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

Structural studies on ASAP,

a conserved, EJC-associated complex

Andrea Giovanni Murachelli

aus Mailand, Italien

2010

Erklärung

Diese Dissertation wurde im Sinne von § 13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 (in der Fassung der vierten Änderungssatzung vom 26. November 2004) von Frau Prof. Dr. Elena Conti betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe erarbeitet.

München, 26.11.2010

Andrea Murachelli

Dissertation eingereicht am: 26.11.2010

1. Gutacher: Prof. Dr. Elena Conti

2. Gutachter: Prof. Dr. Karl Peter Hopfner

Mündliche Prüfung am: 10. Januar, 2011

Abstract:

ASAP (Apoptosis and Splicing Associated Protein complex) is a protein complex comprising three members: Acinus, SAP18 and RNPS1. In multicellular eukaryotes, ASAP is found in spliceosome purifications and is known to interact with the exon junction complex (EJC), a protein complex with diverse functions in splicing, mRNA localisation and nonsense mediated decay. These associations suggest that ASAP might play an as yet unknown role in mRNA maturation. To shed light on this issue, in this thesis I characterised the minimal core of ASAP and showed that Acinus interacts with its partners through a novel domain, the ABM (ASAP binding motif). I solved the crystallographic structure of the core ASAP and used the structural knowledge to identify ABMs in other eukaryotic proteins. Additionally, I solved two crystallographic structures of SAP18. This allowed the identification of a conserved interaction surface on the protein, which might mediate contacts with evolutionary conserved partners. Based on the structural data presented here, I suggest new avenues of research that might shed light on the cellular roles of SAP18 and ASAP.

Table of Contents

1 Introduction

1.1 RNA processing in eukaryotes

In prokaryotes, translation occurs co-transcriptionally: as soon as the newly-synthesised mRNA emerges from the RNA polymerase, it is engaged by ribosomes and decoded into proteins (Gowrishankar and Harinarayanan, 2004). In eukaryotes, the genetic material and the transcriptional apparatus reside in the nucleus and are segregated from the ribosomes; as a consequence, the RNA, be it mRNA or rRNA, needs to be exported from the nucleus before it can be used in the cytoplasm. This physical separation is thought to be linked to the evolution of eukaryotic RNA metabolism, whereby almost all primary transcripts in the cell will need to undergo extensive processing before they can be put to use. Furthermore, it is becoming clear that the processing steps involved in RNA maturation, far from being a simple, sequential cascade of events, are rather a non-linear and interwoven mesh, in which any step influences and regulates all the others (see review by Moore and Proudfoot, 2009).

1.2 Splicing: a key stage in the life of the mRNA

A key step in the life of a eukaryotic mRNA is splicing, the process during which the non-coding regions (introns) of the primary transcript (pre-mRNA) are removed, and the coding regions (exons) are joined to form the mature transcript. In all eukaryotes, splicing is performed by virtue of a complex ribonucleoprotein assembly called the spliceosome, which catalyses the two trans-esterification reactions required to join two consecutive exons into a spliced product. First, the 2′-hydroxyl group of a specific adenine contained in the intron is used as a nucleophile to attack the phosphodiester bond at the 5′ splice site (SS) (i.e., the 3′ of the first intron), thereby generating a circularised lariat structure. Second, the 3′ hydroxyl

group of the first intron becomes the nucleophile, forming a phosphodiester bond with the 5′ base of the second exon and expelling the branched intron, which is subsequently degraded (Figure 1-1).

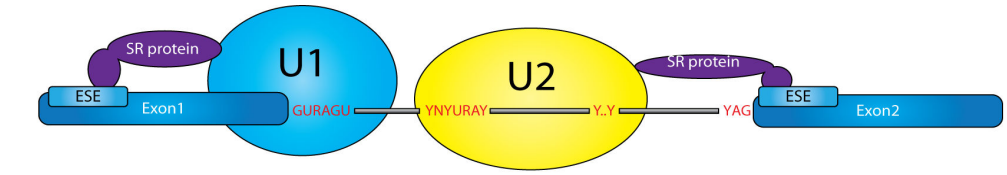
Extensive biochemical studies have shown that the spliceosome is assembled over the mRNA in a stepwise manner and undergoes several remodelling steps during the catalytic cycle, as schematised in Figure 1-1. First, the U1 and U2 spliceosomal nucleoproteins (snRNPs) recognise the 5′ SS and the branching point adenine (BPA) through specific base pairing and with the help of several SR proteins, thereby forming a complex known as the pre-spliceosome, or complex A. Second, the preformed complex formed by the U4, U5 and U6 snRNPs is recruited to form complex B, the pre-catalytic spliceosome. Through conformational rearrangements, U1 and U4 snRNPs are released, leading to complex B*. Finally, after further rearrangement, the spliceosome catalyses the first trans-esterification step, yielding complex C, followed by the second step and the release of the spliced product as well as the intron lariat and the snRNPs. The cycle is repeated for every exon-exon junction (see Wahl et al., 2009 for review).

Key to the fidelity of the splicing reaction is the ability of the spliceosome to correctly recognise the 3′ SS, the 5′ SS and the BPA. This is especially true in higher eukaryotes, in which the majority of pre-mRNAs can produce multiple protein isoforms and / or RNA products through the selective inclusion or exclusion of exons during splicing (Blencowe, 2006). The recognition of the splicing site is determined by a complex balance of counteracting influences. First, some RNA components of the spliceosome can recognise the SS and BPA consensus sequences through base pairing. Second, two classes of proteins are thought to play opposite roles in splice selection: SR-repeat-containing (SR) proteins usually act to increase the usage of a splicing site, whereas hnRNPs (heteronuclear ribonucleoparticles) generally have the opposite function, repressing splicing site usage. Both classes of proteins are thought to find their target by means of RNA-binding domains, which bind to splicing enhancers or suppressors present in exons or introns. The final exonic composition of the

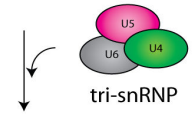
mRNA is determined by the synergistic action of multiple factors, which exert their function through various mechanisms: examples include physical blockage of the BPA (e.g. the hnRNP Sexlethal in *Drosophila melanogaster*), enhancement of 3´ SS recognition by U2 (e.g. the *Drosophila melanogaster* SR protein Tra) or enhancement of 5´SS recognition by U1 (e.g. ASF/SF2) (reviewed in Black, 2003).

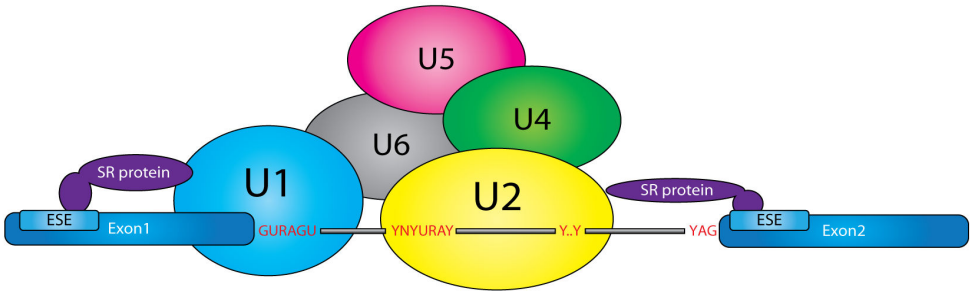
Far from being an isolated event, splicing is tightly coupled to all the other stages of life of the mRNA: for example, splicing is connected to capping and poly-adenylation and has been shown to stimulate translation by the ribosome (Nott et al., 2004; Wiegand et al., 2003). Splicing is thought to exert these far reaching influences by decorating the mRNA with several protein complexes (Fuller-Pace and Ali, 2008; Huang and Steitz, 2005; Merz et al., 2007), the best studied of which is the exon junction complex (EJC).

Figure 1-1 (next page): Schematic representation of the splicing process: A. The U1 and U2 mRNPs recognise the splicing site: U2 binds to the branching point adenine (BPA) and to the polypyrimidine tract, whereas U1 recognises the 3' splicing site (SS). SR-repeat-containing proteins favor splice site recognition by binding exonic splicing enhancer (ESE) sequences on exons through their RRM motif, and by contacting the snRNPs through the SR repeats. The vertebrate consensus sequence for SSs and BPA is indicated. Sequence code: Y = pyrimidine; R = purine; N= any nucleotide. **B.** The tri-snRNP, composed by U4, U5, U6 and associated proteins is recruited to the pre-mRNA to form complex B. **C.** Through an ATP dependent conformational rearrangement, U1 and U4 are released and the spliceosome is ready for catalysis. **D.** The first trans-esterification step produces an intron lariat, in which the BPA forms a 2' phosphodiester bond with the 5' guanosine of the 5' SS. The 3' exon remains associated with the spliceosome and is left with a free 3' OH. According to the current consensus, the EJC core, with the exception of Barentsz, is already present on exon 1 by this step of splicing (see section 1-3). **E.** In the second trans-esterification step, the 3' OH of exon 1 is ligated to the 5' phosphate of the first base of the second exon. The spliced mRNA dissociates from the spliceosome, which can be recycled. The lariat is subsequently degraded.

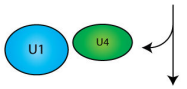


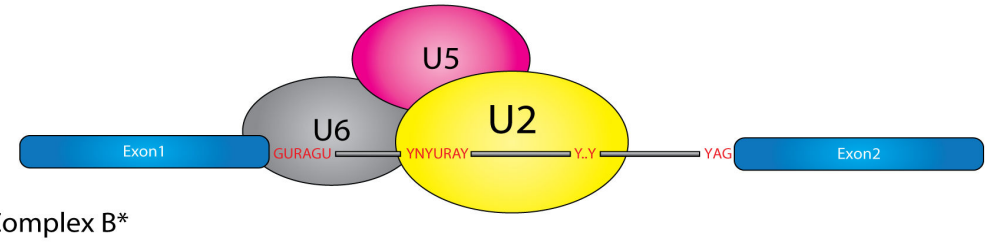
Complex A



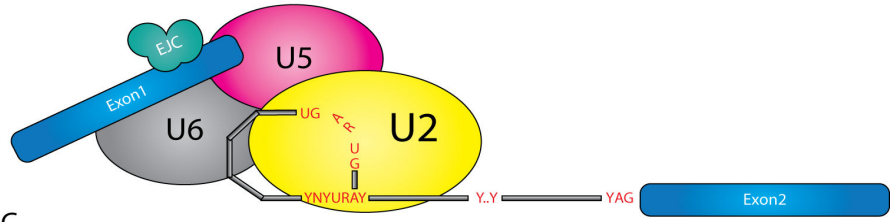


Complex B

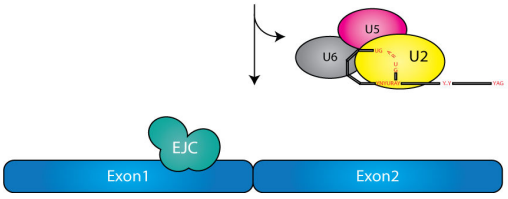




Complex B*



Complex C



1.3 The EJC, a versatile marker of RNA processing

The EJC is deposited 20 to 24 nucleotides 5′ of exon-exon junctions (Le Hir et al., 2000). It comprises a core scaffold, formed by EIF4AIII, Barentsz/MLN51, Magoh and Y14, that is deposited on the pre-mRNA during splicing and which remains associated with the mRNA until translation (Le Hir et al., 2001; Le Hir et al., 2000; Tange et al., 2005). The core EJC is assembled on the mRNA stepwise, with EIF4AIII being found already in the spliceosome A complex and Mago/Y14 being most likely deposited by the activated spliceosome C complex. Barentsz/MLN51 can only be detected on mature mRNAs, suggesting that it is added to the EJC only in the last splicing step or immediately prior to export (Gehring et al., 2009).

Several proteins or protein complexes bind onto the core EJC, the number and identity of which changes during the life of the mRNA (Tange et al., 2005). It is currently thought that by reading or specifically altering the composition of the EJC on the mRNA, the cell can determine at which stage of maturation the mRNA is and what its subsequent fate should be.

The EJC fulfils different roles among eukaryotes: in vertebrates and plants the EJC has been implicated in nonsense mediated decay, a quality control process whereby mRNAs harbouring premature termination codons (PTCs) are recognised and degraded by cellular ribonucleases. NMD functions in preventing aberrant mRNAs from giving rise to potentially dangerous truncated protein products. Additionally, it is thought to be a physiological means of controlling the expression levels of endogenous transcripts (Chang et al., 2007b). In mammals, normal termination codons usually reside in the last exon of an mRNA. Any stop codon that resides more than 55 nucleotides 5′ to an EJC will be recognised as premature and elicit degradation of the carrier mRNA (Le Hir et al., 2001; Le Hir et al., 2000).

In *Drosophila melanogaster* (*D.m.*), the EJC is involved in sub-cellular mRNA localisation of the *oskar* mRNA, which is normally deposited at the posterior pole of the developing embryo. This localisation is dependent on the presence of the EJC on the mRNA. The mechanisms behind this function are still being elucidated but all components of the core EJC are necessary for this process (Hachet and Ephrussi, 2001; Hachet and Ephrussi, 2004). Intriguingly, the EJC might also be involved in mRNA localisation in mammals, where quiescent, EJC-decorated mRNAs are found in dendritic granules (Giorgi et al., 2007).

1.4 ASAP, a conserved complex of unknown function

Among the complexes that associate with the core EJC is the apoptosis and splicing associated protein complex (ASAP), which comprises SAP18, RNPS1 and Acinus (Schwerk et al., 2003; Tange et al., 2005).

SAP18 (Sin3 Associated Protein of 18 KDa) is a ubiquitin-like (McCallum et al., 2006), nuclear protein that was originally identified as a component of the mouse Sin3 histone deacetylase complex (HDAC) (Zhang et al., 1997). Sin3-HDAC is a conserved eukaryotic complex that functions in gene repression by deacetylating histones H3 and H4 (Grzenda et al., 2009). In vertebrates SAP18 has been shown to bind directly to Sin3, the main scaffold onto which the HDAC complex assembles (Zhang et al., 1997). SAP18 has also been shown to bind directly to transcription factors such as Bicoid (Zhu et al., 2001), Krüppel (Matyash et al., 2009) and GAGA (Espinás et al., 2000) in *D.m.*, Suppressor of Fused (Cheng and Bishop, 2002), Mohawk (Anderson et al., 2009), and Hairy1 (Sheeba et al., 2007) in vertebrates, and ERF4/5 (Song and Galbraith, 2006), SOC1 (Liu et al., 2009) and AGL24 (Liu et al., 2009) in *Arabidopsis thaliana (A.t.)*. In most cases, it was also shown that the presence of SAP18 and/or the recruitment of the HDAC complex are necessary for these transcription factors to repress their gene targets. These data have lead to a model

whereby SAP18 is the means by which transcription factors can recruit the HDAC complex and repress their target genes. It should be noted, however, that Sin3 is also able to bind directly to transcription factors such as Mohawk (Anderson et al., 2009), implying that SAP18 might merely facilitate the interaction between the HDAC and the transcription factors. SAP18 has also been reported to interact with *D.m.* E (Z) (Wang et al., 2002a), a member of the polycomb repressor complex 2 (PRC2), which mediates long term repression of homeobox genes through trimethylation of lysine 27 on histone H3 (Muller and Kassis, 2006). This interaction is probably transient as SAP18 cannot be biochemically purified along with the PRC2 (Nekrasov et al., 2007). However, it hints at a possible role of SAP18 in mediating the action of different complexes that act in transcriptional repression. A very recent report also implicates SAP18 in splicing regulation: artificial tethering of SAP18 can induce alternative splicing of reporter pre-mRNA substrates (Singh et al., 2010). Finally, SAP18 has been shown to bind directly to *D.m.* Pnn (Costa et al., 2006), a component of the EJC and the spliceosome. The mouse homolog of Pnn, Pinin, is essential for embryonic development (Joo et al., 2007; Joo et al., 2010), probably because of its involvement in the regulation of the Wnt signal transduction pathway (Alpatov et al., 2008). Other reports also suggest that Pinin can modulate splicing and mRNA export through its interaction with RNPS1 (see below) (Li et al., 2003; Wang et al., 2002b).

RNPS1 (RNA binding protein S1) is a member of the wide class of SR proteins that cooperate with the U1 and U2 snRNPs in splicing site recognition. RNPS1 has a S-rich region at its N-terminus, followed by an RRM (RNA Recognition Motif) domain and by a C-terminal region that contains SR repeats (Figure 1-2). RNPS1 was originally isolated as a general enhancer of splicing with a potential role in selecting alternative 3′ splice sites (Mayeda et al., 1999). Subsequent studies have confirmed that RNPS1 enhances the formation of the spliceosomal A complex *in vitro* and this activity is dependent on the phosphorylation of Ser 53 by casein kinase 2 (Trembley et al., 2005). Furthermore, it has been shown that the modulation of alternative splicing by RNPS1 is dependent on its RRM and RS domains

(Sakashita et al., 2004). At least in vertebrates, RNPS1 is also an NMD factor: when artificially tethered to an mRNA downstream of a stop codon, it can induce the degradation of the messenger (Lykke-Andersen et al., 2001), similarly to other *bona fide* NMD factors. The NMD functions of RNPS1 are carried out in concert with the EJC, to which RNPS1 presumably associates during splicing (Gehring et al., 2009; Sakashita et al., 2004) and with which RNPS1 shuttles to the cytoplasm (Lykke-Andersen et al., 2001). In order to induce NMD, the S-rich domain of RNPS1 must be intact (Gehring et al., 2005). Interestingly, like SAP18, RNPS1 has been shown to directly interact with *D.m.* Pnn (Li et al., 2003; Sakashita et al., 2004); the interaction occurs between the RRM domain of RNPS1 and the conserved coiled coil domain of Pinin.

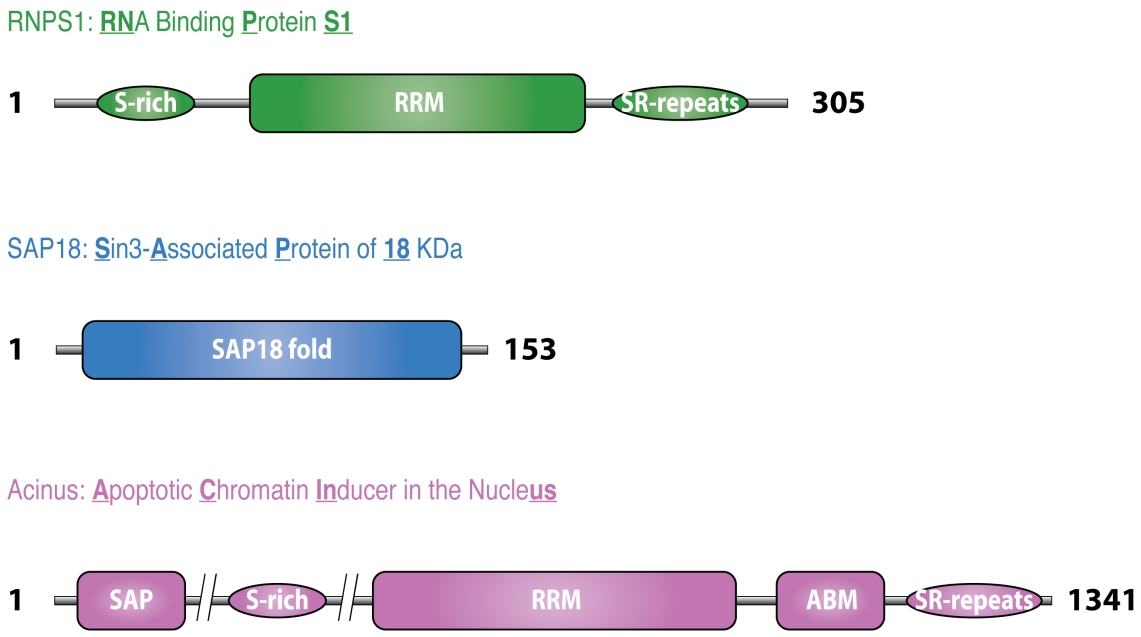


Figure 1-2: The domain organisation of RNPS1, SAP18 and Acinus. The domain organisation of the components of ASAP is depicted here and is derived from Pfam (Finn et al., 2010). The ABM (ASAP Binding Motif) is not recognised by any of the domain databases and is described for the first time in this work. Rectangles indicate folded domains, ellipses represent motif that are thought to be unstructured. To preserve the relative scale of the domains, two non-conserved and disordered stretches on Acinus have been omitted (//).

Acinus (Apoptotic chromatin inducer in the nucleus) is 1341 amino acids long in humans. It has the typical domain organisation of a splicing protein with an S-rich domain, an RRM domain and a C-terminal series of SR repeats. In addition, it contains a conserved SAP (SAF-A/B, Acinus and PIAS) domain, thought to be involved in DNA binding, and a conserved C-terminal motif of unknown function that, based on the results of the present work, I named ABM, for ASAP binding motif (Figure 1-2). Human Acinus is alternatively spliced in three isoforms: the reference isoform Acinus L (Large) and two smaller version, Acinus S (Small) and Acinus S', which share residues 770-1341 with Acinus L but possess two distinct short exons at their N-termini. All three isoforms are capable of associationg with RNPS1 and SAP18 to form the ASAP complex and their functional differences are unknown (Sahara et al., 1999). During apoptosis Acinus is targeted by caspase 3, which cleaves it close to the boundaries of the RRM domain. The cleaved product is necessary for chromatin condensation, a hallmark feature of apoptosis (Hu et al., 2005; Joselin et al., 2006; Sahara et al., 1999). In a similar process, caspase 3 cleaves Acinus during terminal erythropoiesis, probably in preparation for terminal enucleation (Nagata, 2005; Zermati et al., 2001). Recent reports implicate Acinus in transcription control, binding to retinoic acid receptors and regulating the expression levels of various genes, including the cell cycle regulator cyclin A1 (Jang et al., 2008). The same reports suggest that the regulatory activity of Acinus on transcription depends on the HDAC complex – hinting at the possible involvement of SAP18 and of the ASAP complex. The role of Acinus in RNA metabolism is still uncharacterised. Its domain organisation, typical of SR proteins, suggests a role in splicing. Indeed, Acinus L associates with spliceosomes (Merz et al., 2007) and seems to weakly stimulate splicing *in vitro* (Schwerk et al., 2003). Acinus is also found in purifications of the EJC but is thought to dissociate from it immediately prior to the export of the RNA (Tange et al., 2005). Finally, a recent report also shows that dosage disregulation of the *D.m.* ortholog of Acinus, (hook-like, hkl) perturbs the EGFR, Notch and autophagocytosis pathways in the eye; this regulation could be an indirect effect and its mechanisms are unknown (Haberman et al., 2010).

Despite the wealth of information available about each single component of ASAP, the complex itself has not been extensively studied since its isolation was reported. ASAP has a weak inhibitory effect on *in vitro* splicing; furthermore, although its over expression cannot induce apoptosis (similarly to Acinus alone), increased levels of ASAP can accelerate staurosporine-induced cell death (Schwerk et al., 2003).

1.5 Aims and scope of the project

The high sequence conservation of ASAP in eukaryotes implies an important role for this complex in the life of the cell. Yet little is known as to what this role might be.

In this PhD thesis I have identified a minimal, core ASAP complex, comprised by SAP18, the RRM domain of RNPS1 and the ABM motif of Acinus. I have solved the crystallographic structure of this core complex, showing how RNPS1 and Acinus associate in a subcomplex, the association of which is necessary for the binding of SAP18. I have then searched for ABM motifs in other proteins, and I have found one in Pinin, an essential transcriptional regulator and spliceosomal component. Additionally, I have solved the structure of two distinct SAP18 constructs. Analysis of these structures shows that SAP18 possesses a conserved, hydrophobic groove, which in the crystal packing accommodates the N-terminus of a symmetry-related molecule. The interaction between the N-terminus and the groove is reminiscent of the interactions between several classes of ubiquitin-like molecules and their ligands, suggesting that the groove on SAP18 might be involved in the interactions of the protein with unknown partners.

The experimental results presented in this thesis are divided into three sections: in the first chapter, the phylogenetic distribution and the conservation of the ASAP complex are examined; in the second section, two crystallographic structures of SAP18 are presented;

the third and final part deals with the characterisation of a minimal ASAP core and the solution of its structure.

The work in the present thesis was carried out in the laboratory of Prof. Dr. Elena Conti, first at the European Molecular Biology Laboratory in Heidelberg (Jan 2007 – Aug 2007) and subsequently at the Max Planck Institute of Biochemistry in Martinsried (Sep 2007 – present). The search for ABM sequences in eukaryotic proteins and for possible SAP18 interactors were performed in collaboration with Dr. Johannes Söding and Dr. Echkhart Guthörlein from the Gene Center in Munich. The final, preparative limited proteolysis experiment and the crystallisation of the core ASAP were performed by Judith Ebert. Preliminary analysis of the binding of ASAP to the EJC has been carried out in collaboration with Dr. Herve' Le Hir and Dr. Lionel Ballut.

2 Results:

2.1 ASAP is conserved in the eukaryotic domain of life.

A central prediction of the evolutionary theory is that in a protein, functionally important residues are phylogenetically conserved. Therefore, as a first step towards elucidating the function of ASAP, I surveyed the distribution of its components in the tree of life and prepared multiple sequence alignments to identify conserved portions of their sequences.

Based on the analysis of sequenced genomes, several phylogenetic classifications of Eukaryotes have been proposed in the last years. According to the prevalent models, 6 kingdoms are present in eukaryotes: Animals, Fungi, Amoebozoa, Plants, Chromalveolates (which include Rhizaria), and the poorly defined Excavates. It is currently accepted that Animals, Fungi and Amoebozoa form a sister group, named Unikonts, whereas Plants and Chromalveolates form another, named Bikonts. However, no consensus has been reached as to the position of Excavata and the root of the eukaryotic tree. One recent model postulates that the eukaryotic tree is rooted between Bikonts on one side and Unikonts and Excavata on the other side; A second and more recent model suggests instead that Bikonts and Unikonts are sister groups, and equally related to Excavata (Cavalier-Smith, 2010a; Stechmann and Cavalier-Smith, 2003) (Figure 2-4). Both models are compatible with the existing data, and currently subject of considerable scientific debate.

2.1.1 Phylogenetic conservation of the ASAP components

SAP18 orthologs can be readily identified in Plants, Animals and Fungi through a BLAST (Altschul et al., 1990) search. The level of conservation is high, with 23% and 53% identity between mouse and *Neurospora Crassa* (*N.c.*) and *Arabidopsis thaliana* (*A.t.*). SAP18 respectively. Notably, no SAP18 homolog is present in *Saccharomyces cerevisiae* (*S.c.*). Using PSI-BLAST (Altschul et al., 1997) or CS-BLAST (Biegert and Soding, 2009), more distant SAP18 homologs can be detected also in Chromoalveolates. In Excavata, a SAP18 homolog can be found in the sequenced genome of *Trichomonas vaginalis* (*T.v.*), but not in those of *Trypanosoma brucei* and *Giardia Lambia*. Since SAP18 is a highly conserved, single-domain protein, this is likely to reflect a *bona fide* absence rather than a false negative. It is therefore unclear whether SAP18 has been acquired by *Trichomonas* through lateral gene transfer, which is common in parasites (Keeling and Palmer, 2008), or whether it has been independently lost in all other Excavates (also a common occurrence in parasites). Figure 2-1 shows a multiple sequence alignment of eukaryotic orthologs of SAP18.

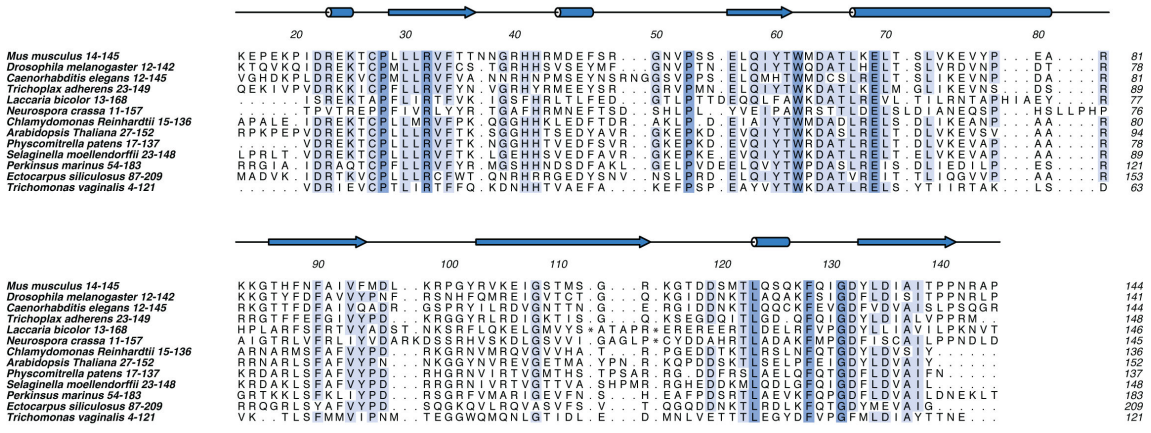


Figure 2-1: Multiple sequence alignment of SAP18 orthologs. [Continued on next page]

RNPS1 orthologs in vertebrates and plants are also readily identified by a BLAST search. Despite considerable sequence divergence, RNPS1 orthologs can also be confidently identified in Fungi and Chromalveolates. In all phyla, the overall domain organisation of RNSP1 is conserved, with a N-terminal S-rich region, a central RRM and a C-terminal SR repeat domain (Figure 1-2 and Figure 2-2). Interestingly, a protein exists in *T.v.,* which contains an RRM with the signature of RNPS1, but no S-rich region or RS-repeats (Figure 2-2). No homolog of RNPS1 is present in *Saccharomyces cerevisiae.*

Bona fide homologs of Acinus are present in Plants and Metazoans: all of these proteins share (1) the same domain architecture, with a N-terminal SAP domain, a middle RRM domain and a C-terminal ABM domain (Figure 1-2) and (2) where available, a similar reported pattern of splicing, with L and S isoforms (where the L isoforms contains the SAP domain). Partial exception to the rule is *D.m.* Acinus (also known as hook-like, hkl), which does not possess a SAP domain. Clear-cut Acinus L homologs can also be found in the chromoalveolate *Ectocarpus Siliculus*, and in the fungi *Aspergillus clavatus* and *Tuber melanosporum* (black truffle) (Figure 2-3). No homologs can be found in Excavates (including *T.v.*) or *Saccharomyces cerevisiae*.

[Continued from previous page] SAP18 orthologs from selected organisms have been identified using CS-BLAST and aligned with MAFFT (Katoh et al., 2005) using the Jalview alignment analysis suite (Waterhouse et al., 2009). For comparison, the SAP18 ortholog found in *T.v.* is also shown. The experimentally determined structure of mouse SAP18 is shown above the sequence. The N- and C-termini, which are not conserved, were omitted for clarity. Extensive insertions present only in one sequence of the ensemble have been removed to improve legibility and this position is marked with the symbol "*". The figure was prepared using ALINE (Bond and Schüttelkopf, 2009) and Adobe Illustrator. Color code: dark blue = 66-100% identity; light blue: 33-66% identity. Phylogenetic distribution: *Mus musculus, Drosophila melanogaster, Caenorhabditis elegans, Trichoplax adhaerens*: phylum Animalia; *Laccaria bicolor, Neurospora crassa*: phylum Fungi; *Arabidopsis thaliana, Physcomitrella patens, Selaginella moellendorfii, Chlamidomonas reinhardtii*: phylum Planta; *Perkinsus marinus, Ectocarpus siliculosus*: phylum Chromoalveolata; *Trichomonas vaginalis*: phylum Excavata. Accession codes, from top to bottom: NP_033145.2; Q9VEX9.1; NP_497833.1; XP_002112892.1; XP_001876400; XP_956672.1; XP_001695133; NP_566050.1; XP_001764568; EFJ15544; XP_002785750.1; CBJ29313.1; XP_001582301.1.

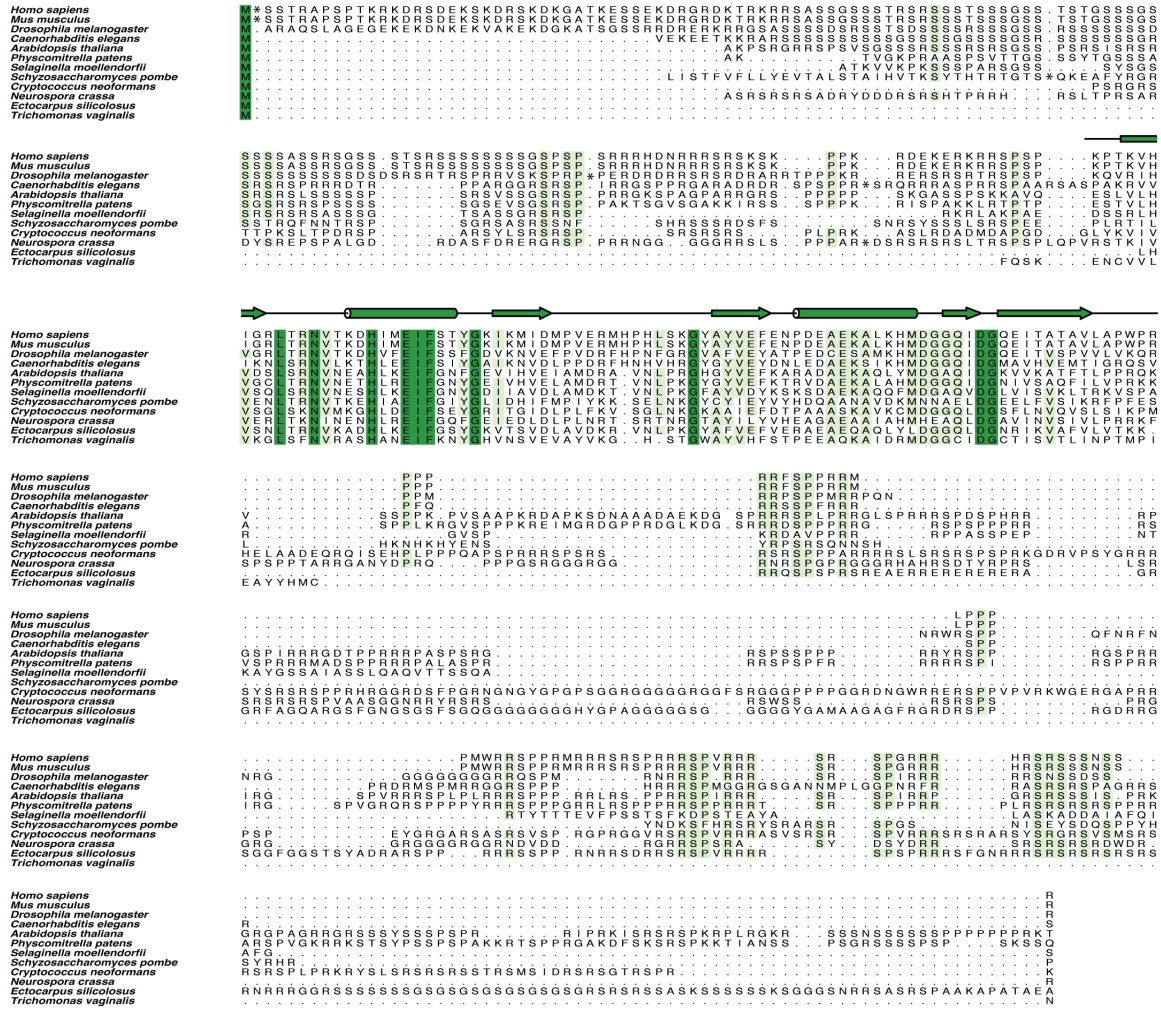


Figure 2-2: Multiple sequence alignment of RNPS1 orthologs. RNPS1 orthologs from selected organisms have been identified through CS-BLAST and aligned with MUSCLE (Edgar, 2004) using the Jalview alignment analysis suite. Sequence conservation outside the RRM domain is poor; however, the N-terminal S-rich region and the C-terminal SR repeats can be clearly distinguished in all the sequences. Note that in *Ectocarpus siliculosus* the S-rich domain is in the C-terminus, after the SR repeats region. Due to the flexibility of both regions, and due to the fact that within the ASAP complex they should be adjacent, (see the position of the C- and N-terminus in Figure 2-28), this translocation is not expected to result in functional differences. The experimentally determined structure of the RRM domain of human RNPS1 is shown above the sequence. Extensive insertions present only in one sequence of the ensemble were removed to improve legibility and their position is marked with the symbol "*". The figure was prepared using ALINE (Bond and Schüttelkopf, 2009) and Adobe Illustrator. Color code: dark green = 66-100% identity; light green: 33-66% identity. Phylogenetic distribution: *Cryptococcus neoformans*: phylum Fungi; for the other organisms, see the legend to Figure 2-1. Accession codes, from top to bottom: Q15287; NP_649903.1; NP_497276.2; NP_173107.1; XP_001759034.1; XP_569824.1; EFJ13110.1; XP_958016.1; CBJ26878.1; XP_001317742.1.

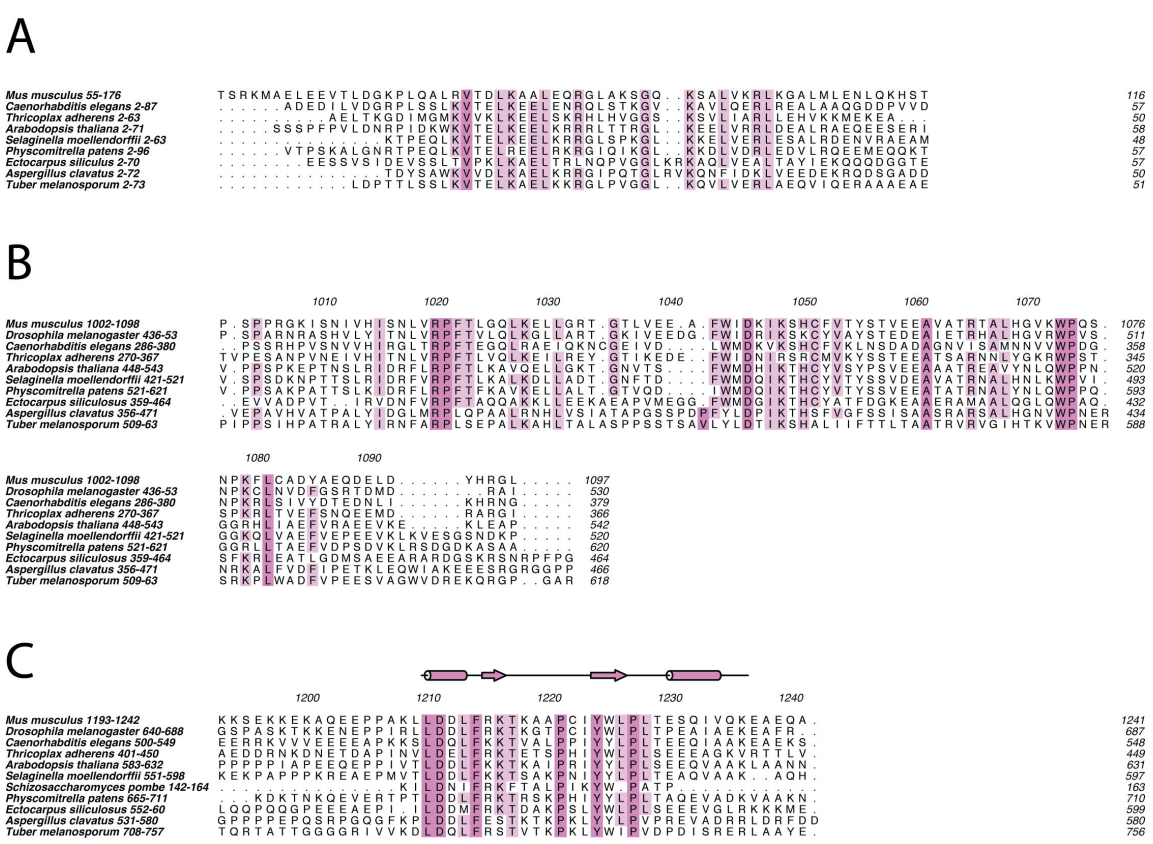


Figure 2-3: Multiple sequence alignment of Acinus orthologs. Acinus orthologs from selected organisms have been retrieved through a CS-BLAST search and aligned with MAFFT (Katoh et al., 2005) using the Jalview alignment analysis suite (Waterhouse et al., 2009). For clarity, only selected parts of the alignment are shown. **A.** Multiple sequence alignment of the SAP domain. For reasons unknown, the domain is absent in *Drosophila melanogaster* and *Schizosaccharomyces pombe* Acinus, which have been removed from the alignment. **B.** RRM domain. This domain is not present in *Schizosaccharomyces pombe* Acinus, in which it seems to have been substituted with a domain that is similar to a portion of glutamate tRNA syntethases. The functional relevance for this substitution is unknown. **C.** ABM domain. The experimentally determined structure of this domain is shown above the sequence. The figure was prepared using ALINE and Adobe Illustrator. Colour code: dark purple = 66-100% identity; pink: 33-66% identity. Phylogenetic distribution: *Aspergillus clavatus*, *Tuber melanosporum*, *Schizosaccharomyces pombe*: phylum Fungi. For the other organisms, see legend to Figure 2-1. Accession codes, from top to bottom of the ABM alignment: CAX15414.1; NP_609935.1; NP_491344.1; XP_002111598.1; NP_195678.1; EFJ23297.1; NP_595942.1; XP_001766987.1; CBJ25459.1; XP_001274049.1; XP_002836771.1.

2.1.2 The phylogenetic distribution of ASAP suggests a role in splicing

All the components of ASAP are present in Unikonts and Bikonts, implying that they were inherited from the concestor (last common ancestor) of these groups. Additionally, Excavates seem to have lost all or some of the ASAP components. Three hypotheses are compatible with this phylogenetic distribution: irrespective of the true rooting of the eukaryotic tree, ASAP could have been present in the last eukaryotic common ancestor (LECA) and subsequently partially or completely lost in Excavate lineages (Figure 2-4 A and B). An alternative, less parsimonious hypothesis is that, if Unikonts and Bikonts form a sister group separated from Excavata, ASAP could have evolved after the divergence of the latter lineage. SAP18 and RNPS1 would have been subsequently transferred laterally to *Trichomonas* (Figure 2-4 C).

The LECA was already competent at splicing and NMD, as demonstrated by the fact that these processes are present and mechanistically conserved in all eukaryotic phyla (Izquierdo and Valcárcel, 2006). Interestingly, Excavates prefer *trans-* to *cis-*splicing (Liang et al., 2003); additionally, both Excavates and *Saccharomyces cerevisiae* have intron-poor genomes (Anantharaman et al., 2007). Altogether, these data could suggest an evolutionary relationship between splicing and the maintenance of ASAP that would hint at a role for the latter in the former process.

Phylogenetic conservation of a complex or a regulatory circuit does not necessarily imply functional conservation (Shubin et al., 2009); however, given the fundamental role of the spliceosome and NMD, to which ASAP is most probably associated, it stands to reason that the functions of ASAP should be mostly conserved within eukaryotes. As a consequence, any insight on ASAP gained in simple model organisms, such as *Schyzosaccharomyces pombe* or *Caenorhabditis elegans*, should be generalisable to higher eukaryotes such as humans.

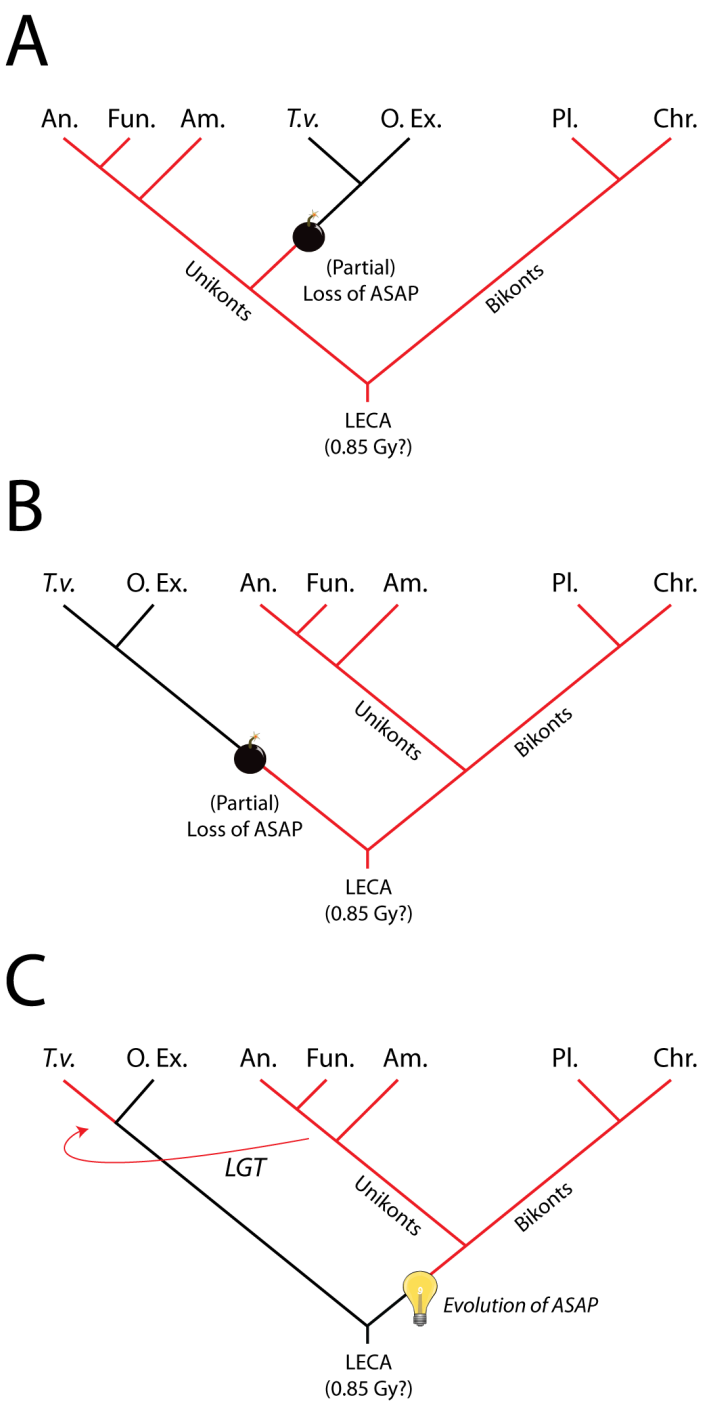


Figure 2-4: Hypotheses on the evolutionary origin of ASAP. A. In the figure a simplified model of the eukaryotic phylogenetic tree is depicted. If the root of the eukaryotic tree is placed between Unikonts and Bikonts + Excavata, Animals and Plants have the last eukaryotic common ancestor (LECA) as their concestor. Therefore, ASAP must have been present in the LECA and has been independently lost in Excavata. **B,C.** Two alternative hypotheses arise if the root of the eukaryotic phylogenetic tree lies between Excavata and the sister group Unikonts + Bikonts: either ASAP was present in the LECA and has been lost in Excavata except *Trichomonas* (B) or ASAP arose in the eukaryotic lineage after the divergence of Excavata, with RNPS1 and SAP18 being laterally transferred to *Trichomonas* (C). The date indicated for the emergence of the LECA is derived from Cavalier-Smith, 2010b and is currently considered speculative. The length of the tree branches is arbitrary and not related to evolutionary distance. Red tree branches in each model indicate presence of ASAP. Abbreviations: LECA: last eukaryotic common ancestor; Gy: billions of years; LGT: lateral gene transfer; An.: animals; Pl.: Plants; Am.: Amoebozoa; Fun.: Fungi; Chr.: Chromalveolates; O.Ex.: Other Excavates; *T.v.*: *Trichomonas Vaginalis*

2.2 Structural studies on SAP18

The structure of SAP18 had been determined previously by NMR (McCallum et al., 2006). Nevertheless, I solved the crystallographic structure of SAP18 in order to obtain a search model suitable for phasing the structure of ASAP by molecular replacement.

2.2.1 Crystallisation of SAP18$^{1-143}$ and SAP18$^{6-143}$

Full length SAP18 is prone to partial degradation, at both the N- and C-terminus, and is therefore not suitable for crystallisation. Indeed, as evidenced by the NMR structure by McCallum *et al.* (McCallum et al., 2006), both portions of the protein are flexible and presumably accessible to proteases. Consequently, on the basis of limited proteolysis and multiple sequence alignments (Chapter 2.3 and Figure 2-1), I designed two constructs for crystallisation: SAP18$^{1-143}$ and SAP18$^{6-143}$. Both constructs were purified according to the protocol detailed in Materials and Methods and set up for crystallisation at the concentration of 45 mg/ml. Both SAP18 constructs yielded crystals within 2-4 weeks in a many crystallisation conditions. Most crystals grew between pH 6.0 and 7.0 using medium weight polyethylene glycol (PEG) as a precipitant (PEG 2000 – 6000). Lithium and potassium were often present as cations. Crystals were also found in conditions containing 2-methyl-2,4-pentanediol (MPD) or Jeffamine M600 as precipitants. After optimisation, the best crystals were obtained from SAP18$^{1-143}$ in 96 well format, mixing 200 nl of protein with 200 nl of crystallisation solution consisting 40% MPD, 0.2 M sodium cacodylate pH 6.2 and 5% PEG 8000 (Figure 2-5). The crystals were harvested with no additional cryoprotectant and a complete dataset was collected at the Swiss Light Source (SLS) beamline PX II, using

10% beam intensity (2 x10^{11} photons / second), 0.7 seconds exposure and 1 degree per frame. The detector was a MarCCD, set at a distance appropriate to capture reflections up to 1.8 Å of maximum resolution. The data were indexed and integrated in spacegroup C2 with iMOSFLM (Leslie, 1992). Subsequent data processing was carried out through the CCP4 interface (Collaborative Computational Project, 1994). The intensities were scaled with Scala (Evans, 2006) to a resolution of 1.9 Å. A 5% subset of reflections was flagged as test set for R_{free} calculation. The quality of the data was assessed with phenix.xtriage. Data processing statistics are shown in Table 2-1. The Matthews coefficient (Kantardjieff and Rupp, 2003; Matthews, 1968) suggested that two SAP18 molecules were present in the asymmetric unit (ASU), with a predicted solvent content of 50%. The self rotation function (Rossmann and Blow, 1962) calculated through Polarrfn (Collaborative Computational Project, 1994) showed four peaks in the κ = 180 section, of which two were due to the crystallographic twofold axis, and two were consistent with the presence of two molecules related by a non crystallographic twofold axis in the ASU (Figure 2-6).

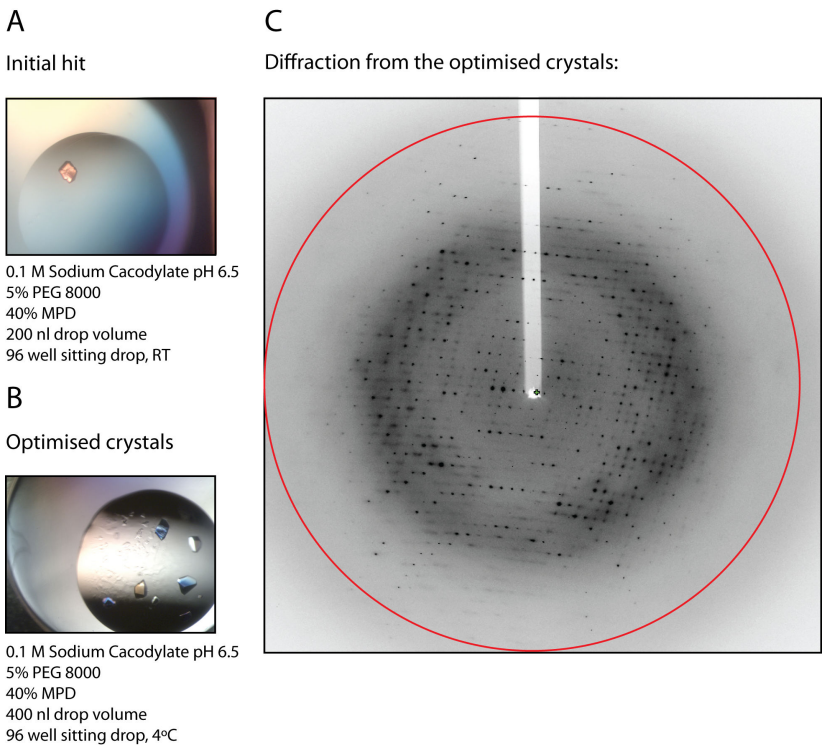


Figure 2-5: Native crystals of SAP18$^{1-143}$.

A. The initial crystal obtained in the crystallisation screen. **B**. The crystals after optimisation. **C.** Diffraction obtained from the crystals in B. The red circle shows the 1.9 Å resolution range. RT: Room temperature.

Table 2-1: Data collection statistics for SAP18$^{1-143}$. Values in parentheses correspond to the highest resolution shell.

Dataset	Native
Space group	C2
Unit cell (Å)	a = 95.42 b = 65.14 c = 61.61 α = γ = 90.00° β = 119.40°
Resolution (Å)	32.57-1.9 (2.0-1.9)
R_{meas} (%)	7.2 (40.1)
R_{pim} (%)	2.5 (14.6)
Observed (Unique)	193020 (28292)
I/σ(I)	14.4 (3.8)
Completeness (%)	98.9 (100.0)
Multiplicity	7.5 (7.5)
Wavelength (Å)	0.978

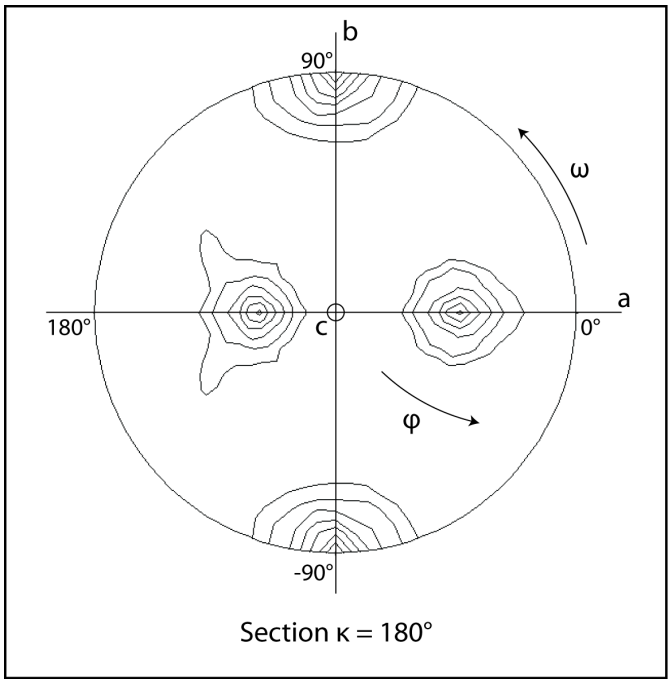


Figure 2-6: Self-rotation function calculated for SAP18$^{1-143}$. The self rotation function was calculated with Polarrfn in the CCP4 suite (Collaborative Computational Project, 1994) and is here shown for the section κ = 180°. Four peaks are clearly recognisable: the peaks at (ω = 90, φ = 90) and (ω = -90, φ = 90) are derived from the 2-fold axis of the C2 spacegroup; the two peaks at (ω = 0, φ = 67.5) and (ω = 180, φ = 157.5) are due to a two-fold, non-crystallographic symmetry (NCS) axis and indicate the presence of a dimer in the ASU.

Given the absence of anomalous scatterers in the crystal, phasing was attempted by molecular replacement, using as a search model either the NMR structure of SAP18 (2HDE) (McCallum et al., 2006) or an ensemble of crystallographic structures of ubiquitin-like molecules. In all cases, disordered loops were manually removed from the search model prior to the search and non-identical side chains were truncated using the program Chainsaw (Stein, 2008). Molecular replacement was attempted with Phaser (McCoy et al., 2007), Balbes (Long et al., 2008) and MrBUMP (Keegan and Winn, 2008), but proved unsuccessful. I therefore decided to attempt experimental phasing by selenomethionine derivatisation of SAP18.

2.2.2 Production of selenomethionine-derivatised SAP18

Selenomethionine-derivatised proteins (SeMet proteins) are usually expressed in an *E.coli* strain auxotrophic for methionine (typically B834) using a synthetic, minimal medium supplemented with SeMet. The yield of SeMet SAP18$^{1-143}$ in these conditions was however too low, about 50 to 100 fold less than the one obtained with Terrific Broth (TB, Sambrook J et al., 1989) in the standard expression conditions. In collaboration with John Weir, a fellow PhD student in the lab, we defined a new medium based on the work published by Sreenath et al., 2005. Sreenath and co-workers supplemented the commonly used M9 minimal growth medium (Sambrook J et al., 1989) with vitamins, trace minerals and sugar. This allowed them to obtain yield of SeMet proteins comparable to the ones obtained for the native expression in TB. We modified their protocol in the following ways: (1) most of the additives used by Sreenath and colleagues are present in similar amounts in the commonly used and commercially available yeast nitrogen base (YNB). Therefore, instead of adding every component separately, we added YNB powder at the same concentration commonly used for *Pichia pastoris* (3.4 g/l); pH was then adjusted to 7.5 by addition of KH_2PO_4. (2) Sreenath *et al.* supplemented their medium with glucose and galactose, with the aims of improving growth rate and yield and endowing the medium with auto inducing capabilities. Being uninterested in the auto induction, we chose fructose as a post-glucose carbon source, since it is cheaply available and allows the highest specific growth rate among common monosaccharides (Lendenmann and Egli, 1995). I will henceforth refer to the new selenomethionine derivatisation medium as LAMM (Levulose augmented minimal medium). Further details about the composition of the LAMM medium are in Materials and Methods. The use of LAMM allowed to obtain yields of SeMet SAP18 comparable to those obtained in TB for the native protein.

SeMet incorporation was 100% as judged by electrospray/ionization (ESI/TOF (Domon and Aebersold, 2006)) total mass measurements (not shown).

SeMet SAP18$^{1-143}$ and SeMet SAP18$^{6-143}$ were prepared according to the same protocol used for the purification of their native counterparts, with the only exception that the concentration of DTT used was increased from two to five mM in order to protect the Se atoms from oxidation; furthermore, 5-10 mM tris (2-carboxy-ethyl) phosphine · hydrochloride, pH 7.5 (TCEP) was added to the protein prior to crystallisation.

2.2.3 Crystallisation of SeMet SAP18

SeMet SAP18$^{1-143}$ and SAP18$^{6-143}$ were set up for crystallisation at the concentration of 30 mg/ml against sparse matrix crystallisation screens. For both protein constructs, crystals were obtained in conditions similar to those that had yielded native crystals. Additionally, crystals of SeMet SAP18$^{6-143}$ grew in presence of isopropanol or tert-butanol at slightly basic pH. All these crystallisation environments were sampled through systematic grid screens; the optimised crystals were also subjected to seeding. Eventually, the best crystals were obtained from SAP18$^{6-143}$ in 30% isopropanol and 0.1 M bicine or 0.1 M HEPES pH 8.0. These crystals were cryoprotected by stepwise addition of MPD to a final concentration of 30%, flash cooled in liquid nitrogen and tested on beamline PXII of SLS (Figure 2-7). Two complete datasets were collected from two different, isomorphous SeMet crystals: one for phasing through Single Anomalous Diffraction (SAD), and one to obtain higher resolution data (see Table 2-2 for details). The SAD dataset was collected at the selenium absorption edge (λ = 0.9802 Å), with exposure time of 0.25 seconds, flux of 8.21 x 10^{10} photons/second, 0.5 degrees oscillation per image and inverted beam setup for a total of

360 degrees in two 180 degrees wedges. The native dataset was collected at the high energy remote (λ = 0.9732 Å) with 0.25 seconds of exposure time, flux of 5.2 x 10^{11} photons/second, 0.5 degrees oscillation per image and 360 consecutive frames. Both datasets were collected using a Pilatus detector. The maximum resolution at the edge of the detector was 1.5 Å for the SAD dataset and 1.2 Å for the high-resolution dataset.

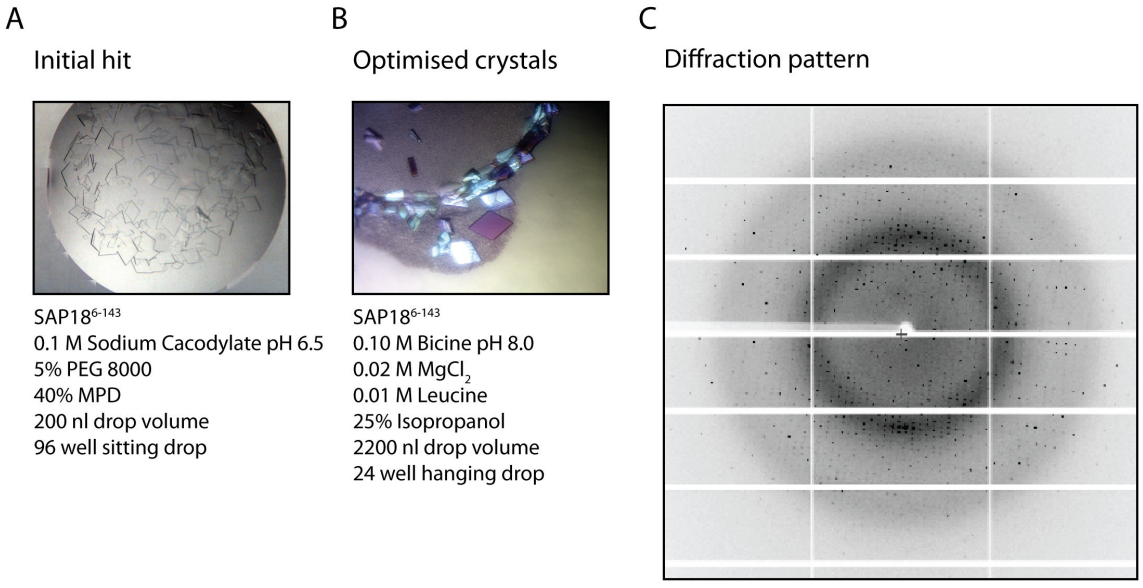


Figure 2-7: Optimisation of the SAP18$^{6-143}$ crystals. A. The initial hit. **B.** The optimised crystals that yielded the SAP18$^{6-143}$ structure. **C.** Diffraction pattern obtained from the crystals in B. The image has been cropped so that its edges correspond to a resolution of 1.7 Å on the detector.

2.2.4 Phasing of the SAP18$^{6-143}$ structure

Indexing with XDS (Kabsch, 2010) suggested a point group 222 and Bravais lattice I. Two enantiomorphic space group fulfil these requirements: I222 and I212121, which share the same general reflection conditions for non-centrosymmetric molecules and are therefore undistinguishable on the basis of intensity measurements alone (Hahn, 2005). A single copy of SAP18 is present in the ASU, with a solvent co-efficient of 40%. The integrated datasets were imported into CCP4 through Combat (Collaborative Computational Project, 1994) and scaled using Scala (Evans, 2006), keeping anomalous pairs separated for merging statistics. The same subset of 5% of the reflections was flagged in each dataset as test set for the calculation of the Rfree. The useful resolution was 1.7 Å and 1.5 Å for the SAD and high-energy remote datasets respectively (see Table 2-2 and Table 2-3). Experimental phasing was carried out with the SHELX suite of programs (Sheldrick, 2008) using the hkl2map graphical interface (Pape and Schneider, 2004). Both the enantiomorphic space groups were used for the experimental phasing, with I222 being successful. After data preparation with SHELXC, SHELXD was used to search for 6 selenium atoms in the ASU; all six anomalous scatterers could be located, with site occupancies ranging from 1 to 0.33. After density modification with SHELXE, an experimental map of excellent quality could be derived (Figure 2-8). An initial model was built into the experimental map with Arp/wARP (Mooij et al., 2009a; Morris et al., 2002; Murshudov et al., 1997; Perrakis et al., 1999). This model was used as a starting point for manual building with Coot (Emsley et al., 2010) and refinement using phenix.refine (Adams et al., 2010) against the high resolution data. B factors were refined isotropically and geometry and torsion restraints were automatically optimised to yield the best Rfree using the wxc_ and wxu_optimise function of phenix.refine. Validation was carried out using Molprobity

(Davis et al., 2007) and the overall model quality was judged by phenix.polygon (Urzhumtseva et al., 2009) (Figure 2-9 and Table 2-4). Some samples of the refined electron density are shown in Figure 2-8. All the residues in SAP18$^{6-143}$ could be traced, including the non-native N-terminal GPDSM sequence derived from cloning and tag cleavage. In addition to the protein residues, 206 ordered water molecules, one MPD molecule and one isopropanol molecule could be placed in the model.

Table 2-2: Data collection statistics for SAP18$^{6-143}$. Values in parentheses correspond to the highest resolution shell.

Data collection statistics:	SAD dataset	High resolution dataset
Space group	I222	I222
Unit cell (Å)	a = 50.35 b = 73.45 c = 86.46	a = 50.37, b = 73.63, c = 86.70
Resolution (Å)	43.5-1.7 (1.8-1.7)	56.12 – 1.5 (1.6-1.5)
R_{meas} (%)	6.9 (59.9)	5.1 (59.3)
R_{pim} (%)	3.7 (31.9)	2.7 (31.2)
Observed	118961 (16713)	170479 (24317)
I/σ(I)	16.7 (3.4)	15.0 (3.7)
Completeness (%)	98.7 (97.6)	98.9 (97.3)
Multiplicity	6.7 (6.6)	6.6 (6.6)
Wavelength (Å)	0.9802	0.9732

Table 2-3: Distribution of the anomalous signal vs. resolution in the SAD dataset.

Resolution (Å)	43.5	8.0	6.0	5.0	4.0	3.5	3.0	2.5	2.2	2.0	1.8
N. of reflections	203	241	304	679	675	1198	2318	2570	2633	3930	6497
<I/σI>	40.6	34.6	38.6	43.5	39.5	34.7	23.1	17.2	11.9	6.8	2.7
Completeness (%)	94.9	93.4	96.8	95.6	97.0	97.2	97.0	97.3	96.9	96.7	94.8
<d´´/σ(d´´)>	2.91	3.27	2.69	2.21	1.96	1.85	1.61	1.3	1.12	0.94	0.82

Table 2-4: Refinement statistics for SAP18$^{6-143}$

Refinement statistics	
R_{work} (%)	18.2
R_{free} (%)	20.9
Model statistics	
Protein residues	143
Most favoured region (%)	99.4
Allowed region (%)	0.6
RMSD bonds (Å)	0.015
RMSD angles (°)	1.45
Average B factor	20.3
Waters	206
Average B factor (solvent)	40.0

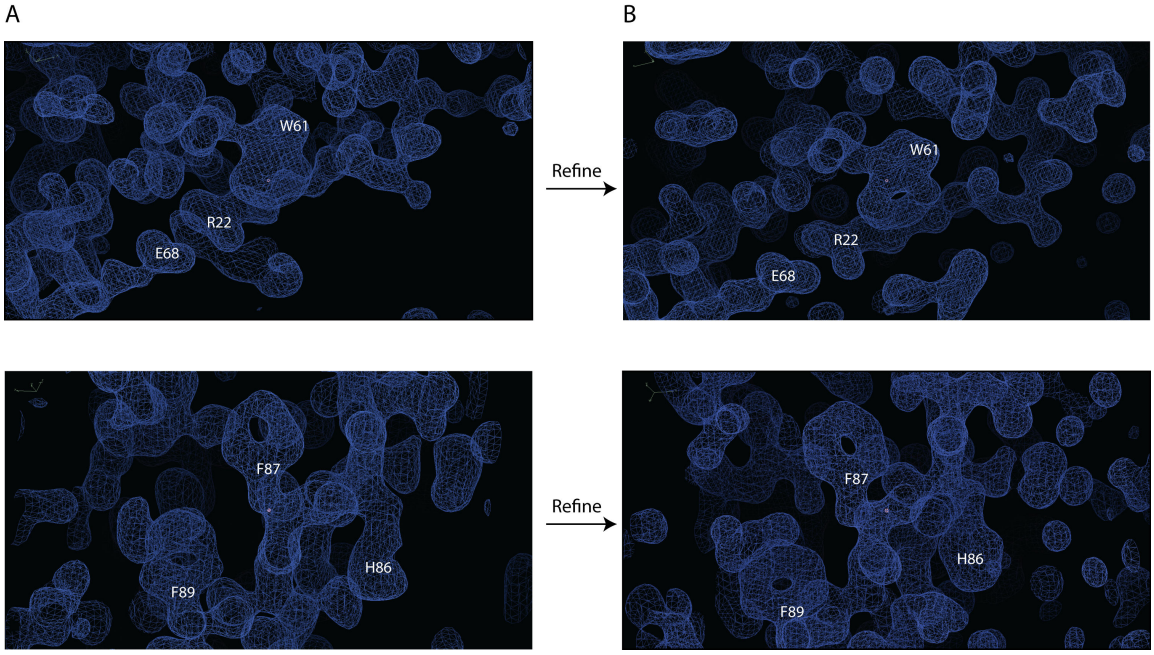


Figure 2-8: Sample of the experimental and refined $2F_o - F_c$ electron density maps for SAP18$^{6-143}$ A. Samples of two regions of the experimental electron densities calculated after density modification by SHELXE and contoured at 1.5σ. **B.** The same regions of electron density after refinement. The residues fitting in the density are indicated in white.

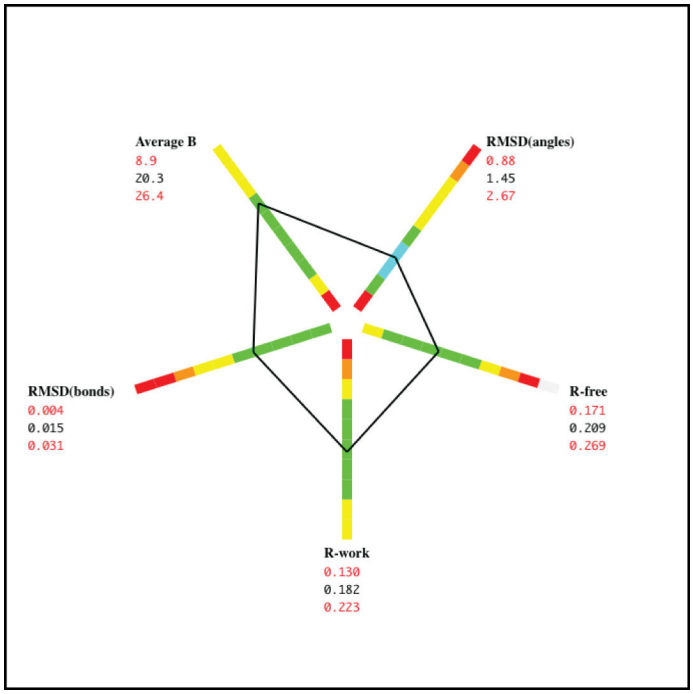


Figure 2-9: Phenix.polygon output for SAP18$^{6\text{-}143}$.

The output of phenix.polygon is presented, which compares the statistics of the structure of SAP18$^{6\text{-}143}$ with those of the structures in the PDB that have the same resolution (1.5 Å). The axes are defined as follows: for each parameter x, the values extracted from the structures in the PDB are plotted on a linear scale. The interval between x_{min} and x_{max} is then divided in equal size bins and the ratio $N_{bin}/N_{uniform}$ is calculated for each bin. N_{bin} is the total number of structures in a given bin; $N_{uniform}$ is the number of structures each bin would contain if the distribution were uniform (i.e., $N_{uniform}$ = total number of structures/number of bins). The ratio $N_{bin}/N_{uniform}$ therefore directly reflects the frequency of structures in the PDB whose value for the parameter x falls in a given bin. If one assumes that the PDB mostly comprises correctly refined structures, it is desirable that for each parameter, a newly refined structure should fall in the most populated bins. The colours on the axis represent the ratio $N_{bin}/N_{uniform}$: white: <0.1; red: <0.25; orange: <0.5; yellow: <1.0; green: <2.0; sky blue: <3.0; blue: <5.0. The black line connects the value measured for SAP18$^{6\text{-}143}$ on each axis.

2.2.5 Structure of SAP18$^{1-143}$

The refined crystallographic model of SAP18$^{6-143}$ was used to solve the structure of SAP18$^{1-143}$ by molecular replacement. A search model in which disordered loops had been removed was used in Phaser (McCoy et al., 2007) to locate two molecules of SAP18$^{1-143}$ in the ASU. The resulting electron map showed clear, continuous density in place of the loops that were missing in the search model, confirming that the molecular replacement solution was correct. Additionally, Cys 26 of each molecule in the ASU was connected by a bridge of electron density to the corresponding residue of a symmetry related mate, indicating the presence of a disulfide bond. After several cycles of manual building with Coot and refinement with phenix.refine, a model with good geometry and an R_{work} / R_{free} of 21.9% / 24.1% could be obtained (see Table 2-5 and Figure 2-10 for refinement statistics). To reduce model bias, after the first and second cycle of manual building, simulated annealing was performed using phenix.refine, to a final temperature of 3000 K and with heating and cooling steps of 100K. Due to intrinsic disorder, not all the residues could be modelled in the final structure: for molecule A residues 1-141 could be confidently traced; for molecule B, only residues 1-12, 16-95 and 100-141 displayed a clear electron density. In addition, several side chains were disordered in the structure: atoms that could not be confidently located in the electron density were set to zero occupancy.

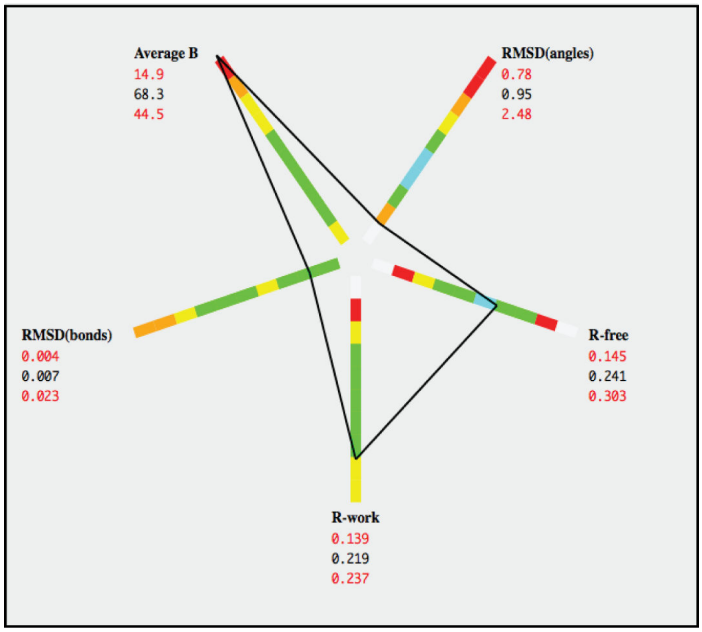


Figure 2-10:
Output of phenix.polygon for SAP18$^{1-143}$.

Table 2-5: refinement statistics for SAP18$^{1-143}$

Refinement statistics	
Rwork (%)	21.9
Rfree (%)	24.1
Model statistics	
Protein residues	286
Most favoured region (%)	99.2
Allowed region (%)	0.8
RMSD bonds (Å)	0.007
RMSD angles (°)	0.95
Average B factor	68.3
Waters	112
Average B factor (solvent)	60

2.2.6 Comparison of the structures of SAP18

The structures of $SAP18^{1-143}$ and $SAP18^{6-143}$ perfectly superpose in the core fold (residues 16-143), with a root mean square deviation (RMSD) over all atoms of 0.47 Å as calculated by the Fold tool (Krissinel and Henrick, 2004) on the PDBe (Velankar et al., 2010) web server (Figure 2-11 A). In each structure, residues 1-16 are differently oriented in space, as dictated by the needs of crystal packing. Comparison with the NMR structure reveals that the core fold is conserved, although the distances and orientation of the secondary structure elements differ slightly, yielding an RMSD of 2.2 Å. Such difference can be ascribed to 'breathing' motions of the molecule in solution, where the secondary structure elements wobble and oscillate due to thermal motion, resulting in an intrinsic conformation variability of the molecule (Lange et al., 2008).

2.2.7 The N-terminus of SAP18 is held in place by a "lock" structure.

In the NMR structure, residues 6-28 are disordered, as highlighted by the superposition of all the structures in the NMR ensemble (Figure 2-11 B). In contrast, in the crystal structures of SAP18 determined in the present work, residues 16-28 fit against the loop between the sheet β2 and the helix α2, in addition to forming a polar interaction with the loop connecting sheets β4 and β5 (Figure 2-12. See Figure 2-26 for a topology scheme of SAP18). This conformation is stabilised by (1) a stacking interaction between Arg^{22}, Trp^{61} and Pro^{27}, which I will refer to as "lock" and (2) a polar interaction between Arg^{22} and Glu^{68}, which cements the N-terminus in its current position. Additionally, Gln^{123} and Asp^{63} form hydrogen bonds with the carbonyl oxygen of Lys^{18} and the nitrogen of Ile^{20} respectively. The interface is then completed

by the interaction between the side chain of Glu^{17} and Thr^{121}. All the residues involved in this interaction are highly conserved in eukaryotes, with the exception of Gln^{123} (Figure 2-1), arguing that their function is crucial for SAP18. Indeed, this locking of the N-terminus is necessary to properly shape the surface of SAP18 that is involved in the interaction with ASAP (see section 2.3.7). This N-terminal lock is not observed in any of the structures of the NMR ensemble: Trp^{61}, Arg^{22} and Glu^{68} are all exposed to the solvent (Figure 2-12). The discrepancy could reflect either true conformational freedom for these residues of SAP18 or the inability of the minimisation protocols used in NMR structure solution to computationally capture the stacking interaction.

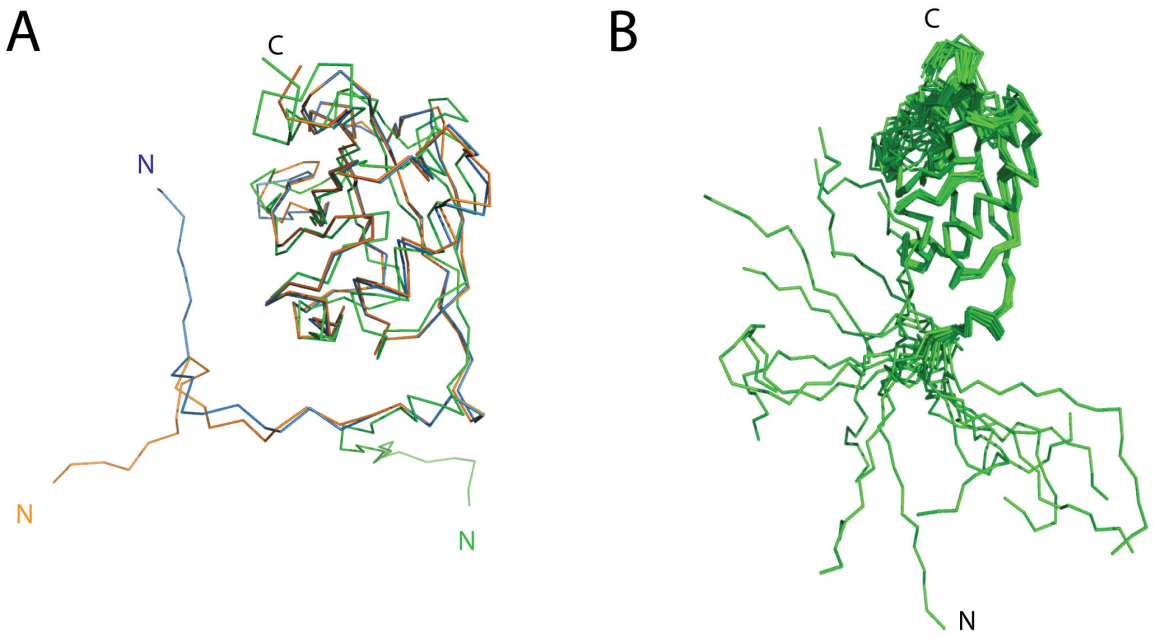


Figure 2-11: The NMR and crystal structures display differences in the N-terminal region. A. Superposition of the Cα traces from the structures of $SAP18^{1-143}$ (blue), $SAP18^{6-143}$ (orange) and the previously reported NMR structure (green). The two crystal structures superpose with RMSD of 0.47 Å in the region 12-143, but diverge in the region 1-12. The NMR structure, while maintaining the same overall fold, superposes with a RMSD of 2.2 Å, due to a higher conformational freedom. N and C indicate the position of the N- and C-termini respectively. **B.** Superposition of all the 20 structures in the NMR ensemble: the NMR structure displays high conformational variability at the N- and C-termini and in the loops between residues 40-54 and 92-103.

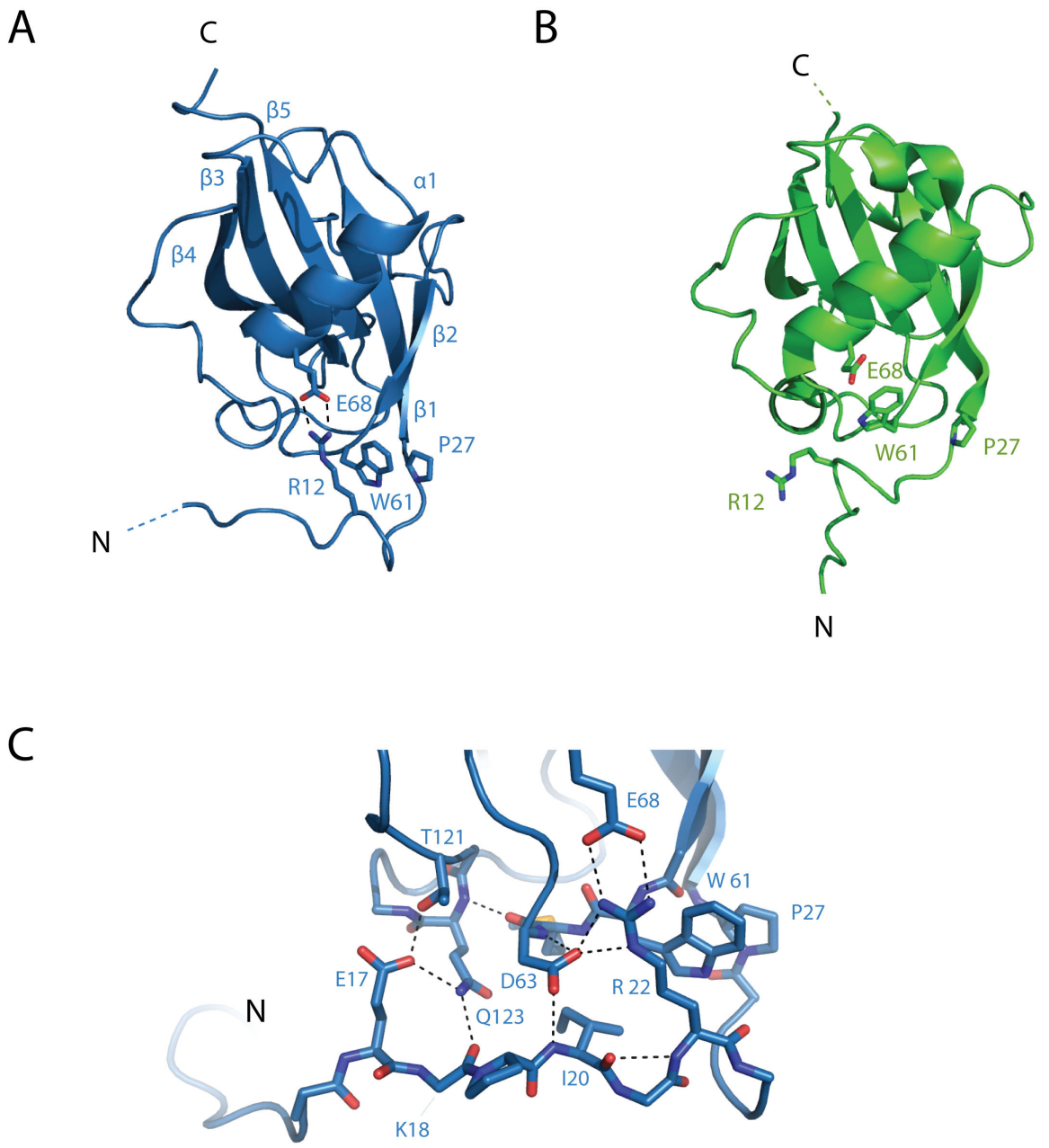


Figure 2-12: The N-terminus of SAP18 is cemented in place by a "lock" structure. A. In both the crystal structures of $SAP18^{6-143}$ and $SAP18^{1-143}$ (depicted), the N-terminus (residues 12-25) is kept folded against the core of the molecule by a "lock" structure (see text for details). **B:** in the NMR structure the ''lock'' is absent. As a consequence, residues 12-20 are disordered in the structure. **C.** Detailed, close up view of the "lock". Dashed black lines indicate polar contacts. The side chain of Lys^{18} is pointing into the solvent and has been truncated in the picture. N and C indicate the position of the N- and C-termini respectively.

2.2.8 SAP18 is likely to be a monomer in solution

2.2.8.1 The disulfide-linked dimer is likely to be an artefact

In the structure of $SAP18^{1-143}$, a covalent dimer is observed in the ASU, with a disulfide bridge formed by Cys^{26} on both molecules (Figure 2-13 A). Physiological, disulfide-bridge-mediated oligomerisation has been reported in the literature (see for example Sanchez-Perez et al., 2003, or Furukawa et al., 2004); additionally, emerging evidence shows that a redox-sensitive proteome exists in the cell, whose members can change their activity levels or function in response to the redox state of their cysteines (Wouters et al., 2010). However, the molecular interface buried between the two disulfide-bonded molecules of SAP18 is small, with a calculated surface of 141 $Å^2$, suggesting a non-physiological assembly. Therefore, while it cannot be excluded that SAP18 might change its oligomeric state in response to redox conditions, this dimer is more likely to reflect an experimental artefact due to the purification and crystallisation process rather than a *bona fide* physiological state. Indeed, in the absence of DTT, stable SAP18 dimers can be observed in analytical ultra centrifugation (AUC) velocity experiments; the dimerisation is however disrupted if fresh DTT is added to the solution (not shown), indicating that SAP18 is a monomer in the physiological conditions normally encountered in the cytoplasm.

2.2.8.2 The N-terminus of SAP18 is probably involved in transient dimer formation

In the structure of SAP18$^{1-143}$, each molecule in the ASU uses its N-terminus to contact a symmetry related mate in the unit cell (which I will refer to as acceptor molecule); in turn, each molecule is also contacted by the N-terminus of a symmetry related mate (donor molecule). The assembly is such that in the crystal the donor and the acceptor are always distinct molecules (Figure 2-13 B). This holds true also for the disulfide-linked dimer present in the ASU. In the crystal, each of its component protomers donates its N-terminus to a different acceptor (Figure 2-14 A). The geometry of the N-terminal interaction prevents the two protomers from swapping their N-termini, as this would require opening of the "lock" (Figure 2-14 B). In the structure of SAP18$^{6-143}$ only one molecule per ASU is present and no disulfide-mediated dimerisation is observed. Nevertheless, the crystal contacts are mainly mediated by the N-termini of the SAP18 molecules and are very similar to the interactions observed in the structure of SAP18$^{1-143}$ (see below). Each molecule of SAP18$^{6-143}$ contacts a neighbouring acceptor, and is contacted by a symmetry related donor, through the N-terminus. Once again, no stable, symmetric dimer can be observed in the crystal.

Overall, the crystal packing of both structures strongly suggests that SAP18 cannot form symmetric dimers through the N-terminus. According to PISA (Krissinel and Henrick, 2007), a single N-terminal interaction between donor and acceptor buries a surface of 664 Å^2 and has an estimate ΔG of -3.5 kcal/mol. A stable interaction is estimated to require $\Delta G < 5$ kcal/mol, suggesting that an asymmetric dimer, in which the two molecules of SAP18 interact through a single donor / acceptor interaction would not be stable in solution. However, such asymmetric interaction could be strong enough to allow the formation of transient dimers, which could explain the N-terminus dependent dimerisation of SAP18 observed in yeast two hybrid studies (Hill et al., 2007; Wang et al., 2002a).

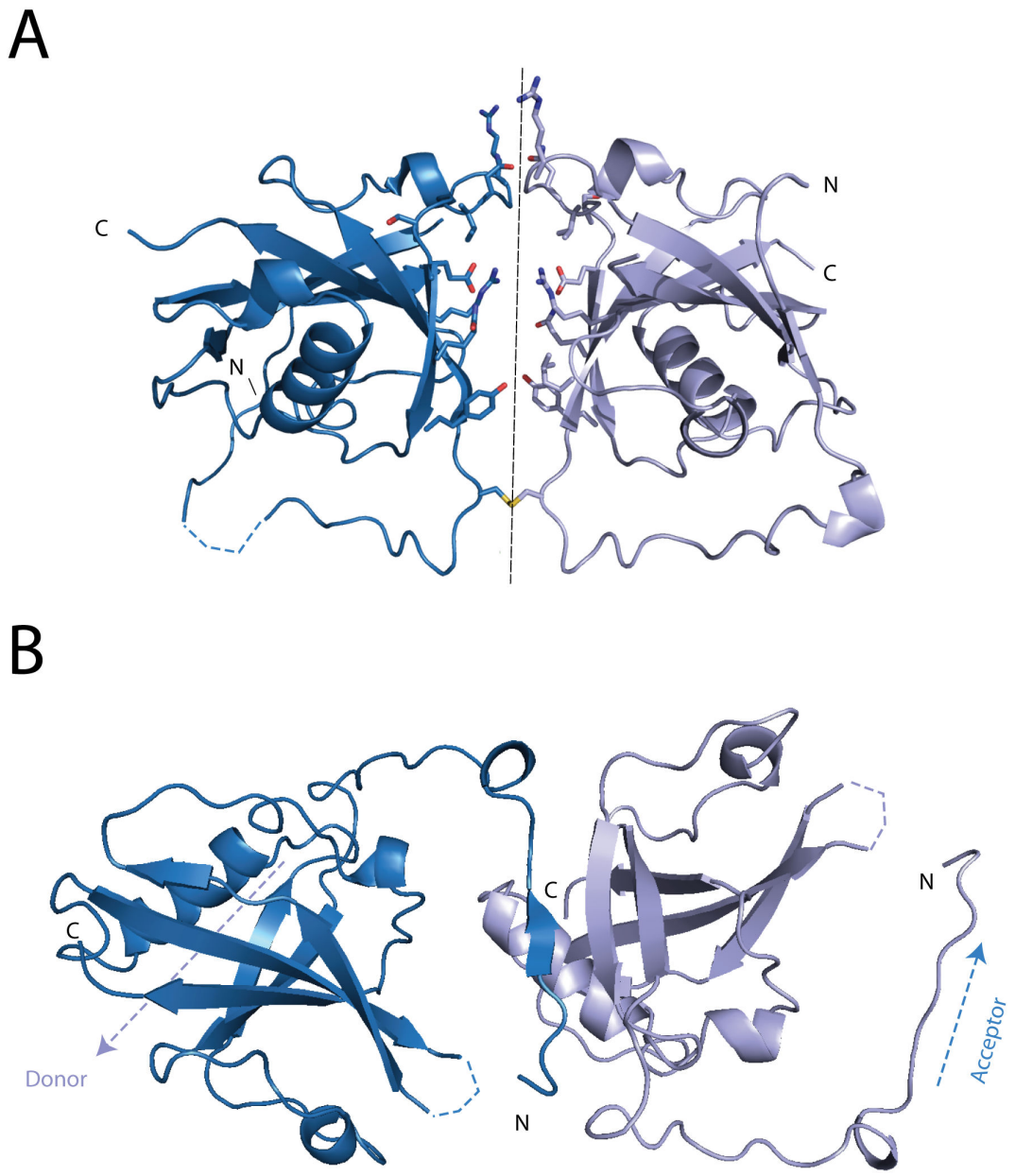


Figure 2-13 Analysis of the crystal packing in SAP18$^{1-143}$. A. In the crystal SAP18$^{1-143}$ forms covalent dimers, disulfide-bridged through Cys^{26}. The residues projecting in the solvent channel between the two subunits are shown in sticks representation. Disordered loops that could not be modelled are indicated with a dashed line. N and C indicate the N- and C-terminus of each molecule. The black, dashed line indicates the approximate position of the non-crystallographic 2-fold axis. **B.** The crystal packing in the SAP18$^{1-143}$ structure is mediated by N-terminal contacts. Any molecule in the ASU (donor) contacts a symmetry related mate (acceptor) through its N-terminus. The arrangement in the crystal is such that each molecule has a different donor and acceptor, thus precluding the formation of a stable, symmetric dimer. The dashed arrows represent the interaction with one further donor and one further acceptor in the crystal that have not been depicted in the figure.

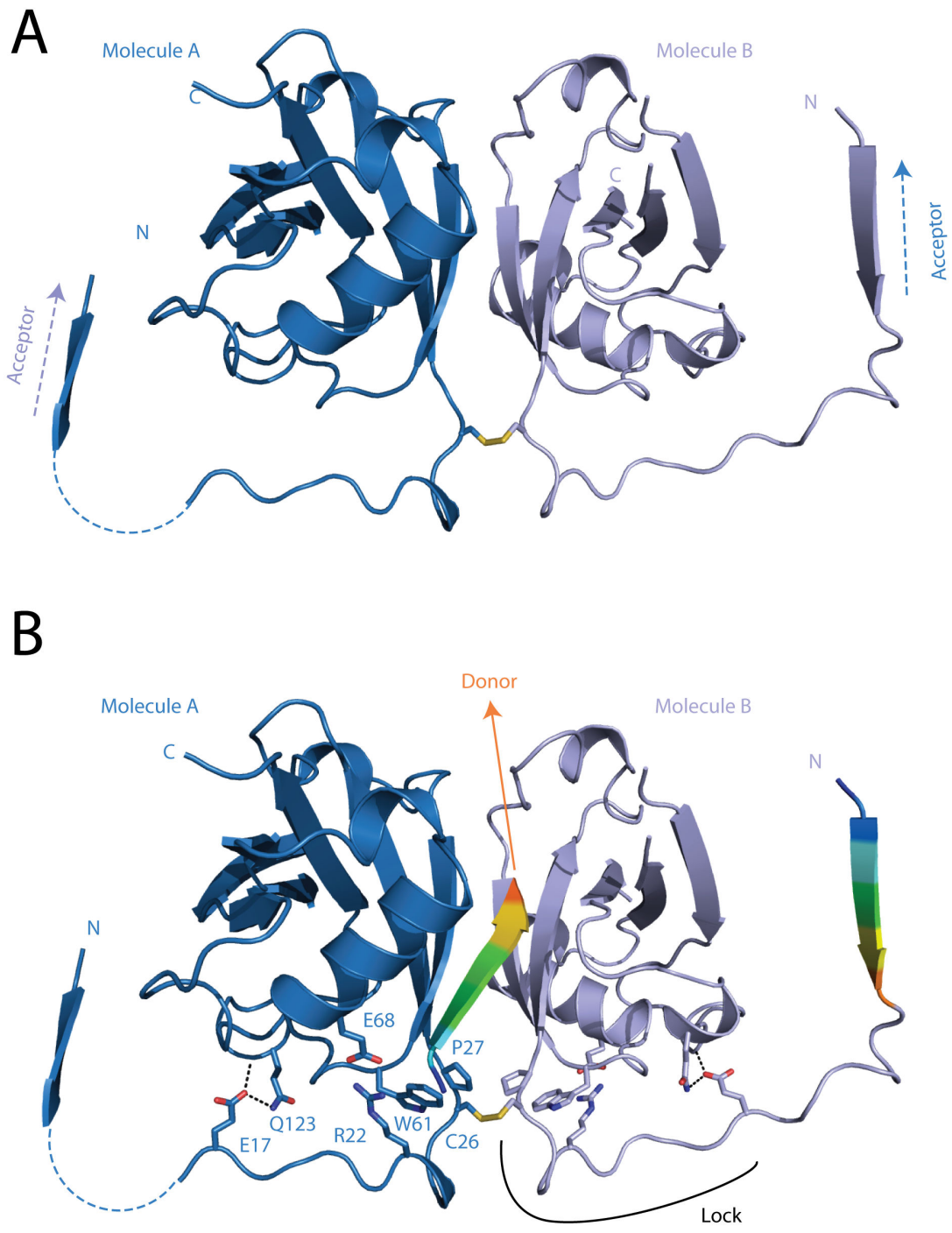


Figure 2-14: The two disulfide-linked molecules in the ASU cannot form a stable dimer through the N-terminus. **A.** In the crystal the two protomers in the disulfide-linked dimer do not swap N-termini, but rather contact two distinct acceptors. **B.** Swapping of the N-termini between the protomers is not compatible with the presence of the lock. The N-terminus of the donor to molecule A is displayed in cartoon representation and rainbow coloured. The N-terminus of molecule B is coloured with the same scheme (i.e. orange corresponds to residue Gln^5 on both molecule B and C). To donate its N-terminus to A, molecule B would need to "open" the lock.

2.2.9 The N-terminus of SAP18 interacts with a conserved surface

In both the structures of $SAP18^{1-143}$ and $SAP18^{6-143}$, crystal packing is mediated by the N-terminus, which protrudes from the donor molecule and extends the beta sheet of the acceptor molecule. Interestingly, in both structures, the N-termini of the donor molecules interact with the acceptors by exploiting the same hydrophobic groove between helix 1 and beta strand 2. The mode of interaction is very similar between the two structures, despite the fact that the N-termini of the two SAP18 constructs differ in sequence. In the structure of $SAP18^{1-143}$ the donor binds to the acceptor through residues 3-9 (Figure 2-15 A). Residues 6-8 of the donor form a beta strand that interacts with the main chain of residues 56-59 in the acceptor. Further donor – acceptor interactions are mediated by side chains. Val^{3} from the donor is buried between the aliphatic portion of the side chains of Glu^{23} and Arg^{22} of the acceptor; additionally, its main chain carbonyl oxygen creates a polar contact with the guanidium group of Arg^{22}. Ser^{5} from the donor is accommodated in a shallow surface pocket on the acceptor and forms hydrogen bonds with the side chains of Thr^{60} and Asp^{63}. In the centre of the binding surface, Val^{7} (donor) is buried in the conserved hydrophobic pocket formed by the side chains of Ile^{58}, Val^{76} and Leu^{72} of the acceptor. Additionally, Thr^{8} from the donor creates a hydrogen bond with the side chain of Gln^{57} from the acceptor. Finally, the aliphatic portion of the side chain of Gln^{9} of the donor interacts with Leu^{56} of the acceptor.

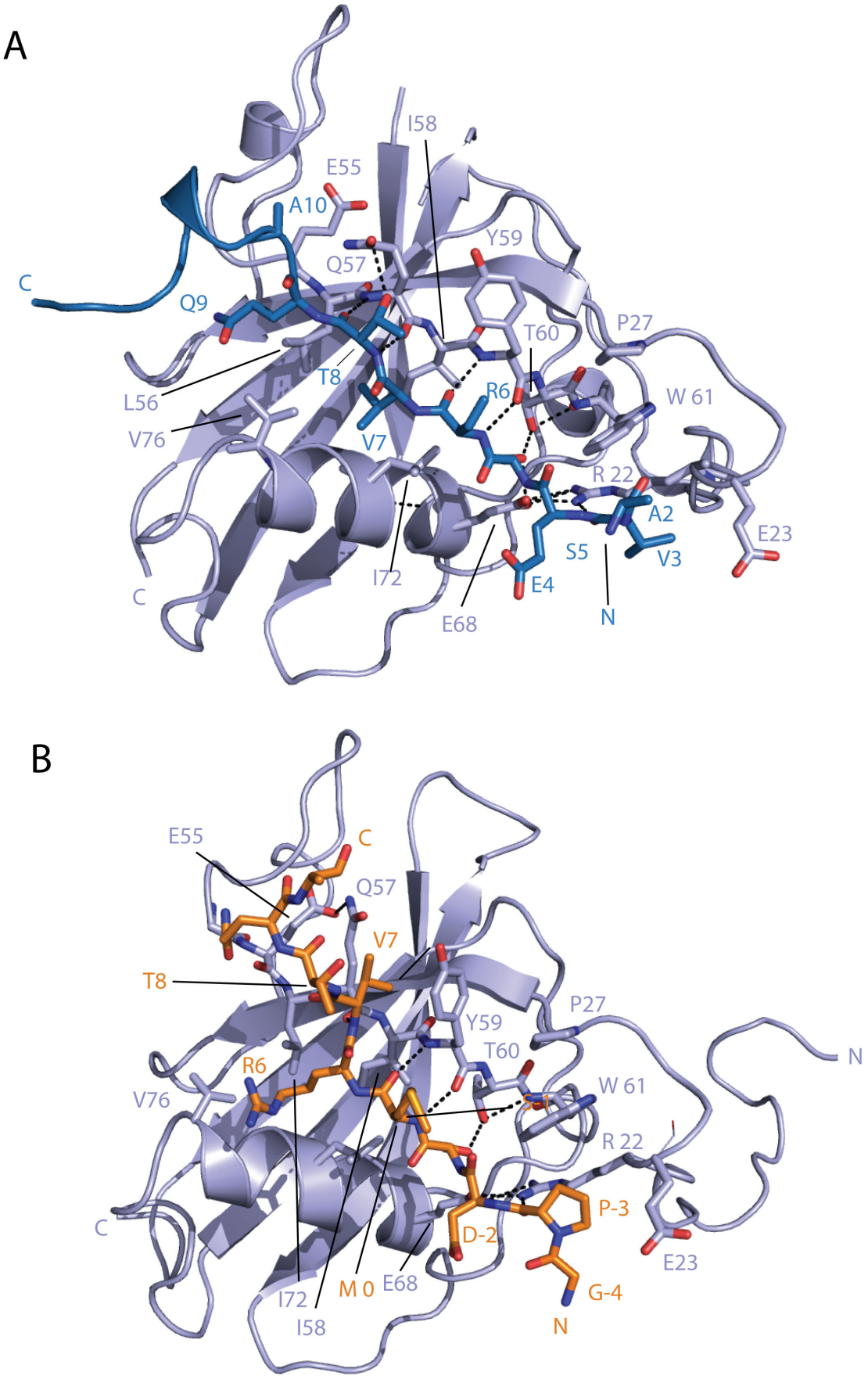

A
I58
E55
A10
Y59
Q57
C
Q9
P27
T60
T8
R6
W 61
L56
V76
V7
R 22
A2
E23
S5
V3
I72
E4
C
E68
N
B
E55
C
Q57
V7
T8
P27
Y59
T60
N
R6
W 61
V76
R 22
P-3
C
D-2
E23
M 0
E68
G-4
I72
I58
N

Figure 2-15: (previous page) The N-termini of SAP18$^{1-143}$ and SAP18$^{6-143}$ bind the acceptor with similar modes. A. Interaction between donor and acceptor in SAP18$^{1-143}$. Residues 1-10 of the donor are depicted in blue; the acceptor is coloured in violet. The side chain of Arg^{6} is disordered in the structure and has been truncated to the carbon beta, which is the last atom visible in the electron density. The acceptor molecule is depicted in violet. Dashed black lines show polar contacts. N and C indicate the N- and C-termini of the molecules, if visible. See main text for a detailed description of the interaction. **B.** Interaction between a donor and acceptor molecules in the structure of SAP18$^{6-143}$. The N-terminus of the donor molecule is depicted in orange; Residues derived from the cleavage of the expression tag and from the cloning site are numbered -4 to 0. The sequence of SAP18 starts with Arg 6.

The N-terminus of SAP18$^{6-143}$ contains a GPDSM sequence derived from cloning and removal of the hexahistidine tag. Due to a fortunate coincidence, this sequence is still able to bind to the groove between helix 1 and beta sheet 2, and does so in a way that mimics that of the native N-terminus (Figure 2-15 B, Figure 2-16). Similarly to what is observed for the native N-terminus, the donor molecule of SAP18$^{6-143}$ extends the beta sheet of the acceptor molecule by forming hydrogen bonds between residues 6-8 (donor) and residues 57-59 (acceptor). The interactions of the side chains are also maintained: in SAP18$^{6-143}$ the role of the native Val^{3} is fulfilled by Pro^{-3}, which can interact with Glu^{23} and Arg^{22} on the acceptor. In the native N-terminus, positions 4 to 6 correspond to the sequence ESR; in SAP18$^{6-143}$ this sequence is substituted with DSM (residues -2 to 0). Asp^{-2} and Met^{0} are exposed to the solvent like their native counterparts Glu^{4} and Arg^{6}, whereas Ser^{-1} recapitulates the interactions of the native Ser^{5}. The role of the native Val^{7} is fulfilled by Arg^{6}, which uses the hydrophobic portion of its side chain to interact with Leu^{56} Ile^{58} and Ile^{72}. The long side chain of the arginine introduces a distortion that pushes the main chain away from the beta strand of the donor, with Val^{7} now tucking between Gln^{57} and Tyr^{59} and the rest of the N-terminus finally protruding away towards the donor molecule.

The residues lining the groove on the acceptor molecule are extremely conserved in all eukaryotes (Figure 2-16), suggesting that *in vivo* the groove might be involved in

functionally critical interactions with as yet unknown binding partners. As shown in section 2.3.9, binding of the N-terminus of SAP18 in the groove would not sterically hinder the formation of the ASAP complex. The possible biological significance of this conserved groove are discussed in section 3.3

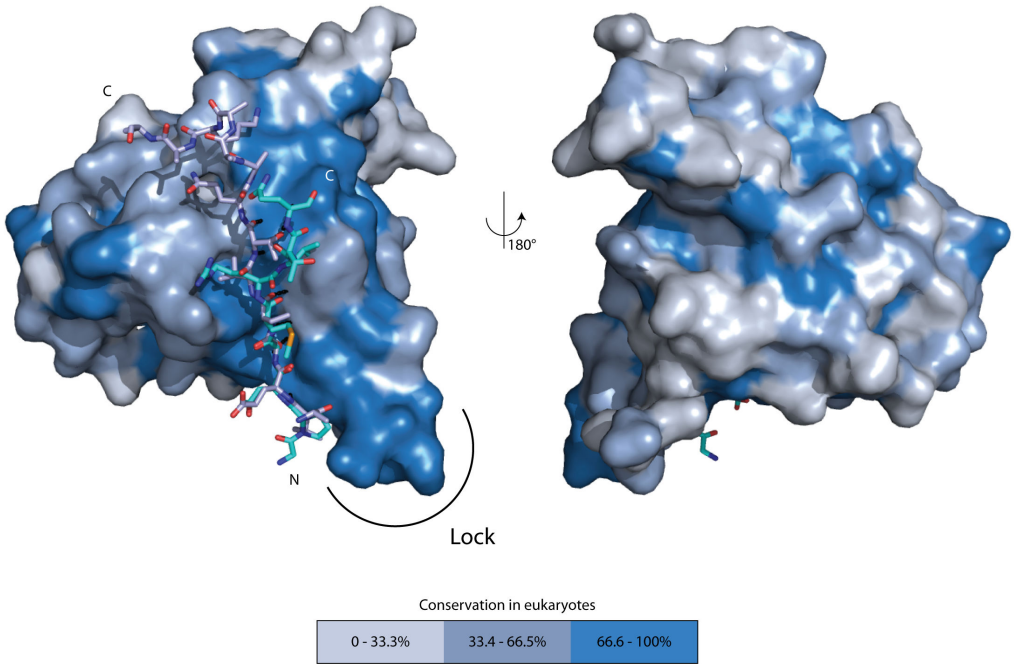


Figure 2-16: The N-terminus of SAP18$^{1-143}$ binds to a highly conserved surface pocket. The acceptor molecules of SAP18$^{1-143}$ have been superposed and are shown in surface representation, coloured according to the conservation of the exposed residues. Two views of the molecule are shown, which are rotated by 180° around the Y axis. The N-terminus of the donor molecules of SAP18$^{1-143}$ and SAP18$^{6-143}$ are shown in ball and stick representation (light blue and cyan respectively). The surface pocket to which the N-terminus binds is part of a highly conserved patch. Also conserved is the lock region, as indicated. Conservation was calculated with the program Consurf (Ashkenazy et al., 2010), based on the multiple sequences alignment shown in Figure 2-1. For the colour code, refer to the graphical legend.

2.3 Structural studies on the ASAP complex

2.3.1 Reconstitution of the ASAP complex

In Metazoans, Acinus is alternatively spliced in three distinct isoforms: Acinus L (the reference isoform), Acinus S and Acinus S'. These splicing isoforms share a common C-terminus (residues 770-1341) and are equally able to form ASAP. Therefore, rather than reconstituting ASAP using full-length Acinus L, I designed a truncation construct of Acinus, which spans residues 770-1250. This construct maintains most of the sequence common to all Acinus isoforms, with the exception of 91 residues at the C-terminus that are poorly conserved and predicted to be disordered. Throughout this work, I will refer to the Acinus$^{770\text{-}1250}$ construct as AcinusΔC. To reconstitute ASAP, bacterially expressed human AcinusΔC and mouse SAP18 were mixed in equimolar amounts with insect cells expressed, full length human RNPS1. The reaction was incubated on ice for 30 minutes and the products were resolved by size exclusion chromatography (Figure 2-17). Despite the C-terminal truncation, AcinusΔC is sufficient to form ASAP *in vitro*.

2.3.2 Identification of the minimal ASAP core

The reconstituted ASAP complex is not suitable for crystallisation, since both Acinus and RNPS1 contain long stretches of residues that are predicted to be disordered. The reconstituted ASAP was therefore subjected to limited proteolysis, with the aim of identifying stable domains within each protein component of the complex. These domains were then tested for their ability to form the ASAP complex.

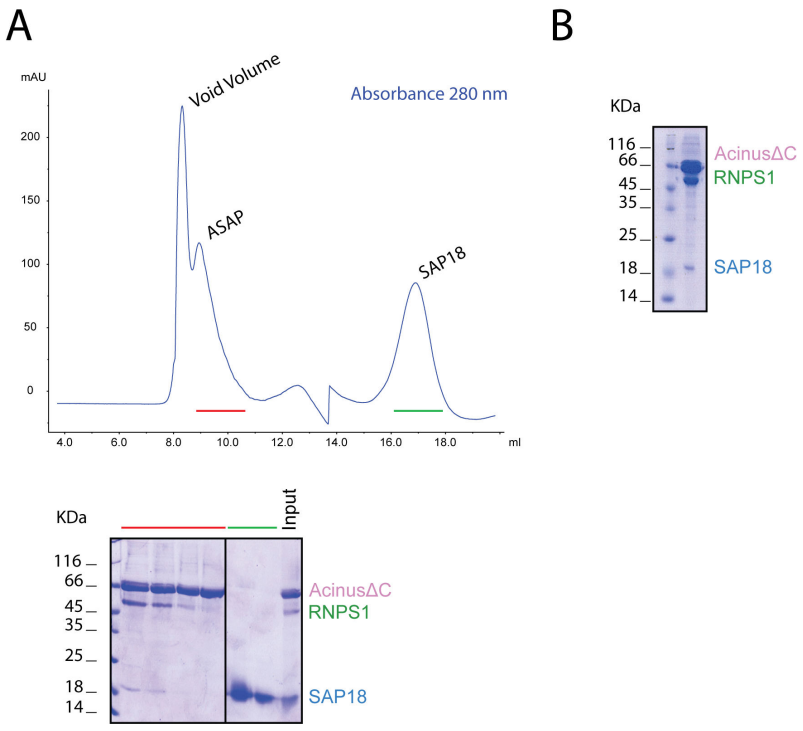


Figure 2-17: Reconstitution of ASAP starting with AcinusΔC. A. Stoichiometric amounts of AcinusΔC and full length RNPS1 were incubated with an excess of SAP18 for 30 minutes on ice and loaded on a S200 10/30 size exclusion column. The chromatogram and the SDS page analysis of the peaks are shown. ASAP is formed and elutes as a single peak. The peak eluting in the void volume is mainly comprised by uncomplexed RNPS1 FL, which is probably aggregated (not shown). **B.** Final concentrated sample.

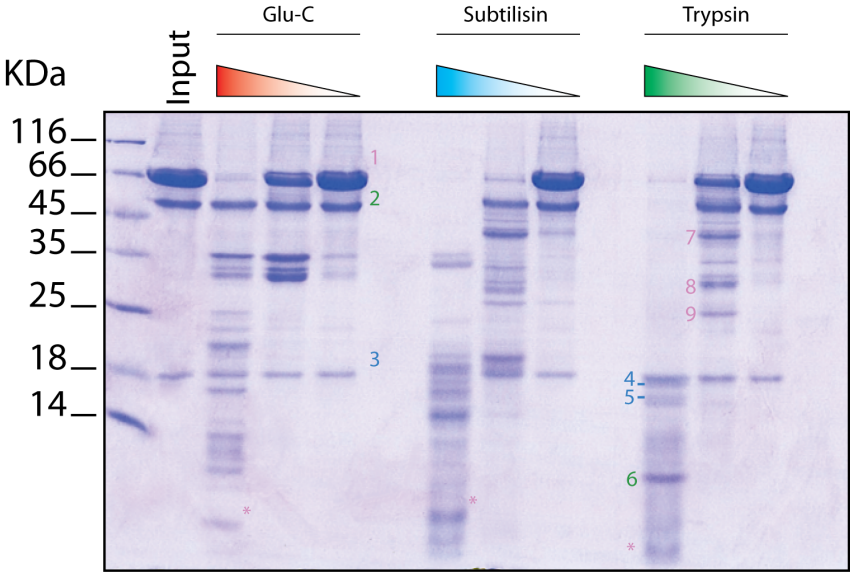


Figure 2-18: Limited proteolysis of ASAP formed with AcinusΔC. The ASAP complex was incubated with different amounts of each protease for 30 minutes on ice. The reaction products were resolved by 14% Tricine PAGE. The identity of numbered bands was determined by N-terminal sequencing and / or tryptic digestion followed by MALDI-TOF mass spectrometry. Numbered bands: 1: AcinusΔC; 2: full-length RNPS1; 3: full-length SAP18; 4: SAP18$^{6\text{-}143}$; 5: SAP18$^{12\text{-}149}$; 6: RNPS1$^{159\text{-}244}$; 7: Acinus$^{940\text{-}1230}$; 8: Acinus$^{846\text{-}1107}$; 9: Acinus$^{1000\text{-}1107}$. The bands marked with asterisk contain the C-terminal conserved region of Acinus, but their boundaries could not be reliably identified.

The limited proteolysis experiment was performed according to the protocol specified in Materials and Methods and its results are shown in Figure 2-18. SAP18 (band 3) was stable in most conditions, as expected for a globular protein; at the highest concentrations of Glu-C and trypsin, partial degradation was observed, yielding SAP18$^{6-143}$ and SAP18$^{12-143}$ (bands 4 and 5 respectively), in which the flexible N- and C-terminus have been digested. RNPS1 (band 2) was degraded by trypsin to RNPS1$^{159-244}$, which corresponded to the predicted RRM domain (band 6). Altogether, these results suggested that the N- and C-termini of RNPS1 and SAP18 are accessible to the proteases and are therefore probably dispensable for binding to the ASAP complex. The proteolytic degradation of Acinus (band 1) did not indicate a predominant, stable construct. Rather, a series of degradation products were identified that were centred on the RRM domain (bands 7-9) or contained the C-terminal conserved region (bands marked with asterisk).

Based on the limited proteolysis results, I cloned RNPS1$^{159-244}$ and SAP18$^{6-143}$. For Acinus, I designed a series of deletion constructs based on sequence alignments and secondary structure prediction. Acinus$^{770-1230}$ contains the N-terminus and the RRM domain, but is truncated within the C-terminal conserved region; Acinus$^{846-1107}$ contains the RRM domain; Acinus$^{1171-1250}$ spans the C-terminal conserved region, but lacks the RRM domain (Figure 2-19 A). The RNPS1, Acinus and SAP18 constructs were expressed in *E. coli* and were tested for their ability to form the ASAP complex in a calmodulin binding peptide (CBP) pull down experiment, in which full-length, CBP-proteinA-tagged RNPS1 (TAP RNPS1) or CBP-tagged RNPS1$^{159-244}$ were used as bait (Figure 2-19 B). Neither the N-terminus nor the RRM of Acinus are needed for forming the ASAP complex (Lanes 1 and 2 respectively); rather, the C-terminal conserved region is necessary and sufficient (Lane 3). For this reason, I decided to name this region ASAP binding motif (ABM). As suggested by the limited proteolysis experiment, the RRM domain of RNPS1 is necessary and sufficient for ASAP formation (Lane 4);

furthermore, the disordered N- and C-termini of SAP18 are dispensable for forming the complex (Lane 5). Lanes 6-9 show the negative controls in which the prey proteins have been incubated in the absence of tagged RNPS1.

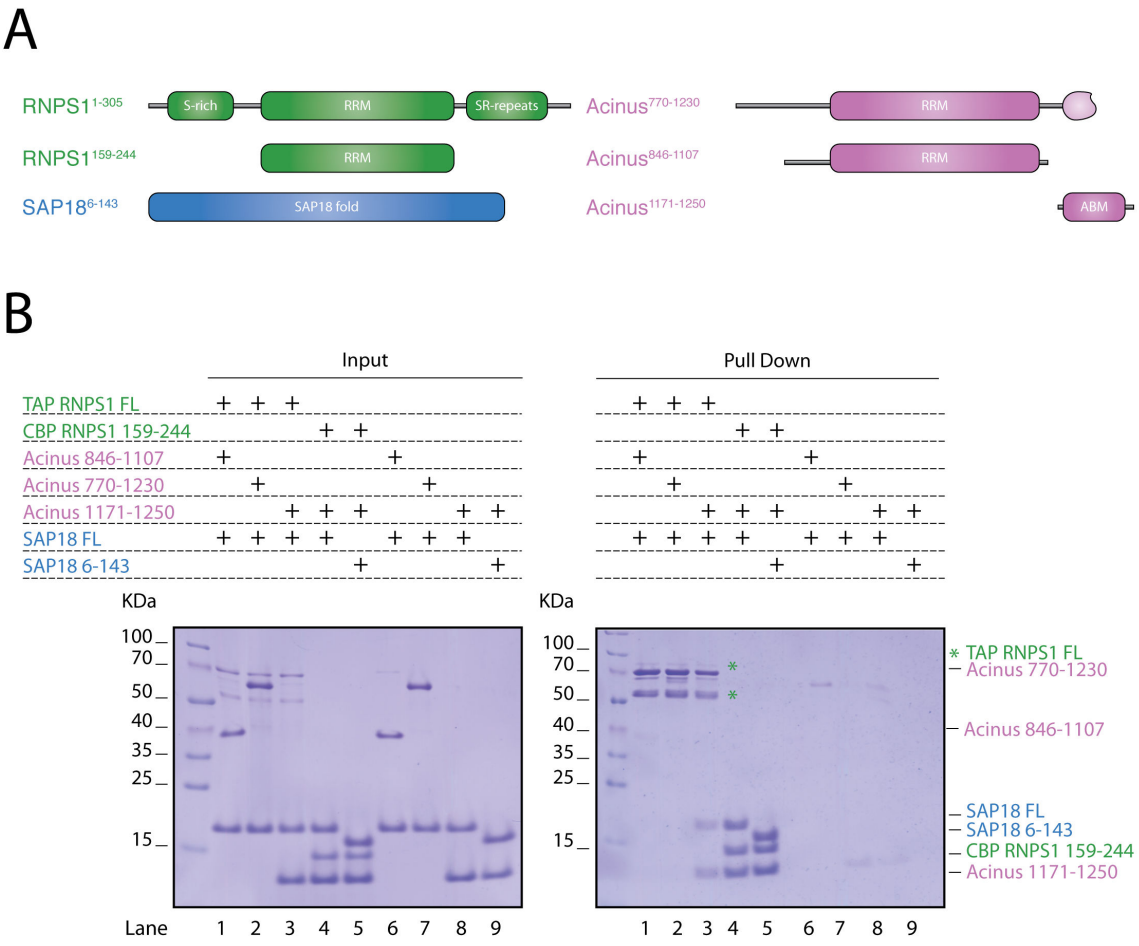


Figure 2-19: Acinus$^{1171-1250}$, RNPS1$^{159-244}$ and SAP18$^{6-143}$ can form a minimal, stable ASAP complex. CBP tagged RNPS1$^{159-244}$ or TAP tagged full length RNPS1 were used as baits for a CBP pull down and tested against several deletion constructs of Acinus and SAP18. **A.** Schematic of the constructs used in the experiment. **B.** Left: input gel; right: pull down. For the description of the results, see main text.

2.3.3 RNPS1 and Acinus form a stable sub complex of ASAP

Using pull downs with TAP tagged proteins, Tange and co-workers had shown that RNPS1 and Acinus form a stable subcomplex, which must be preformed for SAP18 to bind (Tange et al., 2005). In order to confirm that this holds true for the core ASAP, a series of gel filtration experiments were performed using Acinus$^{1171-1250}$, RNPS1$^{159-244}$ and SAP18$^{6-143}$. As shown in Figure 2-20, Acinus$^{1171-1250}$ could stably bind RNPS1$^{159-244}$. Conversely, SAP18$^{6-143}$ could not bind either Acinus$^{1171-1250}$ or RNPS1$^{159-244}$ alone. As a positive control, the ASAP complex could be formed when all three components were present. The same results were obtained if full-length proteins (or AcinusΔC) were used (not shown).

These data confirm that Acinus and RNPS1 can form a stable sub-complex, and their association is required for the binding of SAP18. This in turn suggests that SAP18 binds to both Acinus and RNPS1 at the same time, presumably via a composite surface.

2.3.4 Large scale purification of the core ASAP complex

The truncated ASAP complex comprising SAP18$^{6-143}$, RNPS1$^{159-244}$ and Acinus$^{1171-1250}$ could be purified on a scale suitable for crystallisation experiments. Briefly, stoichiometric amounts of each protein were mixed and incubated 30 minutes on ice to allow complex formation. ASAP was then separated from any excess of reagents by size exclusion chromatography (Figure 2-21). Despite being apparently homogeneous and proteolytically stable, the truncated ASAP complex did not yield any crystal. The sub-complex between RNPS1$^{159-244}$ and Acinus$^{1171-1250}$ was similarly unsuccessful in crystallisation experiments.

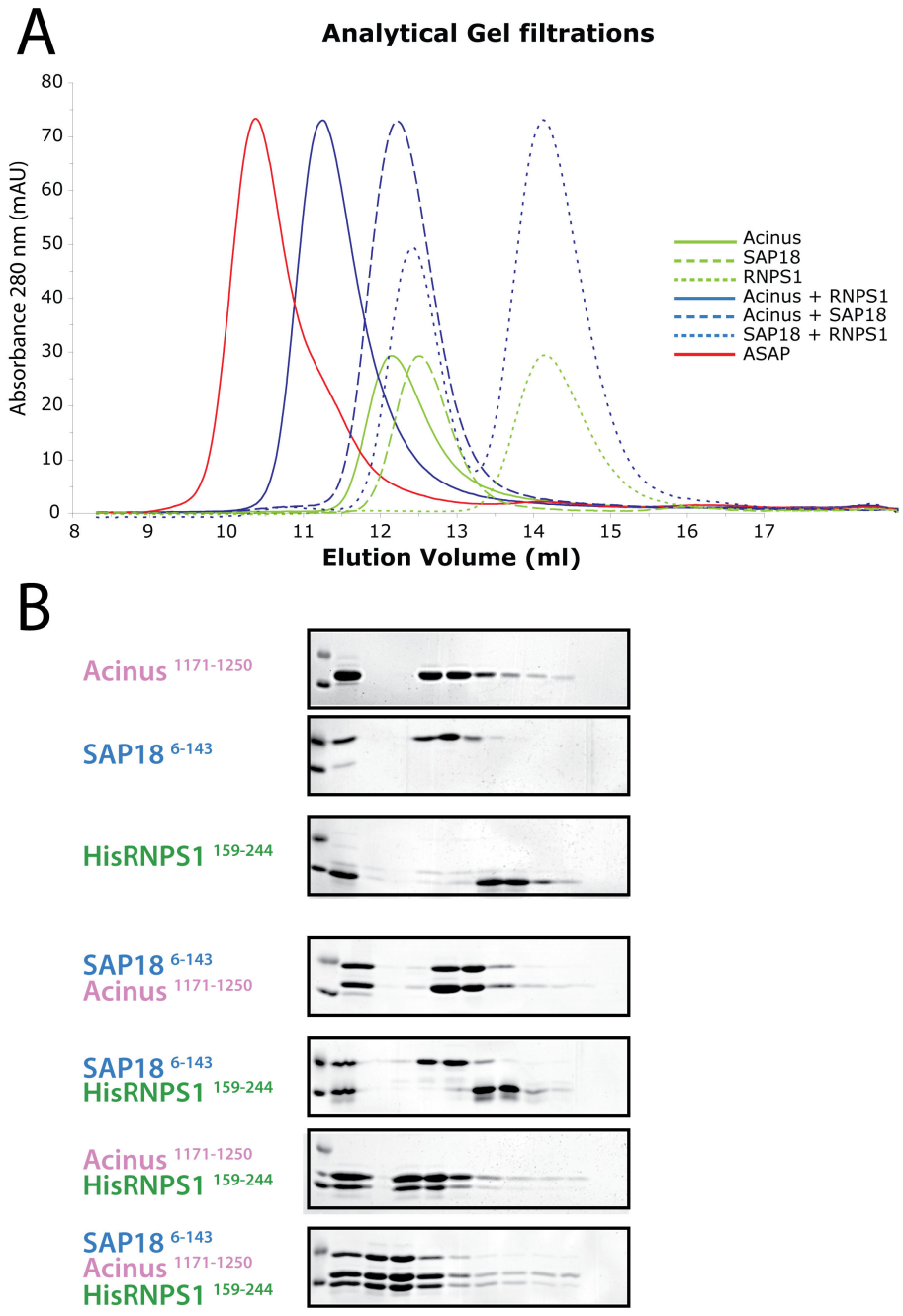


Figure 2-20: RNPS1 and Acinus can form a stable subcomplex of ASAP.

A. Individually purified proteins were mixed, incubated on ice for 30 minutes and injected on a size exclusion column (S75 10/30) in the indicated combinations. **B.** The same fractions for each run were then analysed by SDS PAGE. Acinus and RNPS1 can bind independently of SAP18. SAP18 cannot bind to either RNPS1 or Acinus alone.

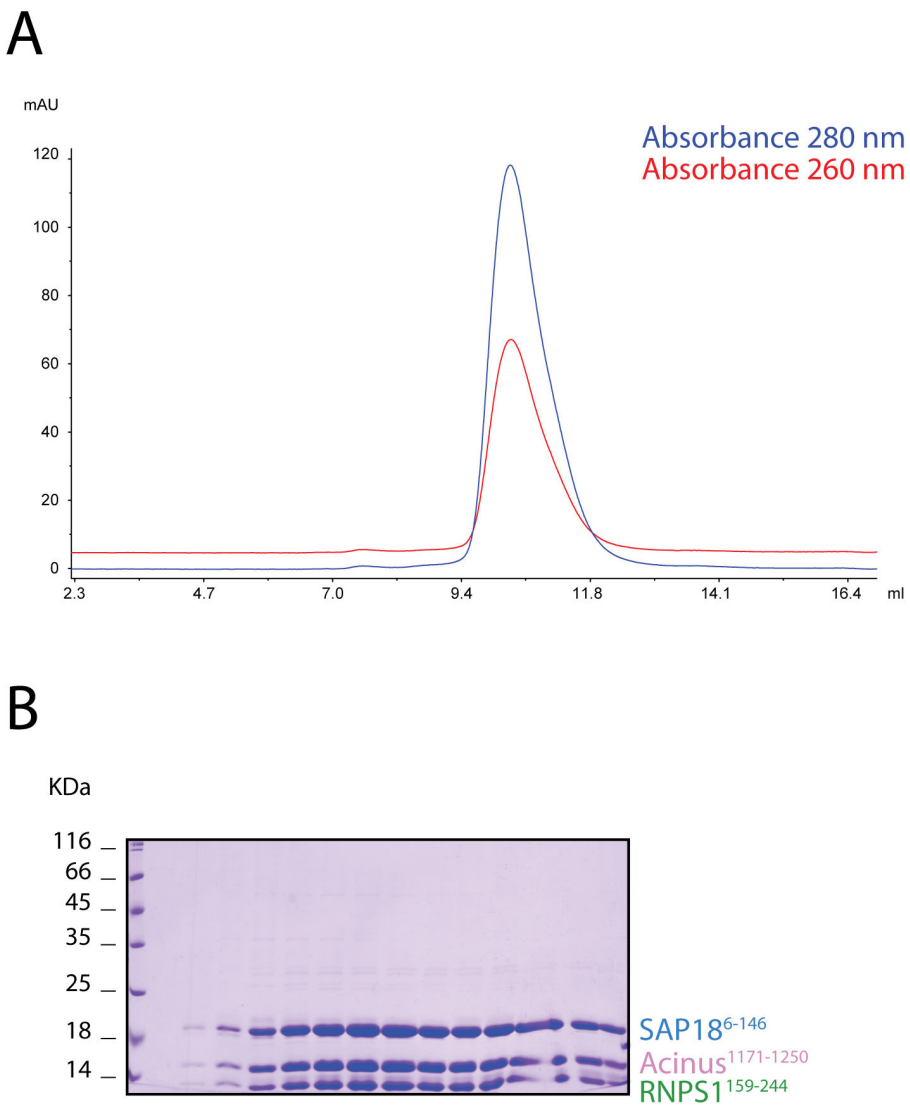


Figure 2-21: Purification of the minimal ASAP core. A. Size exclusion profile of the complex. **B.** SDS PAGE analysis of the peak.

2.3.5 Preparation of a chimeric ASAP complex

In a protein crystal, surface residues mediate contacts between neighbouring molecules. In solution, these residues usually enjoy a large conformational freedom; in the crystal, they must assume a stable, relatively rigid conformation that is conductive to crystal contacts. This transition carries high entropy costs, which are only partially offset by the enthalpy of the new crystal contact. Indeed, it has been shown that surface residues with high conformational freedom, such as Lys and Glu, are significantly underrepresented in crystal contacts of solved crystallographic structures (Cieslik and Derewenda, 2009; Price et al., 2009). One possible way to promote crystal contact formation in proteins is to modify their high-entropy surface residues. This can be done by mutagenesis (with one or more surface Lys, Glu, Gln and Asn being mutated to alanine), or by chemical modification (usually lysine methylation). Both methods have been reported successful for proteins that were unable to crystallise (Cieslik and Derewenda, 2009). Another commonly employed strategy to tame recalcitrant proteins or protein complex into crystallising is to change their organism of origin: surface residues that are not involved in interaction interfaces are generally poorly conserved and can vary a lot among orthologs of the same protein or protein complex. Changing organisms is equivalent to changing most of the surface properties of the protein at once, without affecting its function and interactions.

A quick glance at the multiple sequence alignment of the Acinus orthologs shows that the ABM is well conserved within Vertebrates (Figure 2-3). In order to achieve a sufficient difference in surface properties, I decided to clone and express Acinus (and consequently the ASAP complex) from the invertebrate *Drosophila melanogaster*. The *D.m.* equivalent of Acinus$^{1171\text{-}1250}$ (Acinus$^{616\text{-}710\ Dm}$) proved to be more stable and expressed at higher levels than its human counterpart. Unfortunately,

D.m. SAP18 and RNPS1 proved to be poorly soluble (not shown). Owing to the phylogenetic conservation of Acinus, however, a chimeric complex could be formed, containing human RNPS1$^{159-244}$, mouse SAP18$^{6-143}$ and Acinus$^{616-710\ Dm}$. Briefly, each protein was separately expressed and purified; stoichiometric amounts of each component were then used to form the chimeric ASAP, which was separated on a size exclusion chromatography column (Figure 2-22).

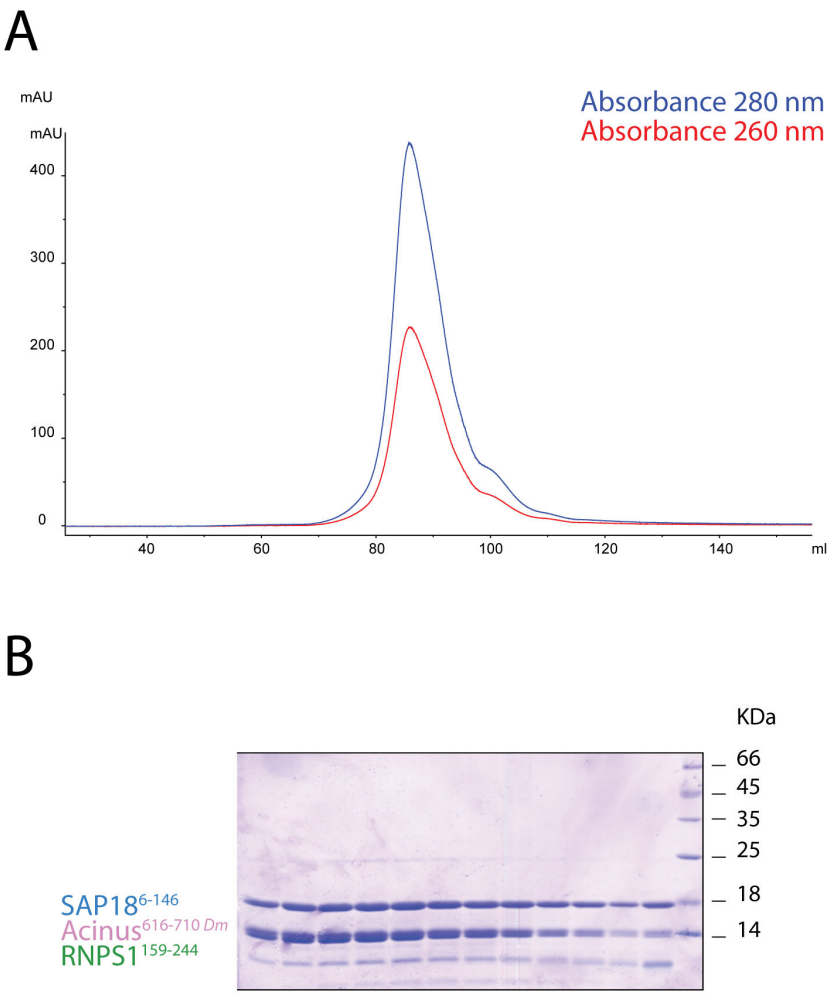


Figure 2-22: Purification of the chimeric ASAP. A. Size exclusion profile of the complex. **B.** SDS PAGE analysis of the elution peak.

If purified in the absence of protease inhibitors, Acinus$^{616\text{-}710\ Dm}$ displays partial degradation (not shown), suggesting some disordered portions are still present in the protein, which could hinder crystallisation. In order to remove these unstructured regions, limited proteolysis was attempted on a preparative scale. The proteolysis and crystallisation experiments described in the next section were performed by Judith Ebert.

The chimeric ASAP, containing SAP18$^{1\text{-}143}$, RNPS1$^{159\text{-}244}$ and Acinus$^{616\text{-}710\ Dm}$, was prepared according to the standard protocol and subsequently incubated with 0.025 mg/ml trypsin for 30 minutes on ice. After blocking the reaction with protease inhibitors, the digestion products were resolved through size exclusion chromatography (Figure 2-23). The composition of the trypsinised ASAP was determined by N-terminal sequencing and ESI/TOF total mass measurements: SAP18$^{1\text{-}143}$ was digested to yield SAP18$^{7\text{-}143}$ and SAP18$^{14\text{-}143}$, RNPS1$^{159\text{-}244}$ was left intact and Acinus$^{616\text{-}710\ Dm}$ was trimmed to yield Acinus$^{648\text{-}688\ Dm}$.

2.3.6 Crystallisation of the chimeric ASAP complex

The core ASAP obtained by limited proteolysis was screened for crystallisation at the concentration of 15 mg/ml. Small, needle-shaped crystals grew in presence of 50 mM tris pH 8.0, 30% PEG 3350 (Figure 2-24). After optimisation, diffracting crystals were obtained from seeding in 24 well sitting drop format, using 1.8 µl drops (0.8 µl protein, 0.8 µl crystallisation solution and 0.2 µl seed stock), with 25% PEG 3350, 50 mM tris pH 8.0.

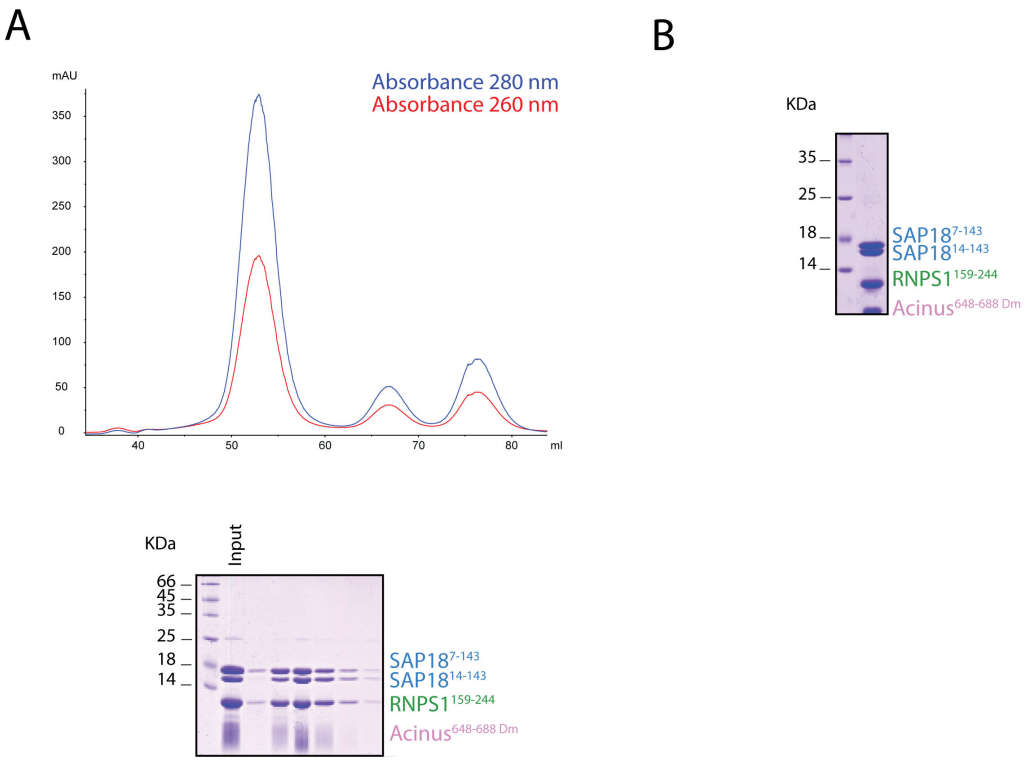


Figure 2-23: Preparative limited proteolysis of the chimeric ASAP. For details of the experimental procedure see text. **A.** Size exclusion chromatography and SDS PAGE analysis of the preparative limited proteolysis. ASAP elutes as a single, symmetric peak. **B.** Final concentrated sample.

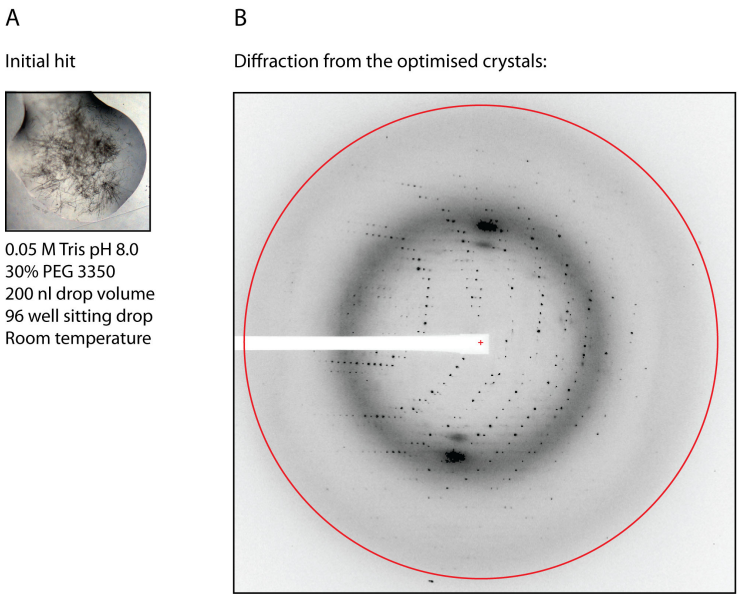


Figure 2-24: Optimisation of the ASAP crystals and final diffraction pattern.

A. Initial hit **B.** Diffraction from the optimised crystals. The red circle in the picture indicates 2.0 Å resolution.

2.3.7 The structure of the chimeric ASAP complex

The crystals were cryoprotected with 10% ethylene glycol and harvested for data collection on the microfocus beamline PXI of SLS. One native dataset was obtained, to a maximum resolution of 1.9 Å (Judith Ebert). iMOSFLM (Leslie, 1992) identified the data as monoclinic primitive (mP); the integrated data was imported into CCP4 through pointless (Evans, 2006) and scaled in space group $P2_1$ with Scala (see Figure 2-26 for data collection statistics). The Matthews coefficient suggests the presence of one complex per ASU, with a solvent content of 45%. The structure was solved by molecular replacement using Phaser (McCoy et al., 2007). Two models were used in the search: $SAP18^{12-143}$, derived from the previously determined structure of $SAP18^{6-143}$ and a homology-based, polyalanine model of $RNPS1^{159-244}$ generated with modeller (Eswar et al., 2008; Fiser and Sali, 2003) on the basis of the RRM domain of Y14 (PDB: 1rk8 Bono et al., 2004). Both models could be confidently placed in the ASU. The resulting electron density was of good quality, with clear density for some of the missing side chains in RNPS1 and a stretch of continuous density corresponding to Acinus present outside of the search models (Figure 2-25). Initial building was carried out with Buccaneer (Cowtan, 2006) and Refmac (Murshudov et al., 1997); the structure was then manually built and refined using Coot (Emsley et al., 2010) and Phenix (Adams et al., 2010), with one TLS (Winn et al., 2001) group for each chain in the model. To reduce model bias, after the first and second round of manual building, simulated annealing was performed using phenix.refine, heating to a maximum temperature of 3000 K and with heating and cooling steps of 100K. In the final structure all residues could be traced for $RNPS1^{159-244}$ and $SAP18^{14-143}$. Of $Acinus^{648-688}$, only residues 656-682 could be located in the electron density. Since the crystals were obtained by limited proteolysis, the missing residues could be either disordered in the

crystal or absent altogether. The model was refined to good overall geometry and R_{work} / R_{free} of 18.2% / 21.2%. For refinement statistics see Figure 2-26.

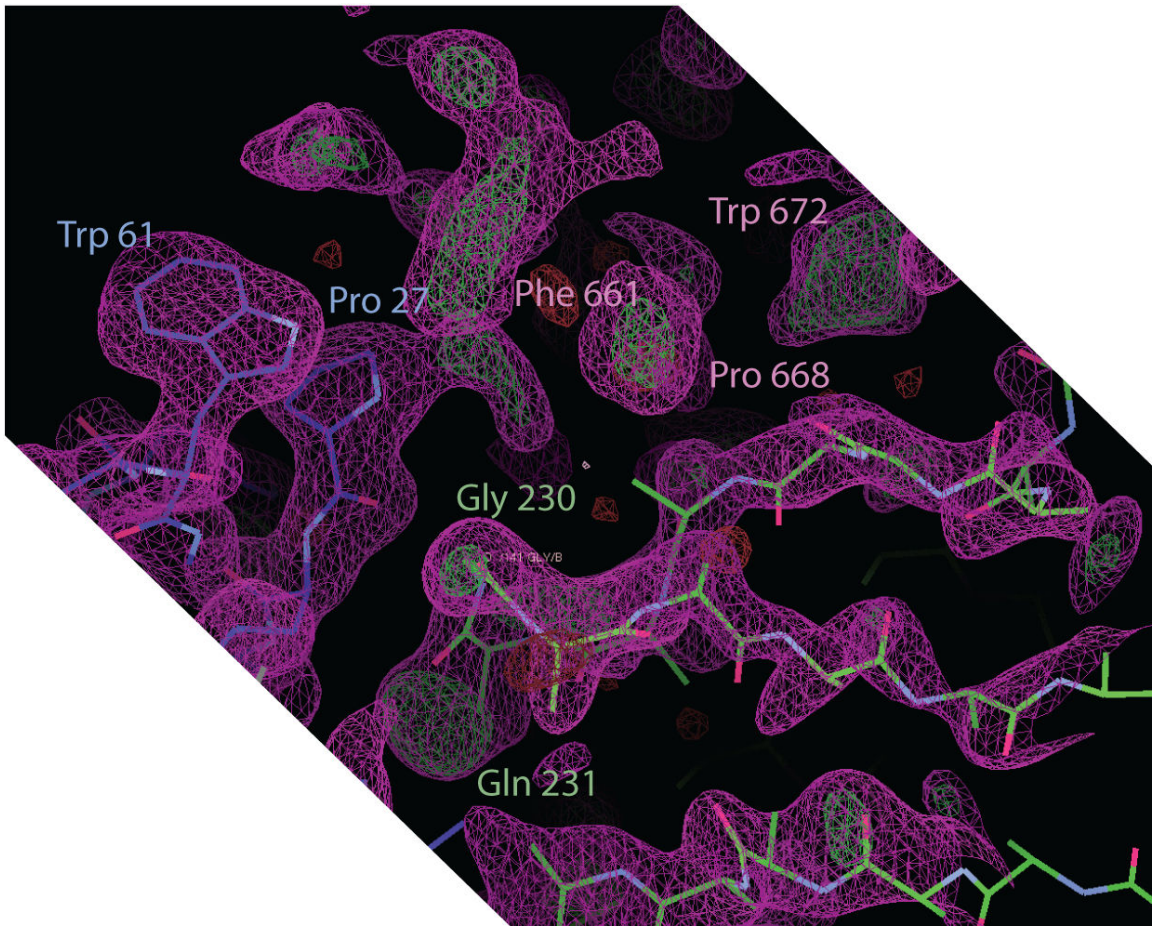


Figure 2-25: Electron density map after molecular replacement. Purple: 2Fo - Fc map; green: Fo – Fc positive density. Contour is at 1σ. The electron density corresponding to some of the missing side chains from RNPS1 and to the backbone of Acinus can be seen.

A

Dataset	Native
Space group	P2
Unit cell (Å)	a = 38.87, α = 90.0° b = 73.33, β = 110.99° c = 50.75; γ = 90.0°
Resolution (Å)	36.67 - 1.9 (2.0 - 1.9)
R_{meas} (%)	7.2 (76.9)
R_{pim} %)	4.0 (28.4)
Observed (Unique)	158835 (21039)
I/σI	10.3 (2.7)
Completeness	100.0 (100.0)
Multiplicity	7.5 (7.2)

B

Refinement statistics:	
R_{work} (%)	18.2
R_{free} (%)	21.2
Model statistics:	
Protein residues	242
Most favoured region (%)	99.2
Allowed region (%)	0.8
RMSD bonds (Å)	0.085
RMSD angles (°)	1.06
Average B factor	41.0
Waters	186
Average B factor (solvent)	42.0

C

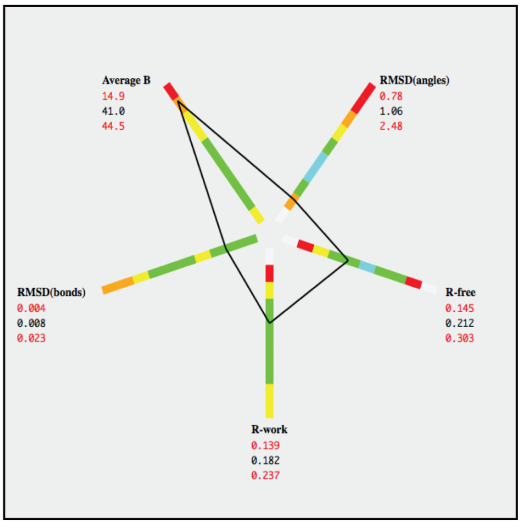


Figure 2-26: Data collection and refinement statistics for the ASAP complex. A. Data collection statistics Values in parentheses correspond to the highest resolution shell. **B.** Refinement statistics. **C.** Overview of the structure quality through phenix.polygon.

2.3.8 Overall structure of the complex

As expected from sequence similarity, in the ASAP complex RNPS1 adopts the typical β-α-β-β-α-β-β topology of an RRM domain; Acinus adopts a hairpin-like fold, with an α-β-turn-β-α topology (Figure 2-27). The hairpin hydrophobically interacts with the surface created by helix 1 and beta strand 4 of RNPS1; additionally, the two β-strands of Acinus augment β-strand 4 of RNPS1. Together, the hairpin formed by Acinus and the loops 1, 3 and 6 of RNPS1 provide a composite surface to which SAP18 binds (see Figure 2-28 for an overall depiction of the complex). In agreement with the biochemistry, PISA identifies two possible quaternary structures: the Acinus:RNPS1 subcomplex and ASAP as a whole. Furthermore, no crystal contact was found, which would suggest ASAP multimerisation. This was confirmed by analytical ultracentrifugation experiments, which show that ASAP is a monomer in solution (not shown). The structure of SAP18 in the complex is essentially indistinguishable from the uncomplexed form, with an RMSD of 0.9 Å over the backbone atoms: the only difference brought about by the binding is the movement of the loop between strands 3 and 4 towards RNPS1 (Figure 2-29).

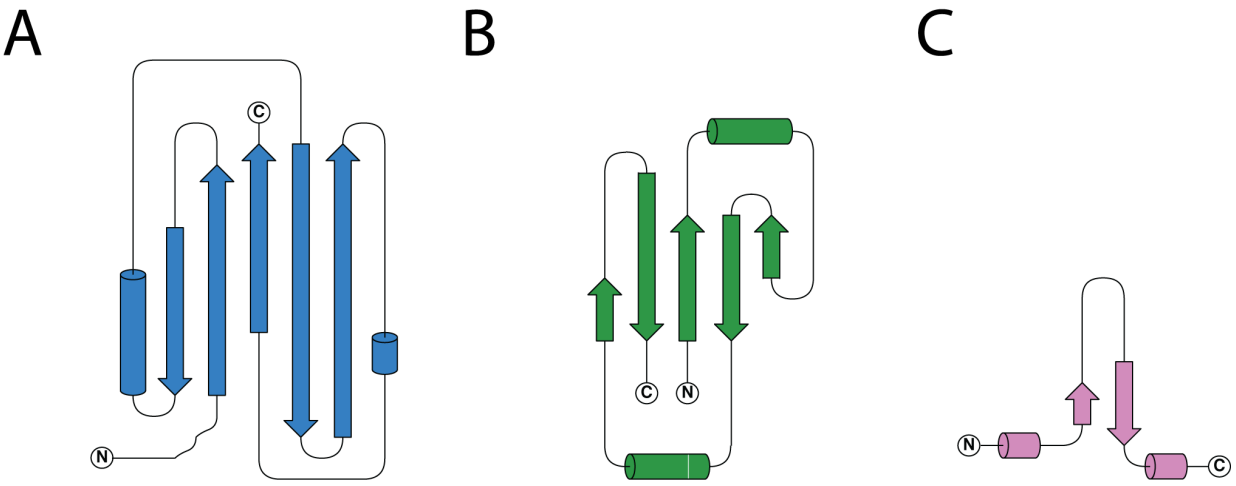


Figure 2-27: Topology of the ASAP components. Schematic drawing of the topology of SAP18 (A), RNPS1 (B) and Acinus (C). The topology of the structure was calculated with stride (Heinig and Frishman, 2004) and schematised using topdraw (Bond, 2003).

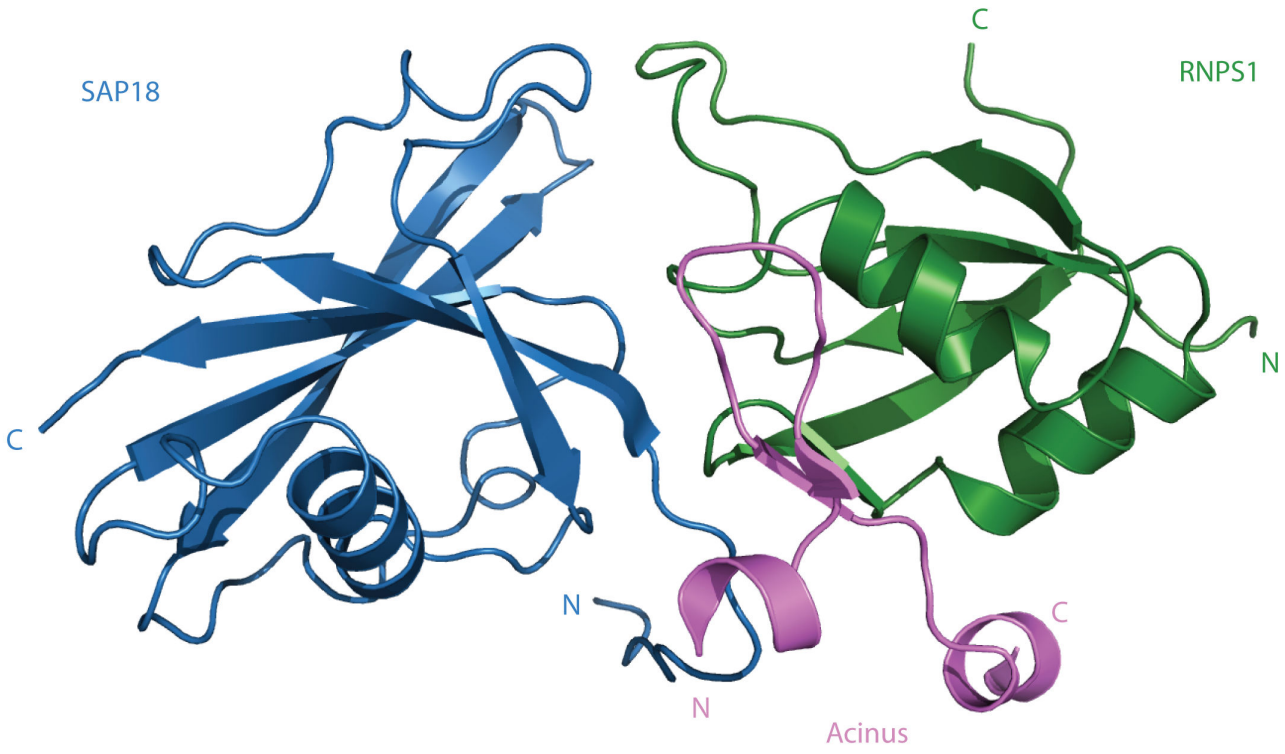


Figure 2-28: Structure of the core ASAP. The complex is shown in cartoon representation, with SAP18 in blue, Acinus in pink and RNPS1 in green.

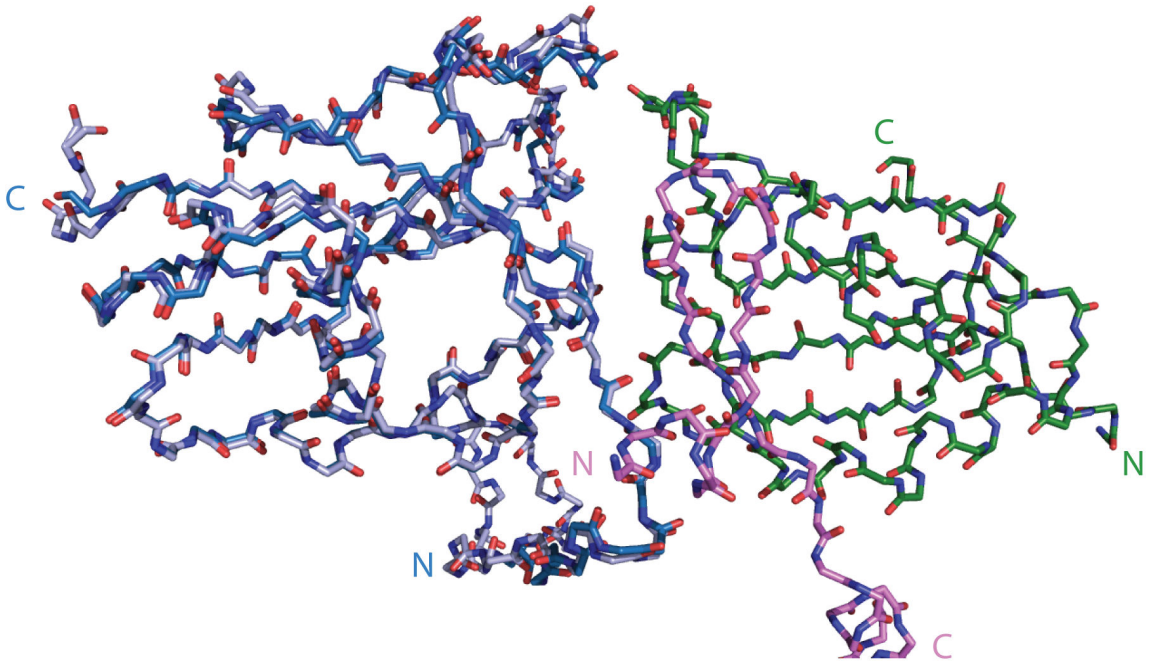


Figure 2-29: No significant conformational change is observed in SAP18 upon binding to the ASAP complex. The structures of apo and complexed SAP18 have been superposed with Pymol and are shown in sticks representation. With the partial exception of the loop between strands 3 and 4, the two structures are perfectly superposed with a RMSD of 0.7 Å. Colours: pink = Acinus; green = RNPS1; dark blue = SAP18$^{14-143}$ (complex); light blue = SAP18$^{6-143}$ (apo).

2.3.8.1 The interaction between Acinus and RNPS1

The major interaction between Acinus and RNPS1 are the hydrophobic contacts between the side chains of Ile^{670}, Trp^{672}, Pro^{674} Leu^{675} of Acinus and those of Val^{172}, Ile^{180}, Met^{223} and Ile^{228} of RNPS1 (Figure 2-30). Notably, all these residues display a high level of conservation in RNPS1 orthologs but not in the other RRM domains (Figure 2-34). The fold of both proteins is crucial for correct positioning of the interacting residues: helix 1 of RNPS1 correctly orients Ile^{180} to pack against Trp^{672}; furthermore, it provides Glu^{179}, Thr^{183} and Tyr^{184}, which enclose the hydrophobic patch within the non-polar portion of their chains. Glu^{179} and His^{176}, which are highly conserved in eukaryotes, are also involved in polar contacts with the main chain of Aci-

nus in the proline turn region; these interactions cement the turn in place and position Ile^{670} of Acinus for a stacking interaction with His^{176}; additionally, Thr^{663} becomes sandwiched between the turn and RNPS1: as a result, the oxygen gamma of the threonine side chain intervenes in strengthening the hydrogen bonding mesh of the turn, whereas the carbon gamma snugly fits in a small crevice between Ile^{670} of Acinus and Ile^{228}, Val^{172}, Glu^{179} and His^{176} of RNPS1. The folding of the turn is also involved in positioning the small beta sheet of Acinus, which interacts with beta strand 5 of RNPS1; as a result, Acinus and RNPS1 form a continuous, extended beta sheet that partially surrounds helix 1 of RNPS1. Finally, the most C-terminal turn of the helix 2 of RNPS1 is embraced by loop 4 and helix 2 of Acinus, with Leu^{675} and Ile^{680} of Acinus interlocking with His^{222} and the polar portion of Lys^{221} on RNPS1. This interaction might help in positioning the invariant Met^{223} in the hydrophobic patch.

2.3.8.2 The interaction between Acinus and SAP18

Two regions of Acinus interact with SAP18: helix 1, which binds to the lock of SAP18, and the C-terminal portion of the turn, which contacts beta strand 1 of SAP18 (Figure 2-31). On the lock, the side chains of Leu^{656}, Leu^{659}, Phe^{660} and Tyr^{671} of Acinus interact hydrophobically with the stacked side chains of Arg^{22}, Pro^{27} and Trp^{61} of SAP18. Above the lock, Leu^{29} and Tyr^{59} of SAP18 interact with Pro^{668} and with the carbonyl oxygen of Cys^{669} respectively. Additionally, the side chain of Cys^{26} from SAP18 is buried beneath beta sheet 2 of Acinus. As a result of these interactions, the lock of SAP18 is tightened in place, and the second beta sheet from SAP18 tightly adheres to the turn region of Acinus. Consistently with the biochemical data, the interface between SAP18 and Acinus is small, burying only 381.2 $Å^2$.

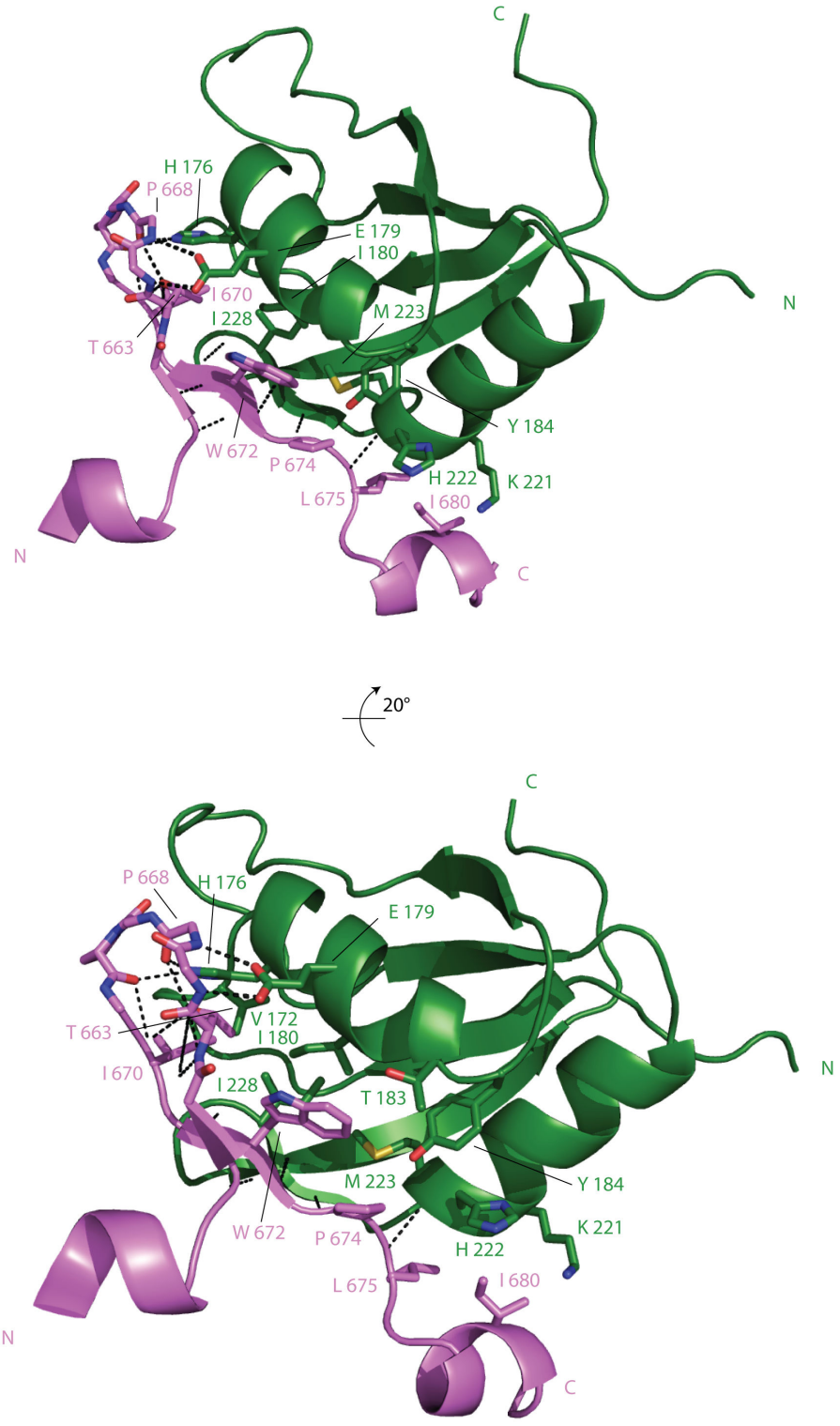


Figure 2-30: The binding interface between Acinus (pink) and RNPS1 (green). N and C indicate the N- and C-termini of the molecules respectively. Two viewing orientations are presented for clarity. Black dashed lines indicate polar contacts as calculated by Pymol.

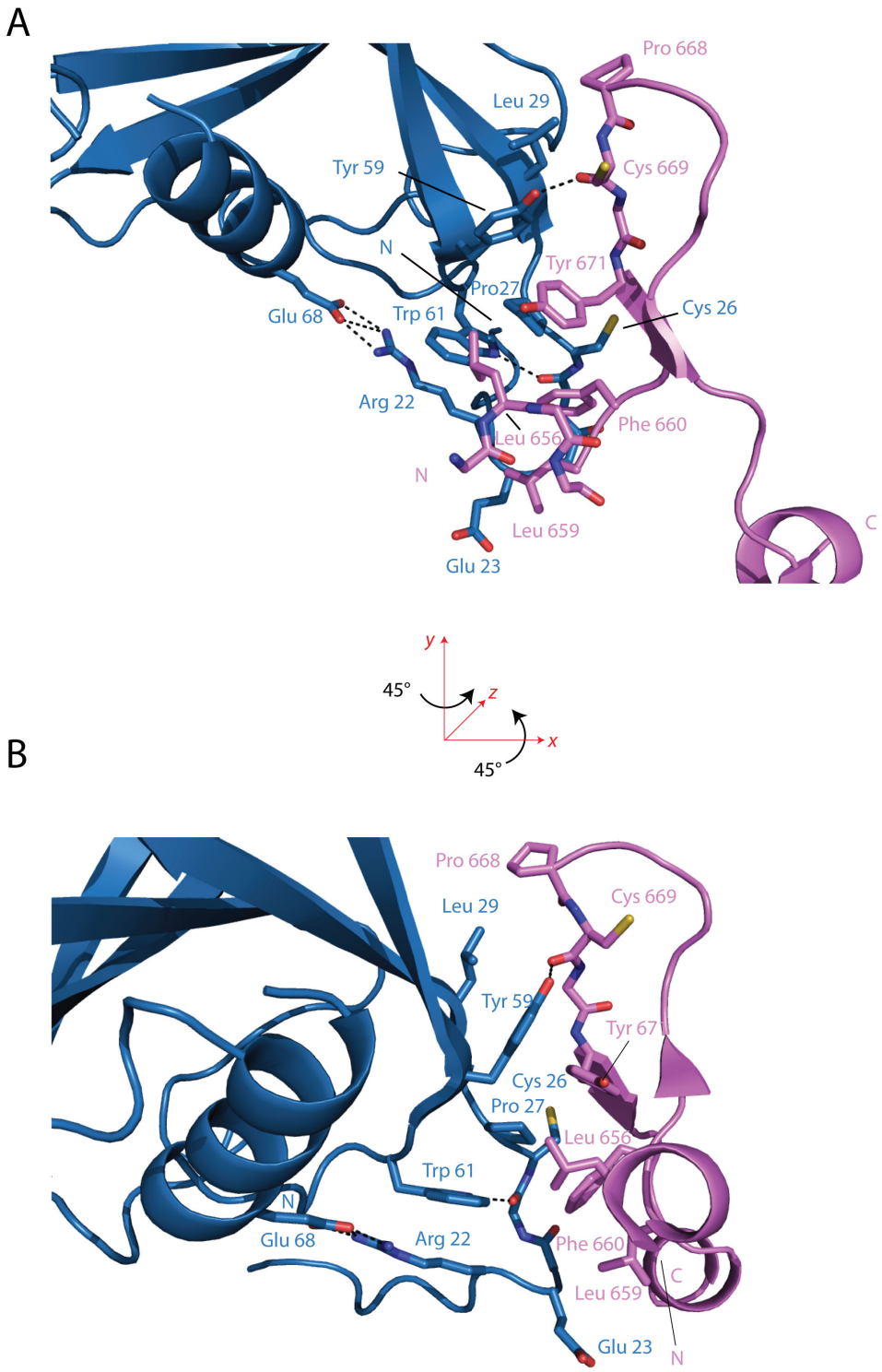


Figure 2-31: The binding interface between SAP18 (blue) and Acinus (pink). N and C indicate the N- and C-termini of the molecules respectively. Two viewing orientations are presented for clarity. Black dashed lines indicate polar contacts as calculated by Pymol.

2.3.8.3 The interaction between SAP18 and RNPS1

RNPS1 interacts with SAP18 through two distinct regions: first, polar contacts are formed with loop 1 and beta strand 2 of SAP18; second, a hydrophobic interaction occurs between loop 3 of RNPS1 and loop 2 of SAP18 (Figure 2-32). In the first interaction, Asp^{229} and Gln^{231} form hydrogen bonds with the main chain atoms of Leu^{29}, Lys^{24} and Cys^{26} of SAP18. Additionally, Asp^{229} uses the second carbonyl oxygen to form intramolecular polar contacts with Thr^{169} and Asn^{171}, thereby helping to position the latter side chain for a hydrogen bond to Leu^{29} on SAP18. All the residues mentioned so far also contribute to forming a flat surface onto which Acinus can bind. Cys^{26} from SAP18 protrudes above this surface and snugly fits between Acinus and Gln^{231} and Gly^{235} of RNPS1. Notable is also the position of Gly^{235} of RNPS1, which is invariant in eukaryotes (Figure 2-2) and which allows loop 4 of RNPS1 to both fit against the lock region of SAP18 and perform the turn that is necessary to position Glu^{235}. The presence of any side chain in this position would probably disrupt the binding to SAP18 and Acinus. The interaction landscape for this patch is completed by a hydrophobic residue of SAP18 (Ile^{129} in mouse), which intervenes to seal in place both Gly^{230} and Thr^{169} of RNPS1. The second interaction surface between RNPS1 and SAP18 is dominated by hydrophobic interactions: Leu^{96}, Met^{94} and Tyr^{101} from SAP18 interlock with His^{198}, Pro^{199}, His^{200} and Leu^{201} of RNPS1. Arg^{170}, whose positioning is facilitated by the hydrogen bonding of Asp^{229}, Asn^{171} and Thr^{169}, seals the hydrophobic patch from below via a stacking interaction. According to PISA, the interaction between SAP18 and RNPS1 buries 575.4 $Å^2$, deemed insufficient for the formation of a stable heterodimer (Krissinel and Henrick, 2007); furthermore, a large part of the interaction is composed by polar contacts, which would be entropically disfavoured in solution.

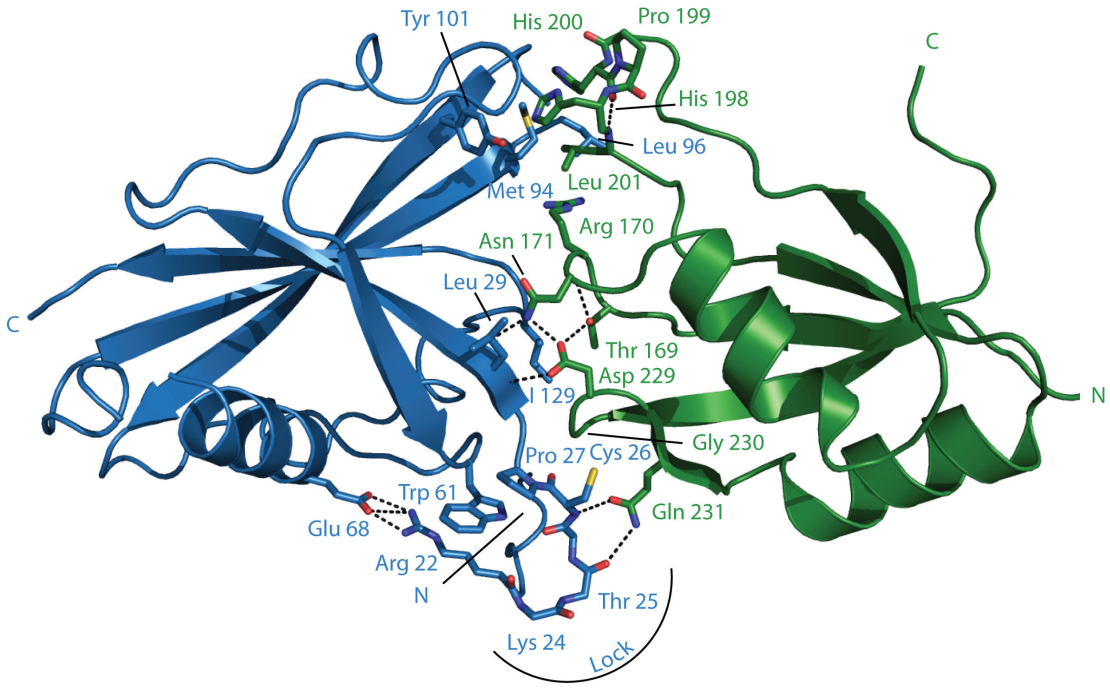


Figure 2-32: The binding interface between SAP18 (blue) and RNPS1 (green). N and C indicate the N- and C-termini of the molecules respectively. Black dashed lines indicate polar contacts as calculated by Pymol.

2.3.9 The interaction surfaces in the ASAP are phylogenetically conserved

The sequence conservation of ASAP can be mapped onto the crystal structure using the program Consurf (Armon et al., 2001) and the multiple alignments in Figure 2-1, 2-2 and 2-3 as input. As shown in Figure 2-33, the surfaces of RNPS1 involved in binding to either SAP18 or Acinus are highly conserved; the sequence conservation observed is not simply due to structural conservation of the RRM domain, since (1) the RRM of RNPS1 shows strong conservation in regions that are not normally conserved in generic RRM domains and (2) these unique regions, and only these, are all involved in direct binding to either SAP18 or Acinus (Figure 2-30, Figure 2-32). A notable ex-

ception to this pattern is the high conservation of an AEKA motif on the solvent exposed face of helix 2. Whereas the conservation of the alanines is observed in the RRM domain consensus, the conservation of the glutamate and the arginine is striking, given that, in the absence of a specific function, any hydrophilic residue is expected to 'fit' in the structure. Whether a binding partner might recognise the surface of this helix is an open question.

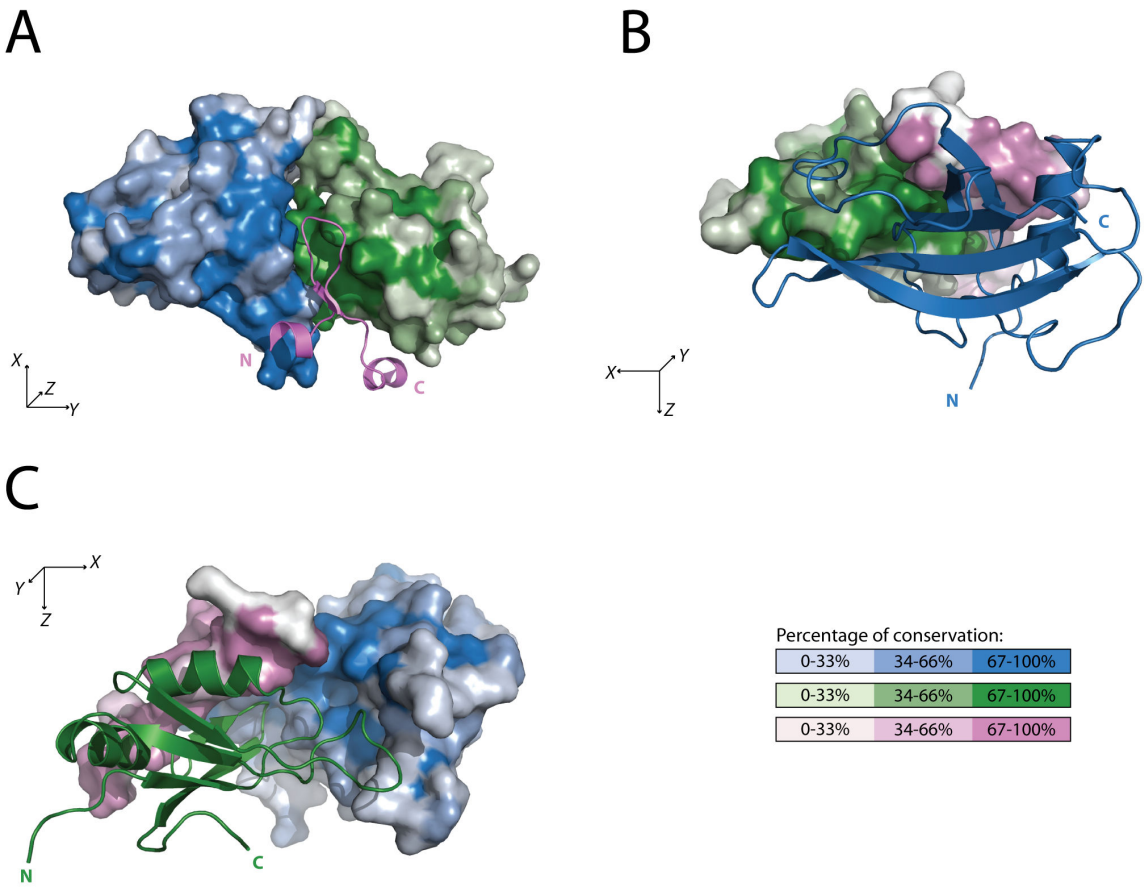


Figure 2-33: The surfaces involved in binding to ASAP are conserved in eukaryotes. The surface conservation of each ASAP component was calculated using Consurf and plotted on the surface representation using Pymol. A, B and C show the binding surfaces recognised by Acinus, SAP18 and RNPS1 respectively. The axes in the lower left corner of each panel indicate the rotations that were performed on the structure to allow informative visualisation. The colour code for the surface conservation is detailed in the legend.

The Consurf plot for Acinus shows high conservation throughout the sequence. The residues involved in binding to ASAP are invariant in eukaryotes, and so are those involved in structuring the turn. The solvent exposed surface of Acinus is also conserved (Figure 3-3), suggesting that it might be involved in interactions with as yet uncharacterised, conserved binding partners. SAP18 also binds Acinus and RNPS1 through a highly conserved surface, centred on the surface of the N-terminal lock.

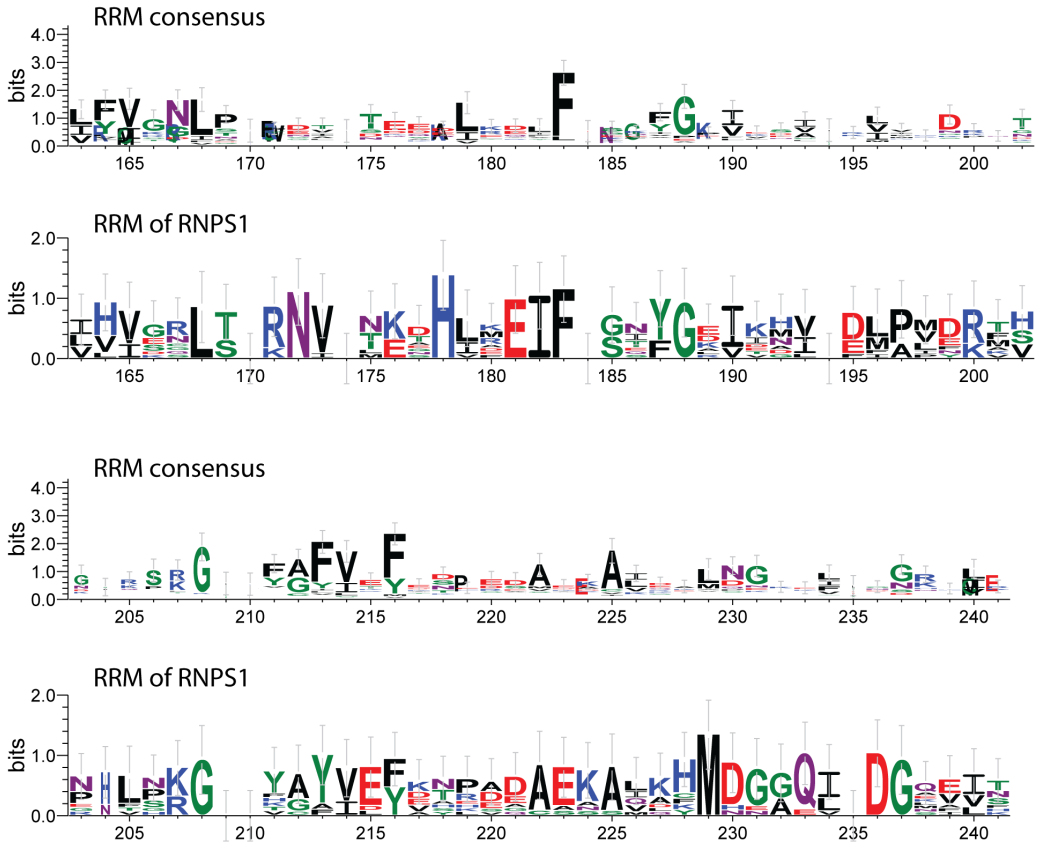


Figure 2-34: The RRM of RNPS1 has a unique conservation profile among RRMs. Sequence profiles for the RRM family and for the RRM of RNPS1 were calculated using weblogo (Crooks et al., 2004). The seed alignments of the Pfam (Finn et al., 2008) family RRM_1 and the multiple sequence alignments in Figure 2-2 were used as queries for the calculations. The Y axes on each profile have different scales for clarity of visualisation.

2.3.10 A contiguous patch of conserved residues is present on the surface of the ASAP complex

Inspection of the surface of ASAP shows that conservation is not equally distributed, but rather a conserved surface patch exists, which spans all three components of the complex (Figure 2-35). Interestingly, the shallow hydrophobic groove on SAP18, which was involved in mediating crystal contacts in the apo structure of SAP18, is part of the conserved patch on ASAP. Overall, these data suggest that (1) this conserved ASAP surface might be involved in binding to an as yet unknown binding partner and (2) this partner might engage the groove on SAP18 with peptides whose sequence resembles that of the N-terminus of SAP18.

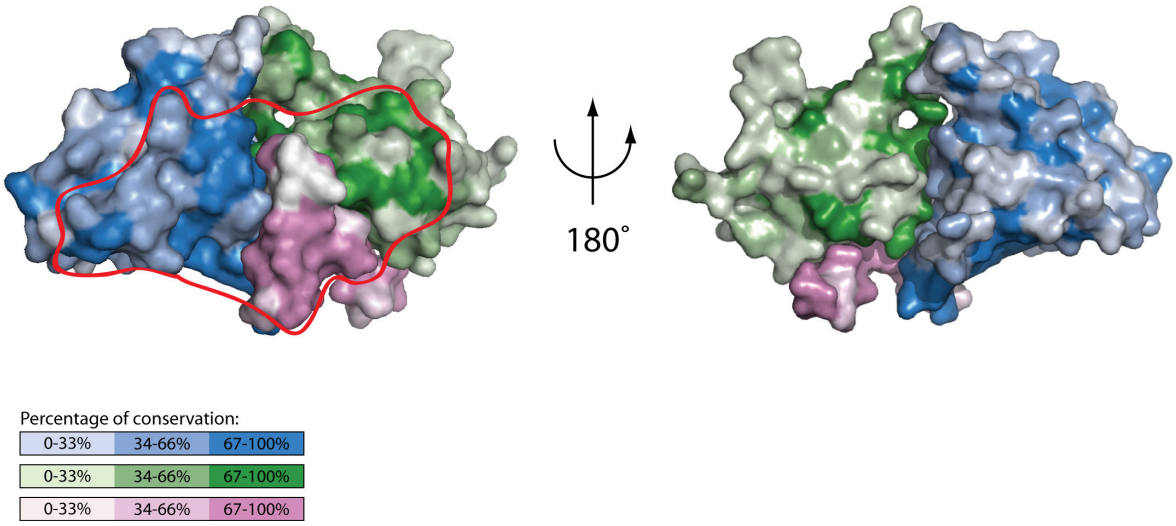


Figure 2-35: Surface conservation of the ASAP. The ASAP complex has been colored by conservation using Consurf. Two views are shown, rotated by 180°. Acinus, SAP18 and RNPS1 form a highly conserved patch (outlined in red), to which an as-yet-unknown binding partner might bind.

2.3.11 Putative ABM motifs can be found in other eukaryotic proteins

Based on the knowledge obtained from both the structure of ASAP and the multiple alignment of Acinus, a database search was attempted in order to locate any protein, which might contain an ABM. In collaboration with Dr. Johannes Söding at the Gene Center in Munich we employed a CSI-BLAST, a context specific, profile-profile comparison algorithm to find putative ABM-containing proteins. ABM motifs were detected in Pinin, Abh6 (Alkb6 homolog 6), LOC148423 and LOC285051 (see Figure 2-36 for a multiple sequence alignment of the putative ABMs).

2.3.11.1 Pinin

The putative ABM of Pinin resides in the conserved C-terminal domain and is conserved in the eukaryotic domain of life. According to secondary structure predictions, the ABM is located in a loop connecting two helixes (Figure 2-37). In HeLa cells, Pinin co-immunoprecipitates RNPS1; additionally, recombinantly expressed GST-Pinin can pull down in-vitro-translated RNPS1. The interaction has been mapped by yeast two hybrid to the C-terminal domain of Pinin (which contains the ABM) and to the RRM of RNPS1 (Sakashita et al., 2004). In Drosophila Pinin co-immunoprecipitates SAP18 in vivo; furthermore, recombinant GST-Pinin binds recombinant SAP18 in vitro pull down experiments (Costa et al., 2006). Taken together, these data strongly suggest that Pinin can use its ABM to form a ternary complex with SAP18 and RNPS1. I propose to name this new putative complex "PSAP" (Pinin, SAP18 and RNPS1). Structural considerations suggest that Acinus and Pinin should be unable to bind concomitantly to RNPS1 and SAP18, i.e. a competition for

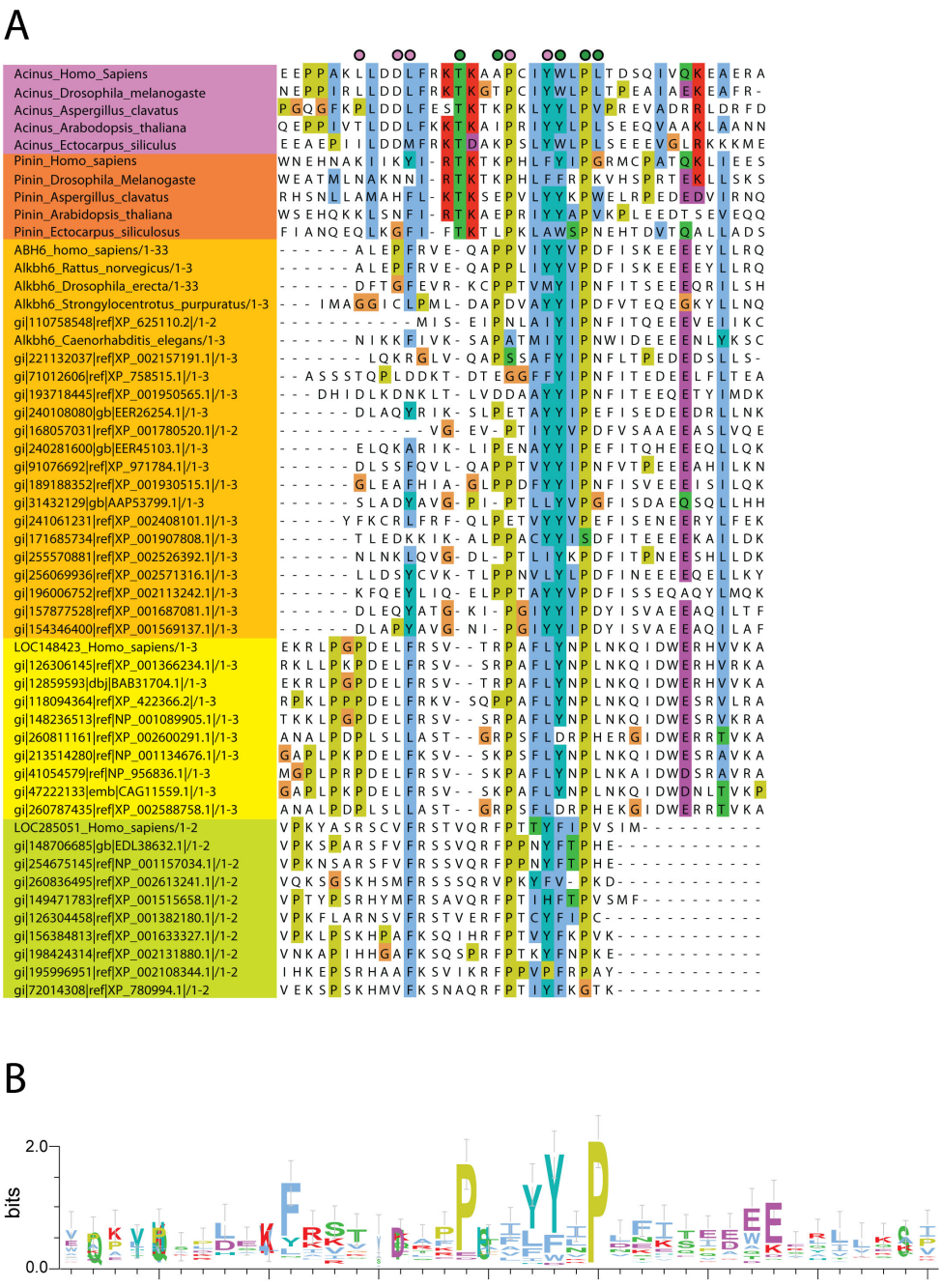


Figure 2-36: ABM motifs can be found in several classes of eukaryotic proteins. A. Multiple alignment of the ABM from Acinus, Pinin, Abh6, LOC148423 and LOC285051. Several orthologs are shown for each protein, with each group separated by colour. The residues interacting with RNPS1 or SAP18 are marked with a green or blue dot respectively. **B.** Sequence profile calculated by Weblogo using the alignment in A as seed. Both the alignment and the sequence profile are coloured according to the ClustalX colouring scheme (Thompson et al., 1997), with conservation cut-off of 15%. The alignment was calculated with MAFFT (Katoh et al., 2005) and displayed with Jalview (Waterhouse et al., 2009).

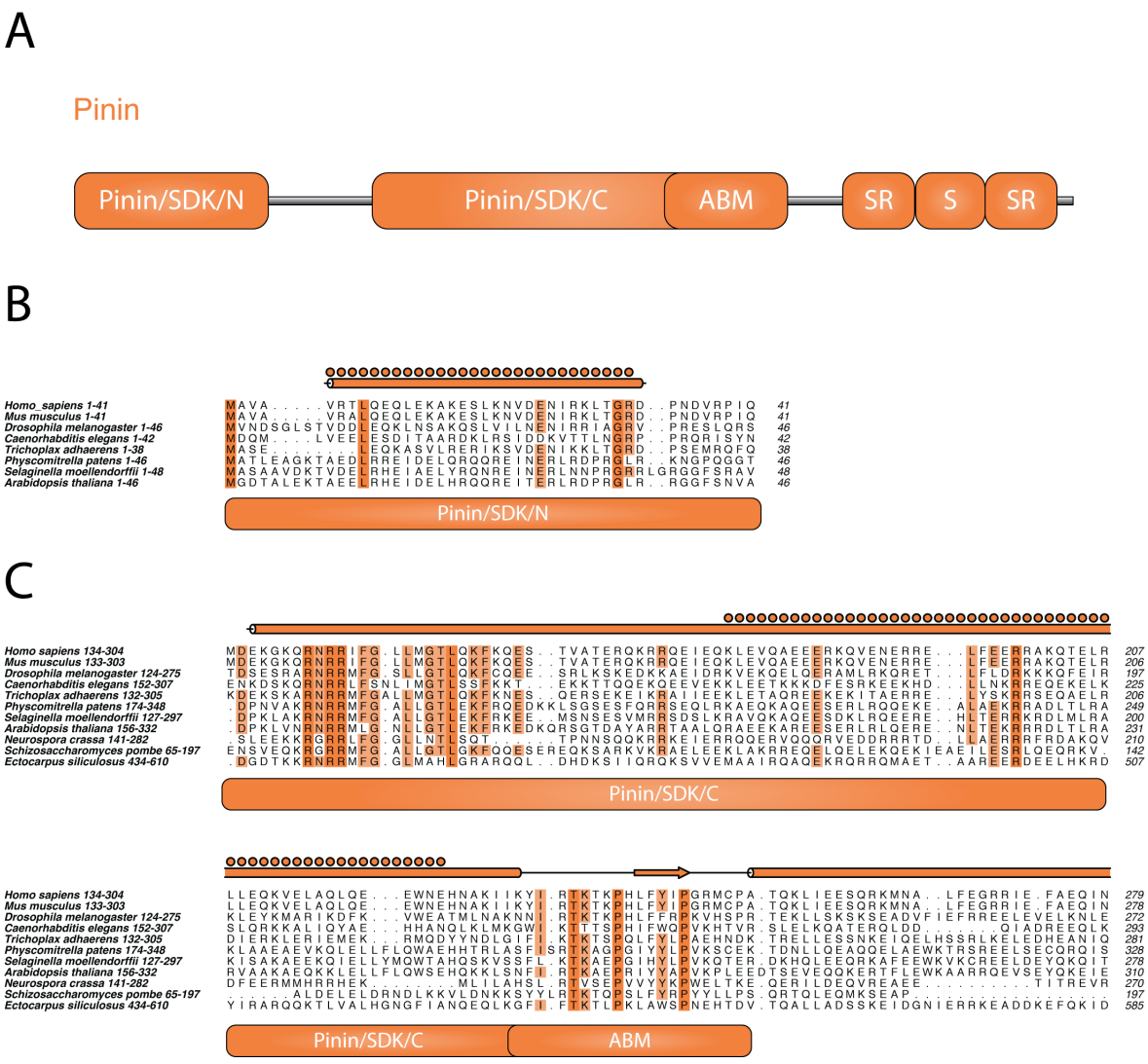


Figure 2-37: Pinin is conserved in eukaryotes and contains an ABM. A. Domain organisation of Pinin as determined by Pfam (Finn et al., 2008) and by my own analysis. Pinin contains two signature domain (Pnn/SDK/N and Pnn/SDK/C) that are only found in this protein – Pnn and SDK being two of the other names under which Pinin has been reported in the literature. **B.** Multiple alignment and secondary structure prediction for the N-terminal conserved domain of Pinin. Conservation outside Metazoan and Plants is poor. **C.** Multiple alignments and secondary structure prediction of the C-terminal conserved domain of Pinin, which contains the putative ABM. This domain is conserved throughout eukaryotes, as evidenced by its presence in the Chromalveolate *Ectocarpus siliculosus* and in the Fungi *Neurospora Crassa and Schizosaccharomyces pombe*. In both alignments, round dots indicate predicted coiled coil regions. The secondary structure prediction was calculated using Jpred (Cole et al., 2008) and human Pinin as input. The multiple alignments were calculated by MAFFT (Katoh et al., 2005) and displayed with Jalview (Waterhouse et al., 2009) and ALINE (Bond and Schüttelkopf, 2009). Accession codes (from top to bottom in alignment C): NP_002678.2, NP_032917.2, CAD56383.1, NP_506247.1, XP_002115763.1, XP_001759199.1, EFJ17467.1, NP_563965.1, XP_957820.1, NP_594893.1, CBJ26439.1.

binding could occur. The possible implications for this discovery are detailed in the Conclusion and Perspectives section.

2.3.11.2 Abh6

E. coli 2-oxy-glutarate-dependent (2-OG-dependent) bi-oxygenase 6 (Alkb6) is a member of the 2OG-Fe(II) oxygenase family, which functions in demethylating 1-methyladenine and 3-methylcytosine and prevents DNA damage by methylating agents (Falnes et al., 2002; Trewick et al., 2002). In humans, eight AlkB homologs (Abh) are known (Abh1-8) (Tsujikawa et al., 2007); of these, Abh1, Abh2 and Abh3 have been shown to demethylate single stranded RNA or DNA *in vitro* and complement for Alkb6 deficiencies in *E. coli* (Lee et al., 2005; Westbye et al., 2008*)*; Abh4, Abh6 and Abh7 lack this activity (Lee et al., 2005). Inspection of the sequence shows that Abh6 possesses an intact catalytic triad and an intact lid (Sundheim et al., 2006), and is therefore expected to be an active enzyme. Recently, several histone demethylases have been discovered, which are 2-OG dependent and contain JMJC (Jumonji C) domains. These enzymes act to demethylate lysines and arginines on histones and are generally associated with transcriptional inactivation (Klose et al., 2006; Tsukada et al., 2006). A member of the family, JMJD6 has also been shown to bind to and to demethylate the spliceosome component U2AF65**, suggesting a possible role in splicing regulation (Chang et al., 2007a; Hong et al., 2010; Webby et al., 2009). Intriguingly, both the JMJC and 2OG-Fe(II) oxygenase domains use 2-OG as a cofactor and share the same fold, hinting at possible functional commonalities. If a model of Abh6 is produced based on the crystal structure of its close homolog Abh3, the ABM can be mapped to the first beta sheet of the structure. Beta sheet 1 interacts with beta sheet 7,

** **U2** small ribonucleoprotein **a**uxiliary **f**actor, **65** KDa subunit

whose sequence is hydrophobic and can be aligned to the N-terminus of SAP18$^{1-143}$. It is tempting to speculate that in presence of SAP18 and RNPS1, the putative ABM could fold into a hairpin, expose beta sheet 7 and allow it to interact with the conserved hydrophobic groove on SAP18 (Figure 2-37). The complex between SAP18, RNPS1 and Abh6 could be involved in histone demethylation or in splicing regulation. Such bold speculation will obviously require experimental confirmation.

2.3.11.3 LOC_148423 and LOC_285051

The remaining hits containing putative ABM, LOC_148423 and LOC_285051 are uncharacterised proteins that are conserved in Vertebrates and in Metazoans respectively. No further information is available for either of these proteins.

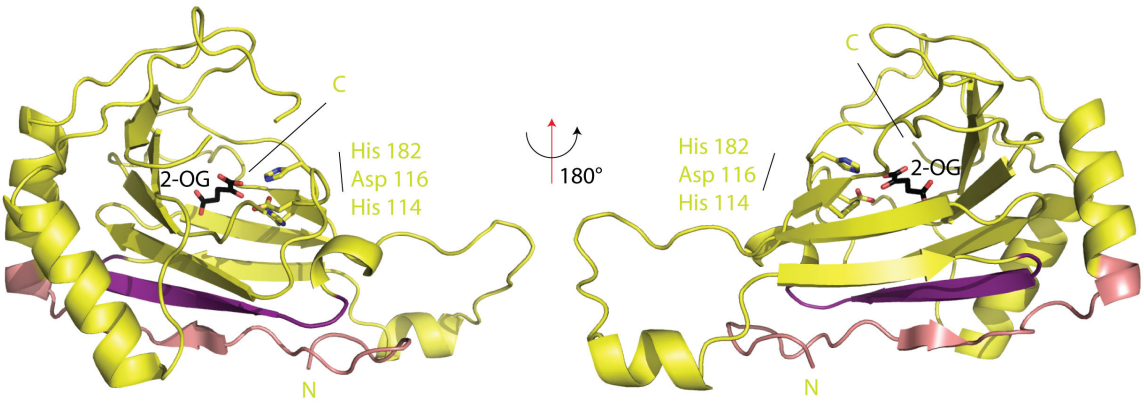


Figure 2-38: Abh6 contains an ABM in its N-terminus. The structural model of Abh6 was calculated by homology modelling based on the structure of the close homolog Abh3. The ABM sequence is coloured in light pink, whereas the beta sheet 7 is coloured in violet. The 2-OG cofactor is shown in black and the catalytic residues are represented as sticks.

3 Conclusions and Perspectives

3.1 Summary

Phylogenetic analysis shows that the ASAP complex is conserved in the whole eukaryotic domain of life, possibly dating back to the last common ancestor (Section 2.1). Such degree of conservation implies a fundamental role for ASAP in cellular metabolism, yet only few hints exist in the literature as to what this role might be. It is known that ASAP can inhibit splicing *in vitro* and that it is a target for the executioner caspase 3 during apoptosis (Schwerk et al., 2003). ASAP has also been shown to associate with the EJC, being deposited by the spliceosome and presumably dissociating before mRNA export, since Acinus is restricted to the nucleus (Tange et al., 2005). Other functions might be inferred based on the roles that have been reported for the isolated components of ASAP. Both Acinus and SAP18 have been implicated in transcriptional control, with SAP18 being part of the Sin3 HDAC complex and Acinus being able to bind to and regulate retinoic acid receptors; RNPS1 (and Pinin) have established roles in splicing, and Acinus can stimulate splicing *in vitro*; RNPS1 is also a known NMD factor. Which of these functions can be ascribed to ASAP, and which are an effect of its separate components moonlighting in other complexes?

To shed light on this question, I have characterised the minimal, core region of ASAP, composed of SAP18, the RRM domain of RNPS1 and the ABM of Acinus, a previously unknown protein-protein interaction motif (Section 2.3.2). I have confirmed that Acinus and RNPS1 form a separate subcomplex *in vitro* (Tange et al., 2005), and that formation of this complex is a prerequisite for the binding of SAP18 (Section 2.3.3). I have solved the crystallographic structure of the core ASAP, showing how RNPS1, Acinus and SAP18 interact within the complex (Section 2.3.8). By analysing the surface of ASAP, I have identified a patch of contiguous, conserved residues that extends over all three components of ASAP and might be involved in binding to

unknown, evolutionary conserved partners (Section 2.3.10). With the confidence afforded by the structural analysis, I have identified putative ABM motifs in several proteins, including Pinin, a component of both the spliceosome and the EJC (Section 2.3.11). The presence of a putative ABM in Pinin suggests that it might be able to form a new, hitherto uncharacterised complex with RNPS1 and SAP18, which I have named PSAP (Pinin, SAP18, RNPS1) (Section 2.3.11.1). Furthermore, I have also solved the crystallographic structures of $SAP18^{1-143}$ and $SAP18^{6-143}$. Both structures differ from the reported NMR structure of SAP18 (McCallum et al., 2006) in that the N-terminus of the molecule partially adheres to the core ubiquitin fold rather than being disordered. The presence of this "lock", whose sequence is invariant in eukaryotes, might have functional implications: the "locked" positioning of the N-terminus is necessary to form the interaction surface to which Acinus and RNPS1 bind (Sections 2.2.7 and 2.3.8). Additionally, in both the SAP18 structures presented here, the crystal packing is mediated by the N-terminus. Residues 1-12 in $SAP18^{1-143}$ and residues 6-12 in $SAP18^{6-143}$ protrude away from the core ubiquitin fold and bind to a hydrophobic groove on a symmetry related molecule. This groove is highly conserved in eukaryotes, suggesting that SAP18 might exploit this interaction surface to bind unknown, conserved functional partners (Section 2.2.8.2 and 2.2.9).

Collectively, the data presented here lead to several predictions, which are amenable to experimental testing and suggest new avenues of research.

3.2 The conserved patch on SAP18: a groove open for interactions

SAP18 is highly conserved in eukaryotes, as shown by the multiple sequences alignment in Figure 2-1. If the protein is coloured by conservation of the surface residues, a contiguous patch of invariant amino acids emerges, which spans the surface of the lock and the shallow hydrophobic groove enclosed between helix 1 and beta sheet 2 (α/β groove, Figure 2-16). In the two structures of SAP18 solved in the present work, this groove is occupied by the N-terminus that protrudes from a neighbouring molecule (Figure 2-16). This particular interaction is driven by the needs of crystal packing and is probably not physiologically relevant (Section 2.2.8.2). However, it is possible that the interaction of the N-terminus of SAP18 might partially mimic the interaction of one or more binding partners of SAP18. In partial support of this hypothesis is the fact that the α/β groove is a common interaction surface in ubiquitin-like molecules such as SUMO (Small Ubiquitin Related Modifier), Atg8 (Autophagy related protein 8), GABARAP (GABA-A-receptor-associated protein) and Elongin B (Babon et al., 2008; Noda et al., 2008; Song et al., 2005; Thielmann et al., 2009) (Figure 3-1). Possible candidates for the binding to the α/β groove would be transcription factors of the Krüppel/Hairy/Knirps/Snail family, to which SAP18 has been shown to bind (Anderson et al., 2009; Matyash et al., 2009; Sheeba et al., 2007); alternatively, Sin3A or other members of the HDAC complex could interact with this surface. The α/β groove of SAP18 is not directly involved in the binding to either Acinus or RNPS1, suggesting that SAP18 might be able to maintain the association with its binding partners upon formation of the ASAP complex (Figure 3-3). The groove is however part of a bigger, contiguous patch of conserved residues that spans all three

components of ASAP (Figure 3-3 and Section 2.3.10). It is therefore also possible that binding partners of ASAP might exploit the α/β groove of SAP18 for their interactions with the complex. To shed light on the issue, in collaboration with Dr. Johannes Söding and Dr. Eckhart Guthöhrlein, we have searched the sequence of several known direct interactors of SAP18 and ASAP for common motifs.

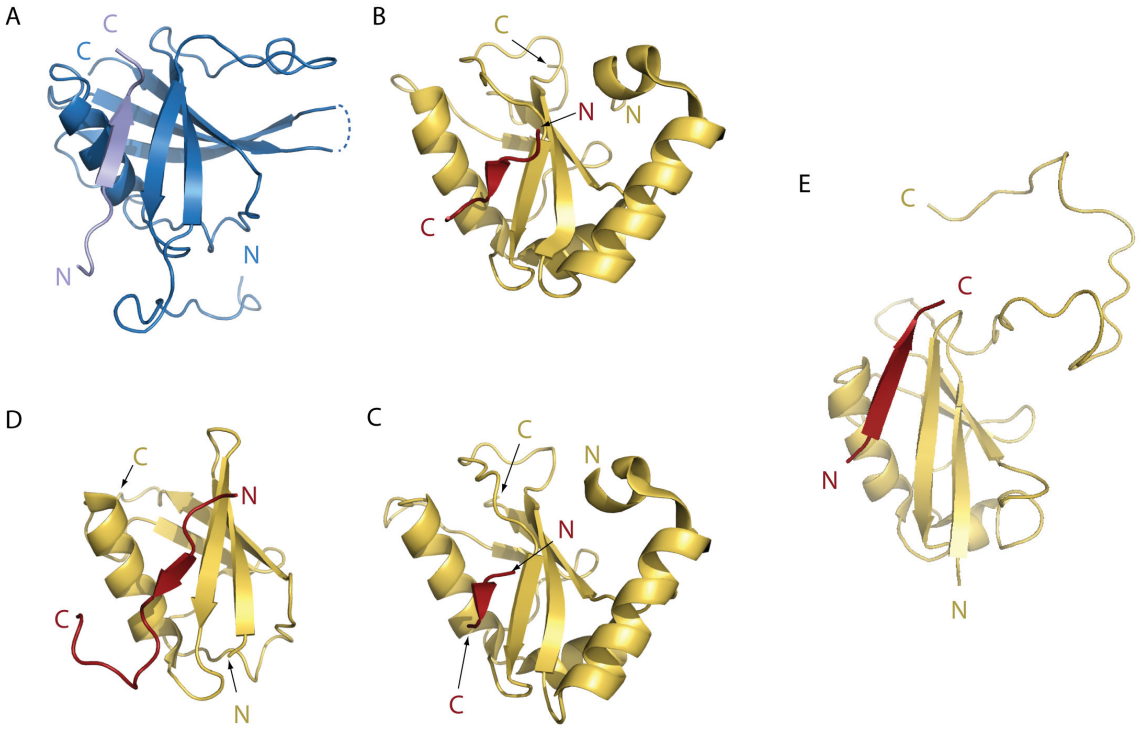


Figure 3-1: The α/β groove is a common interaction surface for ubiquitin-like molecules. A. The interaction of the N-terminus of SAP18$^{1-143}$ in the crystal packing of the crystallographic structure presented in this work. Dark blue: SAP18$^{1-143}$ acceptor molecule. Light blue: N-terminus of the donor SAP18$^{1-143}$ molecule. B. Structure of the GABA A-receptor-associated protein (gold) with a peptide from calreticulin (red) PDB:3DOW (Thielmann et al., 2009). C. Structure of Atg8 (gold) in complex with a peptide from Atg19 (Autophagy-related protein 19, red) PDB: 2ZPN (Noda et al., 2008). D. Structure of SUMO1 (gold) in complex with a peptide from PIAS (Protein inhibitor of Acitveted STAT, red) PDB ID: 2ASQ (Song et al., 2005). E. Structure of Elongin B (gold) in complex with Elongin C (red). PDB ID: 2JZ3 (Babon et al., 2008). For ease of visualisation, only the N-terminal β-strand of Elongin C is shown.

The dataset was manually compiled from the available literature, selecting for those interactions that are most likely direct because they have been validated by pull-down between recombinant or *in vitro* translated proteins. The algorithm used to identify candidate motifs has been designed for unbiased search on yeast two hybrid datasets, which contain possibly indirect interactions (E. Guthöhrlein, unpublished data). Two motifs with significant E-value were found. Their sequence logo representation is shown in Figure 3-2. The first motif is identified with higher confidence (i.e. lower estimated E-value), but is present only in several of the transcription factors used for the search. The second motif is shorter and has a correspondingly lower statistical confidence, but is present in all of the seed sequences. We have synthesised peptides corresponding to each of these motif and we are currently testing their binding to SAP18 by fluorescence anisotropy. Another possible unbiased approach to finding partners of SAP18 and ASAP could be to compare the interactomes of wild type proteins and α/β groove mutants (e.g. in yeast two hybrid or tandem affinity purification + mass spectrometry strategies).

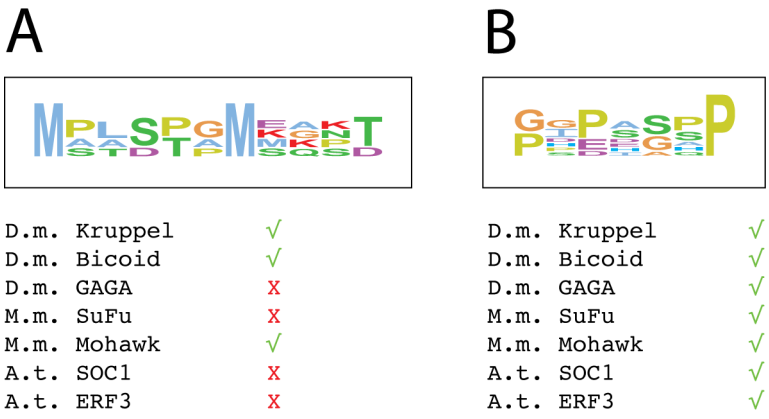


Figure 3-2: Candidate motifs for the interaction with SAP18. The motifs are represented with weblogo-like graphs. The common names of the interactors used in the search are reported underneath the motifs. √ or X indicate whether the motif is present or absent in the given protein.

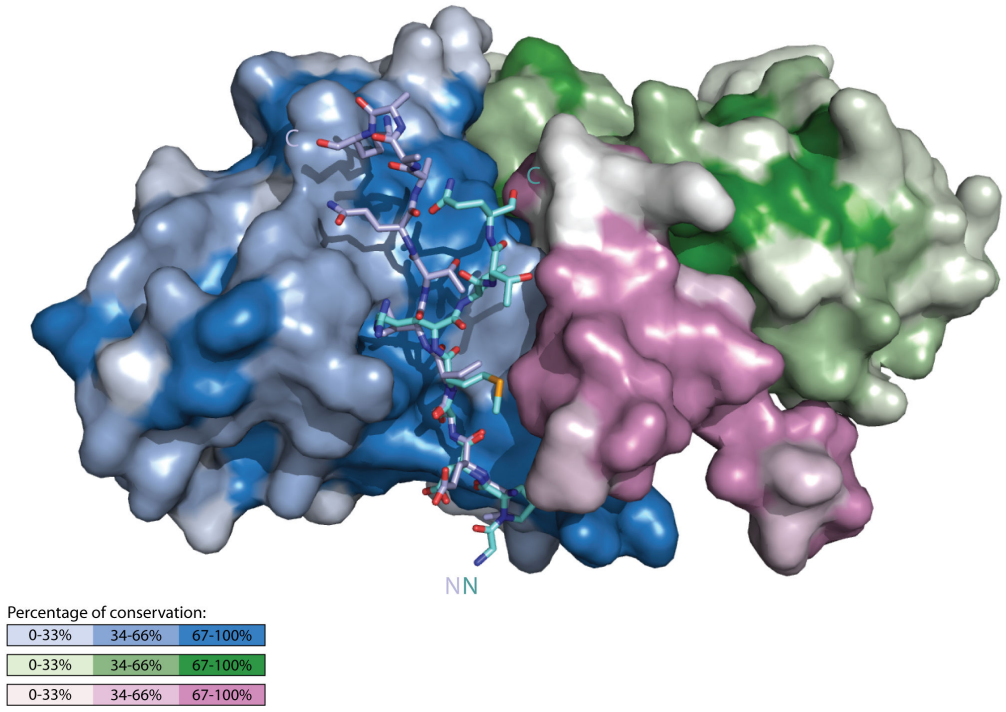


Figure 3-3: The conserved hydrophobic groove of SAP18 can bind amphipatic peptides. In the apo structures of SAP18, each molecule of SAP18 (acceptor) is contacted by the N-terminus of a neighbour in the crystal packing (donor). Here, the ASAP complex has been superposed to the apo structures and is shown in surface representation. The N-termini of donor molecules from both apo structures bind to a conserved hydrophobic patch on SAP18; their binding surface does not collide with the one occupied by Acinus and RNPS1, suggesting that SAP18 can bind at the same time to the ASAP complex and to as yet unspecified, evolutionarily conserved partners. The α/β groove of SAP18 is adjacent to a conserved patch of residues on Acinus and RNPS1.The colour code of the surface representation is detailed in the legend. The N-termini of the donor $SAP18^{1-143}$ and $SAP18^{6-143}$ are coloured in lavender and teal respectively.

3.3 The putative PSAP complex: possible functions

Pinin contains a putative ABM, and is likely to associate with SAP18 and the RRM of RNPS1 to form the PSAP complex (Figure 2-37). Structural considerations suggest that Pinin and Acinus are likely to bind to the same binding surfaces on SAP18 and RNPS1, implying that a competition could occur and that the formation of the PSAP and ASAP could be mutually exclusive. Several independent lines of evidence strongly support the existence of the PSAP: (1) the ABM is present in all Pinin orthologs in the eukaryotic lineages (Figure 2-36); (2) within the ABM of Pinin, the residues that are expected to bind RNPS1 and SAP18 are conserved (Figure 2-36); (3) In HeLa cells, Pinin has been shown to co-immunoprecipitate RNPS1. Additionally, *in vitro*, recombinantly expressed GST Pinin can pull down *in-vitro*-translated RNPS1. The interaction has been mapped by yeast two hybrid to the RRM of RNPS1 and to the C-terminal domain of Pinin (which contains the ABM) (Sakashita et al., 2004); (4) In *Drosophila* Pinin co-immunoprecipitates SAP18 *in vivo.* Furthermore, recombinant SAP18 and recombinant GST-Pinin form a complex in pull down experiments (Costa et al., 2006). To conclusively demonstrate the existence of the PSAP complex, pull-down experiments and/or co-immunoprecipitation will be necessary. Additionally, pull down experiments and isothermal titration calorimetry could also be used to investigate a possible competition between Acinus and Pinin for the binding to RNPS1 and SAP18.

What could be the function of the PSAP complex (were it shown to exist)? Data in the literature hint at a role of the PSAP in transcriptional regulation. SAP18 has been reported to bind to short-range transcription factors belonging to the Krüppel/Hairy/Knirps/Snail family (Matyash et al., 2009; Sheeba et al., 2007). This interaction is thought to result in transcriptional repression. Pinin has been shown to act on

the same class of transcription factors, but with the opposite effect of relieving transcriptional repression (Alpatov et al., 2004; Alpatov et al., 2008). It is tempting to speculate that Pinin might at least partially fulfil its role by acting on SAP18, e.g. by sequestering it in the PSAP or by hindering its interaction with the HDAC complex.

Pinin has also been shown to associate with the EJC and the spliceosome (Merz et al., 2007; Tange et al., 2005) and to functionally interact with RNPS1 in splicing modulation (Li et al., 2003). These data hint at a function of PSAP in splicing, possibly in competition or synergy with ASAP. Both the hypotheses presented above could be addressed by using mutants of Pinin, RNPS1 and SAP18 that selectively disrupt the formation of the PSAP. The structure of the core ASAP presented in this work could serve as a homology modelling template for the design of such mutants.

3.4 ASAP might be involved in splicing

Various lines of evidence suggest that the ASAP complex might play a role in splicing modulation. In humans and *Drosophila*, ASAP co-purifies with active spliceosomes (Herold et al., 2009; Merz et al., 2007) and can modulate the efficiency of splicing *in vitro* (Schwerk et al., 2003); RNPS1 is known to enhance splicing *in vitro* and induce alternative splicing *in vivo* (Sakashita et al., 2004); Acinus and RNPS1 present the typical domain organisation of SR-containing proteins, which are thought to modulate the recognition of splicing sites by the spliceosome (Figure 1-2, Wahl et al., 2009); finally, in eukaryotes, loss of ASAP correlates with intron-poor genomes (Section 2.1.2). Interestingly, a recent report showed that tethering of SAP18 to an intron-containing artificial mRNA promotes alternative splicing by inducing exon inclusion (Singh et al., 2010). Based on multiple sequence alignments with ubiquitin,

Singh and co-workers designed point mutants of SAP18 that are unable to induce alternative splicing in tethering assays. Despite not being localised to the ASAP binding interface, these point mutations also impaired the ability of SAP18 to co-immunoprecipitate RNPS1 and Acinus, suggesting that the ability of the artificially tethered SAP18 to modulate splicing might depend on the recruitment of ASAP. Altogether, the data presented above hint at a role of ASAP in alternative splicing. Further evidence suggests that this role is performed in the context of the EJC: two recent reports showed that depletion of the core EJC components Y14, Mago and eIF4AIII led to a decrease in the levels of MAPK (mitogen activated protein kinase) in *Drosophila* (Ashton-Beaucage et al., 2010; Roignant and Treisman, 2010). This depletion was due to the aberrant splicing of a long intron in the primary MAPK transcript (Ashton-Beaucage et al., 2010). The effect of EJC depletion is not limited to the MAPK gene: both studies showed that depletion of Mago reduces mRNA levels for genes that contain long introns. Interestingly, the effects of the depletion of the EJC core can be recapitulated by the simultaneous knockdown of RNPS1 and Srm160, another SR protein associated with the EJC. Single depletion of either RNPS1 or Srm160 led to a milder effect, suggesting that the two proteins might have partially redundant functions in this context. Altogether, these data suggest that the EJC plays a role in the recognition of long introns in *Drosophila*, and this role is partially dependent on RNPS1.

Taken together, the data discussed above prompt to speculate that the EJC might exert its influence on splicing by means of the ASAP complex. In this scenario, the EJC and ASAP would be deposited onto the maturing mRNA by the spliceosome. The presence of an EJC:ASAP complex upstream of a long intron would help the spliceosome recognise the intron, leading to the inclusion of the downstream exons (Figure 3-4). In this context, the artificial tethering of SAP18 to splicing reporter constructs by Singh and co-workers would artificially mimic the recruitment of ASAP by the EJC. With the knowledge afforded by the structure presented in this work, it should be pos-

sible to design point mutants of RNPS1, SAP18 and Acinus that impair the formation of the ASAP complex. If indeed the EJC fulfils its role in splicing by means of ASAP, then these mutants should recapitulate the effects of EJC depletion on the splicing of *Drosophila* genes that contain long introns. Similarly, in artificial tethering systems like the one employed by Singh and co-workers (Singh et al., 2010), disruption of the ASAP complex should impair exon inclusion.

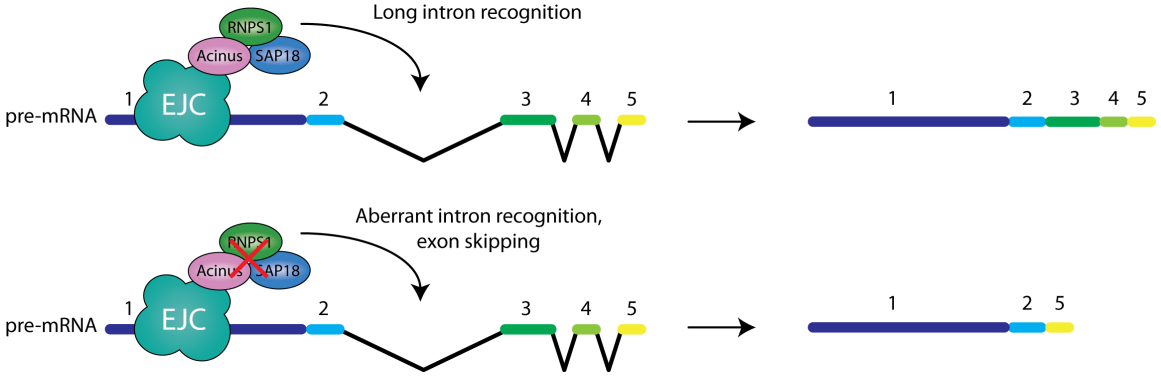


Figure 3-4. ASAP might function in EJC-mediated alternative splicing: I propose that the effect of the EJC on splicing might be mediated by ASAP. **A.** Upon splicing, the EJC is deposited on exon-exon junctions, along with ASAP. The presence of the 5′ EJC helps the spliceosome recognise the long intron, leading to the inclusion of the 3′ exons. **B.** If the role of the EJC is dependent on ASAP, disruption of the latter by structure-based rational mutagenesis would result in misrecognition of long introns and exon skipping.

3.5 Further open questions and outlook

The functions of ASAP in mRNA metabolism seem to be intimately connected to those of the EJC (Section 3.4, (Gehring et al., 2005; Lykke-Andersen et al., 2001; Tange et al., 2005)). Determining how ASAP interacts with the EJC is therefore going to be crucial to advancing our understanding of the functions of both complexes. Pull-down studies by our collaborators Herve' Le Hir and Lionel Ballut have already shown that full-length ASAP can interact with the EJC directly, and this interaction is mediated by Acinus (not shown). We will expand this preliminary result to characterise the minimal EJC-binding regions of ASAP. The deletion constructs of Acinus, RNPS1 and SAP18 produced during this work will likely prove useful in this context.

SAP18 seems to fulfil two separate, disconnected roles: on one hand, it is involved in chromatin condensation and transcription regulation; on the other hand, it participates in the ASAP complex and possibly splicing. The connection between the two functions is at present poorly understood. The data presented in this work suggest several possible approaches to identifying this missing link. First, structural and evolutionary considerations suggest that SAP18 could interact with its partners through the α/β groove. Bioinformatic analysis and mutational studies could allow the identification of these partners and potentially shed insight into how SAP18 performs its multiple roles. The putative PSAP complex might also constitute one of the missing links between transcription and mRNA metabolism. Understanding its functions and its connection with the ASAP complex would definitely advance our knowledge of mRNA metabolism. Finally, the ability to design structure-based mutants of the ASAP complex will likely help understanding the role of SAP18 in splicing.

4 Materials and Methods

4.1 Cloning procedures.

4.1.1 PCR amplification

4.1.1.1 Template DNA

Constructs encoding full length human Acinus L, full length human RNPS1 and full length mouse SAP18 were a generous gift from Dr. Herve' Le Hir. These were used as a template for the amplification of all the truncation constructs.
Drosophila melanogaster Acinus, RNPS1 and SAP18 constructs were amplified from cDNA, kindly provided by Dr. Christian Benda.

4.1.1.2 Primer design:

Amplification primers were designed either manually, using the software ApE (Advanced Plasmid editor, Davis, 2009) for alignment and the website Oligocalc (Kibbe, 2007) for the calculation of melting temperature or automatically using the CCD (Crystallisation Construct Designer) java applet developed by the Netherlands Cancer Institute (Mooij et al., 2009b). For the calculation of the melting temperature of the primers, only the sequence that specifically anneals to the target was considered (i.e. not the restriction enzyme or LIC overhangs). All the primers have melting temperatures between 58 and 60 degrees Celsius.

4.1.1.3 Amplification protocol:

All the construct used in this work were amplified using a standard PCR protocol:

PCR reaction			Amplification protocol	
Primers	100 pM	0.5 µl each		
$MgCl_2$	100 mM	1 µl	**18-25 cycles:**	
Phusion HF buffer	5x	10 µl	30 sec	95°C
DNA template	0.1-0.2 g/l	0.5 µl	60 sec	58°C
Phusion DNA Polymerase	0.5 U/µl	1 µl	60" for each kb	72°C
dNTPs	100 µM	1 µl	**Final extension:**	
ddH_2O		36 µl	180"	72°C

Human Acinus constructs were amplified using a modified protocol, which reduced non -specific amplification due to internal sequence repeats and GC content:

PCR reaction			Amplification protocol	
			5 cycles:	
Primers	100 pM	0.5 µl each	30 sec	95°C
$MgCl_2$	100 mM	1 µl	60 sec	55°C
Phusion GC buffer	5x	10 µl	60 sec for each kb	72°C
DNA template	0.1-0.2 g/l	0.5 µl	**18 cycles:**	
Phusion DNA Polymerase	0.5 U/µl	1 µl	30 sec	95°C
dNTPs	100 µM	1 µl	60 sec	60°C
ddH_2O		36 µl	60 secfor each kb	72°C
			Final extension:	
			180 sec	72°C

4.1.2 Cloning in pGEX vectors:

These constructs were cloned using standard molecular biology techniques (Sambrook J et al., 1989). Briefly, the desired construct was amplified with primers containing 5′ and 3′ extensions coding for the appropriate restriction enzymes. The insert was then gel purified and digested with restriction enzymes according to the manufacturer instructions. After digestion, the insert was purified with commercial PCR purification kits (Qiagen GmbH). The pGEX vectors (GE healthcare) were digested with the appropriate restriction enzymes in a total volume of 20 µl and according to the enzyme manufacturer instructions; After completion of the digestion, 3 µl of 10X phosphatase buffer, 6 µl of ddH_20 and 1 µl of Arctic phosphatase (NEB) were added to restriction reaction and allowed to incubate for further 45′ at 37°C. The dephosphorylated vector was then gel purified using commercially available kits. Digested vector and insert were then mixed to a molar ratio of 1:5, for a total DNA concentration in the ligation reaction of 5 µg/µl; 1 µl T4 ligase (NEB), 2 µl of 10X ligase buffer and water to a final volume of 20 µl were added, and the reaction was allowed to incubate for 30 minutes at room temperature. As a negative control, a reaction lacking the insert was also set up. 2 µl of the ligation reaction or the negative control were then used to transform electrocompetent *E. coli* XL1Blue or chemically competent *E.coli* TOP10 (Invitrogen). The success in cloning was assessed by visually estimating the ratio of transformants generated by the ligation and those generated by the negative control. If the ratio was higher than ~3, three to five positive clones were picked and amplified for plasmid DNA extraction and sequencing. In case of less favourable ratios (i.e. < 3), DNA was extracted from 5-10 clones and tested by restriction analysis prior to sequencing.

4.1.3 Cloning in pEC vectors

The pEC series of vectors was generated by Jerome Basquin and Florence Martin in our laboratory. It is based on the skeleton of a pET vector and is specifically designed for Ligation Independent Cloning (LIC) (Aslanidis and de Jong, 1990)

The LIC method relies on two principles: (1) In addition to the 5′-3′ polymerase activity, T4 DNA Polymerase possesses a 3′-5′ exonuclease activity, which is normally used to excide nucleotides erroneously incorporated during the replication. In the absence of deoxynucleotides triphosphates (dNTPs), the polymerase activity is inhibited and T4 DNA polymerase is able to selectively degrade the 3′-5′ filament in a DNA duplex. If only one dNTP is present, for instance dATP, the polymerase will degrade the 3′-5′ filament until it encounters an adenosine in the sequence, at which point it will engage in a futile cycle of removal and addition of the adenosine until either the dATP runs out or the polymerase ceases to function. (2) A linearised vector and an insert that display compatible cohesive ends longer than 10 base pairs can form a stable circularised plasmid even in the absence of ligase activity. The discontinuities in the resulting plasmid do not hinder bacterial transformation and can be repaired by *E. coli* (Aslanidis and de Jong, 1990; Aslanidis et al., 1994). In the LIC system, every construct is amplified using oligonucleotide primers that contain specific 12 bp long overhangs outside of the regions complementary to the template. This results in an amplification product that displays double stranded overhangs. The amplification product is then treated with T4 DNA polymerase (Novagen) in the presence of a single dNTP (in this case dATP), resulting in the formation of 12 bp long sticky ends. The destination vector is linearised by digestion with a restriction enzyme that cuts in the LIC cassette (SacII in the pEC system). Subsequently, cohesive ends are produced by T4 DNA polymerase treatment in presence of dTTP. The 5′ and 3′ cohesive ends on the vector

are not compatible, blocking its recircularisation. The cohesive ends on the insert and the pEC vector are compatible and can anneal, forming a stable, circular product. Vector and insert are therefore mixed and incubated at room temperature for 10 minutes. The mixture is then treated with EDTA and used to transform competent *E.coli*. Due to the very low rate of plasmid recircularisation, the number of false positives is usually low.

LIC Protocol:

Vector digestion:

pEC vector		2 µg
SacII (NEB)	20 U/µl	60 U
NEBuffer 4	10x	10 µl
ddH_2O		to 100 µl

Digest 2h at 37˚C

Gel purify the linearised vector.

Insert T4 digestion		
PCR product		600 ng
T4 Pol. buffer	10x	2 µl
dATP	25 mM	2 µl
DTT	100 mM	1 µl
T4 DNA pol.	2.5 U/µl	0.4 µl
ddH_2O		to 20 µl

Vector T4 digestion		
linearised vector		450 ng
T4 Pol. buffer	10x	3 µl
dTTP	25 mM	3 µl
DTT	100 mM	1.5 µl
T4 DNA pol.		0.6 µl
ddH_2O	2.5 U/µl	to 30 µl

Incubate 30 minutes at room temperature.

Heat inactivate the polymerase for 20 minutes at 75˚C

Mix 2 µl of insert with 1 µl of vector; incubate 10 minutes at room temperature

Add 1 µl of 100 mM EDTA; incubate 10 minutes at room temperature

Use 2 µl to transform competent *E.coli*.

4.2 E. coli and baculovirus expression constructs

Construct	Organism	Tag	Resistance	Vector
Acinus$^{770\text{-}1250}$	*H.s.*	GST	Amp	pGEX2T-TEV
Acinus$^{770\text{-}1230}$	*H.s.*	6xHis	Amp	pEC-His-TEV
Acinus$^{846\text{-}1107}$	*H.s.*	GST	Amp	pEC-GST-TEV
Acinus$^{1107\text{-}1250}$	*H.s.*	GST	Amp	pEC-GST-TEV
Acinus$^{1171\text{-}1250}$	*H.s.*	CBP-6xHis	Kan	pEC-CBP-TEV
Acinus$^{616\text{-}710\ Dm}$	*D.m.*	6xHis	Kan	pEC-His-3C
Acinus$^{648\text{-}688\ Dm}$	*D.m.r*	6xHis	Kan	pEC-His-3C
SAP18	*M.m.*	6xHis	Amp	pEC-His-TEV
SAP18$^{1\text{-}143}$	*M.m.*	6xHis	Kan	pEC-His-3C
SAP18$^{6\text{-}143}$	*M.m.*	6xHis	Kan	pEC-His-3C
SAP18$^{14\text{-}143}$	*M.m.*	6xHis	Kan	pEC-His-3C
SAP18$^{12\text{-}149}$	*M.m.*	6xHis	Kan	pEC-His-3C
RNPS1$^{159\text{-}244}$	*H.s.*	6xHis	Kan / Strp /Amp	pEC-His-TEV
RNPS1$^{159\text{-}244}$	*H.s.*	CBP-6xHis	Kan	pEc-His-TEV

Construct	Organism	Tag
RNPS1	*H.s.*	TAP-6xHis
RNPS1	*H.s.*	6xHis

Legend: Kan = Kanamycin; Amp = Ampicillin; Strp = Streptomycin. *H.s.: Homo sapiens; D.m.: Drosophila melanogaster; M.m.: Mus musculus.*

Acinus648-688 was cloned by Petra Birle and Tatyana Krywcun. The baculovirus constructs were a kind gift of Dr. Herve' Le Hir and Dr. Catherine Tommasetto

4.3 Protein purification:

4.3.1 Chromatography materials

Unless otherwise specified, all chromatography material were produced by GE health-care (formerly Amersham)

4.3.2 Commonly used buffers:

This section details the composition of the buffers commonly used during the purification procedures discussed in the following sections. Since most of the purification steps were carried out on AKTA purification systems, the washing steps were carried using mixtures of buffers A and B. In the following sections, the buffer mixture will be noted in percentage of buffer B (e.g. the notation 20% buffer B implies that 4 parts of buffer A and 1 part of buffer B were premixed by the AKTA and applied on the column.)

Lysis buffer:
300 mM NaCl, 10% glycerol, 20 mM trisCl pH 7.5, 2 mM DTT,
DNAse I 1 μg/ml

HisBuffer A:
300 mM NaCl, 10% glycerol, 20 mM trisCl pH 7.5

HisBuffer B:
300 mM NaCl, 10% glycerol, 20 mM TrisCl, 500 mM imidazole pH 7.5

Buffer 7A

100 mM NaCl, 10% glycerol, 20 mM HEPES pH 7.5

Buffer 7B

1000 mM NaCl, 10% glycerol, 20 mM trisCl pH 7.5

Buffer6A

100 mM NaCl, 10% glycerol, 20 mM MES pH 6.0, 2 mM $MgCl_2$

Buffer 6B

1000 mM NaCl, 10% glycerol, 20 mM MES pH 6.0, 2 mM $MgCl_2$

4.4 E.coli cultures:

4.4.1 Expression tests

For all the constructs in the present work, expression was tested in the strains BL21 (DE3) PLysS, BL21 (DE3) Rosetta II PLysS and BL21 (DE3), using several induction temperatures (18°, 22° and 37°) and induction times (3h, overnight). Expression levels were assessed by small-scale pull down experiments with the appropriate affinity resin. Most of the constructs express better with overnight induction of 0.5 mM IPTG at 18° in BL21 (DE3) PLysS, and this condition has been used for large-scale expression unless otherwise noted.

4.4.2 Expression media

Normal protein expression was carried out in Terrific Broth (Sambrook J et al., 1989). Selenomethionine derivatised proteins were expressed in LAMM (Levulose Augmented Minimal Medium) medium, which was developed by the author in collaboration with John Weir, a fellow PhD student in the lab. The LAMM is a modification of the recipe published by Sreenath et al, 2005.

LAMM medium (1L)

980 ml LAMM base

20 ml 50x sugar stock

50 mg Selenomethionine

Antibiotic as needed.

LAMM Base:

17 g Yeast Nitrogen Base (YNB, Invitrogen)

5 g amino acids mix (obtained by mixing 25 g of each amino acid except methionine)

1.15 g KH2PO4

64.1 g Na2HPO4:2H20

Autoclave 120°, 20 minutes

50x Sugar stock:

100 g Glucose

300 g Fructose

Sterile filter 0.22 μm.

4.4.3 Transformation

Chemically competent or electrocompetent *E. coli* was transformed with 0.5 µl of plasmid DNA (~50 ng) using standard protocols (Sambrook J et al., 1989) and grown on agarose plates overnight at 37˚C. Transformants were selected with chloramphenicol (34 µg/ml), kanamycin (50 µg/ml), ampicillin (110 µg/µl) or streptomycin (50 µg/ml), as required.

4.4.4 Large scale culture

Expression was typically carried out in 2.5 litre Tunair shake flasks (Sigma-Aldrich), each containing 0.5 litres of Terrific Broth (TB) (Sambrook J et al., 1989). The appropriate antibiotics were added at the same concentration used in the agar plates. The *E. coli* BL21 (DE3) PLysS strain was used for all the expressions. The culture was inoculated as follows: 12 ml of TB were poured on the freshly transformed agar plate, and the bacteria were resuspended using a sterile cell scraper. The resuspended bacteria were then collected from the plate and aliquoted in the culture, typically 1 ml per flask (i.e. a 1:500 dilution) The culture was grown at 37°C under continuous agitation (220 rpm) until it reached an optical density of 2.0 – 2.5, at which point the temperature was reduced to 18°C; After complete cooling (~30 minutes), the expression of the recombinant protein was induced with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) for a duration of 16-20 hours. Cells were then harvested by centrifugation (8000xg, 10') and either used directly or flash-frozen in liquid nitrogen and stored at -80 °C for future use. Selenomethionine derivatised proteins were expressed similarly to the native proteins, with the exceptions that: (1) LAMM medium was used instead of

TB; (2) the *E.coli* methionine auxotrophic strain B834 was used and (3) induction was performed at OD 0.5 – 0.7.

4.5 Insect cells cultures

Hi5 T*richoplusia ni* cells (Invitrogen) were grown in suspension at 25˚C, in serum free medium, until they reached the density of $5x10^6$ cells / ml. 100 ml of cells were then transferred to a new flask, supplemented with 50 ml of P3 viral stock and diluted with fresh serum-free medium to a final volume of 500 ml and a final concentration of $1x10^6$ cells / ml. The culture was grown for 72 hours at 25˚C with 80 rpm shaking. The cells were then harvested by centrifugation (20 minutes at 1000xg) and flash frozen with liquid nitrogen.

4.6 Protein purification protocols:

4.6.1 Purification of SAP18 and its deletion constructs.

This purification protocol applies to all the SAP18 constructs used throughout the present work, namely full length SAP18, $SAP18^{1-143}$ and $SAP18^{6-143}$. Recombinant, hexahistidine-tagged SAP18 was produced in *E.coli* BL21(DE3) PLysS according to the standard protocol detailed above, except for the fact that the induction temperature was 22˚C. Typically 3 litres of cultures were used for each preparation. Bacterial pellets were resuspended in lysis buffer and lysed by French press (Avestin C3, 2 passes at 15000 psi). The lysate was clarified by centrifugation (1h, 75000 x G) and applied onto two 5 ml HisTrap FF columns connected in series. The columns were subse-

quently transferred to a ÄKTA prime system, washed with 10 Column volumes (CV) of HisBuffer A, followed by 10 CV of 14% HisBuffer B. The protein was eluted with 80% HisBuffer B, transferred to a dialysis bag (Spectrapor; Molecular weight cut-off 8000 Da) and dialysed overnight against buffer 6A + 2mM DTT, 1 mM EDTA. The hexahistidine tag was removed by 3c protease cleavage ($SAP18^{1-143}$ and $SAP18^{6-143}$) or TEV cleavage (full length SAP18). The dialysate was subsequently collected and centrifuged 10' at 3220 x G to remove the copious precipitate that formed during dialysis. In the case of full length SAP18, the soluble fraction would be further incubated for one hour at room temperature to increase TEV cleavage efficiency The cleavage mixture was then supplemented with HEPES pH 7.5 to a final concentration of 50 mM, followed by a passage on a 5 ml HisTrap to remove any remaining uncleaved fusion protein. For $SAP18^{1-143}$ and $SAP18^{6-143}$ this was not necessary, owing to the much higher activity of 3c protease. The soluble supernatant was then loaded onto a MonoS 4.6/10 Cation exchanger column coupled to an ÄKTA purifier system. The unbound contaminants were removed with 5 CV of buffer 6A. SAP18 could then be eluted using a 12 CV gradient to 100% buffer 6B. The resulting highly pure SAP18 was then concentrated by ultrafiltration (Millipore Amicon Ultra, 10 KDa Molecular Weight Cut-off) and further purified using a Superdex 75 16/60 size exclusion column (buffer 6A + 2 mM DTT). The resulting pure protein was then concentrated to 30-45 mg/ml by ultrafiltration. The final concentration was measured with a nanodrop spectrophotometer (Thermo Scientific), assuming a calculated extinction coefficient at 280 nm of 12950 M^{-1} cm^{-1}. Purity of the sample was assayed by SDS gel electrophoresis and electrospray ionization – time of flight (ESI-TOF) mass spectrometry. Notably, if the histidine tag is not removed, HisSAP18 will display three peaks on ESI-TOF: one corresponding to the calculated molecular mass, the other two displaying an excess mass of 258 and 178 Daltons. These correspond to spontaneous α-N-6-

Phosphogluconoylation or α-N-gluconoylation of the hexahistidine tag (Geoghegan et al., 1999)

4.6.2 Purification of Acinus$^{770-1250}$ (Acinus ΔC)

GST Acinus$^{770-1250}$ was expressed in Tuneair flasks, with 1 litre of Terrific Broth per flask, with induction lasting 3.5 hours instead of overnight. The strain used was *E.coli* BL21 (DE3) PLysS. Bacterial pellets were resuspended in lysis buffer supplemented with 1 mM 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF) and 0.5x complete protease inhibitor (Roche) and lysed with two passes in an Avestin Emulsiflex C3, with pressures of 15000-20000 psi. The lysate was clarified by centrifugation (45´, 75000xg) and flowed twice through a gravity column filled with 10 ml of GSH sepharose resin pre-equilibrated in HisBuffer A. The resin was washed with 50 ml of buffer 7B and subsequently re-equilibrated with 50 ml of buffer7A. The washed resin was then resuspended in buffer 7A + 1mM DTT and 1 mM EDTA, then transferred to a 50 ml falcon tube and supplemented with 1-2 mg of TEV for tag removal overnight at 4˚C under gentle agitation. The cleaved Acinus was eluted with 50 ml buffer 7A in a gravity column and then loaded onto a 5 ml HiTrap Heparin column. The resin was washed 50% buffer 7B until flattening of the A_{280} baseline. Acinus ΔC was eluted with either directly 5 column volumes of buffer 7B or with a 12 CV, 50-100% gradient of buffer 7B. The fraction corresponding to full length AcinusΔC were then collected, concentrated by ultrafiltration and flash frozen for further use. AcinusΔC suffers from C-terminal degradation, losing about 20 amino acids at the C-terminus, down to the stable construct Acinus$^{770-1230}$. This degradation is already apparent during expression and is more pronounced for longer expression times.

4.6.3 Purification of Acinus $^{770-1230}$:

Bacterial pellets were resuspended in 200 ml lysis buffer supplemented with 0.5 mM AEBSF and lysed with two passes through an Emulsiflex C3 homogeniser (Avestin Inc.) at pressures of 15000-20000 psi. The lysate was subsequently clarified by centrifugation (1 hour, 75000 x g) and the soluble fraction was applied through a peristaltic pump on a 5 ml HisTrap FF column. Through an ÄKTA prime system, the column was washed with 10 CV of HisBuffer A, followed by 10 CV of 12% HisBuffer B. Elution was carried out with 80% HisBuffer B. The eluted protein was dialysed overnight against buffer 7A supplemented with 100 mM DTT. The hexahistidine tag was removed by TEV cleavage. The dialysate was then applied onto a heparin column, washed with 30% buffer 7B and eluted with a 30-100% gradient of Buffer 7B. The purification was completed by size exclusion chromatography (Superdex 200) carried out in buffer 7A

4.6.4 Purification of Acinus$^{1171-1250}$:

The purification protocol is similar to the one outlined for Acinus $^{770-1250}$, with the exception that instead of a heparin column, a MonoS 4.6/10 Cation exchanger column was used. Elution was carried out with a 12 CV gradient (0-100%) of buffer 7B. Using a gradient allows the separation of Acinus $^{1107-1250}$ from its degradation products.

4.6.5 Purification of Acinus$^{846-1107}$ and Acinus$^{616-710\ Dm}$

These proteins were purified using the same protocol detailed for the SAP18 constructs. All the constructs contain a 3C-cleavable hexahistidine tag.

4.6.6 Purification of His RNPS1 or TAP His RNPS1

RNPS1 is unstable. It is therefore necessary to carry out the purification in the shortest time possible, ideally one single day. Typically, the pellet resulting from 3L of insect cells culture was resuspended in lysis buffer supplemented with 2 mM DTT and 1x complete protease inhibitor (Roche). Lysis was performed by sonication on ice (VS70T tip, 50% power, intermittent output with 0.5" sonication and 1" rest, 2 cycles of 5 minutes each). The lysate was clarified by 45 minutes of centrifugation at 75000xg. Using a peristaltic pump, the soluble fraction was applied to a 5 ml HisTrap column, after which the column was transferred to an AKTA prime. The column was washed with HisBuffer A until baseline flattening, followed by a further wash with 14% HisBuffer B (corresponding to 80 mM imidazole). The protein was eluted with 80% HisBuffer B, immediately diluted 1:5 with buffer 7A and loaded on a 5 ml HiTrap heparin column. The column was washed with buffer 7A until baseline flattening, followed by a wash with 50% buffer 7B (500 mM final NaCl concentration). Elution was carried out with 100% buffer 7B. The eluted protein is pure as assayed by SDS and can be directly concentrated and stored at -80˚C.

4.6.7 Purification of His RNPS1$^{159-244}$:

Expression is carried out in BL21 PLysS according to the standard protocol. Typically 2L of culture are used per protein preparation. The bacterial pellets were resuspended in 200 ml lysis buffer, lysed with two passes in an Avestin Emulsiflex C3, with pressures of 15000-20000 psi. The lysate was clarified by centrifugation (45´, 75000xg), and then applied onto two 5 ml HisTrap columns by means of a peristaltic pump. The columns were then transferred to an AKTA prime system, washed with 10 CV HisBuffer A, followed by 10 CV 14% HisBuffer B. Elution was carried out in a step with 80% HisBuffer B. The eluate was dialysed overnight against buffer 6A supplemented with 2 mM DTT 1 mM EDTA. TEV protease was used for cleavage of the His tag. To remove any excess-tagged protein, the cleavage mixture was supplemented with 50 mM HEPES pH 7.5 and passed on two 5 ml HisTrap columns (Buffer 7A). The eluted untagged protein was further purified with a MonoS 4.6/10 cation exchanger column, followed by size exclusion chromatography (Superdex 75 16/60, buffer 6A). The purified protein was then concentrated by ultra filtration, aliquoted, flash frozen and stored at -80° C.

4.6.8 Purification of ASAP and of chimeric ASAP

For both ASAP and the chimeric ASAP, equimolar amounts of Acinus, RNPS1 and SAP18 were mixed and incubated 30' on ice. If necessary, salt concentration was adjusted to 250 mM NaCl, 50 mM HEPES pH 7.5 and 10% glycerol by addition of an appropriate 2x buffer. After the incubation, ASAP was concentrated to 1 ml by ultrafiltration and resolved by size exclusion chromatography (Superdex 75 16/60 for the core ASAP and core chimeric ASAP; Superdex 200 10/30 for the full length ASAP)

4.7 Crystallisation procedures

All the crystallisation experiments in 96 well plates were kindly set up by Karina Valer Saldana, Sabine Pleyer or Monica Prothmann from the MPI crystallisation facility.

4.7.1 Preliminary, large scale crystallisation screens

New constructs were screened for crystallisation using the commercially available screens listed below. In case the amount of protein was limiting, sparse matrix screens (Index, Classics, JCSG+ and PACT) were usually prioritised over systematic grid screens. Each screen was set up both at 4° and room temperature. Pipeting was carried out with a Phoenix pipeting robot, with 100 nl of reservoir solution being mixed with 100 nl protein solution in 96 well plates. The plates were imaged at regular intervals with an automatic imaging system in the crystallisation facility.

Screen	Manufacturer	Type
Index	Hampton research	Sparse matrix
Classics	Qiagen	Sparse matrix
JSCG+	Qiagen	Sparse matrix
PACT	Qiagen	Sparse matrix
Magic I	MPI facility	Sparse matrix / grid screen
Magic II	MPI facility	Sparse matrix / grid screen
PEGs	Qiagen	Grid screen
MPDs	Qiagen	Grid screen
Cations	Qiagen	Grid screen
Anions	Qiagen	Grid screen

4.7.2 Seeding

Seed stock solutions were prepared by resuspending crystals in 10 µl of stabilizing solution and by manually crushing them with a loop. Surviving macro crystal were then pelleted by centrifugation, and the stock was diluted as appropriate. The stabilising solution had the same formulation as the reservoir solution, with the addition of 5% more precipitant agent. Seeding was performed in 24 well plate, hanging drop format, by adding 0.2 µl of seed solution to a freshly set-up crystallisation drop (1µl of protein and 1 µl of well solution). The drop was not mixed upon adding the seeds.

4.7.3 Cryoprotection

Crystals were cryoprotected by gradual increase of the concentration of cryoprotectant in the original drop. Based on the original well solution, a series of cryoprotectant solution with 5% increment of cryoprotectant was prepared. 1 µl of the cryoprotectant solution with the lowest concentration of cryoprotectant was added to the 2 µl crystallisation drop; after mixing, 1 µl was removed from the drop. The cycle was repeated 3 times for each solution of the cryoprotectant series. The crystals were then flash frozen by plunging them in liquid nitrogen.

4.8 Analytical limited proteolysis

ASAP was diluted to 0.6 g/l with 50 mM NaCl, 20 mM HEPES pH 7.5, mixed with 4µl of protease solution and incubated on ice for 30'. For each protease, three different concentrations were used (0.1 g/l, 0.01 g/l, 0.001 g/l). The reaction was stopped by adding 1 µl of 100 mM 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF). 10 µl of the degradation product were loaded on SDS PAGE and stained with coomassie brilliant blue for visual inspection of the results and / or peptide finger-printing; the remainder of the reaction was either loaded on SDS PAGE and blotted on PVDF membrane for N-terminal sequencing or directly analysed by ESI / TOF mass spectrometry to determine the total mass of the fragments in solution. The ESI/TOF analyses were performed by Elisabeth Weyher-Stingl (MPI core facilities). N-terminal sequencing was performed by Dr. Reinhard Mentele. Peptide fingerprinting experiments were performed by Dr. Cyril Boulegue (MPI core facilities).

4.9 Preparative limited proteolysis of the chimeric ASAP

The chimeric ASAP, containing SAP18$^{1-143}$, RNPS1$^{159-244}$ and Acinus$^{616-710\ Dm}$, was prepared according to the standard protocol. The purified complex was concentrated to A_{280} = 1.0 and incubated with trypsin for 30 minutes on ice (4.450 ml of complex + 1.335 ml of 0.1 g/l trypsin solution). The reaction was then blocked with the addition of 2 mM AEBSF, concentrated to 500 µl and loaded on a Superdex 75 10/30 size exclusion column. The resulting peak was concentrated to 15.1 g/l and screened for crystallisation.

4.10 CBP Pull down experiments

Beads pre-equilibration:

1 ml of 50% calmodulin beads slurry was washed 3 times with 900 µl of 250 mM NaCl, 50 mM HEPES pH 7.5, 0.1% NP40 (blocking buffer) and resuspended in 500 µl of 550 mM NaCl, 50 mM HEPES pH 7.5, 0.1% NP40, 0.08 g/l glycogen carrier, 0.08 g/l tRNA and 0.8 g/l bovine serum albumin. After an overnight incubation, the beads were washed 3 times with 500 µl blocking buffer and stored at 4°C.

Pull down:

1.5 µg of each protein were mixed and 2x buffer was added to a total volume of 60 µl. Typically, the protein stocks were highly concentrated (> 10 g/l) and stored in buffer containing 0.1 - 1M NaCl, 20 mM HEPES / MES pH 7.5/6.0 and 10% glycerol. The composition of the 2x buffer was adjusted on a sample by sample basis in order to keep the condition of the reaction as close as possible to 250 mM NaCl, 50 mM HEPES 7.5, 10% glycerol, 2 mM Mg acetate, 2 mM $CaCl_2$, 2 mM imidazole, 2 mM DTT and 0.05% NP40 (calmodulin binding buffer). The reaction was incubated on ice for 30', then 250 µl of calmodulin binding buffer were added, along with 12 µl of pre-equilibrated calmodulin beads. The CBP tagged proteins were allowed to bind to the resin for 1h at 4°C under constant gentle agitation, after which the resin was washed 3 times with 500 µl calmodulin binding buffer. The protein bound to the resin was then resuspended in 15 µl of calmodulin binding buffer lacking $CaCl_2$ and supplemented with 20 mM EGTA. After 15 minutes of incubation at room temperature under gentle agitation, the eluate was collected, mixed with 5 µl 2x SDS loading

buffer, concentrated by vacuum centrifugation and loaded on a gel. 4-20% tris-glycine NuPAGE precast gradient gels (Invitrogen) and MOPS-SDS running buffer (Invitrogen) were typically used to allow optimal separation in the 70 to 8 KDa range.

4.11 Gel filtration binding experiments

SAP18$^{6-143}$ was purified by gel filtration and stored in 100 mM NaCl, 10% glycerol and 20 mM MES pH 6.0. His RNPS1$^{159-244}$ and Acinus$^{1171-1250}$ were purified up to the cation exchange step and were stored in 500 mM NaCl, 10% glycerol and 20 mM MES pH 6.0. Equimolar amounts of each ASAP component were mixed in a total volume of 500 µl. Salt concentration was adjusted to 100 mM by the addition of appropriate amounts of 2x buffer. The proteins were incubated on ice for 1h and then loaded on a size exclusion column (Superdex 75). The run was performed in 100 mM KCl, 10% glycerol and 20 mM MES pH 6.0. The slightly acidic pH does not influence the formation of the complex, as identical results were observed in previous experiments at pH 7.5. The same fractions from each run were collected and run on a 15% SDS PAGE gel for display.

5 Bibliography:

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Alpatov, R., Munguba, G.C., Caton, P., Joo, J.H., Shi, Y., Shi, Y., Hunt, M.E., and Sugrue, S.P. (2004). Nuclear speckle-associated protein Pnn/DRS binds to the transcriptional corepressor CtBP and relieves CtBP-mediated repression of the E-cadherin gene. Mol Cell Biol *24*, 10223-10235.

Alpatov, R., Shi, Y., Munguba, G.C., Moghimi, B., Joo, J.-H., Bungert, J., and Sugrue, S.P. (2008). Corepressor CtBP and nuclear speckle protein Pnn/DRS differentially modulate transcription and splicing of the E-cadherin gene. Mol Cell Biol *28*, 1584-1595.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402.

Anantharaman, V., Iyer, L.M., and Aravind, L. (2007). Comparative genomics of protists: new insights into the evolution of eukaryotic signal transduction and gene regulation. Annu Rev Microbiol *61*, 453-475.

Anderson, D., Beres, B., Wilson-Rawls, J., and Rawls, A. (2009). The homeobox gene Mohawk represses transcription by recruiting the sin3A/HDAC co-repressor complex. Dev Dyn *238*, 572-580.

Armon, A., Graur, D., and Ben-Tal, N. (2001). ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. J Mol Biol *307*, 447-463.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res *38 Suppl*, W529-533.

Ashton-Beaucage, D., Udell, C.M., Lavoie, H., Baril, C., Lefrancois, M., Chagnon, P., Gendron, P., Caron-Lizotte, O., Bonneil, E., Thibault, P., *et al.* (2010). The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in Drosophila. Cell *143*, 251-262.

Aslanidis, C., and de Jong, P.J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res *18*, 6069-6074.

Aslanidis, C., de Jong, P.J., and Schmitz, G. (1994). Minimal length requirement of the single-stranded tails for ligation-independent cloning (LIC) of PCR products. PCR Methods Appl *4*, 172-177.

Babon, J.J., Sabo, J.K., Soetopo, A., Yao, S., Bailey, M.F., Zhang, J.G., Nicola, N.A., and Norton, R.S. (2008). The SOCS box domain of SOCS3: structure and interaction with the elonginBC-cullin5 ubiquitin ligase. J Mol Biol *381*, 928-940.

Biegert, A., and Soding, J. (2009). Sequence context-specific profiles for homology searching. Proc Natl Acad Sci U S A *106*, 3770-3775.

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem *72*, 291-336.

Blencowe, B.J. (2006). Alternative splicing: new insights from global analyses. Cell *126*, 37-47.

Bond, C.S. (2003). TopDraw: a sketchpad for protein structure topology cartoons. Bioinformatics *19*, 311-312.

Bond, C.S., and Schüttelkopf, A.W. (2009). ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. Acta Crystallogr D Biol Crystallogr *65*, 510-512.

Bono, F., Ebert, J., Unterholzner, L., Güttler, T., Izaurralde, E., and Conti, E. (2004). Molecular insights into the interaction of PYM with the Mago-Y14 core of the exon junction complex. EMBO Rep *5*, 304-310.

Cavalier-Smith, T. (2010a). Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. Biol Lett *6*, 342-345.

Cavalier-Smith, T. (2010b). Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. Biol Direct *5*, 7.

Chang, B., Chen, Y., Zhao, Y., and Bruick, R.K. (2007a). JMJD6 is a histone arginine demethylase. Science *318*, 444-447.

Chang, Y.-F., Imam, J.S., and Wilkinson, M.F. (2007b). The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem *76*, 51-74.

Cheng, S.Y., and Bishop, J.M. (2002). Suppressor of Fused represses Gli-mediated transcription by recruiting the SAP18-mSin3 corepressor complex. Proc Natl Acad Sci USA *99*, 5442-5447.

Cieslik, M., and Derewenda, Z.S. (2009). The role of entropy and polarity in intermolecular contacts in protein crystals. Acta Crystallogr D Biol Crystallogr *65*, 500-509.

Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Res *36*, W197-201.

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

Costa, E., Canudas, S., Garcia-Bassets, I., Pérez, S., Fernández, I., Giralt, E., Azorín, F., and Espinás, M.L. (2006). Drosophila dSAP18 is a nuclear protein that associates with chromosomes and the nuclear matrix, and interacts with pinin, a protein factor involved in RNA splicing. Chromosome Res *14*, 515-526.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr *62*, 1002-1011.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S., *et al.* (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res *35*, W375-383.

Davis, W.M. (2009). ApE, Advanced Plasmid Editor.

Domon, B., and Aebersold, R. (2006). Mass spectrometry and protein analysis. Science *312*, 212-217.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

Espinás, M.L., Canudas, S., Fanti, L., Pimpinelli, S., Casanova, J., and Azorín, F. (2000). The GAGA factor of Drosophila interacts with SAP18, a Sin3-associated polypeptide. EMBO Rep *1*, 253-259.

Eswar, N., Eramian, D., Webb, B., Shen, M.Y., and Sali, A. (2008). Protein structure modeling with MODELLER. Methods Mol Biol *426*, 145-159.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr *62*, 72-82.

Falnes, P.O., Johansen, R.F., and Seeberg, E. (2002). AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. Nature *419*, 178-182.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K.*, et al.* (2010). The Pfam protein families database. Nucleic Acids Res *38*, D211-222.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L.*, et al.* (2008). The Pfam protein families database. Nucleic Acids Res *36*, D281-288.

Fiser, A., and Sali, A. (2003). Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol *374*, 461-491.

Furukawa, Y., Torres, A.S., and O'Halloran, T.V. (2004). Oxygen-induced maturation of SOD1: a key role for disulfide formation by the copper chaperone CCS. EMBO J *23*, 2872-2881.

Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., and Kulozik, A.E. (2005). Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. Mol Cell *20*, 65-75.

Gehring, N.H., Lamprinaki, S., Hentze, M.W., and Kulozik, A.E. (2009). The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. PLoS Biol *7*, e1000120.

Geoghegan, K.F., Dixon, H.B., Rosner, P.J., Hoth, L.R., Lanzetti, A.J., Borzilleri, K.A., Marr, E.S., Pezzullo, L.H., Martin, L.B., LeMotte, P.K.*, et al.* (1999). Spontaneous alpha-N-6-phosphogluconoylation of a "His tag" in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins. Anal Biochem *267*, 169-184.

Giorgi, C., Yeo, G.W., Stone, M.E., Katz, D.B., Burge, C., Turrigiano, G., and Moore, M.J. (2007). The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression. Cell *130*, 179-191.

Gowrishankar, J., and Harinarayanan, R. (2004). Why is transcription coupled to translation in bacteria? Mol Microbiol *54*, 598-603.

Grzenda, A., Lomberk, G., Zhang, J.-S., and Urrutia, R. (2009). Sin3: master scaffold and transcriptional corepressor. Biochim Biophys Acta *1789*, 443-450.

Haberman, A.S., Akbar, M.A., Ray, S., and Kramer, H. (2010). Drosophila acinus encodes a novel regulator of endocytic and autophagic trafficking. Development *137*, 2157-2166.

Hachet, O., and Ephrussi, A. (2001). Drosophila Y14 shuttles to the posterior of the oocyte and is required for oskar mRNA transport. Curr Biol *11*, 1666-1674.

Hachet, O., and Ephrussi, A. (2004). Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. Nature *428*, 959-963.

Hahn, T., ed. (2005). International Tables for Crystallography (Springer).

Heinig, M., and Frishman, D. (2004). STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Res *32*, W500-502.

Hill, K., Wang, H., and Perry, S. (2007). A transcriptional repression motif in the MADS factor AGL15 is involved in recruitment of histone deacetylase complex components. Plant J.

Hong, X., Zang, J., White, J., Wang, C., Pan, C.H., Zhao, R., Murphy, R.C., Dai, S., Henson, P., Kappler, J.W.*, et al.* (2010). Interaction of JMJD6 with single-stranded RNA. Proc Natl Acad Sci U S A *107*, 14568-14572.

Izquierdo, J.-M., and Valcárcel, J. (2006). A simple principle to explain the evolution of pre-mRNA splicing. Genes Dev *20*, 1679-1684.

Jang, S.-W., Yang, S.-J., Ehlén, A., Dong, S., Khoury, H., Chen, J., Persson, J.L., and Ye, K. (2008). Serine/arginine protein-specific kinase 2 promotes leukemia cell proliferation by phosphorylating acinus and regulating cyclin A1. Cancer Res *68*, 4559-4570.

Joo, J.-H., Lee, Y.J., Munguba, G.C., Park, S., Taxter, T.J., Elsagga, M.Y., Jackson, M.R., Oh, S.P., and Sugrue, S.P. (2007). Role of Pinin in neural crest, dorsal dermis, and axial skeleton development and its involvement in the regulation of Tcf/Lef activity in mice. Dev Dyn *236*, 2147-2158.

Joo, J.-H., Taxter, T.J., Munguba, G.C., Kim, Y.H., Dhaduvai, K., Dunn, N.W., Degan, W.J., Oh, S.P., and Sugrue, S.P. (2010). Pinin modulates expression of an intestinal homeobox gene, Cdx2, and plays an essential role for small intestinal morphogenesis. Dev Biol.

Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr *66*, 125-132.

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

Katoh, K., Kuma, K.-i., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res *33*, 511-518.

Keegan, R.M., and Winn, M.D. (2008). MrBUMP: an automated pipeline for molecular replacement. Acta Crystallogr D Biol Crystallogr *64*, 119-124.

Keeling, P.J., and Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet *9*, 605-618.

Kibbe, W.A. (2007). OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res *35*, W43-46.

Klose, R.J., Kallin, E.M., and Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. Nat Rev Genet *7*, 715-727.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr *60*, 2256-2268.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. Journal of Molecular Biology *372*, 774-797.

Lange, O.F., Lakomek, N.A., Fares, C., Schroder, G.F., Walter, K.F., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C., and de Groot, B.L. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science *320*, 1471-1475.

Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M.J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J *20*, 4987-4997.

Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. EMBO J *19*, 6860-6869.

Lee, D.H., Jin, S.G., Cai, S., Chen, Y., Pfeifer, G.P., and O'Connor, T.R. (2005). Repair of methylation damage in DNA and RNA by mammalian AlkB homologues. J Biol Chem *280*, 39448-39459.

Lendenmann, U., and Egli, T. (1995). Is Escherichia Coil Growing in Glucose-Limited Chemostat Culture Able to Utilize Other Sugars without Lag. Microbiology-Sgm *141*, 71-78.

Leslie, A. (1992). Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography *26*.

Li, C., Lin, R.-I., Lai, M.-C., Ouyang, P., and Tarn, W.-Y. (2003). Nuclear Pnn/DRS protein binds to spliced mRNPs and participates in mRNA processing and export via interaction with RNPS1. Molecular and Cellular Biology *23*, 7363-7376.

Liang, X.H., Haritan, A., Uliel, S., and Michaeli, S. (2003). trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. Eukaryot Cell *2*, 830-840.

Liu, C., Xi, W., Shen, L., Tan, C., and Yu, H. (2009). Regulation of floral patterning by flowering time genes. Dev Cell *16*, 711-722.

Long, F., Vagin, A.A., Young, P., and Murshudov, G.N. (2008). BALBES: a molecular-replacement pipeline. Acta Crystallogr D Biol Crystallogr *64*, 125-132.

Lykke-Andersen, J., Shu, M.D., and Steitz, J.A. (2001). Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. Science *293*, 1836-1839.

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

Matyash, A., Singh, N., Hanes, S.D., Urlaub, H., and Jaeckle, H. (2009). SAP18 Promotes Kruppel-dependent Transcriptional Repression by Enhancer-specific Histone Deacetylation. J Biol Chem *284*, 3012-3020.

Mayeda, A., Badolato, J., Kobayashi, R., Zhang, M.Q., Gardiner, E.M., and Krainer, A.R. (1999). Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. EMBO J *18*, 4560-4570.

McCallum, S.A., Bazan, J.F., Merchant, M., Yin, J., Pan, B., de Sauvage, F.J., and Fairbrother, W.J. (2006). Structure of SAP18: a ubiquitin fold in histone deacetylase complex assembly. Biochemistry *45*, 11974-11982.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

Merz, C., Urlaub, H., Will, C.L., and Lührmann, R. (2007). Protein composition of human mRNPs spliced in vitro and differential requirements for mRNP protein recruitment. RNA *13*, 116-128.

Mooij, W.T.M., Cohen, S.X., Joosten, K., Murshudov, G.N., and Perrakis, A. (2009a). "Conditional Restraints": Restraining the Free Atoms in ARP/wARP. Structure *17*, 183-189.

Mooij, W.T.M., Mitsiki, E., and Perrakis, A. (2009b). ProteinCCD: enabling the design of protein truncation constructs for expression and crystallization experiments. Nucleic Acids Res *37*, W402-405.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688-700.

Morris, R.J., Perrakis, A., and Lamzin, V.S. (2002). ARP/wARP's model-building algorithms. I. The main chain. Acta Crystallographica Section D-Biological Crystallography *58*, 968-975.

Muller, J., and Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in Drosophila. Curr Opin Genet Dev *16*, 476-484.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica Section D-Biological Crystallography *53*, 240-255.

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Kocher, T., Cohen, A., Stunnenberg, H.G., Wilm, M., and Muller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. EMBO J *26*, 4078-4088.

Noda, N.N., Kumeta, H., Nakatogawa, H., Satoo, K., Adachi, W., Ishii, J., Fujioka, Y., Ohsumi, Y., and Inagaki, F. (2008). Structural basis of target recognition by Atg8/LC3 during selective autophagy. Genes Cells *13*, 1211-1218.

Nott, A., Le Hir, H., and Moore, M.J. (2004). Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. Genes Dev *18*, 210-222.

Pape, T., and Schneider, T.R. (2004). HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. Journal of Applied Crystallography *37*, 843-844.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nature Structural Biology *6*, 458-463.

Price, W.N., Chen, Y., Handelman, S.K., Neely, H., Manor, P., Karlin, R., Nair, R., Liu, J., Baran, M., Everett, J.*, et al.* (2009). Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. Nat Biotechnol *27*, 51-57.

Roignant, J.Y., and Treisman, J.E. (2010). Exon Junction Complex Subunits Are Required to Splice Drosophila MAP Kinase, a Large Heterochromatic Gene. Cell *143*, 238-250.

Rossmann, M.G., and Blow, D.M. (1962). The detection of sub-units within the crystallographic asymmetric unit. Acta Crystallographica *15*, 24-31.

Sahara, S., Aoto, M., Eguchi, Y., Imamoto, N., Yoneda, Y., and Tsujimoto, Y. (1999). Acinus is a caspase-3-activated protein required for apoptotic chromatin condensation. Nature *401*, 168-173.

Sakashita, E., Tatsumi, S., Werner, D., Endo, H., and Mayeda, A. (2004). Human RNPS1 and its associated factors: a versatile alternative pre-mRNA splicing regulator in vivo. Mol Cell Biol *24*, 1174-1187.

Sambrook J, Fritsch EF, and T, M., eds. (1989). Molecular cloning: a laboratory manual (Cold Spring Harbor Laboratory Press).

Sanchez-Perez, G.F., Gasset, M., Calvete, J.J., and Pajares, M.A. (2003). Role of an intrasubunit disulfide in the association state of the cytosolic homo-oligomer methionine adenosyltransferase. J Biol Chem *278*, 7285-7293.

Schwerk, C., Prasad, J., Degenhardt, K., Erdjument-Bromage, H., White, E., Tempst, P., Kidd, V.J., Manley, J.L., Lahti, J.M., and Reinberg, D. (2003). ASAP, a novel protein complex involved in RNA processing and apoptosis. Mol Cell Biol *23*, 2981-2990.

Sheeba, C., Palmeirim, I., and Andrade, R. (2007). Chick Hairy1 protein interacts with Sap18, a component of the Sin3/HDAC transcriptional repressor complex. BMC Developmental Biology 2007 7:83 *7*, 83.

Sheldrick, G.M. (2008). A short history of SHELX. Acta Crystallogr A *64*, 112-122.

Shubin, N., Tabin, C., and Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. Nature *457*, 818-823.

Singh, K.K., Erkelenz, S., Rattay, S., Dehof, A.K., Hildebrandt, A., Schulze-Osthoff, K., Schaal, H., and Schwerk, C. (2010). Human SAP18 mediates assembly of a splicing regulatory multiprotein complex via its ubiquitin-like fold. RNA.

Song, C.-P., and Galbraith, D.W. (2006). AtSAP18, An Orthologue of Human SAP18, is Involved in the Regulation of Salt Stress and Mediates Transcriptional Repression in Arabidopsis. Plant Mol Biol *60*, 241-257.

Song, J., Zhang, Z., Hu, W., and Chen, Y. (2005). Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. J Biol Chem *280*, 40122-40129.

Sreenath, H.K., Bingman, C.A., Buchan, B.W., Seder, K.D., Burns, B.T., Geetha, H.V., Jeon, W.B., Vojtik, F.C., Aceti, D.J., Frederick, R.O., *et al.* (2005). Protocols for production of selenomethionine-labeled proteins in 2-L polyethylene terephthalate bottles using auto-induction medium. Protein Expr Purif *40*, 256-267.

Stechmann, A., and Cavalier-Smith, T. (2003). The root of the eukaryote tree pinpointed. Curr Biol *13*, R665-666.

Stein, N. (2008). CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. Journal of Applied Crystallography *41*, 641-643.

Sundheim, O., Vagbo, C.B., Bjoras, M., Sousa, M.M., Talstad, V., Aas, P.A., Drablos, F., Krokan, H.E., Tainer, J.A., and Slupphaug, G. (2006). Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage. EMBO J *25*, 3389-3397.

Tange, T.Ø., Shibuya, T., Jurica, M.S., and Moore, M.J. (2005). Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. RNA *11*, 1869-1883.

Thielmann, Y., Weiergraber, O.H., Mohrluder, J., and Willbold, D. (2009). Structural framework of the GABARAP-calreticulin interface--implications for substrate binding to endoplasmic reticulum chaperones. FEBS J *276*, 1140-1152.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res *25*, 4876-4882.

Trembley, J.H., Tatsumi, S., Sakashita, E., Loyer, P., Slaughter, C.A., Suzuki, H., Endo, H., Kidd, V.J., and Mayeda, A. (2005). Activation of pre-mRNA splicing by human RNPS1 is regulated by CK2 phosphorylation. Mol Cell Biol *25*, 1446-1457.

Trewick, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T., and Sedgwick, B. (2002). Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. Nature *419*, 174-178.

Tsujikawa, K., Koike, K., Kitae, K., Shinkawa, A., Arima, H., Suzuki, T., Tsuchiya, M., Makino, Y., Furukawa, T., Konishi, N.*, et al.* (2007). Expression and sub-cellular localization of human ABH family molecules. J Cell Mol Med *11*, 1105-1116.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.

Urzhumtseva, L., Afonine, P.V., Adams, P.D., and Urzhumtsev, A. (2009). Crystallographic model quality at a glance. Acta Crystallogr D Biol Crystallogr *65*, 297-300.

Velankar, S., Best, C., Beuth, B., Boutselakis, C.H., Cobley, N., Sousa Da Silva, A.W., Dimitropoulos, D., Golovin, A., Hirshberg, M., John, M.*, et al.* (2010). PDBe: Protein Data Bank in Europe. Nucleic Acids Res *38*, D308-317.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Wang, L., Ding, L., Jones, C.A., and Jones, R.S. (2002a). Drosophila Enhancer of zeste protein interacts with dSAP18. Gene *285*, 119-125.

Wang, P., Lou, P.-J., Leu, S., and Ouyang, P. (2002b). Modulation of alternative pre-mRNA splicing in vivo by pinin. Biochem Biophys Res Commun *294*, 448-455.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.

Webby, C.J., Wolf, A., Gromak, N., Dreger, M., Kramer, H., Kessler, B., Nielsen, M.L., Schmitz, C., Butler, D.S., Yates, J.R., 3rd*, et al.* (2009). Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. Science *325*, 90-93.

Westbye, M.P., Feyzi, E., Aas, P.A., Vagbo, C.B., Talstad, V.A., Kavli, B., Hagen, L., Sundheim, O., Akbari, M., Liabakk, N.B.*, et al.* (2008). Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. J Biol Chem *283*, 25046-25056.

Wiegand, H.L., Lu, S., and Cullen, B.R. (2003). Exon junction complexes mediate the enhancing effect of splicing on mRNA expression. Proc Natl Acad Sci USA *100*, 11327-11332.

Winn, M.D., Isupov, M.N., and Murshudov, G.N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr D Biol Crystallogr *57*, 122-133.

Wouters, M.A., Fan, S.W., and Haworth, N.L. (2010). Disulfides as Redox Switches: From Molecular Mechanisms to Functional Significance. Antioxidants & Redox Signaling *12*, 53-91.

Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (1997). Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. Cell *89*, 357-364.

Zhu, W., Foehr, M., Jaynes, J.B., and Hanes, S.D. (2001). Drosophila SAP18, a member of the Sin3/Rpd3 histone deacetylase complex, interacts with Bicoid and inhibits its activity. Dev Genes Evol *211*, 109-117.

6 Abbreviations and acronyms:

A.t.	***A**rabidopsis **t**haliana*
2-OG	**2**-**o**xy-**g**lutarate
Abh	**A**lk**b**6 **h**omolog
Acinus	**A**poptosis and spli**c**ing associated protein **i**n the **n**ucle**us**
AEBSF	4-(2-**A**mino**e**thyl) **b**enzene**s**ulfonyl **f**luoride hydrochloride
Alkb6	**al**pha-**k**etoglutarate-dependent **b**ioxygenase **6**
ASAP	**A**poptosis and **s**plicing **a**ssociated **p**rotein complex
ASU	**As**ymmetric **u**nit
Atg8 / Atg19	**A**u**t**opha**g**y related protein **8** / **19**
C.e.	***C**aenorhabditis **e**legans*
C.n.	***C**ryptococcus **n**eoformans*
C.r.	***C**hlamidomonas **r**einhardtii*
cAMP	**C**yclic **a**denosine **m**ono**p**hosphate
CCD	**C**oiled **c**oil **d**omain
CREB	**c**AMP **r**esponsive **e**lement **b**inding protein
CtBP	**C**-**t**erminus **b**inding **p**rotein
DTT	**D**i**t**hio**t**hreitol
EJC	**E**xon **j**unction **c**omplex
ESI/TOF	**E**lectro**s**pray **I**onization / **t**ime **o**f **f**light
FA	**F**lourescence **a**nisotropy
GABARAP	**GABA** A **r**eceptor-**a**ssociated **p**rotein
Gy	**G**iga **y**ears (billions of years)
HEPES	4-(2-**h**ydroxy**e**thyl)-1-**p**iperazine**e**thane**s**ulfonic acid
hkl	**H**oo**k**-**l**ike (the *Drosophila melanogaster* homolog of Acinus)
HMM	**H**idden **M**arkov **M**odel
hnRNP	**H**etero**n**uclear **r**ibo**n**ucleo**p**article

IPTG	**I**so**p**ropyl β-D-1-**t**hio**g**alactopyranoside
ITC	**I**sothermal **t**itration **c**alorimetry
JMJC	**J**u**m**on**j**i **C**
L.b.	***L****accaria* ***b****icolor*
LAMM	**L**evulose **a**ugmented **m**inimal **m**edium
LB	**L**ysogeny **b**roth
LECA	**L**ast **e**ukaryotic **c**ommon **a**ncestor
M.m.	***M****us* ***m****usculus*
MAD	**M**ultiple **a**nomalous **d**iffraction
MES	2-(N-**m**orpholino)**e**thane**s**ulfonic acid
MIB	2:3:3 molar ratio of sodium **m**alonate:**i**midazole:**b**oric acid
MIRAS	**M**ultiple **i**somorphous **r**eplacement with **a**nomalous **s**cattering
MME	**M**ono**m**ethyl **e**ther
MPD	2-**M**ethyl-2,4-**p**entane**d**iol
MR	**M**olecular **r**eplacement
mRNA	**M**essenger **RNA**
N.c.	***N****eurospora* ***c****rassa*
NADH	**N**icotinamide **a**denine **d**ineucleotide, reduced form (**H** in the acronymdenotes an extra hydrogen atom as opposed to the oxydised form NAD^+)
NCS	**N**on-**c**rystallographic **s**ymmetry
Ni-NTA	**Ni**ckel-**n**itrilo**t**riacetic **a**cid
NMD	**N**onsense **m**ediated **d**ecay
P.p.	***P****hyscomitrella* ***p****atens*
PAGE	**P**oly**a**crylamide **g**el **e**lectrophoresis
PEG	**P**oly**e**thilene **g**lycol

PISA	Search and analysis of **P**rotein **i**nterfaces, **s**urfaces and **a**ssemblies
PMSF	**P**henyl**m**ethane**s**ulfonyl**f**luoride
RBD	**R**NA **b**inding **d**omain (synonym of RRM)
RMSD	**R**oot **m**ean **s**quare **d**eviation
RNPS1	**RN**A binding **p**rotein **S1**
RRM	**R**NA **r**ecognition **m**otif (synonym of RBD)
RT	**R**oom **t**emperature
S.m.	***S**elaginella **m**oellendorfii*
S.p.	***S**chizosaccharomyces **p**ombe*
SAD	**S**ingle **a**nomalous **d**iffraction
SAP18	**S**in3a **a**ssociated **p**rotein of **18** KDa
SDS	**S**odium **d**odecyl **s**ulfate
SeMet	**Se**leno **met**hionine
SIRAS	**S**ingle **i**somorphous **r**eplacement with **a**nomalous **s**cattering
SLS	**S**wiss **L**ight **S**ource
snRNP	**S**mall **n**uclear **r**ibo**n**ucleo**p**rotein
SRR	**S**ensitivity to **r**ed light **r**educed
SUMO	**S**mall **u**biquitin-like **mo**difier
T.a.	***T**richoplax **a**dhaerens*
TB	**T**errific **B**roth
TCEP	**T**ris(2-**c**arboxy**e**thyl)**p**hosphine
Tris	**Tris**(hydroxymethyl)aminomethane
U2AF65	**U2** small ribonucleoprotein **a**uxiliary **f**actor, **65** KDa subunit
YNB	**Y**east **n**itrogen **b**ase

Acknowledgements

This work would not have been possible without the help and support of many people.

My first and foremost 'thank you' goes to Elena: thanks for giving me the opportunity to work on this challenging, sometimes frustrating but always stimulating project. Thanks for the outstanding mentoring and support, and thanks for creating such a superb working environment, both at EMBL and at the MPI. A huge thank you goes also to Judith, whose great ability and patience were key to obtaining the crystals (and the structure) of ASAP. I am also indebted to our collaborators Johannes Söding and Eckhart Guthörlein, who contributed to this work with the bioinformatic searches of the ABM containing proteins and of the interactors of SAP18, and to the members of my TAC Karl Peter Hopfner, Anne-Claude Gavin and Anne Ephrussi for the helpful suggestions.

Even though their name does not appear anywhere in the results section, a lot of other people contributed to this work by sharing their expertise and providing help with reagents and equipment. Thanks to Petra and Tatjana for the cloning of several constructs and to Michaela for the help with insect cells expression. Thanks to Peter for maintaining the AKTAs in perfect conditions (which is as easy as herding cats) and for always being open to suggestions and special needs. Thanks to Claire and Jörg for the byophisics measurements. Thanks to Jerôme for setting up a superb crystallisation facility, and thanks to Karina and Sabine for running it like clockwork. I want also to thank Debora, JP and Atlanta for teaching me how to turn ugly crystals into beautiful electron densities – and for patiently putting up with my endless questions. Thanks to Christian for setting up the crystallographic environment and for being always there to solve whatever problem arose. Thanks to Rajan, John, Justine, Atlanta and JP for proofreading this thesis.

I would also like to add a special, heartfelt thank you to every member of the Conti department, past and present, for making the lab such an incredibly supportive, friendly and relaxed place.

Finally, these years would have been much harder without my friends, in Milan, Heidelberg and Munich, and without the unconditional support from my family. Thank you all for everything.