left.

B: Agarose gel showing 1µl and 5µl of the sample used for crystal setups. The RNA was stained with ethidium bromide.

11 References

Afonine, P. V., R. W. Grosse-Kunstleve and P. D. Adams (2005). "A robust bulk-solvent correction and anisotropic scaling procedure." Acta Crystallogr D Biol Crystallogr **61**(Pt 7): 850-5.

Aldridge, P., R. Paul, P. Goymer, P. Rainey and U. Jenal (2003). "Role of the GGDEF regulator PleD in polar development of Caulobacter crescentus." Mol Microbiol **47**(6): 1695-708.

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.

Amikam, D. and M. Y. Galperin (2006). "PilZ domain is part of the bacterial c-di-GMP binding protein." Bioinformatics **22**(1): 3-6.

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., K. L. Carroll, J. M. Ames and J. L. Corden (2006). "Regulation of yeast NRD1 expression by premature transcription termination." Mol Cell **21**(5): 641-51.

Ariyoshi, M., T. Nishino, H. Iwasaki, H. Shinagawa and K. Morikawa (2000). "Crystal structure of the holliday junction DNA in complex with a single RuvA tetramer." Proc Natl Acad Sci U S A **97**(15): 8257-62.

Baker, D. A. and J. M. Kelly (2004). "Structure, function and evolution of microbial adenylyl and guanylyl cyclases." Mol Microbiol **52**(5): 1229-42.

Barreau, C., L. Paillard and H. B. Osborne (2005). "AU-rich elements and associated factors: are there unifying principles?" Nucleic Acids Res **33**(22): 7138-50.

Bemeleit, D. (2008). Structural charaterization of the DNA repair protein complex SbcC-SbcD of Thermotoga maritima. Fakultät für Chemie und Pharmazie. München, LMU. **PhD**.

Benard, L., K. Carroll, R. C. Valle, D. C. Masison and R. B. Wickner (1999). "The ski7 antiviral protein is an EF1-alpha homolog that blocks expression of non-Poly(A) mRNA in Saccharomyces cerevisiae." J Virol **73**(4): 2893-900.

Bernstein, J., D. N. Patterson, G. M. Wilson and E. A. Toth (2008). "Characterization of the essential activities of Saccharomyces cerevisiae Mtr4p, a 3'->5' helicase partner of the nuclear exosome." J Biol Chem **283**(8): 4930-42.

Blow, D. (2002). Outline of Crystallography for Biologists, 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.

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.

Burkard, K. T. and J. S. Butler (2000). "A nuclear 3'-5' exonuclease involved in mRNA degradation interacts with Poly(A) polymerase and the hnRNA protein Npl3p." Mol Cell Biol **20**(2): 604-16.

Büttner, K. (2005). Structural Framework for the Mechanism of Archaeal Exosomes in RNA processing. Fakultät für Chemie und Pharmazie. München, LMU München. **PhD:** 163.

Büttner, K., K. Wenig and K. P. Hopfner (2005). "Structural framework for the mechanism of archaeal exosomes in RNA processing." Mol Cell **20**(3): 461-71.

Camilli, A. and B. L. Bassler (2006). "Bacterial small-molecule signaling pathways." Science **311**(5764): 1113-6.

Carneiro, T., C. Carvalho, J. Braga, J. Rino, L. Milligan, D. Tollervey and M. Carmo-Fonseca (2007). "Depletion of the yeast nuclear exosome subunit Rrp6 results in accumulation of polyadenylated RNAs in a discrete domain within the nucleolus." Mol Cell Biol **27**(11): 4157-65.

Carroll, K. L., D. A. Pradhan, J. A. Granek, N. D. Clarke and J. L. Corden (2004). "Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts." Mol Cell Biol **24**(14): 6241-52.

CCP4 (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr D Biol Crystallogr **50**(Pt 5): 760-3.

Chacon, P., F. Moran, J. F. Diaz, E. Pantos and J. M. Andreu (1998). "Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm." Biophys J **74**(6): 2760-75.

Chan, C., R. Paul, D. Samoray, N. C. Amiot, B. Giese, U. Jenal and T. Schirmer (2004). "Structural basis of activity and allosteric control of diguanylate cyclase." Proc Natl Acad Sci U S A **101**(49): 17084-9.

Chang, A. L., J. R. Tuckerman, G. Gonzalez, R. Mayer, H. Weinhouse, G. Volman, D. Amikam, M. Benziman and M. A. Gilles-Gonzalez (2001). "Phosphodiesterase A1, a regulator of cellulose synthesis in Acetobacter xylinum, is a heme-based sensor." Biochemistry **40**(12): 3420-6.

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.

Christen, M., B. Christen, M. Folcher, A. Schauerte and U. Jenal (2005). "Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP." J Biol Chem **280**(35): 30829-37.

D'Argenio, D. A. and S. I. Miller (2004). "Cyclic di-GMP as a bacterial second messenger." Microbiology **150**(Pt 8): 2497-502.

de Rooij, J., F. J. Zwartkruis, M. H. Verheijen, R. H. Cool, S. M. Nijman, A. Wittinghofer and J. L. Bos (1998). "Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP." Nature **396**(6710): 474-7.

Deutscher, M. P. and Z. Li (2001). "Exoribonucleases and their multiple roles in RNA metabolism." Prog Nucleic Acid Res Mol Biol **66**: 67-105.

Doma, M. K. and R. Parker (2006). "Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation." Nature **440**(7083): 561-4.

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.

Edmonds, M. (2002). "A history of poly A sequences: from formation to factors to function." Prog Nucleic Acid Res Mol Biol **71**: 285-389.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-32.

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.

Fedorov, N. A. (1976). "[Cyclic guanosine monophosphate (cGMP): metabolism and its biological role]." Usp Sovrem Biol **82**(4): 34-46.

Francis, S. H. and J. D. Corbin (1999). "Cyclic nucleotide-dependent protein kinases: intracellular receptors for cAMP and cGMP action." Crit Rev Clin Lab Sci **36**(4): 275-328.

Fraser, R. D. B., MacRae, T. P. and Suzuki, E. (1978). "An improved method for calculating the contribution of solvent to the X-ray diffraction pattern of biological molecules." J. Appl. Cryst. **11**: 693-694.

Frischmeyer, P. A., A. van Hoof, K. O'Donnell, A. L. Guerrerio, R. Parker and H. C. Dietz (2002). "An mRNA surveillance mechanism that eliminates transcripts lacking termination codons." Science **295**(5563): 2258-61.

Fujisawa, T., T. Uruga, Z. Yamaizumi, Y. Inoko, S. Nishimura and T. Ueki (1994). "The hydration of Ras p21 in solution during GTP hydrolysis based on solution X-ray scattering profile." J Biochem **115**(5): 875-80.

Galperin, M. Y. (2005). "A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts." BMC Microbiol **5**: 35.

Gherzi, R., K. Y. Lee, P. Briata, D. Wegmuller, C. Moroni, M. Karin and C. Y. Chen (2004). "A KH domain RNA binding protein, KSRP, promotes ARE-directed mRNA turnover by recruiting the degradation machinery." Mol Cell **14**(5): 571-83.

Grunberg-Manago, M. (1999). "Messenger RNA stability and its role in control of gene expression in bacteria and phages." Annu Rev Genet **33**: 193-227.

Guinier, A. (1939). "La diffraction des rayons X aux tres petits angles; application a l’etude de phenomenes ultramicroscopiques." Ann Phys **12**: 166-237.

Hall-Stoodley, L. and P. Stoodley (2005). "Biofilm formation and dispersal and the transmission of human pathogens." Trends Microbiol **13**(1): 7-10.

Hargreaves, D., D. W. Rice, S. E. Sedelnikova, P. J. Artymiuk, R. G. Lloyd and J. B. Rafferty (1998). "Crystal structure of E.coli RuvA with bound DNA Holliday junction at 6 A resolution." Nat Struct Biol **5**(6): 441-6.

Hartung, S. and K. P. Hopfner (2007). "The exosome, plugged." EMBO Rep **8**(5): 456-7.

Hartwell, L. H. and T. A. Weinert (1989). "Checkpoints: controls that ensure the order of cell cycle events." Science

246(4930): 629-34.

Heller, W. T. (2005). "Influence of multiple well defined conformations on small-angle scattering of proteins in solution." Acta Crystallogr D Biol Crystallogr **61**(Pt 1): 33-44.

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): 51-9.

Holliday, R. (1964). "A mechanism for gene conversion in fungi." Genet. Res. Camb. **5**: 282–304.

Holm, L. and C. Sander (1997). "Dali/FSSP classification of three-dimensional protein folds." Nucleic Acids Res **25**(1): 231-4.

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

Houseley, J. and D. Tollervey (2006). "Yeast Trf5p is a nuclear poly(A) polymerase." EMBO Rep **7**(2): 205-11.

Huang, B., C. B. Whitchurch and J. S. Mattick (2003). "FimX, a multidomain protein connecting environmental signals to twitching motility in Pseudomonas aeruginosa." J Bacteriol **185**(24): 7068-76.

Ishii, R., O. Nureki and S. Yokoyama (2003). "Crystal structure of the tRNA processing enzyme RNase PH from Aquifex aeolicus." J Biol Chem **278**(34): 32397-404.

Jenal, U. (2004). "Cyclic di-guanosine-monophosphate comes of age: a novel secondary messenger involved in modulating cell surface structures in bacteria?" Curr Opin Microbiol **7**(2): 185-91.

Jenal, U. and J. Malone (2006). "Mechanisms of cyclic-di-GMP signaling in bacteria." Annu Rev Genet **40**: 385-407.

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

Kadaba, S., A. Krueger, T. Trice, A. M. Krecic, A. G. Hinnebusch and J. Anderson (2004). "Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae." Genes Dev **18**(11): 1227-40.

Kadaba, S., X. Wang and J. T. Anderson (2006). "Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA." Rna **12**(3): 508-21.

Kastenmayer, J. P. and P. J. Green (2000). "Novel features of the XRN-family in Arabidopsis: evidence that AtXRN4, one of several orthologs of nuclear Xrn2p/Rat1p, functions in the cytoplasm." Proc Natl Acad Sci U S A **97**(25): 13985-90.

Kaupp, U. B. and R. Seifert (2002). "Cyclic nucleotide-gated ion channels." Physiol Rev **82**(3): 769-824.

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

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

Kornberg, A. (2005). DNA replication, University Science Books.

Kowalczykowski, S. C. (2000). "Initiation of genetic recombination and recombination-dependent replication." Trends Biochem Sci **25**(4): 156-65.

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

Kushner, S. R. (2004). "mRNA decay in prokaryotes and eukaryotes: different approaches to a similar problem." IUBMB Life **56**(10): 585-94.

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.

Lammens, A., A. Schele and K. P. Hopfner (2004). "Structural biochemistry of ATP-driven dimerization and DNA-stimulated activation of SMC ATPases." Curr Biol **14**(19): 1778-82.

Lejeune, F., X. Li and L. E. Maquat (2003). "Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities." Mol Cell **12**(3): 675-87.

Lemon, K. P. and A. D. Grossman (1998). "Localization of bacterial DNA polymerase: evidence for a factory model of replication." Science **282**(5393): 1516-9.

Li, Z. and M. P. Deutscher (1996). "Maturation pathways for E. coli tRNA precursors: a random multienzyme process in vivo." Cell **86**(3): 503-12.

Lipfert, J. and S. Doniach (2007). "Small-angle X-ray scattering from RNA, proteins, and protein complexes." Annu Rev Biophys Biomol Struct **36**: 307-27.

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., J. Basquin, R. Tomecki, A. Dziembowski and E. Conti (2008). "Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family." Mol Cell **29**(6): 717-28.

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 **8**(5): 470-6.

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 **12**(7): 575-81.

Lucas, K. A., G. M. Pitari, S. Kazerounian, I. Ruiz-Stewart, J. Park, S. Schulz, K. P. Chepenik and S. A. Waldman (2000). "Guanylyl cyclases and signaling by cyclic GMP." Pharmacol Rev **52**(3): 375-414.

Matsumoto, Y., G. Sarkar, S. S. Sommer and R. B. Wickner (1993). "A yeast antiviral protein, SKI8, shares a repeated amino acid sequence pattern with beta-subunits of G proteins and several other proteins." Yeast **9**(1): 43-51.

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

Miller, J. H. (1972). Experiments in molecular genetics. . Cold Spring Harbour, Cold Spring Harbour Laboratory.

Milligan, L., C. Torchet, C. Allmang, T. Shipman and D. Tollervey (2005). "A nuclear surveillance pathway for mRNAs with defective polyadenylation." Mol Cell Biol **25**(22): 9996-10004.

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., E. Petfalski and D. Tollervey (1996). "The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism." Genes Dev **10**(4): 502-13.

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 functions both as a 3' right-arrow 5' exonuclease and a poly(A) polymerase in Escherichia coli." Proc Natl Acad Sci U S A **97**(22): 11966-71.

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.

Navarro, M. V., C. C. Oliveira, N. I. Zanchin and B. G. Guimaraes (2008). "Insights into the mechanism of progressive RNA degradation by the archaeal exosome." J Biol Chem **283**(20): 14120-31.

Oddone, A., E. Lorentzen, J. Basquin, A. Gasch, V. Rybin, E. Conti and M. Sattler (2007). "Structural and biochemical characterization of the yeast exosome component Rrp40." EMBO Rep **8**(1): 63-9.

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.

Paul, R., S. Abel, P. Wassmann, A. Beck, H. Heerklotz and U. Jenal (2007). "Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization." J Biol Chem **282**(40): 29170-7.

Paul, R., S. Weiser, N. C. Amiot, C. Chan, T. Schirmer, B. Giese and U. Jenal (2004). "Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain." Genes Dev **18**(6): 715-27.

Peng, W. T., M. D. Robinson, S. Mnaimneh, N. J. Krogan, G. Cagney, Q. Morris, A. P. Davierwala, J. Grigull, X. Yang, W. Zhang, N. Mitsakakis, O. W. Ryan, N. Datta, V. Jojic, C. Pal, V. Canadien, D. Richards, B. Beattie, L. F. Wu, S. J. Altschuler, S. Roweis, B. J. Frey, A. Emili, J. F. Greenblatt and T. R. Hughes (2003). "A panoramic view of yeast noncoding RNA processing." Cell **113**(7): 919-33.

Petoukhov, M. V. a. S., D.I. (2003). "New methods for domain structure determination of proteins from solution scattering data." J. Appl. Cryst. **36**: 540-544.

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.

Putnam, C. D., M. Hammel, G. L. Hura and J. A. Tainer (2007). "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution." Q Rev Biophys **40**(3): 191-285.

Rafferty, J. B., S. E. Sedelnikova, D. Hargreaves, P. J. Artymiuk, P. J. Baker, G. J. Sharples, A. A. Mahdi, R. G. Lloyd and D. W. Rice (1996). "Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction." Science **274**(5286): 415-21.

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.

Regnier, P. and C. M. Arraiano (2000). "Degradation of mRNA in bacteria: emergence of ubiquitous features." Bioessays **22**(3): 235-44.

Rhee, S. K., T. Icho and R. B. Wickner (1989). "Structure and nuclear localization signal of the SKI3 antiviral protein of Saccharomyces cerevisiae." Yeast **5**(3): 149-58.

Roe, S. M., T. Barlow, T. Brown, M. Oram, A. Keeley, I. R. Tsaneva and L. H. Pearl (1998). "Crystal structure of an octameric RuvA-Holliday junction complex." Mol Cell **2**(3): 361-72.

Romling, U. and D. Amikam (2006). "Cyclic di-GMP as a second messenger." Curr Opin Microbiol **9**(2): 218-28.

Rott, R., G. Zipor, V. Portnoy, V. Liveanu and G. Schuster (2003). "RNA polyadenylation and degradation in cyanobacteria are similar to the chloroplast but different from Escherichia coli." J Biol Chem **278**(18): 15771-7.

Ryjenkov, D. A., M. Tarutina, O. V. Moskvin and M. Gomelsky (2005). "Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain." J Bacteriol **187**(5): 1792-8.

Sambrook, J. a. R., D.W. (2001). Molecular Cloning: A Laboratory Manual, 3rd edition, Cold Spring Harbour Laboratory Press.

Sands, W. A. and T. M. Palmer (2008). "Regulating gene transcription in response to cyclic AMP elevation." Cell Signal **20**(3): 460-6.

Schmidt, A. J., D. A. Ryjenkov and M. Gomelsky (2005). "The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains." J Bacteriol **187**(14): 4774-81.

Siede, W. K., Y. W. and Doetsch P. W. (2006). DNA damage recognition, Taylor & Francis.

Slomovic, S., D. Laufer, D. Geiger and G. Schuster (2005). "Polyadenylation and degradation of human mitochondrial RNA: the prokaryotic past leaves its mark." Mol Cell Biol **25**(15): 6427-35.

Slomovic, S., V. Portnoy, S. Yehudai-Resheff, E. Bronshtein and G. Schuster (2008). "Polynucleotide phosphorylase and the archaeal exosome as poly(A)-polymerases." Biochim Biophys Acta **1779**(4): 247-55.

Slomovic, S. P., V.; Liveanu, V; Schuster G. (2006). "RNA polyadenylation in prokaryotes and organelles; different tails tell different tails." Crit Rev Plant Sci **25**: 65-77.

Sokolova, A. V., Volkov, V.V. and Svergun, D.I. (2003). "Prototype of a database for rapid protein classification based on solution scattering data." J. Appl. Cryst. **36**: 865-868.

Sommer, J. M. and A. Newton (1989). "Turning off flagellum rotation requires the pleiotropic gene pleD: pleA, pleC, and pleD define two morphogenic pathways in Caulobacter crescentus." J Bacteriol **171**(1): 392-401.

Stead, J. A., J. L. Costello, M. J. Livingstone and P. Mitchell (2007). "The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein." Nucleic Acids Res **35**(16): 5556-67.

Steinmetz, E. J. and D. A. Brow (1998). "Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association." Proc Natl Acad Sci U S A **95**(12): 6699-704.

Steinmetz, E. J., N. K. Conrad, D. A. Brow and J. L. Corden (2001). "RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts." Nature **413**(6853): 327-31.

Sutherland, E. W. (1972). "Studies on the mechanism of hormone action." Science **177**(47): 401-8.

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.

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

Svergun, D. I. (1999). "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing." Biophys J **76**(6): 2879-86.

Svergun, D. I., Barberato, C. and Koch, M.H.J. (1995). "CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates." J. Appl. Cryst. **28**: 768-773.

Svergun, D. I., M. V. Petoukhov and M. H. Koch (2001). "Determination of domain structure of proteins from X-ray solution scattering." Biophys J **80**(6): 2946-53.

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 **8**(11): 1215-26.

Toh, E. A., P. Guerry and R. B. Wickner (1978). "Chromosomal superkiller mutants of Saccharomyces cerevisiae." J Bacteriol **136**(3): 1002-7.

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.

Tsutakawa, S. E., G. L. Hura, K. A. Frankel, P. K. Cooper and J. A. Tainer (2007). "Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography." J Struct Biol **158**(2): 214-23.

van Hoof, A., P. A. Frischmeyer, H. C. Dietz and R. Parker (2002). "Exosome-mediated recognition and degradation of mRNAs lacking a termination codon." Science **295**(5563): 2262-4.

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.

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

Wang, H. W., J. Wang, F. Ding, K. Callahan, M. A. Bratkowski, J. S. Butler, E. Nogales and A. Ke (2007). "Architecture of the yeast Rrp44 exosome complex suggests routes of RNA recruitment for 3' end processing." Proc Natl Acad Sci U S A **104**(43): 16844-9.

Wang, Z., J. V. Shah, M. W. Berns and D. W. Cleveland (2006). "In vivo quantitative studies of dynamic intracellular processes using fluorescence correlation spectroscopy." Biophys J **91**(1): 343-51.

Warner, J. R. (1999). "The economics of ribosome biosynthesis in yeast." Trends Biochem Sci **24**(11): 437-40.

Widner, W. R. and R. B. Wickner (1993). "Evidence that the SKI antiviral system of Saccharomyces cerevisiae acts by blocking expression of viral mRNA." Mol Cell Biol **13**(7): 4331-41.

Witte, G., S. Hartung, K. Buttner and K. P. Hopfner (2008). "Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates." Mol Cell **30**(2): 167-78.

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." Cell **121**(5): 725-37.

12 Curriculum Vitae

Personal Details

Name	Sophia Astrid Hartung
Date of Birth	July 18, 1980
Place of Birth	Würzburg, Germany
Nationality	german
Marital Status	single

Education

1986-1990	Grundschule Dürrbachgrund
1990-1999	Mozart-Gymnasium, Würzburg

Academic Training

1999 – 2004	University of Bayreuth studies in biochemistry (Diplom)
2003	University of York erasmus exchange program, supervisor: Dr. Jim G. Hoggett
2003 – 2004	University of Bayreuth, department of biochemistry diploma Thesis: “Strukturbiologische Untersuchungen zur Faltung und Katalyse von Ribonuklease T1” supervisor: Prof. Dr. Jochen Balbach
2004	Sanofi-Aventis, Frankfurt internship
since January 2005	LMU Munich, Gene Center / Institute of Biochemistry PhD thesis in structural biology Supervisor: Prof. Dr. Karl-Peter Hopfner

13 Acknowledgments

I would like to thank everybody, who supported me during the last years and made this work possible:

First of all, I would like to thank my supervisor Prof. Dr. Karl-Peter Hopfner for giving me the opportunity to work in his group and for his guidance and many fruitful scientific discussions. I appreciate especially his encouragement to learn new methods like SAXS.

The atmosphere in the Hopfner lab is without doubt exceptionally pleasant. I would like to express my gratitude especially to Harald for introducing me so well to the labs working techniques. The close collaboration with Kathi, Gregor and Katja was always very enjoyable and instructive. Of course many thanks to the rest of the former and present Hopfner lab, especially Alex, Diana, Nora, Alfred and Christian for help and advice. I had a wonderful time both during work and outside the lab – from Amsterdam to Vienna, from sport events and holidays to regular visits to the beer garden.

I want to thank my practical and bachelor students Marianne, Alessandro and Christian for their help in the lab.

In addition I would like to thank all the other groups and the technical staff of the Gene Center for the nice atmosphere, for helpful advice and frequent exchange of knowledge and material.

I am grateful for the discussions with Prof. Dr. Jens Michaelis and his help with data analysis. Scientific discussions with my collaborator Dr. Gwangrog Lee and the possibility to visit his lab was extremely stimulating.

Thanks to Aaron for help with the mass spec analysis and for being an adopted Happy Hopfner.

Without the help of my sister Marianne I could not have done the MATLAB analysis. Thank you for answering all my questions, whether you had time or not.

Many thanks to all the beamline staffs at the SLS, the ESRF and the DESY. I am specially indebted to Dr. Greg Hura and Dr. Michal Hammel from the SIBYLS beamline at the ALS for their patience in introducing me to the SAXS technique.

Thanks also to the Söding group and Achim who supplied me with a quiet place for writing.

Finally I would like to thank my parents and Wolfgang for permanent support, patience and advice.