# Structural basis of protein-nucleic acid interactions

*Group leader:* *Dietrich Suck*
*Postdoctoral fellows:* *Eric Ennifar, Claude Sauter**
*PhD students:* *Christian Biertümpfel, Hannes Simader*, Stéphane Thore*
*Technician:* *Jérôme Basquin*
*Visitors:* *Lynda Souvannavong Bilal Bham, Julie De Azevedo*
*Indicates part of the year only

We combine X-ray crystallography with various biochemical techniques to study protein-nucleic acid interactions. We are interested in the structural basis of the recognition process and the catalytic mechanism of nucleases/recombinases and proteins binding or modifying RNA. During the year we have continued our studies on structure-selective nucleases focussing in particular on Holliday junction resolvases. Another major focus of our research has been on Sm-related proteins from Archaea and bacteria.

## Structure-selective nucleases

Structure-selective nucleases recognise particular structural features of their substrates while showing little sequence-specificity (Suck, 1997). The level of their structural specificity ranges from simple discrimination between single- and double-stranded substrates to the recognition of specialized structures like flap DNA or Holliday junctions. Among the nucleases we have studied are the single-strand specific fungal enzymes P1 and S1 (Romier *et al*., 1998); the double-strand specific DNase I (Suck, 1997); T5 5' nuclease exhibiting a pronounced selectivity for flap DNA (Ceska *et al*., 1996); the λ integrase-type recombinase Cre, which can be considered a structure-specific nuclease, since a Holliday junction is a reaction intermediate and also a substrate of the enzyme (Suck *et al*., 1999); and the junction-resolving enzymes T4 Endonuclease VII (Raaijmakers *et al*., 1999; 2001) and *A.fulgidus* Hjc.

## Holliday junction resolvases: T4 endonuclease VII, *Archaeoglobus fulgidus* and SIRV2 Hjc

*(Christian Biertümpfel, Eric Ennifar, Jerome Basquin, Claude Sauter; collaboration with Börries Kemper, University of Cologne)*

Phage T4 endonuclease VII (Endo VII), the first enzyme shown to resolve Holliday junctions, exhibits a strong, but rather broad selectivity for branched as well as mismatched DNA with almost no sequence selectivity. The crystal structure revealed a novel fold not seen in any of the known junction resolvases and a very unusual domain-swapped dimer architecture, which stabilizes the open fold of the monomers (Raaijmakers *et al*., 1999). Further analysis and comparison of the wild-type and mutant structures in different crystal environments showed, that Endo VII possesses unexpected intrinsic conformational flexibility, which is thought to be of functional significance for its broad substrate specificity (Raaijmakers *et al*., 2001).

By solving the X-ray structures of complexes with various 4-way DNA junctions and simple mismatched DNA duplexes we want to determine how Endo VII recognizes and cleaves such structurally widely different substrates. Several crystal forms of 4-way DNA junction-Endo VII complexes have been obtained, however, so far none of them diffracted to high resolution. By systematically changing the arm length of the junction we are trying to improve the crystal packing and thereby increase the resolution of the crystals.

In contrast to Endo VII, the archaeal *A. fulgidus* Hjc (Holliday junction cleaving) protein specifically recognizes and cleaves 4-way DNA junctions. We have obtained well diffracting (~1.8Å) trigonal crystals of the *A. fulgidus* Hjc protein (a=b=37.5Å, c=272.1Å) and we are presently trying to solve the structure by molecular replacement using the coordinates of the previously solved *S. solfataricus* or *P. furiosus* proteins. Cocrystallizations with DNA 4-way junctions are under way. Very recently we obtained first crystals of the Holliday junction resolving enzyme from an archaeal virus (SIRV2-Hjc).

## RNA binding/modifying proteins

Projects in this category include Sm-related proteins, Pop2, a component of the major yeast deadenylase complex, and Arc1, a protein necessary for the efficient aminoacylation of Met and Glu tRNA in yeast.

### *Arc1*

*(Jerome Basquin, Hannes Simader; collaboration with George Simos)*

Arc1 associates *in vivo* with the Met and Glu aminoacyl-tRNA synthetases and stabilizes their interaction with their cognate tRNAs (Deinert *et al*., 2001). The C-terminal tRNA-binding domain (TRBD) of Arc1 shares homology with the TRBD of p43, an OB-fold protein that is part of the multi-synthetase complex of higher eukaryotes. Aim of the project is to study the role of Arc1 in the conformational stabilization of the interactions in this multi-component complex. Crystallization experiments of the ternary complex of Arc1 and the two aminoacyl-tRNA synthetases are under way and will be extended to the quaternary complexes once sufficient amounts of the cognate tRNAs are available from *in vitro* transcription.

### *Pop2*

*(Stéphane Thore; collaboration with Bertrand Séraphin, CNRS, Gif sur Yvette)*

Pop2, also known as Caf1, is part of the major deadenylase complex in yeast and together with Ccr4 is required for the efficient mRNA degradation *in vivo*. Recently, a non-canonical RNase D signature sequence has been identified in Pop2 and a recombinant C-terminal domain was shown to possess poly(A)-specific 3'-5'-exonuclease activity (Daugeron *et al*., 2001). We have crystallized this domain of Pop2 ($P2_12_12_1$, a=77.7Å, b=78.6Å, c=102.2Å) and have collected a 2.8Å native data set. Heavy atom searches are under way and very recently a 3.15Å MAD data set of a seleno-methionine labeled crystal has been collected. A first electron density map based on these data is presently being interpreted.

## Sm-related proteins

Sm-related proteins are present in all three domains of life. Originally detected as components of the small nuclear ribonucleoprotein particles (snRNPs) of eukaryotic cells, Sm proteins were later also identified in Archaea and bacteria (Salgado-Garrido, 1999; Zhang *et al*., 2002). They share a common, bi-partite signature sequence, approximately 70 amino acids long, known as the Sm domain, and are associated with U-rich RNA sequences. Eukaryotic Sm and Sm-like (Lsm) proteins are involved in a variety of RNA processing events including pre-mRNA splicing, rRNA processing, mRNA degradation and tRNA maturation. In contrast, very little is known about the function of archaeal Sm proteins.

In Eukarya seven distinct Sm proteins form the core domain of the spliceosomal snRNPs and bind to the so-called Sm-site, a uridine-rich single-stranded region of the snRNA. Archaea encode a maximum of two distinct Sm-related proteins belonging to two subfamilies we refer to as Sm1 and Sm2. As part of a structural and functional analysis of Sm-related proteins we have solved the structures of several archaeal Sm complexes and have very recently solved the structure of the *E. coli* Hfq protein.

### *Archaeal Sm proteins*

*(Stéphane Thore, Jerome Basquin, Imre Törö, Claudine Mayer; collaboration with Bertrand Séraphin, CNRS, Gif sur Yvette and Christa Schleper, University of Darmstadt)*

We reported previously the structures of the *Archaeoglobus fulgidus* Sm1 (AF-Sm1) and Sm2 (AF-Sm2) proteins, and the AF-Sm1/$U_5$ complex (Törö *et al*., 2001; 2002). Recently we have determined the structures of the *Pyrococcus abyssi* protein (PA-Sm1) and its complex with a heptameric oligo(U) (Thore *et al*., 2002) (see below). These structures show that the Sm monomer fold as well as the architecture of the Sm core domain has been conserved during evolution. The

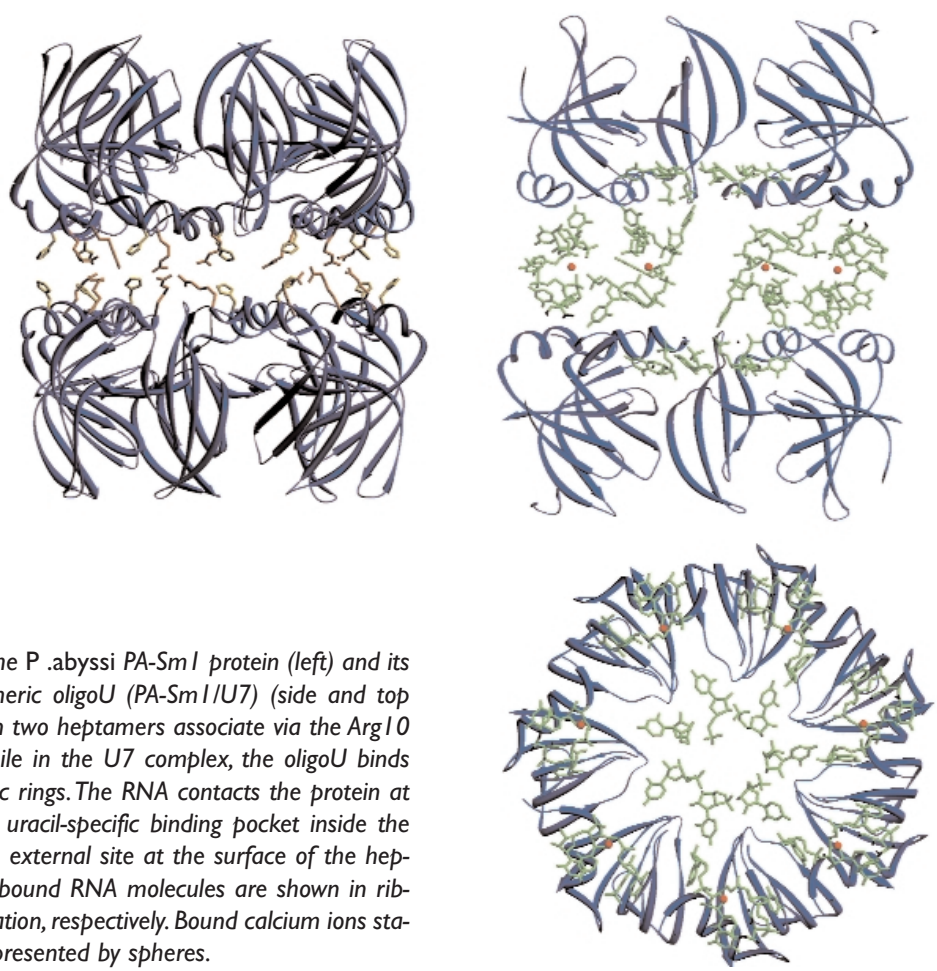

***Figure 1***. *Structure of the* P .abyssi *PA-Sm1 protein (left) and its complex with a heptameric oligoU (PA-Sm1/U7) (side and top view). In the free protein two heptamers associate via the Arg10 and His10 residues, while in the U7 complex, the oligoU binds between two heptameric rings. The RNA contacts the protein at two different sites: in a uracil-specific binding pocket inside the central cavity and at an external site at the surface of the heptamer. The protein and bound RNA molecules are shown in ribbon and stick representation, respectively. Bound calcium ions stabilizing the RNA are represented by spheres.*

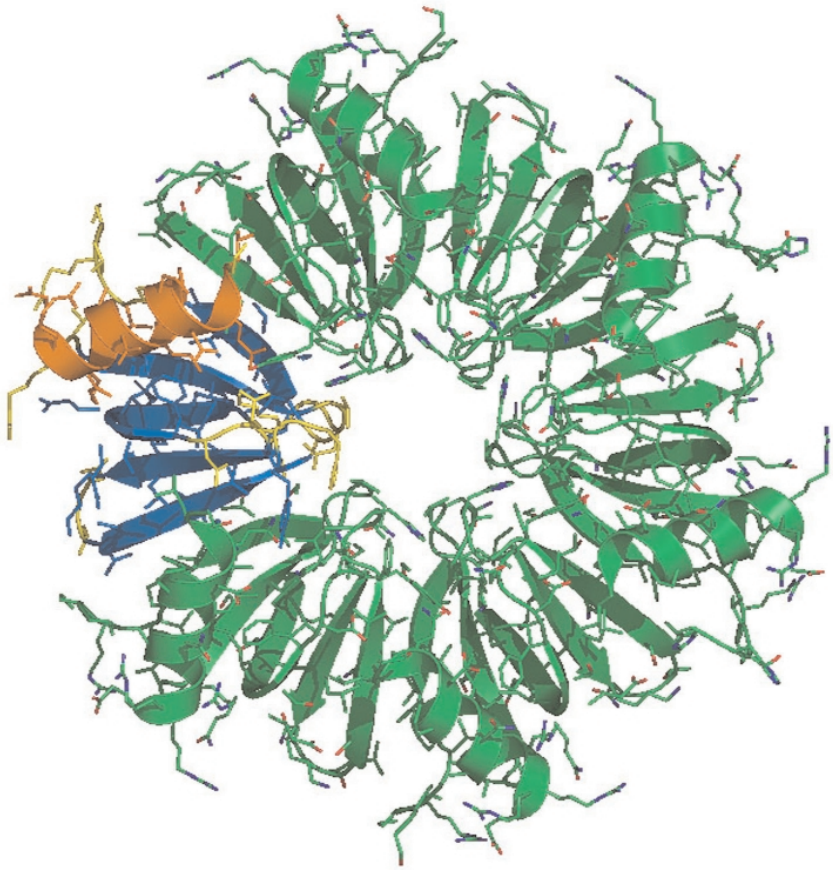

***Figure 2.*** *The structure of the E.coli Hfq (residues 1-72) hexamer.*

homo-heptameric ring-structures seen in the archaeal Sm1 proteins resemble the model proposed for the human snRNP core domain by Nagai and coworkers (Kambach *et al.*, 1999) and may therefore represent a primitive form of the Sm core domain of eukaryotic RNPs. Furthermore, a conserved uracil-binding pocket indicates a common RNA binding mode for archaeal and eukaryotic Sm proteins.

The biological significance of the hexameric AF-Sm2 complex (Törö *et al.*, 2002) is still unclear. The differences in the intersubunit contacts provide an explanation for the distinctly different olgomerization behaviour and stability of the archaeal Sm2-type proteins. The Sm1 complexes are very stable under a wide range of conditions (in this respect resembling the eukaryotic canonical Sm proteins), while the oligomerization of the AF-Sm2 protein strongly depends on the pH and the presence of RNA, more like the eukaryotic Lsm proteins.

We have shown earlier using immunoprecipitation experiments, that the AF-Sm1 and Af-Sm2 proteins from A.fulgidus both associate with RNase P RNA *in vivo* suggesting a possible role in tRNA maturation (Törö *et al.*, 2001). In order to test whether this interaction is conserved in other archaeal species and whether additional RNA species or proteins associate with the Sm proteins we have started a collaboration with Christa Schleper (U. Darmstadt) aimed at defining the *in vivo* partners of the Sm proteins from the crenarchaeon *S.solfataricus*.

## *Crystal structure of the PA-Sm1/U7 complex suggests a model for RNA binding in the eukaryotic Sm core*

We have recently solved and refined the crystal structures of the *P. abyssi* PA-Sm1 protein and its complex with a heptameric oligo(U) at 1.9 and 2.1Å resolution, respectively (Figure 1; Thore *et al.*, 2002). The overall structure of the protein complex, a heptameric ring with a central cavity, is similar to that proposed for the eukaryotic Sm core complex and found for other archaeal Sm proteins.

RNA molecules bind to the protein at two different sites. They interact inside the ring with three highly conserved residues, defining the uridine-specific binding pocket, as seen previously in the AF-Sm1/$U_5$ complex. In addition, nucleotides also interact on the surface formed by the N-terminal $\alpha$-helix as well as a conserved aromatic residue in $\beta$-strand 2 of the PA-Sm1 protein. Mutation of this conserved aromatic residue shows the importance of this second site for the discrimination between RNA sequences (Thore *et al.*, 2002).

Based on this structure as well as on the available biochemical data, we propose a model of the eukaryotic Sm core proteins bound to a 5-mer RNA representing the sequence directly upstream of the U1 snRNA Sm site. According to this model, which is consistent with cryo EM, chemical modification and cross-linking studies, the SmE protein would

serve as the binding site for the U1 snRNA leaving SmF, SmD2 and SmD1 free for unspecific interactions with the pre-mRNA. This model suggests that the Sm core represents a platform for interactions between pre-mRNA and snRNA and explains how the Sm proteins achieve their function of modulating RNA/RNA interactions.

## E. coli *Hfq*

### *(Claude Sauter, Jerome Basquin; collaboration with Toby Gibson)*

*E.coli* Hfq, also known as host factor I (HF-I) for phage Qß RNA replication is a 102 amino acid protein, which has recently been found to interact with many small RNAs (sRNA) involved in regulating mRNA translation and stability (Wassarman *et al*., 2001). It binds to A/U-rich sites in these sRNAs and thereby promotes the interaction with their mRNA targets. Structure-based sequence alignments (Sauter *et al*., unpublished results) suggested some homology with Sm proteins. EM investigations indicate that it forms a multisubunit ring structure, most likely a homo-hexamer. In addition, Hfq was shown to preferentially bind to oligoU *in vitro*. Hfq therefore has the characteristic features of eukaryotic and archaeal Sm proteins.

We first obtained tetragonal and hexagonal crystals of wild-type *E. coli* Hfq yielding data to 2.7 Å resolution at beamline ID14-1 (ESRF). These two crystal forms took several weeks or months to grow and were difficult to reproduce. Suspecting a proteolytic degradation at the C-terminus we engineered a shorter Hfq monomer based on sequence alignments and mutation-deletion data from the literature. This short form encompassing the first 72 residues of Hfq crystallized in a new hexagonal lattice (P61, a=61.3 Å, c=166.1Å) which yielded complete data at 2.15 Å. The structure was solved by molecular replacement using the structure of the *Staphylococcus aureus* homologue (Schumacher *et al*., 2002) as a search model.

The structure of the *E. coli* Hfq protein (Figure 2) is very similar to the Hfq hexamer found in *S. aureus*, as reflected by the rmsd of 1Å of the superimposed structures (based on 6x57 Ca positions in the hexamer core; Sauter *et al*., 2003). This confirms the oligomeric state described in early biochemical data and more recent EM studies, and strongly suggests that the hexameric state is a characteristic of the bacterial Hfq family. Cocrystallization experiments are under way with various *E. coli* sRNAs to unravel the mechanism by which Hfq promotes the interactions of these RNAs with their mRNA targets.

# Publications during the year


Biertumpfel, C., Basquin, J., Suck, D. & Sauter, C. (2002). Crystallization of biological macromolecules using agarose gel. Acta Crystallogr. D Biol. Crystallogr., 58, 1657-1659

Thore, S., Mayer, C., Sauter, C., Weeks, S. & Suck, D. (2003). Crystal structures of the *Pyrococcus abyssi* Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. J. Biol. Chem., 278, 1239-1247

Törö, I., Basquin, J., Teo-Dreher, H. & Suck, D. (2002). Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile *Archaeoglobus fulgidus*. J. Mol. Biol., 320, 129-142


# Other references


Ceska, T.A., Sayers, J.R., Stier, G. & Suck, D. (1996). A helical arch allowing single-stranded DNA to thread through T5 5'- exonuclease. Nature, 382, 90-93

Daugeron, M.C., Mauxion, F. & Seraphin, B. (2001). The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation. Nucleic Acids Res., 29, 2448-2455

Deinert, K., Fasiolo, F., Hurt, E.C. & Simos, G. (2001). Arc1p organizes the yeast aminoacyl-tRNA synthetase complex and stabilizes its interaction with the cognate tRNAs. J. Biol. Chem., 276, 6000-6008

Kambach, C., Walke, S., Young, R., Avis, J.M., de la Fortelle, E., Raker, V.A., Luhrmann, R., Li, J. & Nagai, K. (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell, 96, 375-387

Raaijmakers, H., Törö, I., Birkenbihl, R., Kemper, B. & Suck, D. (2001). Conformational flexibility in T4 endonuclease VII revealed by crystallography: implications for substrate binding and cleavage. J. Mol. Biol., 308, 311-323

Raaijmakers, H., Vix, O., Törö, I., Golz, S., Kemper, B. & Suck, D. (1999). Crystal structure of T4 endonuclease VII – a DNA junction resolvase with a novel fold and unusual domain-swapped dimer architecture. EMBO J. 18,1447-1458

Romier, C., Dominguez, R., Lahm, A., Dahl, O. & Suck, D. (1998). Recognition of single-stranded DNA by nuclease P1: high resolution crystal structures of complexes with substrate analogs. Proteins, 32, 414-424

Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S. & Seraphin, B. (1999). Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. EMBO J., 18, 3451-3462

Sauter, C., Basquin, J. & Suck, D. (2003). Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. Nucleic Acids Res., 31, 4091-4098

Schumacher, M.A., Pearson, R.F., Moller, T., Valentin-Hansen, P. & Brennan, R.G. (2002). Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. EMBO J., 21, 3546-3556

Suck, D. (1997). DNA recognition by structure-selective nucleases. Biopolymers, 44, 405-421

Suck, D., Buchholz, F., Dreher, M., Stewart, F. & Meyer, J.E.W. (1999). Crystal structure of a synaptic Cre recombinase - loxP complex. Biochimie, 81, s279

Törö, I., Thore, S., Mayer, C., Basquin, J., Seraphin, B. & Suck, D. (2001). RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. EMBO J., 20, 2293-2303

Wassarman, K.M., Repoila, F., Rosenow, C., Storz, G. & Gottesman, S. (2001). Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev., 15, 1637-1651

Zhang, A., Wassarman, K.M., Ortega, J., Steven, A.C. & Storz, G. (2002). The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. Mol. Cell, 9, 11-22