Structural analysis of the spliceosomal RNA helicase Prp28 from the thermophilic eukaryote *Chaetomium thermophilum*

Marcel J. Tauchert and Ralf Ficner*

Department of Molecular Structural Biology, Institute for Microbiology and Genetics, GZMB, Georg-August-University Göttingen, Justus-von-Liebig Weg 11, 37077 Göttingen, Germany. *Correspondence e-mail: rficner@gwdg.de

Received 19 February 2016

Accepted 11 April 2016

Edited by N. Sträter, University of Leipzig, Germany

Keywords: spliceosome; RNA helicase; DEAD-box protein; U5-100kD; DDX23.

PDB reference: the RNA-helicase Prp28 from *Chaetomium thermophilum* bound to ADP, 5dtu

Supporting information: this article has supporting information at journals.iucr.org/f

Prp28 (pre-mRNA-splicing ATP-dependent RNA helicase 28) is a spliceosomal DEAD-box helicase which is involved in two steps of spliceosome assembly. It is required for the formation of commitment complex 2 in an ATP-independent manner as well as for the formation of the pre-catalytic spliceosome, which in contrast is ATP-dependent. During the latter step, Prp28 is crucial for the integration of the U4/U6·U5 tri-snRNP since it displaces the U1 snRNP and allows the U6 snRNP to base-pair with the 5'-splice site. Here, the crystal structure of Prp28 from the thermophilic fungus *Chaetomium thermophilum* is reported at 3.2 Å resolution and is compared with the available structures of homologues.

1. Introduction

The spliceosome-mediated removal of non-coding introns from eukaryotic genes is a task of crucial importance during the process of mRNA maturation. For each intron to be removed, the spliceosome assembles *de novo* and catalyses two consecutive transesterifications, resulting in two ligated exons and an intron lariat. During one cycle of splicing, numerous conformational as well as compositional changes are undergone by the spliceosome, which require the tight and accurate orchestration of each rearrangement step. Defects or inaccuracies during the assembly or the rearrangement steps have been reported to be connected to diverse hereditary diseases, such as several forms of β -thalassaemia (Higgs *et al.*, 2012) and spinal muscular atrophy (D'Amico *et al.*, 2011).

In brief, the U1 snRNP binds to the 5'-splice site of the pre-mRNA, which results in the formation of complex E (Will & Lührmann, 2011). The U2 snRNP binds to the branch-point sequence and the pre-spliceosome (complex A) is formed. Subsequent to this, a tri-snRNP comprising of the U4/U6·U5 snRNPs joins, which results in the formation of the pre-catalytic spliceosome, also denoted complex B. The transition from complex B to the B^{act} complex (catalytically inactive) is caused by intensive rearrangements of RNA and protein networks, effecting the destabilization and the release of the U1 snRNP and the U4 snRNP as well as changes in the base-pairing of the remaining U2·U5·U6 tri-snRNP (Wahl *et al.*, 2009). Following this, the B* complex is formed, which is capable of catalysing the first transesterification step. Further remodelling results in complex C, which performs the second catalytic reaction of the spliceosome. Terminally, the occurring post-spliceosomal complex is dissociated into its individual components, the mature mRNA and the intron lariat spliceosome, which is subsequently dismantled into the intron

Table 1
Crystallization.

Method	Vapour diffusion
Plate type	Sitting drop
Temperature (K)	277
Protein concentration (mg ml ⁻¹)	4
Buffer composition of protein solution	20 mM Tris pH 9.0, 150 mM NaCl, 5% (v/v) glycerol, 2 mM dithiothreitol
Composition of reservoir solution	200 mM NaCl, 20% (w/v) PEG 4000, 100 mM Tris pH 8.0
Volume and ratio of drop	0.25 µl; 1:1 ratio
Volume of reservoir (µl)	35

lariat, the separated U2/U5/U6 snRNPs and numerous auxiliary proteins.

Every single one of the abovementioned rearrangement steps features the involvement of at least one RNA helicase (Staley & Guthrie, 1998; Schwer, 2001; Cordin *et al.*, 2012). The helicases involved in spliceosomal business all belong to helicase superfamily 2 (SF2), which includes the three sub-families DExD/H-box and Ski2-like (Jankowsky & Fairman, 2007; Linder & Jankowsky, 2011). Although they share all of the functional motifs, I, II, V, VI (nucleotide binding), Ia, Ib, IV (nucleic acid substrate binding) and III (coupling of both processes) (Cordin & Beggs, 2013), they strongly differ in their degree of processivity, their nucleotide specificity or the proposed mechanism of unwinding.

The protein examined in this study, pre-mRNA-splicing ATP-dependent RNA helicase 28 (Prp28), belongs to the DEAD-box proteins (the eponymic amino-acid sequence is in motif II). They contain an additional N-terminal extension as well as family-specific motifs, namely the Q, GG and QxxR motifs (Hilbert *et al.*, 2009), which provoke, among others, specificity towards ATP. To date, two functions of Prp28 have been reported. The first is involvement in the formation of spliceosomal commitment complex 2 (Price *et al.*, 2014), which is the second step in complex E formation. The other is its classically described function in the spliceosome, which is involvement in the displacement of the U1 snRNP from the 5'-splice site, thereby allowing the U6 snRNP to base-pair with this region (Staley & Guthrie, 1998; Chen *et al.*, 2001; Ismaili *et al.*, 2001) as well as a proofreading function during this step (Yang *et al.*, 2013). Although the classically described function of Prp28 is ATP-dependent, in contrast to the previously mentioned function, no (Strauss & Guthrie, 1994; Lagerbauer *et al.*, 1998; Yang *et al.*, 2013; Möhlmann *et al.*, 2014) or only an extremely low (Jacewicz *et al.*, 2014) ATPase activity could be shown for Prp28 *in vitro*.

Interestingly, the homologues of Prp28 from different organisms vary greatly in the mechanism of how they are recruited to their target substrate. Prp28 from *Homo sapiens* (DDX23; referred to here as hsPrp28) is stably integrated into the U5 snRNP after the phosphorylation of its RS-like domain in the N-terminal extension by the kinase SRPK2 (Mathew *et al.*, 2008; Xiang *et al.*, 2013). The homologue from *Saccharomyces cerevisiae* (scPrp28) does not stably associate with the U5 snRNP or the tri-snRNP (Strauss & Guthrie, 1991; Teigelkamp *et al.*, 1997) and exhibits only a vestigial

N-terminal extension. The two crystal structures available of Prp28, both of which are N-terminally truncated, from *H. sapiens* (PDB entry 4nho; Möhlmann *et al.*, 2014) and *S. cerevisiae* (PDB entry 4w7s; Jacewicz *et al.*, 2014), have unveiled the overall structure of the helicase core. For the human homologue it was presumed that the linker which connects the two RecA-like domains fixes these two domains in an unproductive orientation to each other, which would explain the marginal ATPase activity *in vitro*. Putative mechanisms of Prp28 stimulation have been proposed and Prp8 (Price *et al.*, 2013) or an RNA substrate (Möhlmann *et al.*, 2014) have been proposed as conceivable candidates.

Here, we report the crystal structure of an N-terminally truncated construct of Prp28 from the thermophilic eukaryotic ascomycete *Chaetomium thermophilum* (ctPrp28) at 3.2 Å resolution. Besides a comparison with the available abovementioned crystal structures of Prp28, we focus on an analysis of the conformation of the RecA-like domains, which was previously reported to be of crucial importance for exhibiting activity (Möhlmann *et al.*, 2014), and we attempt to investigate whether the observed conformations are crystallization artefacts or might reflect the actual conformation in solution.

2. Materials and methods

2.1. Macromolecule production

After identification of the *C. thermophilum* Prp28 homologue via a BLAST search (Altschul *et al.*, 1990), the CTHT_0054430 gene was cloned from *C. thermophilum* var. *thermophilum* DSM 1495 cDNA as a truncated variant which lacks 237 N-terminal amino acids into pENTRY-IBA51 according to the IBA StarGate manual (version Oct 2013) and subsequently transferred into pPSG-IBA25.

Expression from pPSG-IBA25-ctPrp28(238–709) was realised in *Escherichia coli* Rosetta 2 (DE3) cells at 17°C for 16 h after induction at an OD₆₀₀ of 0.8 with 500 µM isopropyl β-D-1-thiogalactopyranoside and supplementation with 4% (v/v) ethanol and 30 mM K₂HPO₄. Cell disruption was carried out by microfluidization (Microfluidizer 110S, Microfluidics) in buffer consisting of 50 mM Tris–HCl pH 9.0, 500 mM NaCl, 5% (v/v) glycerol, 2 mM dithiothreitol followed by ultracentrifugation at 35 000g for 30 min. The filtered supernatant was loaded onto 15 ml GSH Sepharose (GE Healthcare), washed with buffer additionally supplemented with 2 M LiCl to remove contaminating nucleic acids and the target protein was subsequently eluted with 30 mM reduced glutathione. Following overnight incubation with PreScission Protease [1:100 (w/w), GE Healthcare] further proteinaceous impurities were removed by size-exclusion chromatography (Superdex 75, GE Healthcare) in 20 mM Tris–HCl pH 9.0, 150 mM NaCl, 5% (v/v) glycerol, 2 mM dithiothreitol. Traces of residual GST were separated via a final affinity chromatography step using GSH Sepharose in gel-filtration buffer. The target protein ctPrp28(238–709) was concentrated to 7 mg ml⁻¹.

Table 2

Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	ID23-1, ESRF
Wavelength (Å)	0.910
Temperature (K)	100
Detector	Pilatus 6M
Crystal-to-detector distance (mm)	611.6
Rotation range per image (°)	0.2
Total rotation range (°)	191.6
Exposure time per image (s)	0.07
Space group	C2
<i>a</i> , <i>b</i> , <i>c</i> (Å)	161.50, 50.40, 65.80
α , β , γ (°)	90.00, 101.40, 90.00
Mosaicity (°)	0.496
Resolution range (Å)	45.77–3.20 (3.30–3.20)
Total No. of reflections	30243 (2660)
No. of unique reflections	8454 (723)
Completeness (%)	95.1 (85.1)
Multiplicity	3.58 (3.68)
CC _{1/2} (%)	98.9 (74.2)
$\langle I/\sigma(I) \rangle$	9.17 (2.45)
<i>R</i> _{meas}	0.147 (0.583)
Overall <i>B</i> factor from Wilson plot (Å ²)	68.94
Molecules per asymmetric unit	1

2.2. Crystallization

The ctPrp28(238–709) construct was incubated with a tenfold molar excess of ADP in the presence of 2 mM MgCl₂ and subjected to crystallization trials at a concentration of 4 mg ml^{−1}. Plate-shaped crystals were obtained after 12 d at 277 K in a crystallization condition composed of 200 mM NaCl, 20%(w/v) PEG 4000, 100 mM Tris pH 8.0. For more detailed information, see Table 1.

2.3. Data collection and processing

Crystals were harvested using a nylon loop and cryoprotected in mother liquor with an additional 15%(v/v) glycerol. A single crystal was employed for data collection at 100 K on beamline ID23-1 at ESRF, Grenoble, France. Data reduction and scaling were performed with the *XDS* package (Kabsch, 2010). Data-collection parameters and data-processing statistics are summarized in Table 2.

2.4. Structure solution, refinement and analysis

The structure of ctPrp28(238–709) was determined by molecular replacement using *Phaser* (McCoy *et al.*, 2007). The crystal structure of Prp28 from *H. sapiens* (PDB entry 4nho; Möhlmann *et al.*, 2014) was used as a search model. To obtain a reasonable molecular-replacement solution, it was crucial to divide the proteinaceous parts of this structure into two independent search models each composed of one RecA-like domain. Subsequent to iterative cycles of refinement in *PHENIX* (Adams *et al.*, 2010) and manual model building in *Coot* (Emsley *et al.*, 2010), the final model of ctPrp28(238–709) was validated *via MolProbity* (Chen *et al.*, 2010) (see Table 3). Crystal structures were visualized and aligned using *PyMOL* (v.1.3; Schrödinger) and crystal contacts were calculated using *PISA* (Krissinel, 2015).

Table 3

Structure solution and refinement.

Values in parentheses are for the outer shell.

Resolution range (Å)	45.77–3.20 (3.45–3.20)
Completeness (%)	95.8
No. of reflections, working + test sets	8446 (1544)
No. of reflections, test set	677 (135)
Final <i>R</i> _{cryst}	0.274 (0.342)
Final <i>R</i> _{free}	0.288 (0.405)
No. of non-H atoms	
Protein	3138
Ion	1
Ligand	27
Total	3166
R.m.s. deviations	
Bonds (Å)	0.006
Angles (°)	0.992
Average <i>B</i> factors (Å ²)	
Protein	65.8
Ion	179.2
Ligand	61.6
Ramachandran plot	
Most favoured (%)	93.9
Allowed (%)	6.1
Outliers (%)	0
Rotamer outliers (%)	0.30
Overall <i>MolProbity</i> score	2.13
PDB code	5dtu

3. Results and discussion

3.1. Overall structure

The N-terminally 237-amino-acid truncated DEAD-box protein Prp28(238–709) from the thermophilic ascomycete *C. thermophilum* crystallized in the monoclinic space group C2, with unit-cell parameters *a* = 161.50, *b* = 50.40, *c* = 65.80 Å, $\alpha = \gamma = 90$, $\beta = 101.40^\circ$ (see Table 2) and with one molecule per asymmetric unit. This corresponds to its state in solution as determined by analytical gel filtration and multi-angle light scattering (data not shown). The reported structure has a resolution of 3.20 Å and was refined to an *R*_{work} of 0.274 and an *R*_{free} of 0.288. 93.9% of all residues are in the most favoured region of the Ramachandran plot and 0% are in disallowed regions (see Table 3). The missing N-terminal region of this protein was predicted to be largely unfolded and our crystallization trials with the full-length protein remained without success. This problem was also encountered by Möhlmann *et al.* (2014) (PDB entry 4nho) and Jacewicz *et al.* (2014) (PDB entry 4w7s), who only could crystallize N-terminally truncated constructs of hsPrp28 (337 deleted N-terminal amino acids) and scPrp28 (126 amino acids removed from the N-terminus), respectively. The structure of ctPrp28 was solved *via* molecular replacement using the structure of hsPrp28 as a search model. Residues 241–340, 346–454, 483–514 and 521–683 of the ctPrp28(238–709) construct used for crystallization were traceable in the electron density. The obtained ctPrp28(238–709) structure can be divided into four different entities, namely the N-terminal extension (NTE), the RecA1 domain and the RecA2 domain as well as a linker connecting the two RecA-like domains. Residues 241–262 are part of the NTE, residues 263–512 belong to the RecA1 domain, the linker is

composed of residues 513–522 but is mainly not resolved in our structure and residues 523–683 are part of the RecA2 domain (see Fig. 1). ctPrp28(238–709) was co-crystallized with ADP and Mg^{2+} , which were both detectable in the electron density. This is the first structure of a Prp28 homologue with bound ADP. Previously, only the apo structure (hsPrp28 and scPrp28) and a structure with bound AMPPNP (scPrp28) were available.

The fold of both ctPrp28 RecA-like domains is highly similar to those of other known structures of SF2 helicases. The N-terminal RecA domain is composed of an eight-stranded β -sheet exhibiting a $\beta 1 \downarrow - \beta 8 \uparrow - \beta 2 \uparrow - \beta 7 \uparrow - \beta 6 \uparrow - \beta 3 \uparrow - \beta 5 \uparrow - \beta 4 \uparrow$ topology and nine α -helices, which pack around this sheet. The C-terminal RecA domain consists of a six-stranded β -sheet ($\beta 9 \uparrow - \beta 14 \uparrow - \beta 13 \uparrow - \beta 10 \uparrow - \beta 12 \uparrow - \beta 11 \uparrow$ topology) with seven α -helices wrapped around it. Between the three larger functional domains of ctPrp28, only a small number of contacts are observable. The RecA1 domain interacts with the N-terminal extension *via* four hydrogen bonds (Arg276–Leu247, Ile264–Lys259, Gly263–Lys259 and Thr266–Lys259) and the RecA2 domain exhibits two hydrogen bonds to the N-terminal extension (Leu608–Arg256 and Gly612–Glu260). The RecA1 and RecA2 domains interact only *via* one hydrogen bond (Ser265–Asp614) and one salt bridge (Glu497–Arg642). The low number of interdomain contacts is an indicator that the helicase is in an unproductive ‘open’ state since the conserved motifs which are required for nucleotide and RNA binding are widely spaced. The conformation of ctPrp28(238–709) resembles the conformations of other inactive DEAD-box proteins, for instance eIF4A-I (PDB entry 1fuu; Caruthers *et al.*, 2000), eIF4A-III (PDB entry 2hxy; Andersen *et al.*, 2006) and Prp5 (PDB entry 4lly; Zhang *et al.*, 2013). The requirement of a ‘closed’ conformation for ATPase

Figure 1
Crystal structure of ctPrp28(238–709) at 3.2 Å resolution in cartoon representation. The heavily truncated N-terminal extension (NTE) (238–262) is presented in purple, the RecA1 domain (263–512) in blue, the linker (513–522), which is mainly not resolved in this crystal structure, in green and the RecA2 domain (523–709) in orange. The bound ADP molecule is presented in ball-and-stick mode with C atoms coloured yellow, N atoms blue, O atoms red, P atoms orange and the Mg^{2+} ion light green. The scale bar indicates the relative size of each domain and the dashed lines highlight the region of the NTE which is not present in the construct used for crystallization.

activity, in which the conserved motifs are tightly arranged and are in close spatial proximity, was deduced from the crystal structures of DEAD-box proteins bound to RNA and non-hydrolysable ATP analogues, such as those of Vasa (PDB entry 2db3; Sengoku *et al.*, 2006), eIF4A-III (PDB entry 2hyi;

Figure 2
Conserved motifs in the ctPrp28(238–709) helicase core. The RecA1 domain is coloured light grey, the RecA2 domain dark grey and the N-terminal extension pale blue. The residues of the linker are not highlighted by additional colouring. Nucleotide-binding motifs are emphasized in blue, RNA-binding motifs in red and the motif which couples both processes is shown in green. Motifs are numbered as described in §3.2. The ADP and the Mg^{2+} ion are coloured as described for Fig. 1.

Figure 3
ADP-binding site of ctPrp28. The ADP molecule is bound to the RecA1 domain of ctPrp28. The adenine moiety is specifically recognized by Gln303 of the Q motif and is stabilized by Tyr296 *via* π -electron stacking. Tyr296 also interacts with the O2' atom of the ribose by hydrogen bonding. The α - and β -phosphates interact with the Mg^{2+} ion as well as the main-chain amides of Gly323, Ser324 and Gly325 and the Thr327 side chain, all of which belong to motif I, which is also denoted the P-loop. For further information, see §3.2.

Andersen *et al.*, 2006) and DDX19 (PDB entry 3g0h; Collins *et al.*, 2009).

3.2. ADP binding of ctPrp28

The abovementioned conserved nucleotide-binding and RNA-binding motifs, typical for all SF2 helicases, are distributed equally over the two RecA-like domains of ctPrp28 (see Fig. 2). The Q motif ($^{278}\text{W}_{16}\text{GYDEPTPIQ}^{303}$), which mediates the specificity towards ATP, motif I ($^{320}\text{AVTGSGKT}^{327}$), motif Ia ($^{360}\text{PTRELVQQ}^{367}$), the GG motif ($^{388}\text{GG}^{389}$), motif Ib ($^{409}\text{TPGRL}^{413}$), motif II ($^{433}\text{DEAD}^{436}$) and motif III ($^{489}\text{TAT}^{491}$) are located in the RecA1 domain (Hilbert *et al.*, 2009). The RecA2 domain provides the interface for motifs IV ($^{553}\text{IVFVNI}^{558}$), QxxR

($^{585}\text{QEQR}^{588}$), V ($^{603}\text{LVATDLAAGRIG}^{614}$) and VI ($^{633}\text{YTHRIGRTGRA}^{643}$). The crystal structure of ctPrp28(238–709) is in an unproductive, presumably post-catalytic state since the reaction product ADP is bound. Only the residues of the Q motif and motif I are involved in ADP binding. Here, the adenine moiety interacts *via* its amino group (N6) with the side chain of the name-giving Gln303 from the Q motif (see Fig. 3). Additionally, the base is stabilized *via* π -electron stacking with Tyr296, which is in close spatial proximity to the Q motif. Furthermore, Tyr296 interacts with O2' of the ribose of the ADP, which is plausible owing to Glu371, which attracts the proton of the OH group from the tyrosine side chain. The α -phosphate and the β -phosphate exclusively interact with residues of motif I. This motif, also denoted the P-loop, is required to neutralize the

Figure 4

Structure comparison of Prp28 from *C. thermophilum* and *H. sapiens*. ctPrp28 is coloured as described in Fig. 1. The human homologue (PDB entry 4nho) is depicted in red. The RecA-like domain of hsPrp28, which was not used for superpositioning, is shown with increased transparency. (a) The RecA1 domains were superposed with an r.m.s.d. value of 1.04 Å for 188 C α atoms and (b) the RecA2 domains were superposed with an r.m.s.d. of 0.79 Å for 113 C α atoms. (c) The insertion in the RecA1 domain of hsPrp28 (amino acids 576–599) is highlighted in green and very likely reflects the position of this α -helical insertion in ctPrp28 (amino acids 460–482), which is not resolved in the reported crystal structure.

negative charge of the phosphate groups. The main-chain amide of Gly325 forms a hydrogen bond to the α -phosphate, and the β -phosphate is also bound *via* non-specific main-chain interactions with Gly323 and Ser324 as well as the side chain of Thr327. It is difficult to determine the exact coordination sphere of the magnesium ion since water molecules were not traceable in the electron density owing to the resolution limit of 3.2 Å, but interactions with both of the phosphate moieties of the ADP can be observed. The magnesium ion is localised at a similar position between the α - and the β -phosphates, which resembles the position of the magnesium ion in the scPrp28 crystal structure with bound AMPPNP. None of the residues from motif II, V or VI show involvement in nucleotide binding even though residues from these motifs have been shown to be crucial for the ATPase activity of Prp28 in mutagenesis studies in *S. cerevisiae* (Chang *et al.*, 1997; Jacewicz *et al.*, 2014). This underlines the extent of the required conformational changes for ctPrp28 to adopt a productive 'closed' conformation.

3.3. Structural comparison to the homologues from *H. sapiens* and *S. cerevisiae*

Despite the fact that the Prp28 homologues clearly differ in size, the protein from *C. thermophilum* exhibits a relatively high sequence identity and similarity to the homologues from human (48 and 50%) and yeast (41 and 43%). By far the least conserved region of this protein is the N-terminal extension, which consists of 380 amino acids in the human homologue and only 156 in the yeast homologue. This difference in size is directly linked to the recruitment mechanism to its target substrate in the spliceosome. The human homologue is stably integrated into the tri-snRNP *via* its NTE after phosphorylation of the RS-like domain (Mathew *et al.*, 2008), whereas the yeast Prp28 acts as a loosely bound factor with a vestigial N-terminal extension (Strauss & Guthrie, 1991). In *C. thermophilum* the NTE comprises the first 262 amino acids and thus occupies an intermediate position with respect to the size of this domain. The exact recruitment mechanism of ctPrp28 is as yet unclear. Owing to its increased size compared with the yeast protein, one might assume that the NTE of ctPrp28 could be used for tri-snRNP integration, but the number of SR sites or phosphorylation mimics such as RE or RD is massively decreased compared with the human Prp28 NTE, which makes a phosphorylation-based mechanism unlikely for ctPrp28.

The RecA1 and RecA2 domains of Prp28 from the three abovementioned organisms exhibit a higher level of sequence identity and similarity compared with the values for the complete proteins. The identity and similarity increase to 57 and 69%, respectively, for the human homologue and to 43 and 52%, respectively, for that from yeast. Despite the high sequence identity, it is not possible to align the complete helicase cores properly with each other. This is primarily not attributable to the conformation of the RecA-like domains of Prp28, which is in principle very similar (Fig. 4a), at least between those from *H. sapiens* and *C. thermophilum*; rather,

this is caused by a difference in the orientation of the second RecA-like domain (Fig. 4b). Aligning the RecA1 or RecA2 domains from *H. sapiens* and *C. thermophilum* separately with each other results in reasonable superpositions, as indicated by r.m.s.d. values of 1.04 Å (for 188 C α atoms) for the RecA1 domains and of 0.79 Å (for 113 C α atoms) for the RecA2 domains. As shown in Fig. 4(a), the RecA2 domains are rotated perpendicularly to the direction of the connecting linker. This figure and the abovementioned alignment scores between human Prp28 and the homologue from *C. thermophilum* additionally reveal that the apo form (hsPrp28) and the ADP-bound state (ctPrp28) are very similar. Furthermore, the position of the linker from ctPrp28 resembles that of hsPrp28, although the linker in our ctPrp28 structure is only fragmentarily resolved. All of the linker residues which are present in the ctPrp28 structure align very well with the corresponding residues in hsPrp28 and one could also expect an increased distance between the two RecA-like domains in ctPrp28 if the linker was more elongated and not in a fixed state. The strong influence on activity of this short RecA-like domain-connecting linker has previously been demonstrated by Low *et al.* (2007), who were able to greatly increase the activity of eIF4A by exchanging its endogenous linker with that from Vasa. The highly similar conformation of the RecA-like domains might hint at the conformation of Prp28 in solution, since it is present in both structures even though hsPrp28 and ctPrp28 crystallized in different space groups (*C*222₁ versus *C*2) and with a clear difference in the number of and the area of crystal contacts (3842 versus 1535 Å², which correspond to 18 and 7% of the total surface, respectively). The assumption that the conformation of the RecA-like domains might be physiologically relevant was proposed by Möhlmann *et al.* (2014) because in the human homologue the position of the linker is fixed by a tight hydrogen-bond network. The necessity for an open conformation has previously been observed for another spliceosomal DEAD-box helicase, scPrp5 (Zhang *et al.*, 2013), which requires this conformation to accomplish a proofreading function. Whether the open conformation of ctPrp28 or hsPrp28 might have a similar function during the formation of commitment complex 2, which is ATP-independent and thus does not necessarily require a productive interface of the two RecA-like domains, remains a subject of speculation because involvement in commitment complex 2 formation has only been reported for scPrp28 to date.

A further feature which underlines the similarity between both homologues is an insertion in the RecA1 domain. This insertion is only present in higher eukaryotes and is unique to Prp28. In hsPrp28 this insertion comprises residues 576–599, which are mainly present as an α -helix. In *C. thermophilum* the insertion is composed of residues 460–482 and exhibits 39% sequence identity and 58% sequence similarity to the human homologue. In the presented ctPrp28 structure this helix was not traceable in the electron density, but as shown in Fig. 4(c) superposition with the human structure can give a clear indication of the position of this insertion in the ctPrp28 structure. The exact function of the helical insertion remains

unclear, but it may be assumed that it is required for the interaction with proteinaceous factors; a link to the enzymatic function is unlikely because no functional motif is located in

Figure 5

Superposition of *C. thermophilum* and *S. cerevisiae* Prp28. ctPrp28 (coloured according to Fig. 1) is superposed on both monomers of the scPrp28 crystal structure (PDB entry 4w7s), shown here in green (chain A) and violet (chain B). The alignment was calculated between (a) the RecA1 domains, with r.m.s.d. values of 0.74 Å for 167 C α atoms (chain A) and 0.91 Å for 169 C α atoms (chain B), as well as (b) the RecA2 domains, with corresponding r.m.s.d. values of 1.30 Å for 119 common C α atoms (chain A) and 1.36 Å for 118 C α atoms (chain B). The RecA-like domain of scPrp28 which was not used in the respective superposition is depicted transparently.

this insertion. In addition to this, the nucleotide-binding and RNA-binding sites are located on the opposite side of the RecA1 domain, as deduced from superpositions with active DEAD-box helicases, for instance Vasa (PDB entry 2db3). Structural analysis of the human tri-snRNP cryo-EM structure (Agafonov *et al.*, 2016) revealed that this insertion is not involved in direct interactions with the tri-snRNP. Owing to this, it seems likely that this α -helical insertion is required for interaction with other components of the human spliceosome.

Apart from the insertion, the RecA1 domain from the yeast homologue shows a high structural similarity to that of ctPrp28. The Prp28 homologue from yeast crystallized with two molecules per asymmetric unit and the RecA1 domains align with the ctPrp28 RecA1 domain with r.m.s.d. values of 0.74 Å for chain A (167 C α atoms) and 0.91 Å for chain B (169 C α atoms) (see Fig. 5). Although AMPPNP is bound to the RecA1 domain of scPrp28 (chain B), the helicase core is presumably in an inactive/post-catalytic state since there are neither RecA1–RecA2 interdomain contacts nor interactions between motif II, V or VI and the AMPPNP molecule. This assumption is also strengthened by an r.m.s.d. value of 0.47 Å (for 205 C α atoms) for the superposition of the RecA1 domains of scPrp28 chains A and B. The alignment of the RecA2 domain of scPrp28 with ctPrp28 results in poorer r.m.s.d. values compared with the superposition of the RecA1 domain [1.30 Å for 119 C α atoms (chain A) and 1.36 Å for 118 C α atoms (chain B)], which is attributable to a difference in the topology of the yeast RecA2 domain. The RecA2 domain of the yeast protein is composed of one additional β -strand.

Since the two molecules of scPrp28 in the asymmetric unit already differ in conformation, it is highly questionable whether any relevant information about the actual conformation of scPrp28 in solution can be deduced from this crystal structure. On the one hand, this may be attributable to crystal contacts because their number and the buried surface of the two monomers in the asymmetric unit is increased to 4465 Å² (chain A) and 4144 Å² (chain B) (20 and 19% of the total surface) compared with the homologues from *C. thermophilum* or *H. sapiens* (1535 and 3842 Å², respectively). On the other hand, it is also possible that the conformation of the two scPrp28 monomers differs because of the yeast Prp28 linker, which is more elongated and does not fix the protein in a certain conformation *via* intra-linker hydrogen bonds. Owing to the fact that the conformation of scPrp28 clearly differs from that of hsPrp28 or ctPrp28, the necessity for the open conformation during the formation of commitment complex 2 is questionable. It is possible that scPrp28 does not require the same conformation as ctPrp28 or hsPrp28 in the inactive state since it is only loosely associated with the tri-snRNP and is a free factor. The conformation of ctPrp28 or hsPrp28 might be a prerequisite for integration into the tri-snRNP, but this remains the subject of speculation at present.

4. Conclusion

Prp28 from the thermophilic ascomycete *C. thermophilum* exhibits higher structural similarity to the homologue from

H. sapiens than to that from *S. cerevisiae* (for the exact values of the superpositions, see §3.3). The conformation of the RecA-like domains strongly resembles that of the human protein, although the orientation of the second RecA-like domain differs. The conformational similarity is very likely to be caused by the linker, which fixes the RecA-like domains in this relative position to each other; thus, this crystal structure might reflect the conformation of ctPrp28 in solution until its activation by further spliceosomal factors.

Acknowledgements

This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG) to RF (SFB 860, TP A2). We are grateful to the ESRF staff at beamline ID23-1 for support during data collection. Our thanks also go to Drs T. Monecke and P. Neumann for the data-collection service.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Agafonov, D. E., Agafonov, D. E., Kastner, B., Dybkov, O., Hofele, R. V., Liu, W. & Urlaub, H. (2016). *Science*, **351**, 1416–1420.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
- Andersen, C. B. F., Ballut, L., Johansen, J. S., Chamieh, H., Nielsen, K. H., Oliveira, C. L. P., Pedersen, J. S., Séraphin, B., Le Hir, H. & Andersen, G. R. (2006). *Science*, **313**, 1968–1972.
- Caruthers, J. M., Johnson, E. R. & McKay, D. B. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 13080–13085.
- Chang, T., Latus, L., Liu, Z. & Abbott, J. (1997). *Nucleic Acids Res.* **25**, 5033–5040.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Chen, J. Y.-F., Stands, L., Staley, J. P., Jackups, R. R. Jr, Latus, L. J. & Chang, T.-H. (2001). *Mol. Cell*, **7**, 227–232.
- Collins, R., Karlberg, T., Lehtiö, L., Schütz, P., van den Berg, S., Dahlgren, L.-G., Hammarström, M., Weigelt, J. & Schüler, H. (2009). *J. Biol. Chem.* **284**, 10296–10300.
- Cordin, O. & Beggs, J. D. (2013). *RNA Biol.* **10**, 83–95.
- Cordin, O., Hahn, D. & Beggs, J. D. (2012). *Curr. Opin. Cell Biol.* **24**, 431–438.
- D'Amico, A., Mercuri, E., Tiziano, F. D. & Bertini, E. (2011). *Orphanet J. Rare Dis.* **6**, 71.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Higgs, D. R., Engel, J. D. & Stamatoyannopoulos, G. (2012). *Lancet*, **379**, 373–383.
- Hilbert, M., Karow, A. R. & Klostermeier, D. (2009). *Biol. Chem.* **390**, 1237–1250.
- Ismaïli, N., Sha, M., Gustafson, E. H. & Konarska, M. M. (2001). *RNA*, **7**, 182–193.
- Jacewicz, A., Schwer, B., Smith, P. & Shuman, S. (2014). *Nucleic Acids Res.* **42**, 12885–12898.
- Jankowsky, E. & Fairman, M. E. (2007). *Curr. Opin. Struct. Biol.* **17**, 316–324.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Krissinel, E. (2015). *Nucleic Acids Res.* **43**, W314–W319.
- Laggerbauer, B., Achsel, T. & Lührmann, R. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 4188–4192.
- Linder, P. & Jankowsky, E. (2011). *Nature Rev. Mol. Cell Biol.* **12**, 505–516.
- Low, W.-K., Dang, Y., Bhat, S., Romo, D. & Liu, J. O. (2007). *Chem. Biol.* **14**, 715–727.
- Mathew, R., Hartmuth, K., Möhlmann, S., Urlaub, H., Ficner, R. & Lührmann, R. (2008). *Nature Struct. Mol. Biol.* **15**, 435–443.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Möhlmann, S., Mathew, R., Neumann, P., Schmitt, A., Lührmann, R. & Ficner, R. (2014). *Acta Cryst.* **D70**, 1622–1630.
- Price, A. M., Görnemann, J., Guthrie, C. & Brow, D. A. (2014). *RNA*, **20**, 46–60.
- Schwer, B. (2001). *Nature Struct. Biol.* **8**, 113–116.
- Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. (2006). *Cell*, **125**, 287–300.
- Staley, J. P. & Guthrie, C. (1998). *Cell*, **92**, 315–326.
- Strauss, E. J. & Guthrie, C. (1991). *Genes Dev.* **5**, 629–641.
- Strauss, E. J. & Guthrie, C. (1994). *Nucleic Acids Res.* **22**, 3187–3193.
- Teigelkamp, S., Mundt, C., Achsel, T., Will, C. L. & Lührmann, R. (1997). *RNA*, **3**, 1313–1326.
- Wahl, M. C., Will, C. L. & Lührmann, R. (2009). *Cell*, **136**, 701–718.
- Will, C. L. & Lührmann, R. (2011). *Cold Spring Harb. Perspect. Biol.* **3**, a003707.
- Xiang, S., Gapsys, V., Kim, H.-Y., Bessonov, S., Hsiao, H.-H., Möhlmann, S., Klaukien, V., Ficner, R., Becker, S., Urlaub, H., Lührmann, R., de Groot, B. & Zweckstetter, M. (2013). *Structure*, **21**, 2162–2174.
- Yang, F., Wang, X.-Y., Zhang, Z.-M., Pu, J., Fan, Y.-J., Zhou, J., Query, C. C. & Xu, Y.-Z. (2013). *Nucleic Acids Res.* **41**, 4660–4670.
- Zhang, Z.-M., Yang, F., Zhang, J., Tang, Q., Li, J., Gu, J., Zhou, J. & Xu, Y.-Z. (2013). *Cell Rep.* **5**, 1269–1278.