

Crystal Structure of the Pml1p Subunit of the Yeast Precursor mRNA Retention and Splicing Complex

Simon Trowitzsch^{1,2}, Gert Weber^{1,2}, Reinhard Lührmann¹
and Markus C. Wahl^{2,3,4*}

¹Zelluläre Biochemie,
Max-Planck-Institut für
Biophysikalische Chemie,
Am Faßberg 11, D-37077
Göttingen, Germany

²Makromolekulare
Röntgenkristallographie,
Max-Planck-Institut für
Biophysikalische Chemie,
Am Faßberg 11, D-37077
Göttingen, Germany

³Röntgenkristallographie,
Universitätsmedizin Georg-
August-Universität Göttingen,
Justus-von-Liebig-Weg 11,
D-37077 Göttingen, Germany

⁴Freie Universität Berlin,
Fachbereich Biologie, Chemie,
Pharmazie, Institut für Chemie
und Biochemie, AG
Strukturbiochemie, Takustrasse
6, D-14195 Berlin, Germany

Received 15 October 2008;
accepted 21 October 2008

Available online
7 November 2008

Edited by R. Huber

The precursor mRNA retention and splicing (RES) complex mediates nuclear retention and enhances splicing of precursor mRNAs. The RES complex from yeast comprises three proteins, Snu17p, Bud13p and Pml1p. Snu17p acts as a central platform that concomitantly binds the Bud13p and Pml1p subunits *via* short peptide epitopes. As a step to decipher the molecular architecture of the RES complex, we have determined crystal structures of full-length Pml1p and N-terminally truncated Pml1p. The first 50 residues of full-length Pml1p, encompassing the Snu17p-binding region, are disordered, showing that Pml1p binds to Snu17p *via* an intrinsically unstructured region. The remainder of Pml1p folds as a forkhead-associated (FHA) domain, which is expanded by a number of noncanonical elements compared with known FHA domains from other proteins. An atypical N-terminal appendix runs across one β -sheet and thereby stabilizes the domain as shown by deletion experiments. FHA domains are thought to constitute phosphopeptide-binding elements. Consistently, a sulfate ion was found at the putative phosphopeptide-binding loops of full-length Pml1p. The N-terminally truncated version of the protein lacked a similar phosphopeptide mimic but retained an almost identical structure. A long loop neighboring the putative phosphopeptide-binding site was disordered in both structures. Comparison with other FHA domain proteins suggests that this loop adopts a defined conformation upon ligand binding and thereby confers ligand specificity. Our results show that in the RES complex, an FHA domain of Pml1p is flexibly tethered *via* an unstructured N-terminal region to Snu17p.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: intrinsically unstructured domain; forkhead-associated domain; pre-mRNA retention and splicing; protein phosphorylation; RES complex

Introduction

During precursor mRNA (pre-mRNA) splicing in eukaryotes, noncoding intervening sequences (introns) are excised and the neighboring coding

regions (exons) are ligated to produce mature mRNA that can serve as a template for protein biosynthesis. Each round of splicing entails two sequential transesterification reactions that are carried out by a large, multifactorial RNA–protein enzyme, the spliceosome.^{1–4} The main task of the spliceosome lies in locating and bringing together the sites at which the cut-and-paste reactions have to proceed with single-nucleotide precision. A spliceosome is assembled *de novo* for each splicing event by the stepwise stable integration of small nuclear ribonucleoprotein particles (snRNPs) and numerous non-snRNP splice factors to accomplish this task.¹ Five snRNAs and more than 150 proteins participate in this process.² After the assembly of all snRNPs, major compositional and conformational rearrangements are required to bring

*Corresponding author. Röntgenkristallographie, Universitätsmedizin Georg-August-Universität Göttingen, Justus-von-Liebig-Weg 11, D-37077 Göttingen, Germany. E-mail address: mwahl@gwdg.de.

Abbreviations used: pre-mRNA, precursor mRNA; RES, retention and splicing; FHA, forkhead associated; snRNP, small nuclear ribonucleoprotein particle; RRM, RNA recognition motif; SAD, single anomalous diffraction; NTA, nitrilotriacetic acid.

about a catalytically competent particle that is correctly positioned on the pre-mRNA substrate.^{1–5}

Many of the spliceosomal constituents are pre-organized as functional modules. Prime among such modules are the snRNPs, each of which comprises a particular snRNA, a set of seven common Sm proteins (Sm-like, LSm, proteins in the case of U6) and a variable number of particle-specific proteins.⁴ In addition, many of the non-snRNP splice factors form multimeric protein complexes. In some

cases, specific functions could be attributed to these modules. For example, the pre-mRNA retention and splicing (RES) complex, a recently discovered spliceosome-associated module in yeast, is involved in retaining unspliced pre-mRNAs in the nucleus and in splicing a subset of pre-mRNAs.^{6–8} The RES complex encompasses proteins Snu17p, Bud13p and Pml1p.⁶ Recently, we have used structural, biochemical and biophysical techniques to work out a detailed interaction map among the RES

Table 1. Data collection and refinement statistics

	Pml1p	Pml1p-Hg	Pml1p ^{51–204}
<i>Data collection</i>			
Wavelength (Å)	0.9802	1.0095	0.9800
Temperature (K)	100	100	100
Space group	<i>P</i> 3 ₂ 21	<i>P</i> 3 ₂ 21	<i>H</i> 32
Unit cell parameters (Å, °)			
<i>a</i> = <i>b</i>	86.89	86.20	100.61
<i>c</i>	96.70	99.11	102.20
Resolution (Å)	40.7–2.4 (2.53–2.40) ^a	30.0–2.98 (3.08–2.98) ^a	29.0–1.8 (2.04–1.80) ^a
Reflections			
Unique	16,896 (2473)	16,703 (1667)	18,621 (5539)
Completeness (%)	99.9 (100)	100 (100)	99.7 (100)
Redundancy	6.3 (6.2)	8.9 (8.7)	10.6 (10.3)
<i>I</i> /σ (<i>I</i>)	19.7 (1.6)	15.4 (3.7)	33.7 (7.7)
<i>R</i> _{sym} (<i>I</i>) ^b	0.083 (0.81)	0.235 (0.590)	0.048 (0.283)
<i>Phasing</i>			
Resolution (Å)		3.3	
Heavy atom sites		8	
Correlation coefficients ^c			
SHELXD <i>CC</i> _{all} / <i>CC</i> _{weak}		34.93/16.22	
<i>CC</i> _{free} left/right hand		62.5%/46.0%	
PATFOM ^d		13.92	
<i>Refinement</i>			
Resolution (Å)	40.7–2.4 (2.46–2.40)		29.0–1.8 (1.85–1.80)
Reflections			
Number	16,862 (1102)		18,614 (1327)
Completeness (%)	99.7 (96.1)		99.6 (98.5)
Test set (%/reflections)	5.1/852		5.1/949
<i>R</i> _{work} ^e	21.8 (31.3)		18.9 (26.5)
<i>R</i> _{free} ^e	26.6 (32.6)		23.6 (29.9)
ESU (Å) ^f	0.21		0.09
Refined atoms			
Protein	2354		1216
Water	60		185
Solute	15		26
Mean <i>B</i> -factors (Å ²)			
Wilson	71.7		30.4
Protein	81.1		33.9
Water	74.1		49.0
Solute	122.3		56.6
Ramachandran plot (%) ^g			
Favored	96.1		98.6
Outliers	0.7		0
RMSD ^h from target geometry			
Bond lengths (Å)	0.007		0.011
Bond angles (°)	1.147		1.277
Protein Data Bank ID	3ELV		3ELS

^a Data for the highest-resolution shell are given in parentheses.

^b $R_{\text{sym}}(I) = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i |I_i(hkl)|$, for *n* independent reflections and *i* observations of a given reflection; $\langle I(hkl) \rangle$ is the average intensity of the *i* observations.

^c $CC = [\sum w E_o E_c \sum w - \sum w E_o \sum w E_c] / \{[\sum w E_o^2 \sum w - (\sum w E_o)^2][\sum w E_c^2 \sum w - (\sum w E_c)^2]\}^{1/2}$, where *w* is weight (see http://shelx.uni-ac.gwdg.de/SHELX/shelx_de.pdf for full definitions).

^d FOM (figure of merit) = $|F(hkl)_{\text{best}}| / |F(hkl)|$; $F(hkl)_{\text{best}} = \sum_{\alpha} P(\alpha) F_{hkl}(\alpha) / \sum_{\alpha} P(\alpha)$.

^e $R = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$; $R_{\text{work}} = hkl \notin T$; $R_{\text{free}} = hkl \in T$; *T* = test set.

^f ESU = estimated overall coordinate error based on maximum likelihood.

^g Calculated with MolProbity (<http://molprobity.biochem.duke.edu/>).

^h RMSD – root-mean-square deviation.

components.⁹ Snu17p harbors an unusual RNA recognition motif (RRM) that presents two separate binding sites for short peptides in the C-terminus of Bud13p and in the N-terminus of Pml1p. This organization leaves the bulk of Bud13p and Pml1p available to engage in other activities.

Sequence analysis suggested that the C-terminal half of Pml1p, which is not involved in Snu17p binding, harbors a forkhead-associated (FHA) domain. Here, we elucidated the crystal structures of full-length Pml1p and a truncated form lacking the first 50 residues. The N-terminal part of Pml1p is invisible in the structure of the full-length protein and exerts no influence on the structure in the C-terminal part. The latter portion exhibits an FHA domain fold that is expanded by a number of noncanonical elements compared with other known FHA domains. In one crystal structure, a sulfate ion binds at a site equivalent to the phosphopeptide-binding site in other FHA domains. Our results support the idea that Pml1p acts as a sensor of phosphorylation marks that is recruited to the spliceosome *via* an intrinsically unstructured N-terminal region.

Results

Structure determination and quality of the models

Although biochemical and biophysical analyses had indicated disorder in the N-terminal region of Pml1p,⁹ well-diffracting crystals were obtained from the full-length protein. Pml1p crystallized in space group $P3_221$ with two monomers per asymmetric unit. We solved its structure at 2.4-Å resolution *via* the single anomalous diffraction (SAD)

method, using the anomalous scattering of a mercury derivative (Table 1). The final refined model exhibited R_{work} and R_{free} values of 21.8% and 26.6%, respectively, with all residues in the favored or additionally allowed regions of the Ramachandran plot (Table 1).

We also determined the crystal structure of Pml1p^{51–204}, a truncated version, lacking the 50 N-terminal residues. Pml1p^{51–204} crystallized in space group $H32$ with one molecule per asymmetric unit. Its structure was refined at 1.8-Å resolution with R_{work} and R_{free} values of 18.9% and 23.6%, respectively, again retaining good stereochemistry (Table 1).

Overall structure

Apart from a long, flexible loop (residues 112–121), amino acids 52–204 could be modeled unequivocally into the electron density map of full-length Pml1p (Fig. 1; because 10 residues in the flexible loop are disordered, it is referred to as the $\Delta 10$ loop in the following). The remaining N-terminal part (residues 1–51) could not be traced, although the crystals contained the intact polypeptide chain (Supplementary Fig. 1a). This finding confirmed the results of our biochemical analyses,⁹ suggesting that the first 51 residues of Pml1p are unstructured in the isolated protein.

A VAST homology search¹⁰ showed that the ordered portion of Pml1p resembles an FHA domain fold (Fig. 1). This portion of the protein comprises a short N-terminal α -helix and a sandwich of two twisted β -sheets (strands $\beta 1/\beta 2/\beta 7/\beta 8/\beta 10/\beta 11$ and strands $\beta 3/\beta 4/\beta 5/\beta 6/\beta 9$; Fig. 1). Strands $\beta 3$ and $\beta 4$ run parallel with each other, and all remaining β -strands are antiparallel. Two short 3_{10} helices are inserted between the N-terminal α -helix and strand $\beta 1$ and between strands $\beta 10$ and $\beta 11$. The

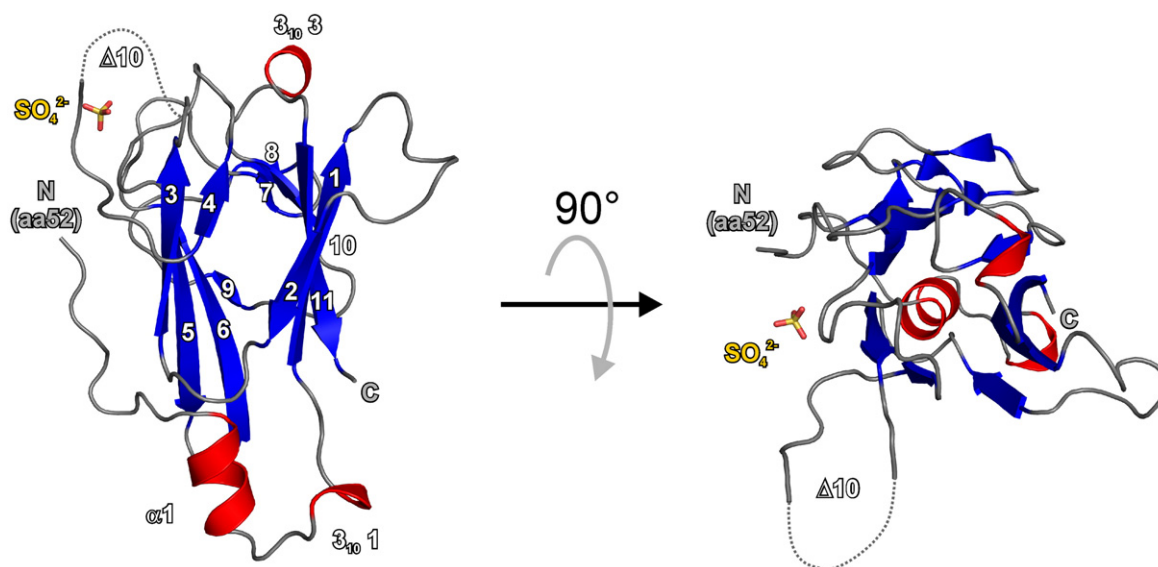


Fig. 1. Structural overview. The structure of full-length Pml1p is shown in cartoon representation. β -Strands are shown in blue; helices, in red; and loops, in gray. Secondary structure elements are numbered. The coordinated sulfate ion is shown as sticks (sulfur is shown in yellow; oxygen, in red). Unstructured regions in the protein showing no electron density are drawn as dashed lines. “C” indicates C-terminus; “N,” N-terminus. The view on the right is rotated by 90° as indicated. All structure figures were prepared using the program PyMOL [<http://www.pymol.org>].

structures of the full-length Pml1p and truncated Pml1p were highly similar (Fig. 1 and Supplementary Fig. 2), exhibiting a root-mean-square deviation (RMSD) of 0.7 Å for 144 C α atoms. The flexible Δ 10 loop of full-length Pml1p, suspended between strands β 3 and β 4, was likewise disordered in the structure of the truncated protein. A third 3_{10} helix was found between β 1 and β 2 in truncated Pml1p.

Crystals of full-length Pml1p contained two molecules in an asymmetric unit, which were related by a local 2-fold axis, forming a V-shaped pseudo-dimer (Supplementary Fig. 1b). Crystals of the truncated protein showed a similar mode of dimerization *via* a crystallographic 2-fold axis. Nevertheless, this apparent higher oligomeric state of Pml1p is most likely a consequence of crystal packing, since gel filtration indicated that the protein is monomeric in solution (Supplementary Fig. 1c).

The Pml1p FHA domain is expanded by noncanonical elements

The Pml1p FHA domain (152 amino acids) is considerably larger than most other FHA domain structures determined so far. For example, FHA domains from the yeast checkpoint protein Rad53p (FHA1¹¹), its human homolog Chk2,¹² the mitotic checkpoint protein Chfr,¹³ mammalian polynucleotide kinase (PNK¹⁴), the antigen identified by monoclonal antibody Ki-67 (Ki67¹⁵), *Arabidopsis thaliana* receptor kinase-associated protein phosphatase (KAPP¹⁶) and the putative transcriptional regulator EmbR¹⁷ encompass between 95 and 123 residues. The increased length of the Pml1p FHA domain is accounted for by several structural insertions and appendices (Fig. 2). Compared with the FHA domains of EmbR, one of the shortest

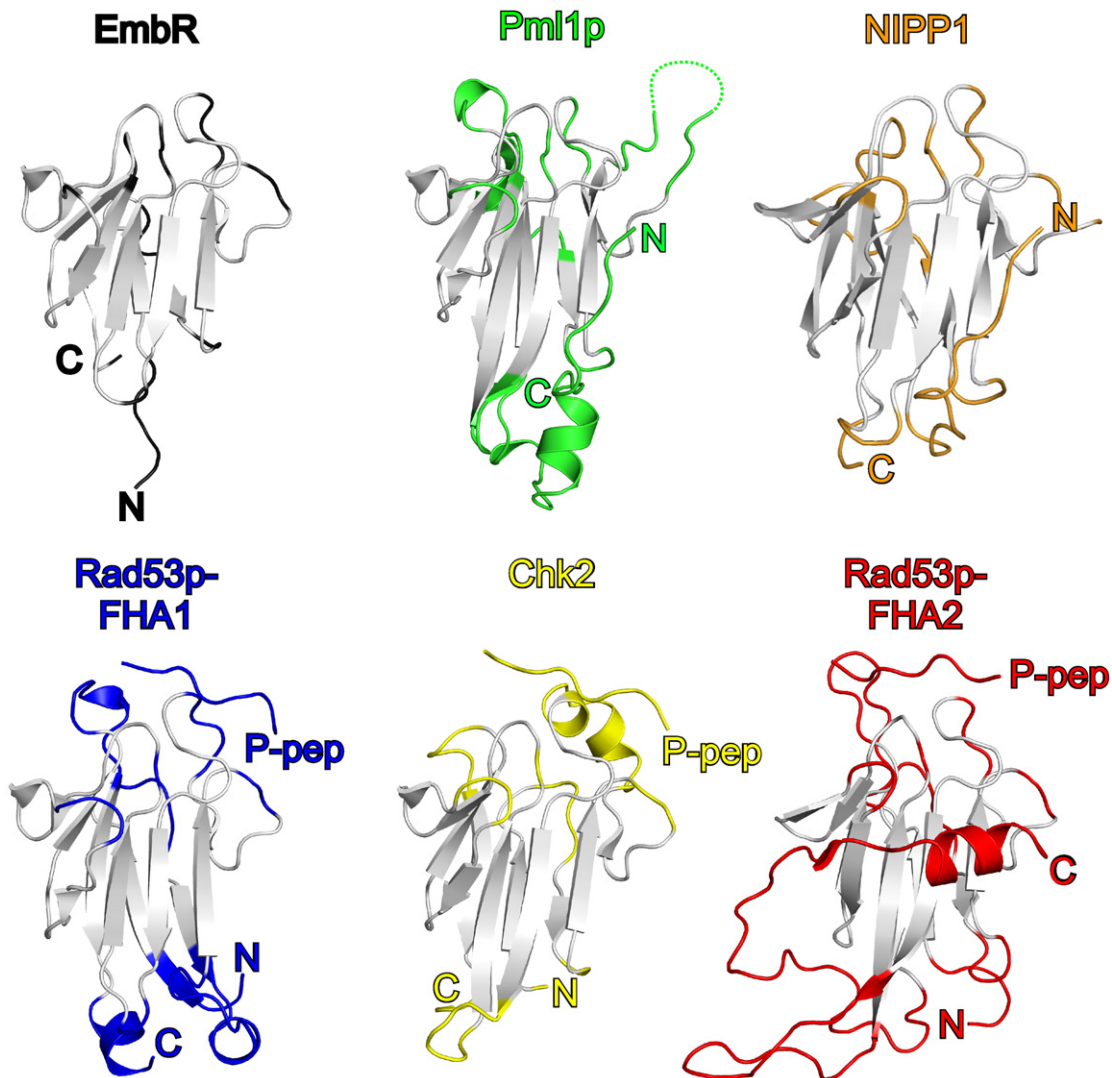


Fig. 2. Global comparison of FHA domains from various proteins. The regions common to the domains according to a VAST structure-based analysis¹⁰ are shown in gray. Protein-specific elements are differentially colored. Secondary structure elements were assigned using the DSSP algorithm (secondary structural elements for NIPP1 were taken from Ref. 18). There are prominent N- and C-terminal extensions in Pml1p, Rad53p FHA1 and Rad53p FHA2 and a helical insertion in Chk2. The small EmbR FHA domain (upper left) has neither extensions nor insertions. The Pml1p and NIPP1 FHA domains are similarly expanded at the N-termini. "C" indicates C-terminus; "N," N-terminus; P-pep, phosphopeptide ligands. Models are rotated 70° around the *x*-axis and 10° around the *y*-axis relative to the orientation in Fig. 1 (left).

known FHA domains, Pml1p exhibits an N-terminal appendix (residues 52–61) that lacks a regular secondary structure, an elongated loop between β 1 and β 2, the long unstructured Δ 10 loop inserted between strands β 3 and β 4 and a helical insertion between β 10 and β 11 (Figs. 1 and 2). In other FHA domains, different structural elements have been found in addition to the core fold (Fig. 2). For example, the second Rad53p FHA domain shows an extended α -helical C-terminal region, which runs across one of the β -sheets,¹⁹ and a long insertion connecting strands β 5 and β 6.

The N-terminal appendix is crucial for the integrity of the Pml1p FHA domain

Residues 52–61 of Pml1p lie like a clamp on top of strands β 3 and β 5 (Figs. 2 and 3a). The peptide thereby positions helix α 1, which in turn stabilizes the loop between β 2 and β 3 by hydrogen bonds. Interestingly, a similar N-terminal expansion is seen in the NIPP1 FHA domain but not in other FHA domain proteins analyzed to date (Figs. 2 and 3a). Crucial residues by which the peptide is fastened to the β -sheet are conserved in NIPP1 and Pml1p (Fig. 4a and b). We tested the importance of the N-terminal peptide for the integrity of the FHA domain by generating a truncated Pml1p variant lacking the first 61 residues. Pml1p^{62–204} exhibited dramatically reduced solubility compared with the full-length protein, Pml1p^{34–204} (Fig. 3b) and Pml1p^{51–204} (data not shown). Therefore, the N-terminal appendix confers critical stability on the Pml1p FHA domain. Similar stabilizing appendices that wrap around a core domain have recently been seen in the Prp8p Jab1/MPN domain.^{20,21}

The Pml1p FHA domain exhibits hallmarks of a phosphopeptide-binding element

Several FHA domains have been shown to function as phosphopeptide-binding modules.¹¹ Six residues that are highly conserved among

phosphopeptide-binding FHA proteins are also found in Pml1p (Gly107, Arg108, Ser137, His140, Asn163 and Asn168). With the exception of Asn168, which is thought to stabilize the structure, these residues lie in surface loops that come together at one edge of the FHA domain (Fig. 5a). Significantly, a sulfate ion is bound to this region in the full-length Pml1p structure, possibly mimicking a phosphoamino acid group (Fig. 5a). The sulfate most likely originates from the Li₂SO₄ contained in the crystallization buffer. Other FHA domains have also been observed to coordinate phosphate mimics in the absence of phosphopeptide ligands, such as a tungstate ion in the case of Chfr¹³ (Fig. 5a).

The conserved Gly107, Ser137, His140 and Asn163 do not directly interact with the sulfate ion but reinforce the productive conformation of the sulfate-binding region. The highly conserved Arg108 in loop β 3– β 4 engages in two salt bridges to sulfate oxygens. The orientation of Arg108 with respect to the sulfate ion resembles the disposition of analogous arginines in phosphopeptide-binding FHA domains.^{13,22} The conformations of the putative phosphopeptide-binding loops of Pml1p are very similar in the full-length and truncated Pml1p structures, although the latter structure lacks a sulfate ion (Fig. 5b; the crystallization buffer of truncated Pml1p did not contain sulfate). Independent of the sulfate, Arg108 seems to be held in place by Ser136.

Work on the FHA domain of *A. thaliana* KAPP suggested that loops forming the phosphopeptide-binding region rigidify upon ligand binding.²³ Similarly, loop regions in the FHA domain of Ki67 have been shown to adopt a fixed structure in the presence of a binding partner.¹⁵ NMR chemical shift changes have indicated a similar situation in NIPP1, where several loops are restructured upon phosphopeptide binding.¹⁸ One of these loops corresponds to the flexible Δ 10 loop of Pml1p, which neighbors the sulfate-binding site (Fig. 6). The Δ 10 loop is rich in acidic residues (Fig. 4a). Similar acidic insertions are only seen in Pml1p orthologs and are

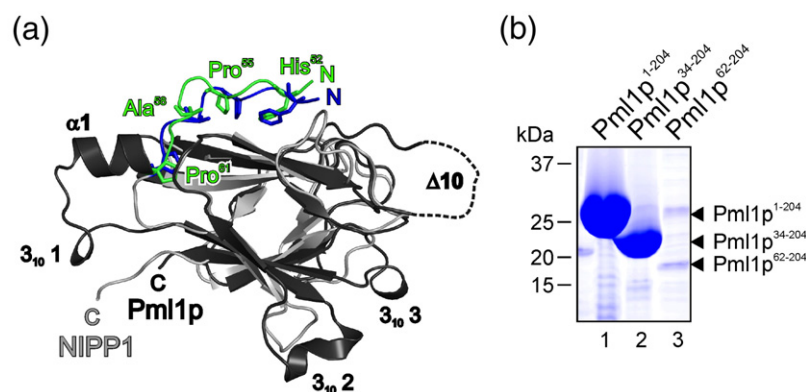


Fig. 3. Structure and function of an N-terminal extension in Pml1p. (a) Superposition of Pml1p (dark gray) and NIPP1 (light gray) reveals a conserved N-terminal extension in the structures. Highly conserved residues are shown as sticks (Pml1p is shown in green; NIPP1, in blue). Some secondary structure elements of Pml1p are labeled for clarity. "C" indicates C-terminus; "N," N-terminus. Models are rotated 100° around the z-axis (into the plane of the paper) compared

with Fig. 1 (left). (b) Full-length Pml1p and N-terminally truncated fragments of Pml1p were produced under identical conditions in *E. coli*. Proteins were captured on Ni²⁺-NTA agarose from soluble extract, and eluted proteins were analyzed via SDS-PAGE. Lane 1, Pml1p^{1–204} (full length); lane 2, Pml1p^{34–204}; lane 3, Pml1p^{62–204}. Removal of the N-terminal extension of Pml1p (residues 51–61) renders the protein insoluble.

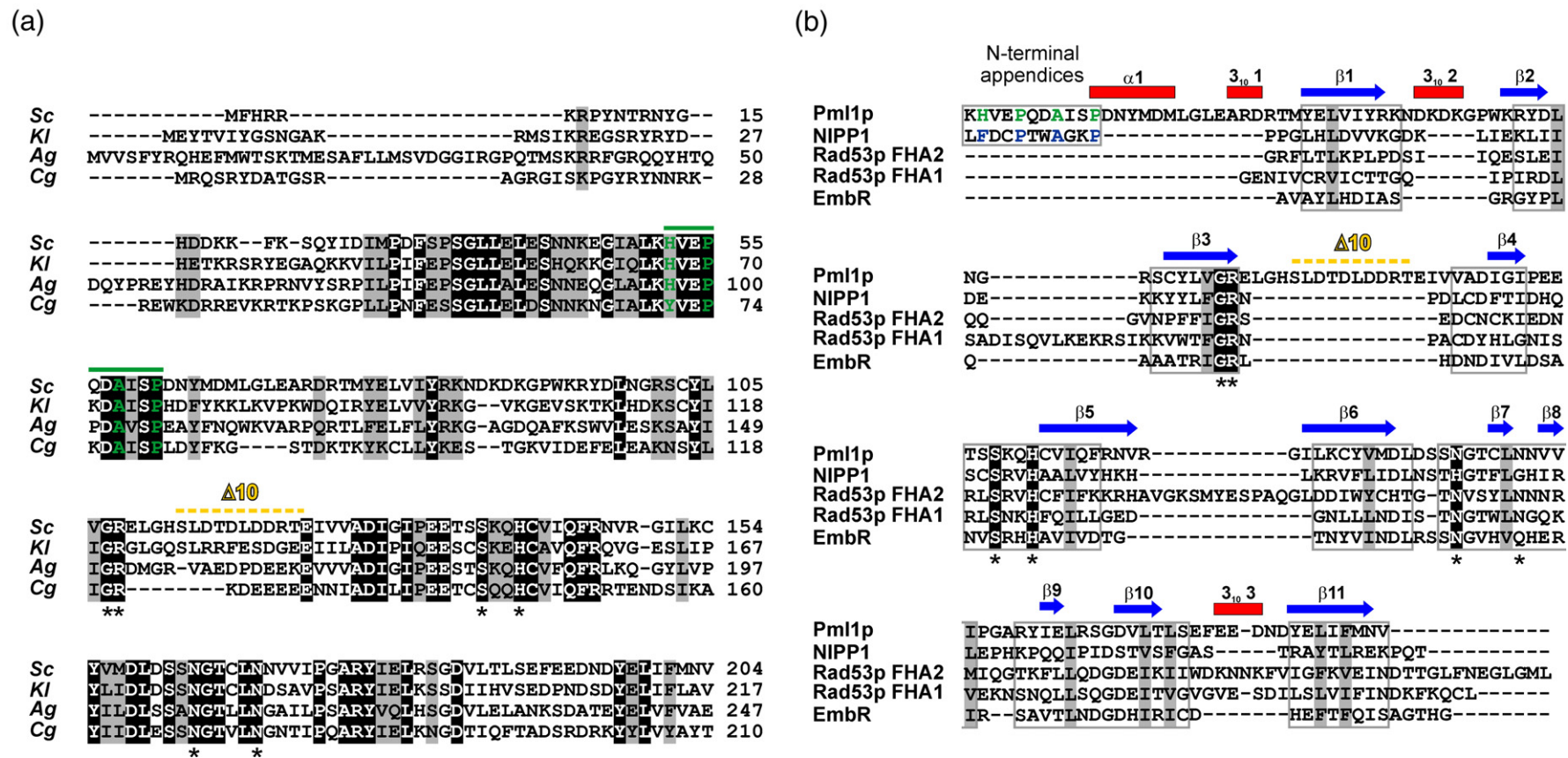


Fig. 4. Multiple-sequence alignments. (a) Multiple-sequence alignment of *Saccharomyces cerevisiae* Pml1p (Sc; accession code NP_013116.1) with orthologs from *Kluyveromyces lactis* (Kl; XP_453486.1), *Ashbya gossypii* (Ag; NP_983442.1) and *Candida glabrata* (Cg; XP_445454.1). Darker background corresponds to higher conservation. The $\Delta 10$ loop of Pml1p is marked by a yellow dashed line. Green characters indicate conserved residues, which seem to be crucial for Pml1p stability. Asterisks mark conserved residues that are important for phosphopeptide binding in other FHA domains. (b) Multiple structure-based sequence alignment of the FHA domain structures of Pml1p (this study), NIPP1 [Protein Data Bank (PDB) ID 2JPE], Rad53p FHA2 (PDB ID 1FHR), Rad53p FHA1 (PDB ID 1G6G) and EmbR (PDB ID 2FF4). Sequences were aligned using the VAST server and adjusted manually. The N-terminal appendices of Pml1p and NIPP1 are indicated. Secondary structure elements of Pml1p⁵¹⁻²⁰⁴ are shown as blue arrows (β -strands) and red boxes (helices) above the alignment. The $\Delta 10$ loop of Pml1p is marked by a yellow dashed line. Structurally aligned residues are boxed (compare also with the gray regions in Fig. 3). The gray background corresponds to highly conserved, hydrophobic residues, whereas the black background corresponds to conserved motifs important for phosphopeptide interaction.

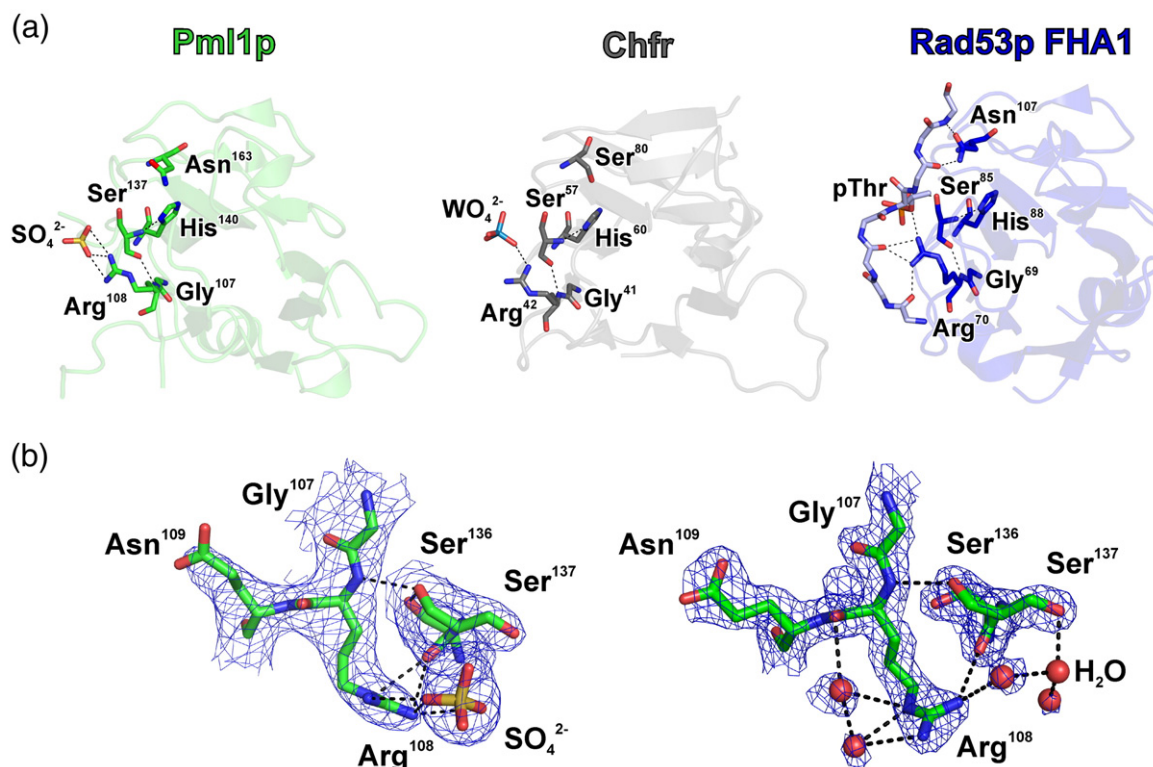


Fig. 5. The putative phosphoamino acid-binding region of Pml1p. (a) Residues of Pml1p (left) and Chfr (center) presumably involved in phosphopeptide binding are shown as sticks and color coded by atom type (carbon is shown in green and in gray, respectively; oxygen, in red, nitrogen, in blue; sulfur, in yellow; and tungstate, in cyan). Residues of Rad53p FHA1 (right; carbon is shown in dark blue) in complex with a phosphopeptide (backbone model with a phosphothreonine residue; carbon is shown in light blue) are shown as sticks. Hydrogen bonds and salt bridges are indicated by black dashed lines. Models are rotated 130° around the y -axis compared with Fig. 1 (left). (b) $2F_o - F_c$ electron density maps of the sulfate-bound (left) and unbound (right) structures of Pml1p contoured at the 1σ level within a 5-Å distance of Arg108. Black dashed lines represent hydrogen bonds and salt bridges. Amino acids are shown as sticks. In both structures, the backbone oxygen of Ser136 is engaged in positioning Arg108. Color coding is the same as in (a).

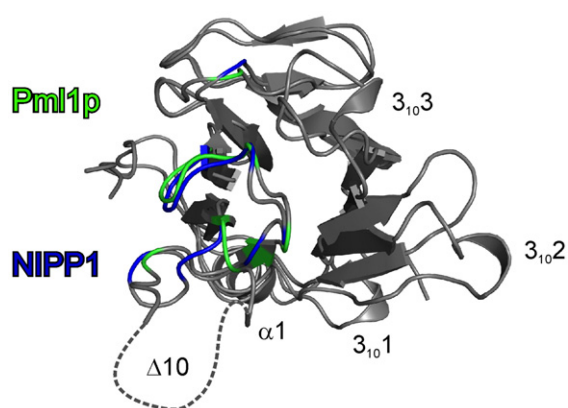


Fig. 6. Comparison of Pml1p and NIPP1 highlighting hotspots for phosphopeptide interaction. Residues in NIPP1, which are affected by binding of a phosphopeptide as judged by NMR chemical shift changes,¹⁸ are highlighted in blue. Corresponding residues in Pml1p are shown in green. An elongated phosphopeptide bound to Pml1p in a similar fashion as to NIPP1 would point toward the unstructured Δ10 loop region. The models are rotated 70° around the x -axis compared with Fig. 1 (left).

absent from the FHA domains of other proteins (Fig. 4a and b). We therefore suggest that the Δ10 loop helps determine ligand specificity by engaging in direct contacts with the phosphopeptide.

Discussion

Functional organization of Pml1p

We have shown by crystal structure analysis that Pml1p exhibits an FHA domain with noncanonical expansions and insertions (Fig. 2) and that the ordered portion of the protein is preceded by an intrinsically unstructured region. A short peptide N-terminal of the canonical FHA domain lies like a clamp across one of the β -sheets and protects the domain against aggregation. A similarly positioned and highly homologous peptide is found only in NIPP1,¹⁸ another FHA domain protein with a function in pre-mRNA splicing (Figs. 3a and 4b).

FHA domain proteins often contain additional functional elements. For instance, in the *A. thaliana* KAPP, a central FHA domain is flanked by an N-

terminal type I membrane anchor and a C-terminal protein phosphatase type 2C catalytic domain.²⁴ A flexible linker probably allows free rotational movement between the FHA and protein phosphatase type 2C domains.²⁵ Similarly, the PNK FHA domain is thought to be flexibly linked to the kinase/phosphatase catalytic domain.¹⁴ Here, we have shown that in Pml1p, a C-terminal FHA domain is associated with an intrinsically unstructured region at the N-terminus. This N-terminal part serves to anchor Pml1p to Snu17p in the RES complex.⁹ Tethering of Pml1p to Snu17p *via* a flexible hinge would allow largely unrestricted movement of the FHA domain with respect to the scaffold, even if Snu17p was docked to a massive spliceosome. The FHA domain could thus reorient in order to interact with binding partners. A similar situation was observed in the lateral L7/12 stalk of the 50S ribosomal subunit, in which L7/12 C-terminal domains are flexibly linked to the stalk base, allowing them to “patrol” neighboring areas in search of elongation factors.²⁶

The Pml1p FHA domain as a putative phosphopeptide-binding element

FHA domains are known to serve as phosphorylation sensors by interacting primarily with phosphorylated threonine residues in ligand proteins.²² As a prominent example, the FHA domain of mammalian PNK targets the enzyme to sites of base

excision repair and nonhomologous end-joining repair by interacting with phosphorylated XRCC1 and XRCC4, which are central components of the respective pathways.^{22,27–29} Phosphopeptide-binding FHA domains are also found in proteins associated with pre-mRNA splicing. NIPP1 harbors an FHA domain and interacts with CDC5L and SF3b155 in a phosphorylation-dependent manner.^{30,31} Dominant negative mutants of NIPP1 that block spliceosome assembly at the step of catalytic activation have been characterized.³²

Several structures of FHA domains alone and in complex with phosphopeptide ligands are available.^{12–16,19,22,33–35} In these structures, phosphopeptides bind in an extended conformation at one side of the domain and the phosphoamino acid-binding site is formed by a set of highly conserved residues. Our data suggest that the Pml1p FHA domain can also act as a phosphopeptide-binding domain. First, interaction of Pml1p with Snu17p leaves the canonical phosphopeptide-binding site on the FHA domain available for further interaction with a phosphoprotein. Second, the Pml1p FHA domain exhibits all residues critical for phosphopeptide binding (Fig. 5a). Third, Pml1p can bind a phosphate-mimicking sulfate ion at the putative phosphoamino acid-binding site (Fig. 5).

Conservation of the structures of full-length Pml1p (sulfate-bound) and truncated Pml1p (lacking sulfate) in the immediate vicinity of the sulfate-binding site suggests that the binding site for the

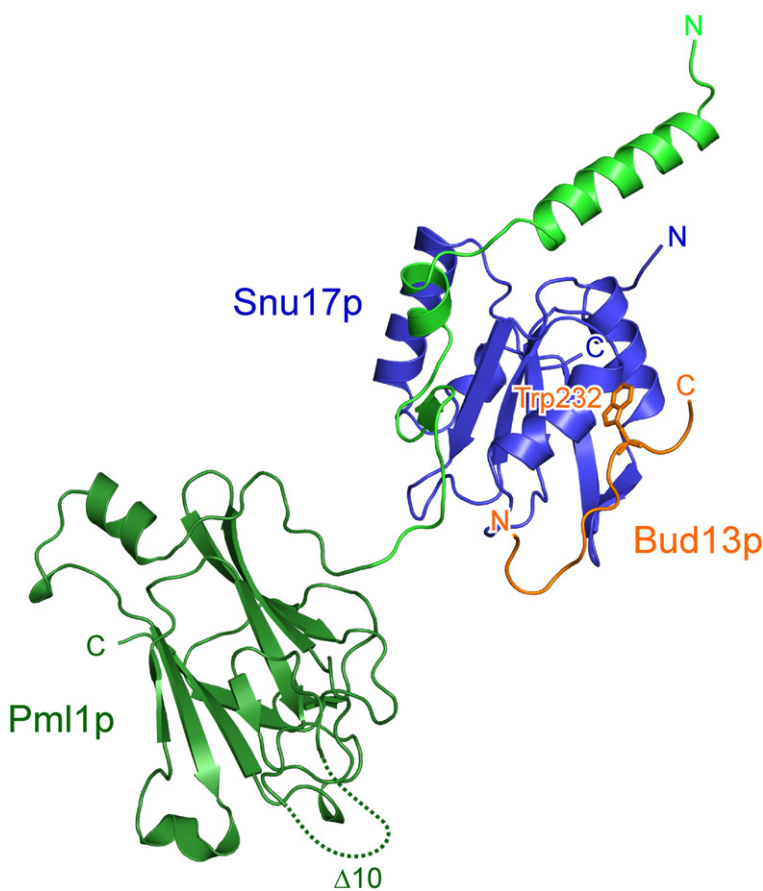


Fig. 7. Global model of the RES complex (see [Materials and Methods](#)). Snu17p, Pml1p and Bud13p are shown in blue, green and orange, respectively. The dark green portion of the Pml1p structure has been determined experimentally (residues 51–204). The light green N-terminal part was modeled using the X-ray crystal structure of human SF3b14a–SF3b155 complex as a template.³⁷ The Snu17p–Bud13p interaction was modeled according to the solution structure of the U2AF65 (RRM3)–SF1 complex.³⁸ Only the highly conserved region of Bud13p involved in the interaction with Snu17p is shown. “N” indicates N-termini; “C,” C-termini.

phosphoamino acid moiety of a putative phosphopeptide ligand is largely preformed in uncomplexed Pml1p. On the other hand, the unstructured $\Delta 10$ loop is ideally positioned to bind neighboring regions of the phosphopeptide ligand by an induced-fit mechanism, thereby increasing specificity. A similar mixed binding strategy has recently been revealed in complexes involving the spliceosomal Prp31 protein.³⁶

Presently, it is not known whether the putative phosphopeptide-binding activity of Pml1p plays a role during pre-mRNA splicing. Further structural and functional studies will be required to characterize the targets recognized by Pml1p and its exact mode of action. One interesting observation is that Snu17p and Pml1p interact with a K_d in the micromolar range, significantly higher than the nanomolar K_d of the Snu17p–Bud13p interaction.⁹ Therefore, it is possible that Snu17p recruits Pml1p intermittently as a flexible phosphorylation sensor during pre-mRNA retention or splicing.

An architectural model of the RES complex

On the basis of the crystal structure of Pml1p (this work) and biochemical analyses,⁹ we have devised a structural model of the RES core complex centered on Snu17p (Fig. 7). In the model, short binding epitopes in the unstructured N-terminal part of Pml1p and in the C-terminal part of Bud13p come to lie on two separate binding sites of Snu17p. We have added the expanded FHA domain structure of Pml1p determined herein to its Snu17p-binding peptide. Our model illustrates how Snu17p serves as a scaffolding molecule that displays a functional element in Pml1p (the C-terminal FHA domain).

Materials and Methods

Purification and characterization of Pml1p

Cloning of the gene encoding yeast Pml1p has been described previously.⁹ Full-length Pml1p was expressed in *Escherichia coli* Rosetta 2(DE3) cells from a pCDFDuet-1 expression vector (Novagen) N-terminally fused to a cleavable His₆ tag. Transformed cells were grown in LB medium supplemented with chloramphenicol/streptomycin to an OD₆₀₀ of 0.8–1.2. After induction with 0.5 mM isopropyl- β -D-thiogalactopyranoside (IPTG), cells were incubated for 3–4 h at 25 °C before harvesting. The target protein was captured from soluble extract *via* Ni²⁺-nitrilotriacetic acid (NTA) resin and eluted by addition of imidazole. The tag was removed by addition of tobacco etch virus protease (protease/protein mass ratio=1:40). After buffer exchange to 50 mM Tris–HCl, pH 7.5, 150 mM NaCl and 1 mM DTT *via* a 16/10 HighPrep desalting column (GE Healthcare), the protein solution was loaded onto CM-Sepharose (GE Healthcare), washed and eluted in a linear gradient to the same buffer plus 500 mM NaCl. Pooled fractions were further purified by size-exclusion chromatography on a Superdex 75 26/60 column (GE Healthcare) in 10 mM Tris–HCl, pH 7.5, 150 mM NaCl and 1 mM DTT. The purified protein was concentrated to 16 mg/ml with spin concentrators

(VIVAscience). The final sample was >98% pure as judged by SDS-PAGE analysis.

An N-terminally truncated version of Pml1p encompassing residues 51–204 (Pml1p^{51–204}), also bearing an N-terminal cleavable His₆ tag, was expressed in *E. coli* Rosetta 2(DE3) cells from a pETM-11 vector (EMBL Heidelberg†, Heidelberg, Germany). The protein was purified as described for full-length Pml1p but loaded directly on a Superdex 75 26/60 column (GE Healthcare) after tobacco etch virus cleavage and was subsequently cycled over preequilibrated Ni²⁺-NTA to remove residual uncleaved fusion protein. The final Pml1p^{51–204} pool was concentrated to 34 mg/ml.

Analytical gel-filtration experiments were performed on a SMART system (GE Healthcare) using a Superdex 75 PC3.2 column equilibrated in 10 mM Tris–HCl, pH 7.5, 150 mM NaCl and 1 mM DTT.

Crystallization of Pml1p

Full-length Pml1p was crystallized by the sitting-drop vapor-diffusion method using a reservoir comprising 100 mM Tris–HCl, pH 7.5, 200 mM Li₂SO₄, 21%–24% PEG (polyethylene glycol) 3350 and IPTG as an additive. For experimental phasing, Pml1p crystals were incubated in reservoir solution containing 1 mM Baker's dimercurial (Hampton Research, Inc.). Soaked crystals were transferred into reservoir solution containing 1 mM Baker's dimercurial and 6% polypropylene glycol and were flash frozen in liquid nitrogen for data collection. Native crystals were transferred into perfluoropolyether (Hampton Research, Inc.), and trace amounts of reservoir were removed before flash freezing in liquid nitrogen.

Pml1p^{51–204} was crystallized by the hanging-drop vapor-diffusion technique using a reservoir comprising 75 mM Hepes, pH 7.1, 17% PEG 10000, 15% glycerol and 6.5% ethylene glycol. Crystals appeared after 5 days at 20 °C. For diffraction experiments, crystals were directly frozen in a liquid nitrogen stream.

Diffraction data collection and crystal structure determination

A native data set of a full-length Pml1p crystal was collected at 100 K on beamline BW6 (DESY, Hamburg, Germany). Crystal symmetry and unit cell dimensions (Table 1) suggested two Pml1p molecules per asymmetric unit. Crystals soaked with Baker's dimercurial were measured at the high-energy side of the Hg–LIII edge on beamline PXII (SLS, Villigen, Switzerland) and beamline 14-2 (BESSY, Berlin, Germany) and were nonisomorphous with the native crystals. Diffraction data were processed using DENZO/Scalepack.³⁹ The structure was solved by the SAD technique using the SHELX program collection.⁴⁰ Scaled diffraction data were analyzed using SHELXC. The heavy atom substructure was solved using SHELXD. Initial phases were calculated using SHELXE and improved by solvent flattening. The SAD electron density map allowed building of β -sheets that could be used as a search model for phasing of the native data set. The model was completed by iterative cycles of manual model building in Coot⁴¹ and automated refinement with the program Refmac5.⁴² During initial refinement cycles, the

† http://www.pepcore.embl.de/cloning/choice_vector/ecoli_embl.html

model was restrained by a 2-fold noncrystallographic symmetry.

Diffraction data of a Pml1p^{51–204} crystal were measured on beamline PXII of SLS and processed as described above. Crystal symmetry and unit cell dimensions (Table 1) suggested one Pml1p^{51–204} molecule per asymmetric unit. The structure was solved by molecular replacement with the program Molrep⁴³ using the structure coordinates of the full-length protein. Automated refinement and building of the water structure were done with Refmac5, and the model was manually adjusted by using Coot.

The final structural models were validated with CNS⁴⁴ by calculating simulated annealing composite omit maps covering the entire asymmetric units of the two crystal forms. Secondary structure elements were assigned with the program DSSP.⁴⁵

Structural modeling

A global architectural model of the RES complex was devised by combining the present crystal structure of Pml1p and the model of the Snu17p RRM in complex with binding peptides of Pml1p and Bud13p.⁹ The peptide representing Pml1p was connected to the crystal structure of the folded portion of Pml1p by manually inserting the corresponding number of alanine residues. The Pml1p FHA domain and the Snu17p RRM were oriented arbitrarily relative to each other. All of the manual modeling was done with Coot.

Protein Data Bank accession code

Coordinates and structure factors have been deposited with the Protein Data Bank with accession numbers 3ELV (full-length Pml1p) and 3ELS (Pml1p^{51–204}).

Acknowledgements

This work was supported by the Max Planck Society (R.L. and M.C.W.). We are grateful to the staff of beamline BW6 at DESY (Hamburg, Germany), beamline PXII at SLS (Villigen, Switzerland) and beamline 14.2 at BESSY (Berlin, Germany) for providing support during diffraction data collection. We also thank Elke Penka for providing excellent technical assistance and Christian Stegmann for helping with the crystallographic analysis.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2008.10.087](https://doi.org/10.1016/j.jmb.2008.10.087)

References

- Brow, D. A. (2002). Allosteric cascade of spliceosome activation. *Annu. Rev. Genet.* **36**, 333–360.
- Jurica, M. S. & Moore, M. J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell*, **12**, 5–14.
- Nilsen, T. W. (2003). The spliceosome: the most complex macromolecular machine in the cell? *BioEssays*, **25**, 1147–1149.
- Will, C. L. & Lührmann, R. (2006). Spliceosome structure and function. In *The RNA World* (Gesteland, R. F., Cech, T. R. & Atkins, J. F., eds), pp. 369–400, 3rd edit. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Staley, J. P. & Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, **92**, 315–326.
- Dziembowski, A., Ventura, A. P., Rutz, B., Caspary, F., Faux, C., Halgand, F. *et al.* (2004). Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *EMBO J.* **23**, 4847–4856.
- Scherrer, F. W., Jr & Spingola, M. (2006). A subset of Mer1p-dependent introns requires Bud13p for splicing activation and nuclear retention. *RNA*, **12**, 1361–1372.
- Spingola, M., Armisen, J. & Ares, M., Jr (2004). Mer1p is a modular splicing factor whose function depends on the conserved U2 snRNP protein Snu17p. *Nucleic Acids Res.* **32**, 1242–1250.
- Trowitzsch, S., Weber, G., Luhrmann, R. & Wahl, M. C. (2008). An unusual RNA recognition motif acts as a scaffold for multiple proteins in the pre-mRNA retention and splicing complex. *J. Biol. Chem.* **283**, 32317–32327.
- Gibrat, J. F., Madej, T. & Bryant, S. H. (1996). Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol.* **6**, 377–385.
- Durocher, D. & Jackson, S. P. (2002). The FHA domain. *FEBS Lett.* **513**, 58–66.
- Li, J., Williams, B. L., Haire, L. F., Goldberg, M., Wilker, E., Durocher, D. *et al.* (2002). Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2. *Mol. Cell*, **9**, 1045–1054.
- Stavridi, E. S., Huyen, Y., Loreto, I. R., Scolnick, D. M., Halazonetis, T. D., Pavletich, N. P. & Jeffrey, P. D. (2002). Crystal structure of the FHA domain of the Chfr mitotic checkpoint protein and its complex with tungstate. *Structure*, **10**, 891–899.
- Bernstein, N. K., Williams, R. S., Rakovszky, M. L., Cui, D., Green, R., Karimi-Busheri, F. *et al.* (2005). The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. *Mol. Cell*, **17**, 657–670.
- Byeon, I. J., Li, H., Song, H., Gronenborn, A. M. & Tsai, M. D. (2005). Sequential phosphorylation and multi-site interactions characterize specific target recognition by the FHA domain of Ki67. *Nat. Struct. Mol. Biol.* **12**, 987–993.
- Lee, G. I., Ding, Z., Walker, J. C. & Van Doren, S. R. (2003). NMR structure of the forkhead-associated domain from the *Arabidopsis* receptor kinase-associated protein phosphatase. *Proc. Natl Acad. Sci. USA*, **100**, 11261–11266.
- Alderwick, L. J., Molle, V., Kremer, L., Cozzone, A. J., Dafforn, T. R., Besra, G. S. & Futterer, K. (2006). Molecular structure of EmbR, a response element of Ser/Thr kinase signaling in *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA*, **103**, 2558–2563.
- Kumeta, H., Ogura, K., Adachi, S., Fujioka, Y., Tanuma, K., Kikuchi, K. & Inagaki, F. (2008). The NMR structure of the NIPP1 FHA domain. *J. Biomol. NMR*, **40**, 219–224.
- Wang, P., Byeon, I. J., Liao, H., Beebe, K. D.,

- Yongkiettrakul, S., Pei, D. & Tsai, M. D. (2000). II. Structure and specificity of the interaction between the FHA2 domain of Rad53 and phosphotyrosyl peptides. *J. Mol. Biol.* **302**, 927–940.
20. Pena, V., Liu, S., Bujnicki, J. M., Luhrmann, R. & Wahl, M. C. (2007). Structure of a multipartite protein–protein interaction domain in splicing factor prp8 and its link to retinitis pigmentosa. *Mol. Cell*, **25**, 615–624.
21. Zhang, L., Shen, J., Guarnieri, M. T., Heroux, A., Yang, K. & Zhao, R. (2007). Crystal structure of the C-terminal domain of splicing factor Prp8 carrying retinitis pigmentosa mutants. *Protein Sci.* **16**, 1024–1031.
22. Durocher, D., Taylor, I. A., Sarbassova, D., Haire, L. F., Westcott, S. L., Jackson, S. P. *et al.* (2000). The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. *Mol. Cell*, **6**, 1169–1182.
23. Ding, Z., Lee, G. I., Liang, X., Gallazzi, F., Arunima, A. & Van Doren, S. R. (2005). PhosphoThr peptide binding globally rigidifies much of the FHA domain from *Arabidopsis* receptor kinase-associated protein phosphatase. *Biochemistry*, **44**, 10119–10134.
24. Stone, J. M., Collinge, M. A., Smith, R. D., Horn, M. A. & Walker, J. C. (1994). Interaction of a protein phosphatase with an *Arabidopsis* serine–threonine receptor kinase. *Science*, **266**, 793–795.
25. Lee, G. I., Li, J., Walker, J. C. & Van Doren, S. R. (2003). ^1H , ^{13}C and ^{15}N resonance assignments of the kinase-interacting FHA domain of *Arabidopsis thaliana* kinase-associated protein phosphatase. *J. Biomol. NMR*, **25**, 253–254.
26. Diaconu, M., Kothe, U., Schlunzen, F., Fischer, N., Harms, J. M., Tonevitsky, A. G. *et al.* (2005). Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell*, **121**, 991–1004.
27. Caldecott, K. W. (2003). DNA single-strand break repair and spinocerebellar ataxia. *Cell*, **112**, 7–10.
28. Koch, C. A., Agyei, R., Galicia, S., Metalnikov, P., O'Donnell, P., Starostine, A. *et al.* (2004). Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. *EMBO J.* **23**, 3874–3885.
29. Loizou, J. I., El-Khamisy, S. F., Zlatanou, A., Moore, D. J., Chan, D. W., Qin, J. *et al.* (2004). The protein kinase CK2 facilitates repair of chromosomal DNA single-strand breaks. *Cell*, **117**, 17–28.
30. Boudrez, A., Beullens, M., Groenen, P., Van Eynde, A., Vulsteke, V., Jagiello, I. *et al.* (2000). NIPP1-mediated interaction of protein phosphatase-1 with CDC5L, a regulator of pre-mRNA splicing and mitotic entry. *J. Biol. Chem.* **275**, 25411–25417.
31. Boudrez, A., Beullens, M., Waelkens, E., Stalmans, W. & Bollen, M. (2002). Phosphorylation-dependent interaction between the splicing factors SAP155 and NIPP1. *J. Biol. Chem.* **277**, 31834–31841.
32. Beullens, M. & Bollen, M. (2002). The protein phosphatase-1 regulator NIPP1 is also a splicing factor involved in a late step of spliceosome assembly. *J. Biol. Chem.* **277**, 19855–19860.
33. Byeon, I. J., Yongkiettrakul, S. & Tsai, M. D. (2001). Solution structure of the yeast Rad53 FHA2 complexed with a phosphothreonine peptide pTXXL: comparison with the structures of FHA2–pYXL and FHA1–pTXXD complexes. *J. Mol. Biol.* **314**, 577–588.
34. Liao, H., Byeon, I. J. & Tsai, M. D. (1999). Structure and function of a new phosphopeptide-binding domain containing the FHA2 of Rad53. *J. Mol. Biol.* **294**, 1041–1049.
35. Yuan, C., Yongkiettrakul, S., Byeon, I. J., Zhou, S. & Tsai, M. D. (2001). Solution structures of two FHA1–phosphothreonine peptide complexes provide insight into the structural basis of the ligand specificity of FHA1 from yeast Rad53. *J. Mol. Biol.* **314**, 563–575.
36. Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Luhrmann, R. *et al.* (2007). Binding of the human Prp31 Nop domain to a composite RNA–protein platform in U4 snRNP. *Science*, **316**, 115–120.
37. Schellenberg, M. J., Edwards, R. A., Ritchie, D. B., Kent, O. A., Golas, M. M., Stark, H. *et al.* (2006). Crystal structure of a core spliceosomal protein interface. *Proc. Natl Acad. Sci. USA*, **103**, 1266–1271.
38. Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A. & Sattler, M. (2003). Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol. Cell*, **11**, 965–976.
39. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology*, Vol. 276: *Macromolecular Crystallography, Part A* (Carter, C. W., Jr & Sweet, R. M., eds), Academic Press, New York, NY.
40. Sheldrick, G. M. (2008). A short history of SHELX. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **64**, 112–122.
41. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
42. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240–255.
43. Vagin, A. & Teplyakov, A. (2000). An approach to multi-copy search in molecular replacement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **56**, 1622–1624.
44. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54**, 905–921.
45. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.