Crystallization and biochemical characterization of the human spliceosomal Aar2–Prp8^{RNaseH} complex

Karine Santos,^{a‡} Marco Preussner,^{b‡} Anna Christina Heroven^a and Gert Weber^{a*}

^aStructural Biochemistry, Free University of Berlin, Takustrasse 6, 14195 Berlin, Germany, and ^bRNA Biochemistry, Free University of Berlin, Takustrasse 6, 14195 Berlin, Germany. *Correspondence e-mail: gweber@chemie.fu-berlin.de

Received 21 August 2015

Accepted 11 October 2015

Edited by M. S. Weiss, Helmholtz-Zentrum Berlin für Materialien und Energie, Germany

‡ These authors made an equal contribution.

Keywords: Aar2; human Aar2–Prp8^{RNaseH}.

In eukaryotes, the removal of nuclear noncoding sequences (pre-mRNA splicing) is catalyzed by the spliceosome, which consists of five ribonucleoprotein particles (U1, U2, U4, U5 and U6 snRNPs, each with a respective snRNA) and a plethora of protein factors that aid spliceosomal maturation, assembly, activation and disassembly. Recently, the U5 snRNP maturation factor Aar2p from baker's yeast has been characterized structurally and biochemically. Aar2p binds to the RNaseH (RH) and Jab1/MPN domains of the highly conserved U5-specific Prp8p, which forms a framework for the spliceosomal catalytic centre. Thereby, Aar2p sterically excludes Brr2p, a helicase essential for the catalytic activation of the spliceosome, from Prp8p binding. At the same time, Aar2p blocks U4/U6 di-snRNA binding to Prp8p. Aar2p therefore prevents premature spliceosome activation and its functions are regulated by reversible phosphorylation. To date, little is known about the hypothetical human Aar2 (hsAar2) orthologue C20ORF4. This study identifies C20ORF4 (i) as part of the HeLa proteome by Western blotting and (ii) as a true Aar2 orthologue which binds to the RH domain (hsRH) of Prp8 and corroborates an evolutionary link between yeast and human Aar2 function. An elaborate strategy was devised to crystallize hsAar2 in complex with hsRH. The analysis of initial weakly diffracting crystals obtained by *in situ* proteolysis and homology modelling guided the design of an hsAar2 construct in which an internal loop was replaced by three serines (hsAar2^{Δloop}). A complex of hsAar2^{Δloop} and hsRH crystallized in space group C2; the crystals diffracted to 2.35 Å resolution and were suitable for structure determination by molecular-replacement approaches. The study presented here suggests a connection between Aar2 and the spliceosome in human cells and paves the way for structural studies of human Aar2.

1. Introduction

The majority of eukaryotic genes and thus precursor messenger RNAs (pre-mRNAs) have coding regions (exons) interspersed with noncoding regions (introns). To yield a continuous reading frame, intron removal and concomitant exon ligation (pre-mRNA splicing) are catalyzed by a multi-megadalton ribonucleoprotein (RNP) complex, the spliceosome. Further modifications, such as 5' capping and 3' polyadenylation of the pre-mRNA, lead to a mature mRNA competent for ribosomal translation. The major U2-type spliceosome includes the small nuclear ribonucleoproteins (snRNPs) U1, U2, U4, U5 and U6, each with a respective U1, U2, U4, U5 or U6 snRNA, seven common Sm proteins [or like-Sm (LSm) in the case of U6 snRNP] and a set of particle-specific proteins (Wahl *et al.*, 2009). Apart from the snRNPs, multiple spliceosomal proteins and protein complexes join, guide and modulate pre-mRNA splicing. Prp8 is one of the most conserved nuclear eukaryotic proteins and coordinates proteins, snRNAs and the pre-mRNA at the spliceosomal

catalytic centre. To this end, Prp8 regulates the helicase Brr2, which unwinds U4/U6 di-snRNAs during catalytic activation of the spliceosome (Galej *et al.*, 2013; Mozaffari-Jovin *et al.*, 2012, 2013). Structural studies in our laboratory and by other groups identified two regulatory pseudo-enzyme domains at the very C-terminus of Prp8 comprising an RNaseH-like (RH) and a Jab1/MPN-like (JM) fold (Pena *et al.*, 2007, 2008; Ritchie *et al.*, 2008; Yang *et al.*, 2008; Zhang *et al.*, 2007).

In *Saccharomyces cerevisiae*, Aar2p was first identified as being involved in splicing of the MAT *a1* cistron and some other genes related to mitotic cell growth. An AAR2 deletion mutant is not viable (Nakazawa *et al.*, 1991; Gottschalk *et al.*, 2001). Aar2p was shown to be part of a cytoplasmic precursor U5 lacking the helicase Brr2p, which triggers spliceosomal activation (Boon *et al.*, 2007). Dissection of the protein-interaction sites of Prp8p by insertion of a transposon and

Figure 1 Design of an hsAar2 loop-truncation mutant based on sequence alignment and structural comparison employing a homology model derived from *HHpred*. (a) The alignment of the human and yeast Aar2 sequences was prepared by *ClustalX2* (Larkin *et al.*, 2007) and shaded with *ALSCRIPT* (Barton, 1993). Conservation is indicated by a grey background and identity by a black background. The numbering above the alignment refers to human Aar2. Below the alignment, secondary-structure elements of yeast Aar2 as detected with *PROMOTIF* (Hutchinson & Thornton, 1996) in the yeast Aar2^{Δloop}-RH-JM complex structure are shown in red. Boxed residues indicate residues that were replaced by three or five serines in human and yeast Aar2^{Δloop}, respectively. Blue triangles above the alignment indicate residues of yeast Aar2^{Δloop} in contact with yeast RH. (b) Superimposition of yeast Aar2 (PDB entry 3sbt; Weber *et al.*, 2011) and a homology model of human Aar2 derived from *HHpred* based on the yeast Aar2p structure (Söding *et al.*, 2005). The inset shows a close-up view of the flexible internal hsAar2 loop which was replaced by three serine residues (3S) bridging residues Arg201 and His169 in hsAar2^{Δloop}.

Table 1

Human Aar2 constructs used in this study for protein production.

Source organism	Human
DNA source	Synthetic
Expression vector	pFL10His
Expression host	<i>Trichoplusia ni</i>
Complete amino-acid sequence of the construct produced for binding studies and <i>in situ</i> proteolysis crystallization (hsAar2) [†]	GAEFMAAVQMDPELAKRLFFEGATVVILNMPKGT- EFGIDYNSWEVGPFRGVKMIPPGIHFHLYSS- VDKANPKEVGPRMGFFLSLHQRGLTVLRWSTL- REEVDLSAPAESEVEAMRANLQELDQFLGPYP- YATLKKWISLTNFISEATVEKLQPENRQICAF- SDVLPVLSMKHTKDRVGNLPRCGIECKSYQE- GLARLPEMKPRAGTEIRFSELPTQMFPEGATP- AEITKHSMDLSYALETVLNKKFPSSPDVLGE- LQFAFVCFLLGNVYEAFAHWKRLNLLCRSEA- AMMKHHTLYINLISILYHQLGEIPADFFVDIV- SQDNFLTSTLQVFFSSACSIAVDATLRKKAKEK- FQAHLTKKFRWDAEPEDCAPVVVELPEGIE- MG
Amino-acid sequence of the construct used for binding studies and crystallization (hsAar2 ^{Δloop}) [†]	GAEFMAAVQMDPELAKRLFFEGATVVILNMPKGT- EFGIDYNSWEVGPFRGVKMIPPGIHFHLYSS- VDKANPKEVGPRMGFFLSLHQRGLTVLRWSTL- REEVDLSAPAESEVEAMRANLQELDQFLGPYP- YATLKKWISLTNFISEATVEKLQPENRQICAF- SDVLPVLSMKHSSRAGTEIRFSELPTQMFPE- RAGTEIRFSELPTQMFPEGATPAEITKHSMDL- SYALETVLNKKFPSSPDVLGELQFAFVCFLL- GNVYEAFAHWKRLNLLCRSEAAAMMKHHTLYI- NLISILYHQLGEIPADFFVDIVSQDNFLTSTL- QVFFSSACSIAVDATLRKKAKEKFQAHLTKKFR- WDAFAEPEDCAPVVVELPEGIEMG

[†] The underlined residues are residual from TEV cleavage.

yeast two-hybrid studies located the Aar2p binding site to the Prp8p C-terminus, coinciding with the Brr2 interaction site on Prp8 (Boon *et al.*, 2006; Liu *et al.*, 2006; Mozaffari-Jovin *et al.*, 2013). Interaction studies with both isolated Prp8p domains in *S. cerevisiae* demonstrated the binding of Aar2p and Brr2p to the Prp8p RH and JM domains, respectively. A larger fragment of Prp8p encompassing RH and JM, termed CTF, was entirely sequestered by Aar2p, thereby excluding association of Brr2p with JM (Weber *et al.*, 2011).

Subsequently, we obtained the structure of *S. cerevisiae* Aar2p in complex with the Prp8p RH and JM domains. The structure explains the mechanism whereby Aar2p binds to RH and employs a C-terminal peptide to tether RH and JM in a conformation that is incompatible with Brr2p binding. The structure of Prp8p CTF in isolation corroborated the domain arrangement of RH and JM, which is consolidated by the Aar2p C-terminus (Weber *et al.*, 2013). Moreover, Aar2p occupies known RNA-binding sites on the RH domain and interferes with U4/U6 di-snRNA binding by RH *in vitro* as shown by band-shift assays (Weber *et al.*, 2013). In a concomitant study, yeast Aar2p was employed to obtain the structure of a larger Prp8p fragment, which confirmed the role of Aar2p as a spliceosomal assembly factor and provided fundamental structural insight into Prp8p architecture (Galej *et al.*, 2013). Aar2p tethers both Prp8p C-terminal domains in a specific conformation so that RH sterically interferes with JM–Brr2p binding in yeast, as deduced from recent structures of JM–Brr2 complexes (Mozaffari-Jovin *et al.*, 2013; Nguyen *et al.*, 2013; Galej *et al.*, 2013; Weber *et al.*, 2013). It is as yet unclear whether a human Aar2 orthologue has a conserved

function. Ser253 of yeast Aar2p was shown to be phosphorylated *in vivo*. Since a phosphomimetic mutation of Aar2p (S253E) abolished the interaction with RH, regulation of Aar2p by reversible phosphorylation was suggested (Weber *et al.*, 2011, 2013). The structure of the Aar2p^{S253E} phosphomimetic protein revealed a major structural rearrangement of the C-terminus of the protein, disrupting a pre-formed Prp8p binding interface (Weber *et al.*, 2013).

In 2001, a human orthologue of yeast Aar2p was suggested, although with a low sequence conservation of 20% identical and 35% similar residues (see also Fig. 1a; Gottschalk *et al.*, 2001). The structure and function of the potential human orthologue have so far remained elusive. In this study, we confirm C20ORF4 to be expressed in HeLa cells by Western blotting and show its interaction with RH *in vitro*. These results consolidate C20ORF4 as a true orthologue of yeast Aar2p. To determine the overall structural conservation of Aar2 between higher and lower eukaryotes and to perform structure-based biochemical studies, we initiated crystallographic analysis of the hsAar2–hsRH complex. The identification of a crystallizable hsAar2 fragment by *in situ* limited proteolysis and prior results from yeast Aar2 guided the design of a truncated hsAar2^{Δloop} protein for co-crystallization with hsRH. The hsAar2^{Δloop}–RH crystals belonged to space group C2 and diffracted to 2.35 Å resolution, and a complete data set was collected.

2. Methods

2.1. Cloning, expression and protein purification

The human Prp8 RNaseH (hsRH; residues 1747–2016) clone used in this study was obtained from Pena *et al.* (2008). A codon-optimized DNA fragment of human Aar2 (coding for amino-acid residues 1–384) was commercially obtained in pFL10His to yield an expression construct for the full-length protein with a cleavable N-terminal His₁₀ tag (hsAar2). This plasmid containing the synthetic gene for human Aar2 was also used to generate hsAar2^{Δloop} (hsAar2 linker residues 170–200 replaced by three serines). All constructs were verified by sequencing (Table 1). HsAar2 and hsAar2^{Δloop} proteins were produced in insect cells as described in Santos *et al.* (2012). For protein expression of hsRH, *Escherichia coli* Rosetta2 (DE3) cells were transformed with the respective plasmid, grown in Terrific Broth to an OD₆₀₀ of 0.6 at 37°C, cooled to 20°C, induced with 0.5 mM IPTG and incubated overnight at 20°C. Cells were harvested by centrifugation and stored at –80°C. Cell pellets from the hsRH, hsAar2 or hsAar2^{Δloop} expression cultures were resuspended in lysis buffer (20 mM Tris–HCl pH 7.5, 300 mM NaCl, 1 mM DTT) in the presence of a protease-inhibitor cocktail (Roche). Cells were lysed using a Sonoplus sonifier (Bandelin) and cell debris was removed by centrifugation. For purification of all hsAar2 and hsRH constructs, the soluble extract was passed over an Ni–NTA gravity-flow column pre-equilibrated with lysis buffer. The beads were washed with lysis buffer containing 20 mM imidazole or 30 mM imidazole for hsRH. Proteins and

Table 2
Crystallization of the hsAar2–hsRH and hsAar2^{Δloop}–hsRH complexes.

	hsAar2–hsRH complex	hsAar2 ^{Δloop} –hsRH complex
Method	Sitting-drop vapour diffusion	
Plate type	Linbro 24-well	
Temperature (K)	277	
Protein concentration (mg ml ^{−1})	20	14
Buffer composition of protein solution	20 mM Tris–HCl pH 7.5, 150 mM NaCl, 0.5 mM DTT	
Composition of reservoir solution	0.1 M HEPES pH 7.5, 10% PEG 3350, 0.1 M L-proline	0.1 M HEPES pH 7.0, 10% PEG 6000, 5% (v/v) 2-methyl-2,4-pentanediol
Volume and ratio of drop	2 μl, 1:1	
Volume of reservoir (μl)	400	

Table 3
Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	P14, PETRA III
Wavelength (Å)	1.2395
Temperature (K)	100
Detector	PILATUS 6M, Dectris
Crystal-to-detector distance (mm)	324
Rotation range per image (°)	0.1
Total rotation range (°)	200
Exposure time per image (s)	1
Space group	C2
<i>a</i> , <i>b</i> , <i>c</i> (Å)	145.28, 57.26, 111.23
α , β , γ (°)	90, 112.74, 90
Mosaicity (°)	0.192
Resolution range (Å)	48.223–2.35 (2.41–2.35)
Total No. of reflections	131459 (9681)
No. of unique reflections	34687 (2502)
Completeness (%)	97.6 (95.6)
Multiplicity	3.79 (3.87)
$\langle I/\sigma(I) \rangle$	15.62 (1.70)
<i>R</i> _{int} (%)	4.6 (92.0)
<i>CC</i> _{1/2} †	99.9 (76.5)
Overall <i>B</i> factor from Wilson plot (Å ²)	66.7

† Correlation factor between random half data sets for reporting results in *XDS* *CORRECT* and *XSCALE* (Karplus & Diederichs, 2012).

mutants were eluted with lysis buffer supplemented with 300 mM imidazole. All hsRH and hsAar2 proteins and protein

fragments were cleaved by TEV protease overnight at 4°C and were loaded onto a Superdex 75 gel-filtration column in lysis buffer. Peak fractions were pooled, passed over an Ni–NTA Sepharose gravity-flow column equilibrated with lysis buffer, concentrated to 20–30 mg ml^{−1} and flash-cooled in liquid nitrogen for storage at −80°C. For production of the hsAar2–hsRH or hsAar2^{Δloop}–hsRH complexes, pre-purified hsAar2 or hsAar2^{Δloop} and a twofold stoichiometric excess of hsRH were mixed and loaded onto a gel-filtration column (Superdex 200, GE Healthcare) in crystallization buffer (20 mM Tris–HCl pH 7.5, 150 mM NaCl, 0.5 mM DTT). Peak fractions of the protein complexes were pooled, concentrated and flash-cooled in liquid nitrogen.

2.2. Analytical gel filtration

Single proteins and protein mixtures were analyzed by analytical size-exclusion chromatography on a Superdex 200 PC3.2 column (GE Healthcare) in size-exclusion buffer (20 mM Tris–HCl pH 7.5, 250 mM NaCl, 0.5 mM DTT) at a flow rate of 50–70 μl min^{−1}. For analysis of complex formation, proteins were mixed in equimolar ratios in a total volume of 60 μl in size-exclusion buffer and were incubated for 30 min on ice. Fractions were supplemented with SDS loading dye

Figure 2
Purification of recombinant hsAar2 and hsAar2^{Δloop} and detection of hsAar2 in HeLa cells by Western blotting (M, marker; P, pellet; S, supernatant; F, flowthrough; W, wash; E, elution). (a) Purification of hsAar2 from insect-cell lysate via Ni–NTA Sepharose and subsequent cleavage of the His₁₀ tag by TEV protease (+T). (b) Purification of hsAar2^{Δloop} from insect-cell lysate via Ni–NTA Sepharose and subsequent cleavage of the His₁₀ tag by TEV protease (+T). (c) Western blot analysis of HeLa cell control (ctrl) and Aar2 siRNA-transfected lysates with an antibody against human Aar2 (αAar2). An antibody against hnRNP L (αhnRNP L) served as a loading control.

and analyzed by SDS-PAGE. Different S200 columns were used for gel-filtration runs of hsAar2 and hsAar2^{Δloop}.

2.3. Crystallization

The hsAar2-hsRH complex crystallized (see Table 2) in the presence of GluC (diluted to a 1:100 mass ratio). Crystals were cryoprotected with 15% PEG 400, 100 mM NaCl. HsAar2^{Δloop}-hsRH complex crystals (see Table 2) were cryoprotected with 10% 2-methyl-2,4-pentanediol, 100 mM NaCl.

2.4. Data collection and processing

Diffraction data were collected on EMBL beamline P14 at the PETRA III synchrotron, Hamburg, Germany and were indexed, processed and scaled with *XDS* (Kabsch, 2010) (Table 3).

2.5. Cell culture and Western blotting

HeLa cells were maintained in DMEM containing 10% FBS and penicillin/streptomycin (Invitrogen). For siRNA transfections, 4×10^4 cells were seeded in 12-well plates. The next day, the cells were transfected with siRNA, either a control (5'-UUCUCCGAACGUGUCACGUTT) or a sequence targeting human Aar2 (a, 5'-GCUGACUUCUUCGUAGACATT; b, 5'-GCCUACACGGCAACCAUAUTT), using Rotifect transfection reagent according to the manufacturer's instructions. 72 h after transfection, the cells were harvested, lysed with

Flag-lysis buffer (60 mM Tris-HCl pH 7.5, 30 mM NaCl, 1 mM EDTA, 1% Triton X-100 with proteinase inhibitors) and analyzed by Western blotting following standard procedures. The antibody against Aar2 was generated using recombinant full-length human Aar2 as an antigen. An antibody against hnRNP L (4D11) served as a loading control.

3. Results and discussion

3.1. Expression and purification of C20ORF4 and detection of the protein in HeLa cells by Western blotting

We have identified Aar2p as a spliceosomal maturation factor with a dual function in *S. cerevisiae*. Remarkably, Aar2p, which is an essential gene in yeast, displays only a low degree of conservation from yeast to human (Fig. 1a). To initiate studies of the potential human Aar2 orthologue and to investigate its functional repertoire on a cellular, structural and biochemical level, we first expressed the protein recombinantly in insect cells. Therefore, a codon-optimized synthetic gene of C20ORF4 was used for baculovirus-based protein production. The potential human Aar2 orthologue was purified *via* an N-terminal His₁₀ tag, which was subsequently cleaved off by TEV protease (Fig. 2a). The protein was further purified by gel filtration and recycling over Ni-NTA agarose (not shown), yielding a homogenous protein fraction (see Fig. 3a). The protein was then used for the immunization of a rabbit to obtain a polyclonal antibody for Western blotting. To

Figure 3
HsAar2 binds to the hsRH domain of Prp8 in gel-filtration analyses. Purified hsAar2 (a) and purified hsRH (b) form a complex (c) that elutes earlier from gel filtration than the single proteins. The absorption at 280 nm of the eluates was recorded and plotted against the elution volume (right panels). A grey box denotes the relevant fractions which were analyzed by SDS-PAGE (left panels; M, marker; IP, input).

test whether the open reading frame C20ORF4 gives rise to expressed protein *in vivo*, we performed a Western blot using HeLa cell extract and detected a single product at approximately 45 kDa; the predicted molecular weight of the full-length product of C20ORF4 in human cells is 43.4 kDa (Fig. 2c). To confirm that this band indeed corresponds to the C20ORF4 gene product, we used two independent siRNAs targeting C20ORF4 and observed a strong reduction of the Western blot signal. This is, to our knowledge, the first evidence for the expression of the C20ORF4 protein in human cells.

3.2. Assignment of C20ORF4 as the human Aar2 orthologue and initial crystallization trials of an hsAar2–RH complex

We next wanted to test whether expressed C20ORF4 binds to the human RH (hsRH) domain as observed in prior studies of yeast Aar2 (Weber *et al.*, 2011). Human RH was expressed and purified as described previously (Pena *et al.*, 2008). We conducted size-exclusion chromatography with both proteins alone and in a stoichiometric mixture. Despite low sequence conservation (20% identity and 35% similarity) between the suggested human Aar2 orthologue and yeast Aar2, especially at the Aar2–RH interface (Fig. 1a), we observed binding of purified C20ORF4 to the hsRH domain of Prp8, as indicated by the earlier elution of the complex compared with the single proteins (Figs. 3a, 3b and 3c). This is the first experimental evidence that C20ORF4 is a true orthologue of yeast Aar2. For crystallization, the human Aar2 orthologue (hsAar2) and

hsRH were expressed and purified on a large scale. HsAar2 was mixed with a twofold molar excess of hsRH and the complex was purified by gel filtration. All initial attempts to crystallize hsAar2 alone or in complex with hsRH failed. For the crystallization of yeast Aar2p and also HeLa U1 snRNP, we successfully applied limited proteolysis approaches (Weber *et al.*, 2010, 2011). Extensive limited proteolysis experiments followed by SDS–PAGE were then applied to hsAar2 alone or in complex with hsRH, in which Aar2 degraded and did not yield any stable fragments, while hsRH was hardly affected by proteolytic treatment (not shown). For the crystallization of native U1 snRNPs, the implementation of *in situ* limited proteolysis was crucial (Dong *et al.*, 2007). Here, gradual proteolysis produced optimally processed particles which were apparently drawn out of solution by crystal formation as a selective process. Notably, differing proteases gave rise to different U1 snRNP crystal morphologies and space groups (Weber *et al.*, 2010). Thus, we employed *in situ* limited proteolysis with the hsAar2–RH complex to remove protein parts (*e.g.* flexible loops) that might hinder crystallization. We tested six proteases, with each protease screen covering 384 crystallization conditions, totalling 2304 crystallization trials. Notably, we obtained crystals of the hsAar2–RH complex only in the presence of endoproteinase GluC in a single crystallization condition (Fig. 4a). The crystals had a tetragonal crystal lattice and only diffracted to 5.4 Å resolution on beamline 14.2 at BESSY, Berlin (Mueller *et al.*, 2012; not shown). All approaches to optimize these crystals to achieve a higher diffraction limit failed and reliable structure solution by

Figure 4

Analyses of the hsAar2–hsRH and hsAar2^{Δloop}–hsRH complex crystals by SDS–PAGE and X-ray diffraction. (a) Crystals of the hsAar2–hsRH complex which appeared upon *in situ* proteolysis with GluC. (b) SDS–PAGE analysis of the crystals shown in (a) after washing and dissolution. The lower migration behaviour of hsRH is a gel artefact possibly arising from traces of the crystallization buffer. (c) Crystals of the hsAar2^{Δloop}–hsRH complex. (d) SDS–PAGE analysis of the crystals shown in (c) after washing and dissolution. (e) A 1° merged diffraction image of hsAar2^{Δloop}–hsRH crystals as shown in (c) was prepared by adding ten 0.1° images using *ALBULA* (Dectris). The inset shows a detector region covering a higher resolution segment with enhanced contrast.

molecular replacement was not possible. SDS–PAGE analysis of dissolved crystals revealed the hsRH domain to be intact and two bands of about 20 kDa representing hsAar2 (Fig. 4*b*). The same proteolysis pattern was observed for subtilisin-treated yeast Aar2p, where the removal of an internal protein loop split the protein into two nearly equivalent moieties and facilitated crystallization (Weber *et al.*, 2011). These observations led to the notion that the overall Aar2 fold may be conserved between yeast and human.

3.3. Crystallization of an hsAar2 loop-truncation mutant (hsAar2^{Δloop}) in complex with hsRH

Subsequently, we devised a homology model of human Aar2 based on the known yeast orthologue by submitting the sequence of full-length human Aar2 to *HHpred* (Söding *et al.*, 2005) implementing *MODELLER* (Šali *et al.*, 1995). The model was utilized to design an expression construct for an hsAar2 variant (hsAar2^{Δloop}) which had the internal loop replaced by three serines (Fig. 1*b*). In yeast, the C-terminal residues of Aar2p contact the RH domain and were therefore retained in the human expression construct (see Fig. 4*a*). HsAar2^{Δloop} was expressed and purified to homogeneity like the full-length hsAar2 protein (Figs. 2*b* and 5*a*). Also, the protocol for complex formation with hsRH was carried out as for full-length hsAar2. As observed for the yeast counterparts, removal of the internal hsAar2 protein loop had no effect on hsRH binding (Figs. 5*a*, 5*b* and 5*c*). The hsAar2^{Δloop}–RH

complex was concentrated and crystallized within 3 d (Fig. 4*c*). SDS–PAGE analysis of dissolved crystals confirmed the presence of both hsAar2^{Δloop} and hsRH (Fig. 4*d*). Complete data were collected with good statistics to a diffraction limit of 2.35 Å (Fig. 4*e*, Table 1).

In this work, we have demonstrated that the C20ORF4 gene encodes a true human Aar2 orthologue, despite its low sequence similarity to yeast Aar2p (Gottschalk *et al.*, 2001). Aar2 is expressed in HeLa cells, and full-length hsAar2 and its variant hsAar2^{Δloop} can be produced recombinantly in insect cells. HsAar2 and hsAar2^{Δloop} bind equally well to the hsRH domain of Prp8, as observed for the yeast proteins, regardless of the lack of conservation of the important Aar2–RH interface residues in hsAar2 (see Fig. 1*a*). SDS–PAGE analysis of hsAar2–hsRH crystals, obtained from a screen in which *in situ* limited proteolysis was employed to promote crystallization, suggests that yeast and human Aar2 share the same structural organization consisting of two equally sized N-terminal and C-terminal domains (Weber *et al.*, 2011). Along with the availability of an hsAar2 antibody, this study paves the way for further *in vitro* and *in vivo* studies to fully elucidate the function of hsAar2, which may be connected to snRNP maturation as in yeast (Weber *et al.*, 2011, 2013). In addition, this work highlights the great potential of *in situ* limited proteolysis to identify and further characterize crystallizable protein fragments (Dong *et al.*, 2007; Weber *et al.*, 2010). Initial crystallization of full-length hsAar2–hsRH complex employing limited *in situ* proteolysis and subsequent analysis

Figure 5
HsAar2^{Δloop} binds to the hsRH domain of Prp8 in gel-filtration analyses. Purified hsAar2^{Δloop} (a) and purified hsRH (b) form a complex (c) that elutes earlier from gel filtration than the single proteins. The absorption at 280 nm of the eluates were recorded and plotted against the elution volume (right panels). A grey box denotes the relevant fractions which were analyzed by SDS–PAGE (left panels; M, marker; IP, input).

of the weakly diffracting crystals guided the design of a crystallizable hsAar2 construct, ultimately yielding complex crystals that diffracted to atomic resolution. We now aim at the structural characterization of hsAar2 and its binding partners to better understand its function in human and to investigate parallels to yeast snRNP maturation.

Acknowledgements

The authors would like to thank R. Lührmann for generously providing the hsAar2 antibody and K. L. Boon for antibody affinity purification. We thank C. Alings and E. Absmeier (FU Berlin) for great support with crystallization. We acknowledge access to beamline BL14.2 of the BESSY II storage ring (Berlin, Germany) via the Joint Berlin MX Laboratory sponsored by the Helmholtz Zentrum Berlin für Materialien und Energie, the Freie Universität Berlin, the Humboldt-Universität zu Berlin, the Max-Delbrück Centrum and the Leibniz-Institut für Molekulare Pharmakologie (Mueller *et al.*, 2012). Parts of this research were carried out at the PETRA III light source at DESY, a member of the Helmholtz Association (HGF). We would like to thank G. Bourenkov for assistance in using beamline P14. The authors would like to thank M. Wahl and F. Heyd for providing infrastructure and great support to this project. KS is supported by the Dahlem International Network Postdocs program.

References

- Barton, G. J. (1993). *Protein Eng.* **6**, 37–40.
- Boon, K. L., Grainger, R. J., Ehsani, P., Barrass, J. D., Auchynnikava, T., Inglehearn, C. F. & Beggs, J. D. (2007). *Nature Struct. Mol. Biol.* **14**, 1077–1083.
- Boon, K. L., Norman, C. M., Grainger, R. J., Newman, A. J. & Beggs, J. D. (2006). *RNA*, **12**, 198–205.
- Dong, A. *et al.* (2007). *Nature Methods*, **4**, 1019–1021.
- Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. (2013). *Nature (London)*, **493**, 638–643.
- Gottschalk, A., Kastner, B., Lührmann, R. & Fabrizio, P. (2001). *RNA*, **7**, 1554–1565.
- Hutchinson, E. G. & Thornton, J. M. (1996). *Protein Sci.* **5**, 212–220.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
- Liu, S., Rauhut, R., Vornlocher, H. P. & Lührmann, R. (2006). *RNA*, **12**, 1418–1430.
- Mozaffari-Jovin, S., Santos, K. F., Hsiao, H.-H., Will, C. L., Urlaub, H., Wahl, M. C. & Lührmann, R. (2012). *Genes Dev.* **26**, 2422–2434.
- Mozaffari-Jovin, S., Wandersleben, T., Santos, K. F., Will, C. L., Lührmann, R. & Wahl, M. C. (2013). *Science*, **341**, 80–84.
- Mueller, U., Darowski, N., Fuchs, M. R., Förster, R., Hellmig, M., Paithankar, K. S., Pühringer, S., Steffien, M., Zocher, G. & Weiss, M. S. (2012). *J. Synchrotron Rad.* **19**, 442–449.
- Nakazawa, N., Harashima, S. & Oshima, Y. (1991). *Mol. Cell. Biol.* **11**, 5693–5700.
- Nguyen, T. H. D., Li, J., Galej, W. P., Oshikane, H., Newman, A. J. & Nagai, K. (2013). *Structure*, **21**, 910–919.
- Pena, V., Liu, S., Bujnicki, J. M., Lührmann, R. & Wahl, M. C. (2007). *Mol. Cell*, **25**, 615–624.
- Pena, V., Rozov, A., Fabrizio, P., Lührmann, R. & Wahl, M. C. (2008). *EMBO J.* **27**, 2929–2940.
- Ritchie, D. B., Schellenberg, M. J., Gesner, E. M., Raithatha, S. A., Stuart, D. T. & MacMillan, A. M. (2008). *Nature Struct. Mol. Biol.* **15**, 1199–1205.
- Šali, A., Potterton, L., Yuan, F., van Vlijmen, H. & Karplus, M. (1995). *Proteins*, **23**, 318–326.
- Santos, K. F., Jovin, S. M., Weber, G., Pena, V., Lührmann, R. & Wahl, M. C. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 17418–17423.
- Söding, J., Biegert, A. & Lupas, A. N. (2005). *Nucleic Acids Res.* **33**, W244–W248.
- Wahl, M. C., Will, C. L. & Lührmann, R. (2009). *Cell*, **136**, 701–718.
- Weber, G., Cristao, V. F., Alves, F. de L., Santos, K. F., Holton, N., Rappsilber, J., Beggs, J. D. & Wahl, M. C. (2011). *Genes Dev.* **25**, 1601–1612.
- Weber, G., Cristão, V. F., Santos, K. F., Jovin, S. M., Heroven, A. C., Holton, N., Lührmann, R., Beggs, J. D. & Wahl, M. C. (2013). *Genes Dev.* **27**, 525–540.
- Weber, G., Trowitzsch, S., Kastner, B., Lührmann, R. & Wahl, M. C. (2010). *EMBO J.* **29**, 4172–4184.
- Yang, K., Zhang, L., Xu, T., Heroux, A. & Zhao, R. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 13817–13822.
- Zhang, L., Shen, J., Guarnieri, M. T., Heroux, A., Yang, K. & Zhao, R. (2007). *Protein Sci.* **16**, 1024–1031.