STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Structural and functional analysis of the RNA helicase Prp43 from the thermophilic eukaryote *Chaetomium thermophilum*

Marcel J. Tauchert,[a] Jean-Baptiste Fourmann,[b] Henning Christian,[a] Reinhard Lührmann[b] and Ralf Ficner[a]*

[a]Department of Molecular Structural Biology, Institute for Microbiology and Genetics, GZMB, Georg-August-Universität Göttingen, Justus-von-Liebig Weg 11, 37077 Göttingen, Germany, and [b]Department of Cellular Biochemistry, Max Planck Institute of Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany. *Correspondence e-mail: rficner@gwdg.de

Received 1 October 2015
Accepted 20 December 2015

Edited by N. Sträter, University of Leipzig, Germany

Keywords: spliceosome; RNA helicase; DEAH-box protein; DHX15.

PDB reference: RNA helicase Prp43 from *Chaetomium thermophilum* bound to ADP, 5d0u

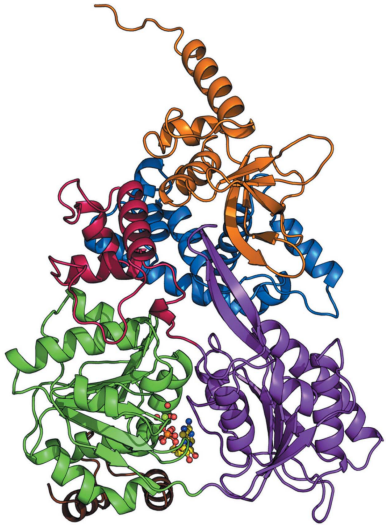
Supporting information: this article has supporting information at journals.iucr.org/f

RNA helicases are indispensable for all organisms in each domain of life and have implications in numerous cellular processes. The DEAH-box RNA helicase Prp43 is involved in pre-mRNA splicing as well as rRNA maturation. Here, the crystal structure of *Chaetomium thermophilum* Prp43 at 2.9 Å resolution is revealed. Furthermore, it is demonstrated that Prp43 from *C. thermophilum* is capable of functionally replacing its orthologue from *Saccharomyces cerevisiae* in spliceosomal disassembly assays.

1. Introduction

RNA helicases are ubiquitously distributed among all domains of life and are of crucial importance for numerous cellular processes such as pre-mRNA splicing, translation initiation, ribosome biogenesis and RNA transport (Cordin *et al.*, 2006; Bleichert & Baserga, 2007; Ozgur *et al.*, 2015). Helicases have been classified into six superfamilies (SFs) based on phylogenetic sequence alignments (Fairman-Williams *et al.*, 2010). The members of SF1 and SF2 share a central helicase core composed of two RecA-like domains. These adjacent domains provide the characteristic interface for DExD/H-box proteins and contain the eight conserved short sequence motifs I, Ia, Ib, II, III, IV, V and VI, in which motif II exhibits the eponymous amino-acid sequence. Mutagenesis and structural studies have unravelled the involvement of motifs I, II, V and VI in NTPase activity. Motifs Ia, Ib and IV are required for RNA binding and motif III couples NTP hydrolysis to RNA unwinding (Cordin *et al.*, 2006; Hilbert *et al.*, 2009). DEAH-box proteins belong to the SF2 helicases, which exhibit an additional N-terminal extension as well as three further C-terminal domains: a winged-helix (WH) domain, a ratchet and an oligosaccharide-binding fold (OB-fold) (He *et al.*, 2010; Walbott *et al.*, 2010). The DEAH-box family is capable of unwinding DNA as well as RNA substrates (Fairman-Williams *et al.*, 2010). Contemporarily, it is assumed that DEAH-box helicases bind a single-stranded overhang of a nucleic acid substrate and unwind this substrate by a continuous movement mediated by the two RecA-like domains, *i.e.* they exhibit a certain level of processivity (Pyle, 2008). In contrast to the DEAD-box and Ski2-like helicases, DEAH-box proteins can utilize all nucleoside triphosphates, at least *in vitro* (Kim *et al.*, 1992; Schwer & Guthrie, 1992; Tanaka & Schwer, 2005, 2006).

Eight conserved SF2 helicases belonging to the DExD/H-box and Ski2-like families are involved in pre-mRNA splicing.



© 2016 International Union of Crystallography

Table 1
Macromolecule production information for ctPrp43(61–764).

Source organism	*C. thermophilum* var. *thermophilum* DSM 1495
Expression vector	pETM-13
Expression host	*E. coli* Rosetta 2 (DE3)
Complete amino-acid sequence of the construct produced	**MA**TTAKQAEAVEDSDINPWTGQRHSERYFKILKA-RRKLPVNKQRQEFLDLYHNNQILVFVGETGSG-KTTQIPQYVLYDELPHQTGKLIACTQPRRVAA-MSVAQRVADELDVKLGEEVGYSIRFENKTSSK-TLLKYMTDGQLLREAMHDRDMSRYSCIILDEA-HERTLATDILMALLKQLSERRKDLKIIVMSAT-LDAQKFQSYFFNAPLLAVPGRTHPVEIFYTPE-AERDYVEAAIRTVLQIHACEPEGDILLFLTGE-EEIEDACRRISLEVDEMIRESDAGPMSVYPLY-GTLPPHQQQRIFEKAPQPFRPGGRPGRKCIVA-TNIAETSLTIDGIVYVVDPGFSKQKIYNPRTR-VESLLVSPISKASAQQRAGRAGRTRPGKCFRL-YTEEAFKKELIEQTYPEILRSNLSNTVLELKK-LGVEDLVHFDLMDPPAPETMMRALEELNYLAC-LDDDGELTPLGNLASEFPLDPALAVMLISSPE-FYCSNEILSITSLLSVPQIWVRPANARKRADE-MKAQFAHPDGDHLTLLNAYHAYKGAEARGEDM-KKWCHEHFLSYRHLSSADNVRAQLKKIMETHG-IELVSTPFHDKNYYTNIRRALLAGFFMQVAMR-ESSNSKVYKTVKDEQLVLIHPSTTVTTPYEWV-VYNEFVLTTKQYVRTVTNIRPEWLLEIAPVYY-DLSTFQKGEIKNALTRVAEKIRRQQAMKASKA-**WSHPQFEK**

These eight helicases are key players in the accurate orchestration of the major compositional and conformational rearrangements which are undergone by the spliceosome during one cycle of intron removal. Prp43 (pre-mRNA processing factor 43) is required for proper disassembly of the yeast intron–lariat spliceosome (ILS) which is composed of the U2·U5·U6 snRNPs (Arenas & Abelson, 1997; Fourmann *et al.*, 2013). To accomplish this function, Prp43 interacts with two cofactors: Ntr1 and Ntr2 (nineteen complex-related proteins 1 and 2) (Tanaka *et al.*, 2007; Tsai *et al.*, 2005, 2007; Boon *et al.*, 2006). Ntr1 contains a G-patch motif (glycine-rich) that stimulates the ATPase and unwinding activities of Prp43 (Tanaka *et al.*, 2007; Christian *et al.*, 2014; Robert-Paganin *et al.*, 2015).

Besides pre-mRNA splicing, Prp43 is additionally involved in ribosome biogenesis, in which it is required for the maturation of 18S and 25S pre-rRNAs (Lebaron *et al.*, 2005; Bohnsack *et al.*, 2009). Thereby, stimulation of Prp43 by the G-patch proteins Sqs1 (squelch of splicing suppression protein 1) and Gno1 (G-patch nucleolar protein 1) is essential. These two activator proteins exhibit no sequence identity to Prp43's spliceosomal activator protein Ntr1 except for the G-patch motif (Aravind & Koonin, 1999; Pertschy *et al.*, 2009).

The crystal structure of Prp43 from *Saccharomyces cerevisiae* (scPrp43) was solved in 2010 by two groups (He *et al.*, 2010; Walbott *et al.*, 2010). Here, we report the crystal structure of Prp43 from the thermophilic ascomycetal fungus *Chaetomium thermophilum* (ctPrp43) at 2.9 Å resolution. After publication of its genome (Amlacher *et al.*, 2011), a continuously increasing number of protein structures (87 deposited in the PDB to date) from *C. thermophilum* have been released (Bock *et al.*, 2014); however, it was not always demonstrated that a putative orthologue from *C. thermophilum* does actually functionally correspond to its mesophilic counterpart. To address this question for ctPrp43, we performed *in vitro* spliceosome disassembly assays which clearly demonstrate that ctPrp43 can fully replace scPrp43 in the yeast spliceosome.

Table 2
Crystallization.

Method	Vapour diffusion
Plate type	Sitting drop
Temperature (K)	293
Protein concentration (mg ml^{-1})	4
Buffer composition of protein solution	10 m*M* Tris–HCl pH 7.5, 100 m*M* NaCl, 2 m*M* $MgCl_2$
Composition of reservoir solution	8%(*w*/*v*) Jeffamine M-2070, 0.17 *M* glycine, 16.7%(*v*/*v*) DMSO, 10 m*M* urea
Volume and ratio of drop	2 µl; 1:1 ratio
Volume of reservoir (µl)	500

2. Materials and methods

2.1. Macromolecule production

The identification of the potential homologue of scPrp43 in *C. thermophilum* was performed using *BLAST* (Altschul *et al.*, 1990) against the complete *C. thermophilum* genome. The highest alignment score was achieved by a protein annotated as 'hypothetical protein CTHT_0005780' and referred to here as ctPrp43.

The gene encoding full-length ctPrp43 was amplified from genomic DNA of *C. thermophilum* var. *thermophilum* DSM 1495 and cloned into pGEX-6P-1 using the EcoRI and SalI restriction sites. All dispensable bases between the PreScission Protease cleavage site and the starting methionine of ctPrp43 were deleted *via* site-directed mutagenesis (QuikChange Site-Directed Mutagenesis Kit, Agilent Technologies). The gene of a truncated ctPrp43 construct comprising amino acids 61–764, generated for crystallization, was amplified from pGEX-6P-1-ctPrp43 and was subsequently cloned into pETM-13 with an additional C-terminal Strep-tag using NcoI and SalI restriction sites (see Table 1).

The fusion proteins GST-ctPrp43 and ctPrp43(61–764)-Strep were expressed from pGEX-6P-1 and pETM-13, respectively, in *Escherichia coli* Rosetta 2 (DE3) cells at 16°C for 18 h after induction with 0.5 m*M* IPTG at an optical density (OD_{600}) of 0.8. Subsequent to cell disruption *via* microfluidization (M-110S Microfluidizer) and the isolation of soluble proteins by ultracentrifugation at 35 000*g* for 30 min, GST-ctPrp43 was loaded onto Glutathione Sepharose 4B (GE Healthcare) and ctPrp43(61–764)-Strep was loaded onto StrepTactin HP Sepharose (GE Healthcare) in 400 m*M* NaCl, 50 m*M* Tris–HCl pH 7.5, 10 m*M* EDTA. After intensive washing with an additional 2 *M* LiCl, target protein elution was realised with 30 m*M* reduced glutathione (GST-ctPrp43) or 2.5 m*M* D-desthiobiotin [ctPrp43(61–764)-Strep]. GST-tag cleavage was realised by the addition of PreScission Protease [1:100(*w*/*w*), GE Healthcare]. Protein samples were purified to homogeneity by size-exclusion chromatography (Superdex 200, GE Healthcare) in 100 m*M* NaCl, 10 m*M* Tris–HCl pH

Table 3
Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	BL14.1, BESSY
Wavelength (Å)	0.918
Temperature (K)	100
Detector	Pilatus 6M
Crystal-to-detector distance (mm)	573.5
Rotation range per image (°)	0.2
Total rotation range (°)	60.6
Exposure time per image (s)	6.0
Space group	$P6_5$
a, c (Å)	152.2, 92.8
α, γ (°)	90.0, 120.0
Mosaicity (°)	0.361
Resolution range (Å)	50.00–2.92 (2.99–2.92)
Total No. of reflections	90272 (13924)
No. of unique reflections	26545 (4202)
Completeness (%)	99.3 (98.5)
Multiplicity	3.4 (3.3)
$CC_{1/2}$ (%)	99.7 (67.8)
$\langle I/\sigma(I)\rangle$	15.4 (2.0)
R_{meas}	0.102 (0.770)
Overall B factor from Wilson plot (Å^2)	63.6
Molecules per asymmetric unit	1

Table 4
Structure solution and refinement.

Values in parentheses are for the outer shell.

Resolution range (Å)	43.94–2.92 (2.99–2.92)
Completeness (%)	99.3
σ Cutoff	$F > 1.36\sigma(F)$
No. of reflections, working set	26538 (1719)
No. of reflections, test set	2022 (143)
Final R_{cryst}	0.190 (0.313)
Final R_{free}	0.231 (0.384)
No. of non-H atoms	
Total	5729
Protein	5661
Magnesium	1
ADP	27
Water	4
DMSO	36
R.m.s. deviations	
Bonds (Å)	0.0043
Angles (°)	0.843
Average B factors (Å^2)	
Overall	76.5
Protein	76.3
Magnesium	66.8
ADP	79.5
Water	72.5
DMSO	103.8
Ramachandran plot	
Most favoured (%)	96.44
Allowed (%)	3.56
Outliers (%)	0
Rotamer outliers (%)	0.32
MolProbity clashscore	7.50
MolProbity overall score	1.65
PDB code	5d0u

7.5, 2 m*M* $MgCl_2$ and concentrated to 40 mg ml^{-1} (Amicon Ultra 50K, Millipore).

scPrp43 was expressed and purified as described previously by Christian *et al.* (2014).

2.2. Crystallization

ctPrp43(61–764)-Strep was diluted to 4 mg ml^{-1} with gel-filtration buffer and incubated with a tenfold molar excess of ADP. The protein was crystallized using the sitting-drop vapour-diffusion method at 293 K with droplets consisting of equal volumes of protein and reservoir [8%(*w*/*v*) Jeffamine M-2070, 0.17 *M* glycine, 16.7%(*v*/*v*) DMSO, 10 m*M* urea] solutions. Rod-shaped crystals with dimensions of up to 20 × 20 × 2000 µm were obtained after 3–4 d. The crystallization procedure is summarized in Table 2.

2.3. Data collection and processing

The crystals obtained were cryoprotected in reservoir solution supplemented with 26%(*v*/*v*) glycerol and flash-cooled in liquid nitrogen prior to data collection. Diffraction data were collected on BL14.1 operated by the Helmholtz-Zentrum Berlin (HZB) at the BESSY II electron-storage ring (Berlin-Adlershof, Germany) (Mueller *et al.*, 2012). Data were processed using the *XDS* package (Kabsch, 2010). Data collection and processing statistics are presented in Table 3.

2.4. Structure solution, refinement and structural analysis

The structure of ctPrp43(61–764) was solved by molecular replacement using *Phaser* (McCoy *et al.*, 2007) with chain *A* of scPrp43 (PDB entry 2xau; Walbott *et al.*, 2010) as a search model. Manual model building was conducted with *Coot* (Emsley *et al.*, 2010) and refinement was performed with *PHENIX* (Adams *et al.*, 2010). During the refinement process, TLS refinement for chain *A* as a single group was performed and the coordination distances for the Mg^{2+} ion were restrained to 2.07 Å for water molecules and the Thr126 (OG1) side chain and to 2.09 Å for the β-phosphate (O2B), allowing a deviation of 0.2σ. After iterative cycles of model building and refinement, the final model (Table 4) was assessed for correctness using *MolProbity* (Chen *et al.*, 2010). Figures were prepared with *PyMOL* (v.1.3; Schrödinger) and *Chimera* (Pettersen *et al.*, 2004). Electrostatic surface potentials were calculated using *PDB2PQR* (Dolinsky *et al.*, 2004) as well as *APBS* (Baker *et al.*, 2001) and the surface conservation was visualized with *AL2CO* (Pei & Grishin, 2001).

2.5. Functional yeast assays

The *in vitro* reconstitution and disassembly assays from purified yeast spliceosomes were performed as described previously and were analysed on a linear 10–30%(*v*/*v*) glycerol gradient (Fourmann *et al.*, 2013).

2.6. Protein melting-point determination

The melting point was determined for ctPrp43 and scPrp43 *via* CD spectroscopy using a Chirascan CD spectrometer (Applied Photophysics) by monitoring the unfolding of α-helical protein regions. Far-UV spectra and melting curves were acquired in 10 m*M* Tris–HCl pH 7.5, 100 m*M* NaCl, 2 m*M* $MgCl_2$. Melting curves were determined at 222 nm for 12 s per °C between 20 and 84°C and far-UV spectra were recorded between 195 and 260 nm before (20°C) and after

(84°C) denaturation of protein samples. The obtained melting curves were fitted according to Pace *et al.* (1998). Raw data were smoothed by a factor of two with *SX-Pro-Data* (Applied Photophysics).

3. Results and discussion

3.1. Overall structure of Prp43 from *C. thermophilum*

ctPrp43 crystallized (see Table 2) in the hexagonal space group $P6_5$ and its structure was refined at a resolution of 2.92 Å with R_{work} and R_{free} values of 0.190 and 0.231, respectively (Tables 3 and 4). Since the crystallization of a full-length ctPrp43 construct remained fruitless, a truncated variant was generated lacking the first 60 N-terminal residues, which are mainly present in loop regions as shown by Walbott *et al.* (2010). After molecular replacement using the yeast structure with PDB code 2xau (Walbott *et al.*, 2010), the remaining 704 residues as well as the first amino acid from the attached C-terminal Strep-tag were traceable in the electron density.

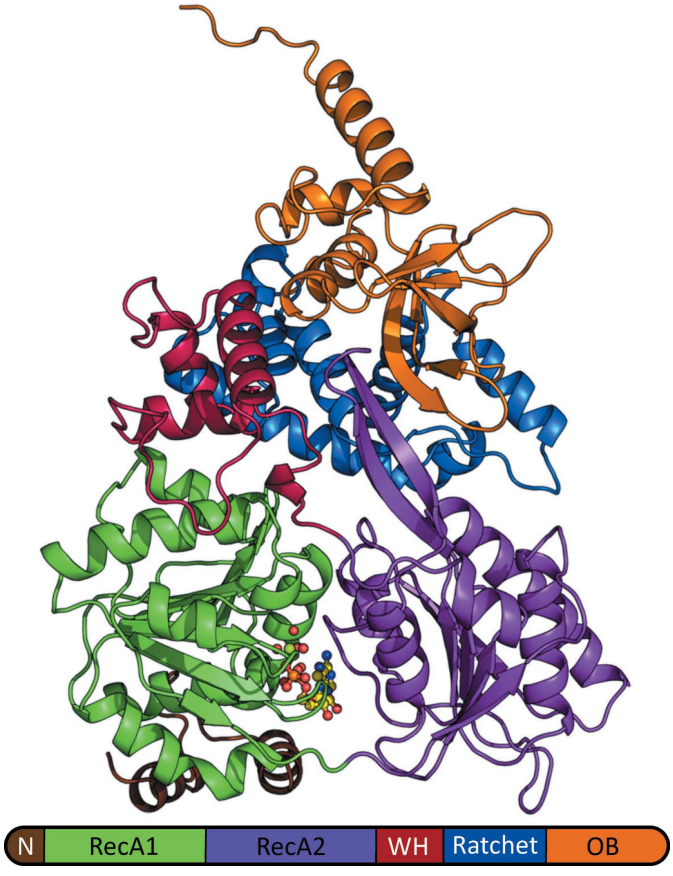


Figure 1
Crystal structure of ctPrp43(61–764) at 2.92 Å resolution. The overall structure of the N-terminally truncated ctPrp43(61–764) is depicted as a cartoon model. The remaining part of the N-terminal extension (amino acids 61–96) is shown in brown, the RecA1 domain (97–273) in light green, the RecA2 domain (274–458) in violet, the WH domain (459–526) in red, the ratchet (527–640) in blue and the OB-fold (641–764) in orange. The bound ADP molecule is shown in ball-and-stick mode and C atoms are coloured yellow, N atoms blue, O atoms red, P atoms orange and the Mg^{2+} ion light green.

The Ramachandran plot reveals that 96.44% of all residues are in the most favoured regions and no residues are in disallowed regions. One monomer of ctPrp43(61–764) is present in the asymmetric unit, which also corresponds to its functional state in solution as determined by analytical gel filtration and multi-angle light scattering (data not shown). These findings are also true for the full-length protein.

Besides the protein ctPrp43(61–764), one ADP molecule, one magnesium ion, four water molecules and nine DMSO molecules are present in the final structure (Fig. 1). ctPrp43 can be divided into six domains, namely the N-terminal extension (residues 1–96), the RecA1 (97–273) and RecA2 (274–459) domains, the WH domain (459–526), the ratchet (527–640) and the OB-fold (641–764). A characteristic feature of all DEAH-box helicases is the presence of a β-hairpin, which is located in the RecA2 domain and comprises residues 401–420 in ctPrp43. While the two RecA-like domains are a typical feature of all SF1 and SF2 helicase members, the three C-terminal domains can only be found in the DEAH helicase family.

The surface representation of ctPrp43 reveals a tunnel inside the molecule which was proposed to be an RNA-binding site (Walbott *et al.*, 2010) by deduction from a superposition with the Ski2-like helicase Hel308 bound to DNA (PDB entry 2p6r; Büttner *et al.*, 2007). From the electrostatic surface potential (Fig. 2), the orientation of an RNA molecule in this tunnel can be predicted. Since the surface of the ratchet domain in this binding tunnel exhibits a highly negative charge, one can assume that the bases of the RNA are oriented in this direction. The RecA1 and RecA2 domains provide basic patches which are compatible with the binding of phosphate groups from the RNA backbone. This binding mode would also explain why Prp43 can bind to RNA in a sequence-independent manner (Tanaka & Schwer, 2006). The surface potential of ctPrp43 also exhibits an additional RNA-binding site which was previously identified biochemically for scPrp43 (Walbott *et al.*, 2010). Mutagenesis or deletion of this region, located in the OB-fold, was shown to drastically decrease the RNA-stimulated ATPase activity as well as the affinity towards RNA.

3.2. ADP binding of ctPrp43

The prototypical domain motifs (Fairman-Williams *et al.*, 2010; Walbott *et al.*, 2010) of all SF2 helicase members were identified for ctPrp43 and are highlighted in Fig. 3. Motifs Ia (^{149}TQPRRVAA^{156}), Ib (^{195}TDGQLLR^{201}) and IV (^{310}LLFLTG^{315}) have been reported to interact with substrate nucleic acid and motif III (^{250}SAT^{252}) couples nucleoside triphosphate hydrolysis to substrate unwinding. Motifs I (^{122}GSGKT^{126}) (also denoted as the P-loop) (Rudolph *et al.*, 2006), II (^{218}DEAH^{221}), V (^{381}TNIAETSLT^{389}) and VI (^{428}QRAGRAGR^{435}) are involved in nucleotide binding. The ADP molecule is sandwiched between the RecA1 and RecA2 domains. Here, the adenine is bound *via* π-electron stacking and cation–π interaction between the side chains of Arg162 and Phe360 (Fig. 4). These two residues are not part of the

conserved motifs but also appear to be relevant for nucleotide binding. In Prp43, as well as all other spliceosomal DEAH-box proteins, the base of the nucleoside triphosphate is not specifically recognized, in contrast to the DEAD-box and Ski2-like helicases, which contain an additional binding motif, the Q-motif, that elicits adenine specificity. In ctPrp43, the ribose forms hydrogen bonds (for bonding distances, see Supplementary Table S1) from O2′ and O3′ to Asp391 (motif V) and Arg435 (motif VI), respectively. The involvement of Arg435 in hydrogen bonding is plausible owing to Gly122, the main-chain carboxyl group of which attracts the proton of the O3′ into its direction, thereby allowing O3′ to form a hydrogen

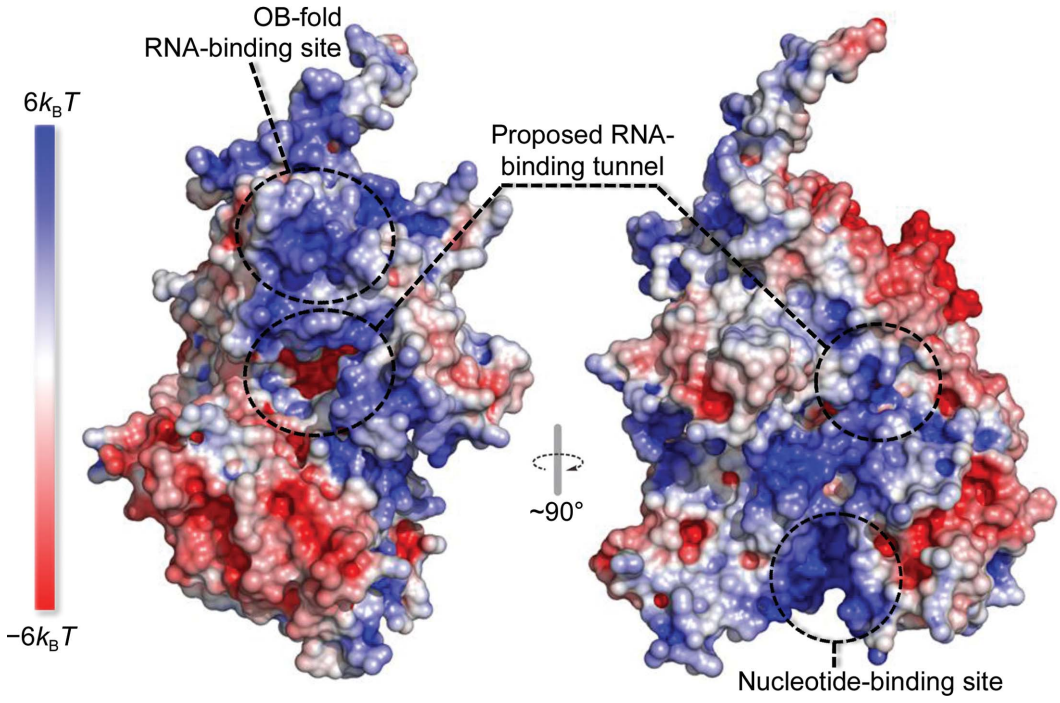


Figure 2
Electrostatic surface potential of ctPrp43(61–764). The electrostatic surface potential was calculated using *APBS* (Baker *et al.*, 2001) and is depicted at a contour level of $\pm 6k_BT$. Blue indicates positive charge and red negative charge. The side (left) and back (right) view of ctPrp43 are shown and (proposed) RNA-binding sites as well as the nucleotide-binding site are indicated by dashed circles. For further information, refer to §3.1.

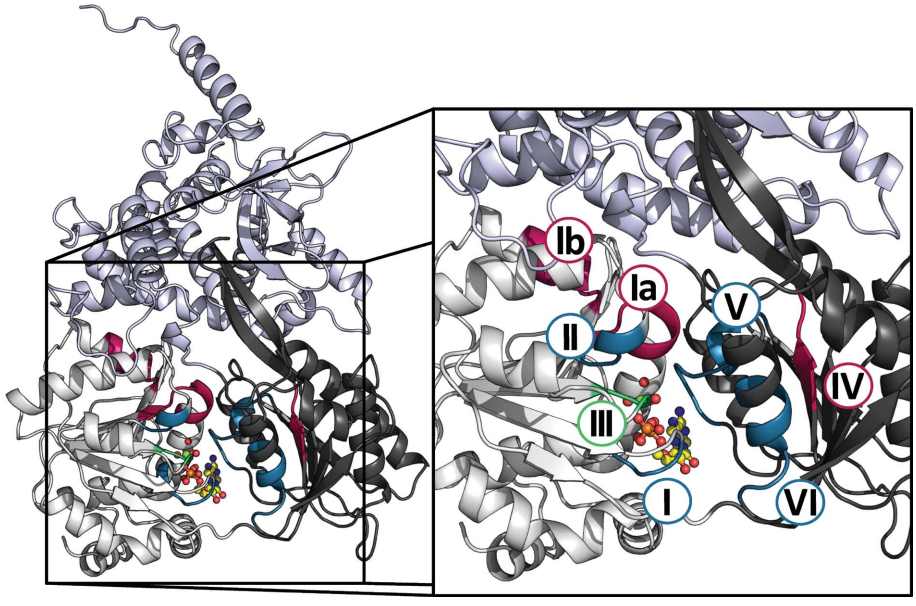
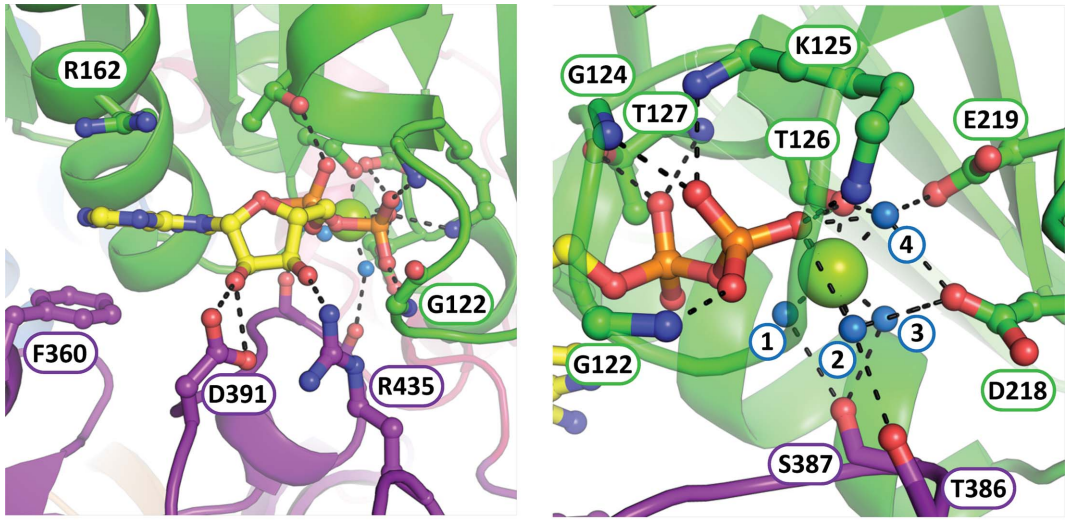


Figure 3
Conserved motifs of ctPrp43. The conserved binding motifs of ctPrp43 are presented in blue (nucleotide binding), red (nucleic acid substrate binding) and green (coupling of NTP hydrolysis to substrate unwinding). The N-terminal extension and the RecA1 domain are visualized in light grey, the RecA2 domain in dark grey and the three C-terminal domains in pale blue. The ADP molecule is coloured according to Fig. 1.

bond to a hydrogen of Arg435. Furthermore, the α-phosphate interacts *via* hydrogen bonds with the main chain and side chain of Thr127, which is adjacent to motif I, and water 1. The β-phosphate exhibits intensive hydrogen bonding to the Lys125 and Thr126 side chains, as well as to Gly122, Gly124 and the Lys125 main-chain amides, all belonging to motif I, and to water molecules 3 and 4. In addition to this, the β-phosphate is involved in the coordination of the Mg^{2+} ion, the hexavalent coordination sphere of which comprises four additional water molecules and the Thr126 side chain (motif I). The residues of eponymic motif II, the DEAH motif, do not exhibit any direct interaction with the ADP molecule. Instead of interacting with the ADP molecule, Asp218 coordinates water molecule 4, whereas Glu219 participates in the coordination of water molecules 2, 3 and 4. At least in the ADP-bound state, Ala220 and His221 do not form any contacts with



(*a*) (*b*)

Figure 4
ADP-binding site of ctPrp43. The ADP molecule is sandwiched between the RecA1 (green) and RecA2 (violet) domains. C atoms are shown in yellow/green/violet, N atoms in blue, O atoms in red, P atoms in orange, the Mg^{2+} ion in light green and water molecules in cyan. Residues which are involved in ADP, water or Mg^{2+} binding are presented in ball-and-stick mode and are labelled according to the ctPrp43 sequence. Polar interactions are visualized as dashed black lines. (*a*) The adenine moiety is bound *via* π-electron stacking and the ribose by hydrogen bonding. (*b*) The α- and the β-phosphates participate intensively in hydrogen bonding. The central Mg^{2+} ion is coordinated by four water molecules, the Thr126 side chain and an O atom of the β-phosphate. The residues Asp218 and Glu219 of motif II are involved in the coordination of water molecules and do not interact directly with the bound nucleotide. For more detailed information, see §3.2 and Supplementary Table S1 for bonding distances.

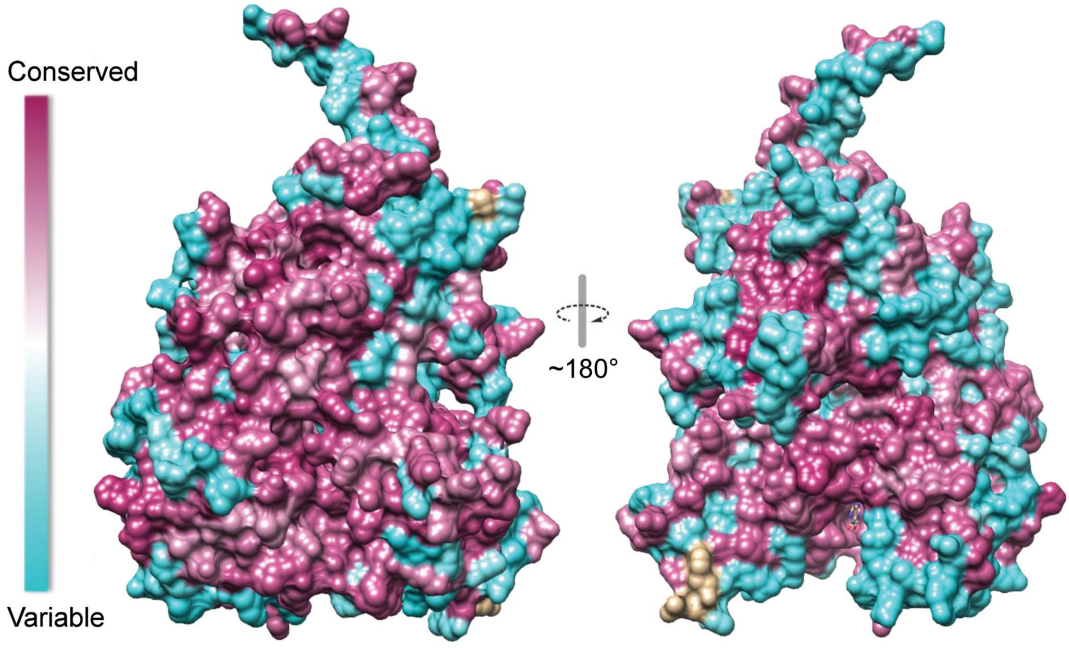


Figure 5
Surface conservation between ctPrp43 and scPrp43. The surface conservation was mapped onto ctPrp43(61–764) using *AL2CO* (Pei & Grishin, 2001) after structure alignment of the ctPrp43 and scPrp43 sequences in *T-Coffee* (Notredame *et al.*, 2000). ctPrp43 is presented as front (left) and back (right) views. Highly conserved regions are coloured magenta, poorly conserved regions light blue and unaligned residues pale wheat.

the bound nucleotide or a water molecule. Moreover, only one residue (Arg435) of motif VI, which comprises eight amino acids, interacts with the bound nucleotide. Here, the lack of interactions might suggest pronounced conformational changes of the RecA-like domains of Prp43 between the ADP-bound and ATP-bound states, which can be also caused by RNA substrate or G-patch protein binding. In contrast to DEAD-box helicases, the conformational changes of which have been extensively studied (for reviews, see Hilbert *et al.*, 2009; Ozgur *et al.*, 2015), the conformational flexibility regarding the relative position and orientation of the two RecA-like domains appears to be restricted by the additional C-terminal domains of the DEAH-box proteins.

3.3. ctPrp43 can functionally replace scPrp43

Prp43 from *C. thermophilum* exhibits high sequence similarity and identity to the homologues from *Homo sapiens* (66.2 and 56.5%, respectively) and *S. cerevisiae* (77.6 and 68.1%, respectively). The explicit degree of conservation of surface-exposed residues between ctPrp43 and scPrp43 is shown in Fig. 5, which illustrates that the RecA1, RecA2 and WH domains are especially highly conserved (each domain exhibits about 90% sequence similarity and 80% identity; for the exact values, see Supplementary Table S2), while the level of conservation is lower in the ratchet and the OB-fold (approximately 70 and 55%, respectively). Moreover, the crystal structure of ctPrp43 superposes very well on that of full-length scPrp43 (PDB entry 2xau) after the removal of its first 57 N-terminal residues, the corresponding residues to which are missing in our crystal structure (see Fig. 6). R.m.s.d. values of 0.90 Å (566 C^{α}) for chain *A* and 0.85 Å (574 C^{α}) for chain *B* were calculated after aligning both structures.

Owing to the high sequence identity and the conservation of a large number of residues located on the surface as well as the virtually identical structure, we wanted to analyse whether ctPrp43 is capable of functionally replacing its yeast homologue in spliceosomal disassembly assays. Proving that ctPrp43 is the authentic orthologue of scPrp43 would increase the importance of spliceosomal RNA helicases from *C. thermophilum* for further crystallographic studies and for the determination of their exact *modus operandi*. The disassembly assays were carried out with purified yeast intron–lariat spliceosomes (ILSs) and recombinant ctPrp43 according to Fourmann *et al.* (2013). When ILSs were incubated solely with ATP and subsequently fractionated on a glycerol gradient, only 5% of intron–lariat RNA was released (Fig. 7*a*). This indicates that ILSs are stable complexes and only a minor amount of RNA is released upon gradient centrifugation. The addition of ctPrp43 to ILSs lead to a slight increase in disassembly (22%; Fig. 7*b*), but in the presence of its spliceosomal activator protein Ntr1 ctPrp43 is now able to dissociate larger amounts of ILSs (45%; Fig. 7*c*). For these experiments, the heterodimer of Ntr1·Ntr2 from yeast was utilized, which indicates that ctPrp43 is also able to interact with Ntr1 from *S. cerevisiae*. Ntr2 is required to recruit Prp43 to its target substrate by binding to Brr2, which is part of the U5 snRNP. Finally, to demonstrate that our employed construct of ctPrp43, which lacks 60 N-terminal amino acids, is still fully functional, we also performed a spliceosome disassembly

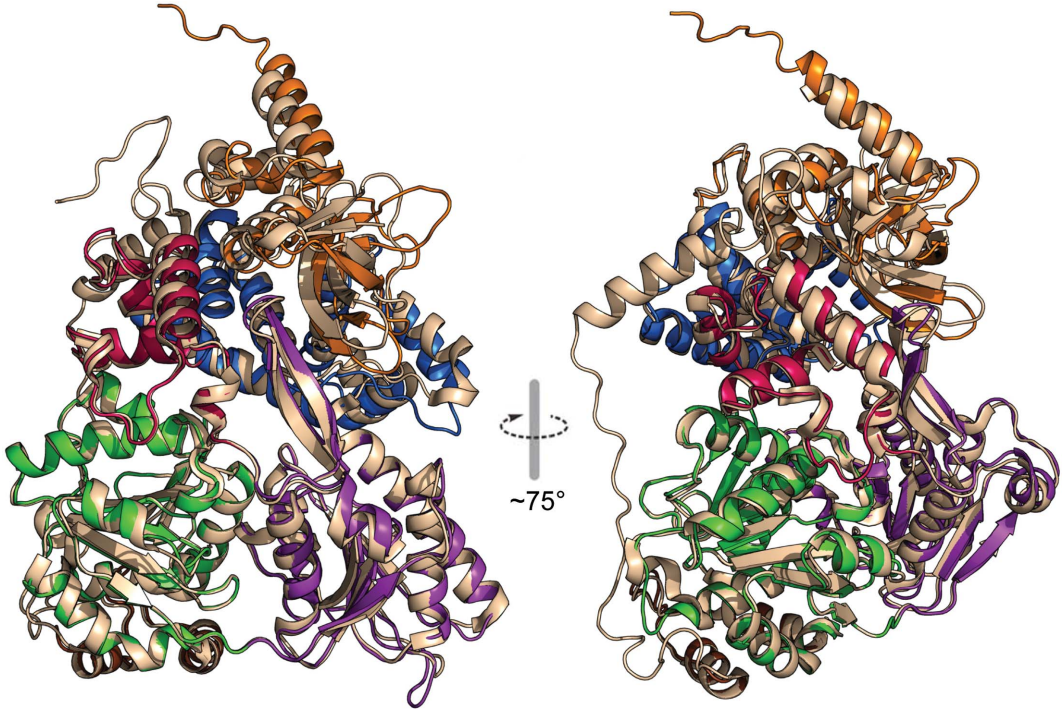


Figure 6
Superposition of *C. thermophilum* and *S. cerevisiae* Prp43. ctPrp43 is coloured as in Fig. 1. scPrp43 (PDB entry 2xau) is shown in pale wheat. The superposition was calculated for chain *B* of scPrp43 with 574 common C^{α} atoms and an r.m.s.d. of 0.85 Å. The front view (left) and side view (right) are presented. The region of scPrp43 which does not superpose on our structure is the N-terminal extension, which is missing in our truncated construct.

assay for this variant. Interestingly, the truncated construct leads to the highest rates of ILS disassembly in the presence of the scNtr1·scNtr2 dimer (70%, Fig. 7*d*) compared with the wild-type protein, which suggests involvement of the N-terminal extension in Prp43 regulation.

3.4. Thermophilic adaptation of ctPrp43

C. thermophilum was introduced to the structural biology community in 2011 by Amlacher and coworkers as a thermophilic eukaryote which grows at temperatures of up to 60°C (Amlacher *et al.*, 2011). We wanted to analyse the thermophilic adaption of ctPrp43 in comparison to its mesophilic counterpart from yeast. For this purpose, we performed CD measurements at 222 nm between 20 and 84°C to monitor the unfolding of α-helices (see Fig. 8). Owing to the highly similar overall structure, direct comparison of the obtained melting curves is possible. The melting points of ctPrp43 (56°C) and scPrp43 (40°C) differ significantly by 16°C. The molecular basis for the thermophilic adaption of ctPrp43 cannot be easily determined since neither the number of salt bridges nor of hydrogen bonds in ctPrp43 is significantly increased compared with scPrp43 nor does the secondary-structure content or the number of solvent-exposed residues change, which would facilitate entropic stabilization of the protein.

4. Conclusions

The presented 2.9 Å resolution crystal structure of Prp43 from *C. thermophilum* exhibits high structural similarity to the homologue from *S. cerevisiae* (for r.m.s.d. values for the superposition, see §3.3). Despite an almost identical overall structure, ctPrp43 shows a thermophilic adaption and exhibits a melting temperature which is elevated by 16°C compared with that of its orthologue from yeast. Nevertheless, ctPrp43 is capable of functionally replacing the orthologue from *S. cerevisiae* in yeast-based spliceosome disassembly assays. Interestingly, the construct we used for crystallization, which lacks the first 60 N-terminal residues, shows an increased capability to dissociate spliceosomes compared with the wild-type ctPrp43, which raises the question of the role of the N-terminal extension in regulation of Prp43. All previously described structural elements, namely the N-terminal extension, the RecA1 and RecA2 domains, the WH domain, the ratchet and the OB-fold (see Fig. 1), are also identifiable in the orthologue from *C. thermophilum* as well as all conserved DEAH-box protein motifs (see Fig. 3; Tanaka & Schwer, 2006; Walbott *et al.*, 2010; He *et al.*, 2010; Cordin *et al.*, 2006). The structure of ctPrp43 bound to ADP is very likely to be in a post-catalytic state (see §3.2), as suggested, among other reasons, by the low number of contacts formed by several motifs to the bound nucleotide, which have been reported to be crucial for the ATPase activity of Prp43. However, the

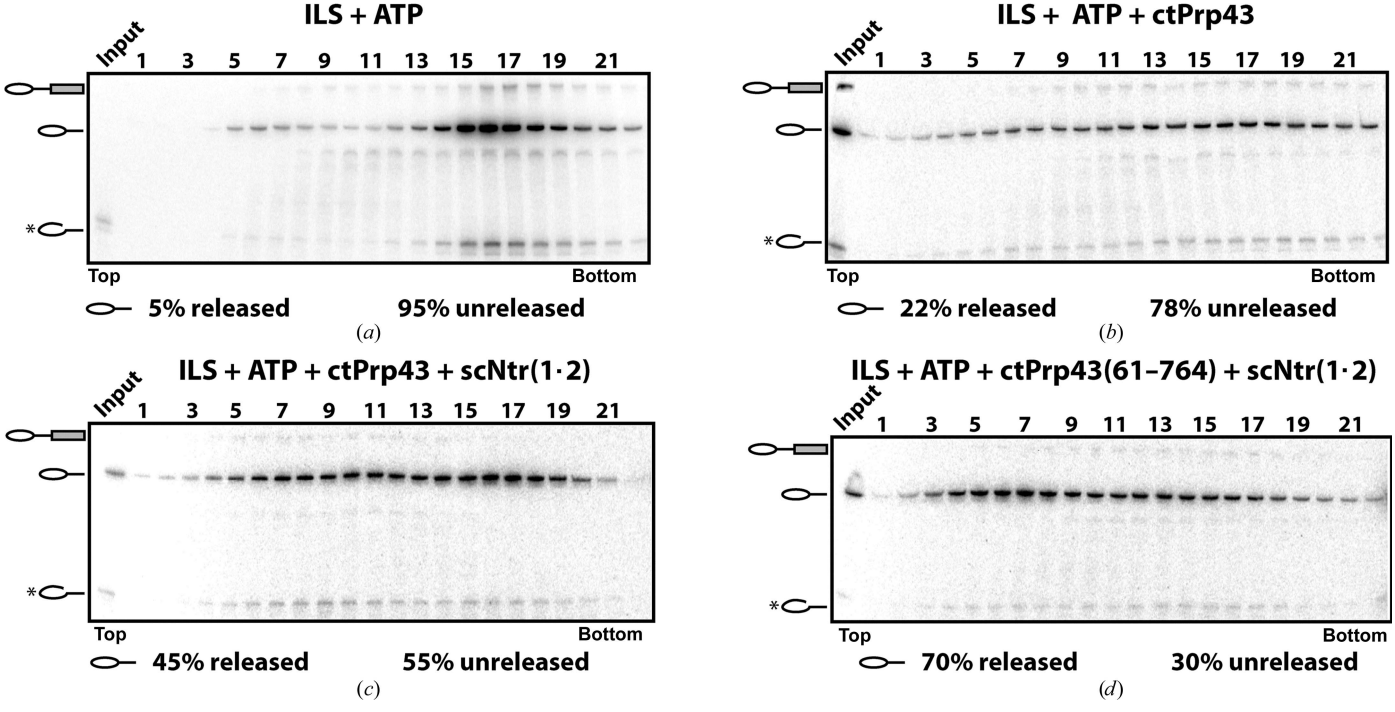


Figure 7
In vitro yeast spliceosomal disassembly assay. The disassembly of the intron–lariat spliceosome was analyzed as described previously (Fourmann *et al.*, 2013). (*a*) The negative control without ctPrp43 revealed 5% dissociated spliceosomes. (*b*) In the presence of ctPrp43 and ATP 22% were disassembled. (*c*) The addition of the activator heterodimer scNtr1–scNtr2 increased the amount of disassembly events to 45%. (*d*) The N-terminally truncated construct ctPrp43(61–764), which was used for crystallization, exhibited an even higher number of dissociated spliceosomes (70%). Quantification was performed with *ImageQuant* (Molecular Dynamics). Numbers represent the percentage of intron–lariat RNA released in the top fractions (sum of fractions 1–11) or remaining associated with the ILS (unreleased; sum of fractions 12–22) relative to the intron–lariat RNA distributed in all 22 fractions, the sum of which was set to 100%.

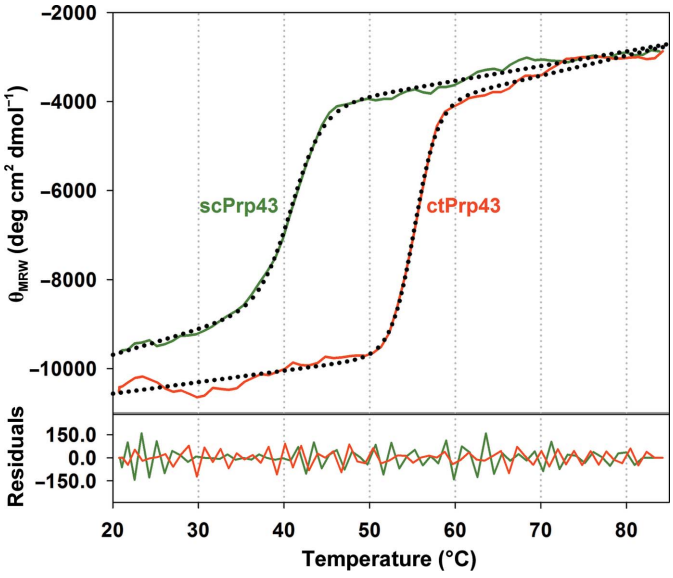


Figure 8
CD spectroscopic melting-point determination of ctPrp43 and scPrp43. Unfolding of α-helices was assayed at 222 nm between 20 and 84°C. Curves were fitted according to Pace *et al.* (1998) and exact melting points were calculated. Far-UV spectra of ctPrp43 and scPrp43 were measured to ensure complete protein denaturation at 84°C (see Supplementary Fig. S3). Residuals indicate smoothing by a factor of two in the corresponding colour to the melting curves.

structure of Prp43 in its activated state with bound ATP, RNA and G-patch proteins is as yet unknown. Owing to this, the mechanism of activation and RNA translocation still remains elusive. The DEAH-box proteins from *C. thermophilum* might provide an alternative source to tackle the structure determination of their functional complexes.

Acknowledgements

We are grateful for beam-time allocation at BESSY II, Berlin, Germany. This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG) to RF and RL (SFB 860, TP A1 and TP A2). We would also like to thank Dr Achim Dickmanns for support during data collection and for critical reading of the manuscript, as well as Marlyn Thölken for technical assistance. Furthermore, we thank Kai Tittmann (University of Göttingen) for access to the CD spectrophotometer and Fabian Rabe von Pappenheim and Viktor Sautner for support during data acquisition.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
Amlacher, S., Sarges, P., Flemming, D., van Noort, V., Kunze, R., Devos, D. P., Arumugam, M., Bork, P. & Hurt, E. (2011). *Cell*, **146**, 277–289.
Aravind, L. & Koonin, E. V. (1999). *Trends Biochem. Sci.* **24**, 342–344.
Arenas, J. E. & Abelson, J. N. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 11798–11802.
Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
Bleichert, F. & Baserga, S. J. (2007). *Mol. Cell*, **27**, 339–352.
Bock, T. *et al.* (2014). *Nucleic Acids Res.* **42**, 13525–13533.
Bohnsack, M. T., Martin, R., Granneman, S., Ruprecht, M., Schleiff, E. & Tollervey, D. (2009). *Mol. Cell*, **36**, 583–592.
Boon, K.-L., Auchynnikava, T., Edwalds-Gilbert, G., Barrass, J. D., Droop, A. P., Dez, C. & Beggs, J. D. (2006). *Mol. Cell. Biol.* **26**, 6016–6023.
Büttner, K., Nehring, S. & Hopfner, K.-P. (2007). *Nature Struct. Mol. Biol.* **14**, 647–652.
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.* D**66**, 12–21.
Christian, H., Hofele, R. V., Urlaub, H. & Ficner, R. (2014). *Nucleic Acids Res.* **42**, 1162–1179.
Cordin, O., Banroques, J., Tanner, N. K. & Linder, P. (2006). *Gene*, **367**, 17–37.
Dolinsky, T. J., Nielsen, J. E., McCammon, A. & Baker, N. A. (2004). *Nucleic Acids Res.* **32**, W665–W667.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fairman-Williams, M. E., Guenther, U.-P. & Jankowsky, E. (2010). *Curr. Opin. Struct. Biol.* **20**, 313–324.
Fourmann, J.-B., Schmitzová, J., Christian, H., Urlaub, H., Ficner, R., Boon, K.-L., Fabrizio, P. & Lührmann, R. (2013). *Genes Dev.* **27**, 413–428.
He, Y., Andersen, G. R. & Nielsen, K. H. (2010). *EMBO Rep.* **11**, 180–186.
Hilbert, M., Karow, A. R. & Klostermeier, D. (2009). *Biol. Chem.* **390**, 1237–1250.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kim, S.-H., Smith, J., Claude, A. & Lin, R.-J. (1992). *EMBO J.* **11**, 2319–2326.
Lebaron, S., Froment, C., Fromont-Racine, M., Rain, J.-C., Monsarrat, B., Caizergues-Ferrer, M. & Henry, Y. (2005). *Mol. Cell. Biol.* **25**, 9269–9282.
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.
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.
Notredame, C., Higgins, D. G. & Heringa, J. (2000). *J. Mol. Biol.* **302**, 205–217.
Ozgur, S., Buchwald, G., Falk, S., Chakrabarti, S., Prabu, J. R. & Conti, E. (2015). *FEBS J.* **282**, 850–863.
Pace, C. N., Hebert, E. J., Shaw, K. L., Schell, D., Both, V., Krajcikova, D., Sevcik, J., Wilson, K. S., Dauter, Z., Hartley, R. W. & Grimsley, G. R. (1998). *J. Mol. Biol.* **279**, 271–286.
Pei, J. & Grishin, N. V. (2001). *Bioinformatics*, **17**, 700–712.
Pertschy, B., Schneider, C., Gnädig, M., Schäfer, T., Tollervey, D. & Hurt, E. (2009). *J. Biol. Chem.* **284**, 35079–35091.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Pyle, A. M. (2008). *Annu. Rev. Biophys.* **37**, 317–336.
Robert-Paganin, J., Réty, S. & Leulliot, N. (2015). *Biomed. Res. Int.* **2015**, 931857.
Rudolph, M. G., Heissmann, R., Wittmann, J. G. & Klostermeier, D. (2006). *J. Mol. Biol.* **361**, 731–743.
Schwer, B. & Guthrie, C. (1992). *EMBO J.* **11**, 5033–5039.
Tanaka, N., Aronova, A. & Schwer, B. (2007). *Genes Dev.* **21**, 2312–2325.
Tanaka, N. & Schwer, B. (2005). *Biochemistry*, **44**, 9795–9803.
Tanaka, N. & Schwer, B. (2006). *Biochemistry*, **45**, 6510–6521.
Tsai, R.-T., Fu, R.-H., Yeh, F.-L., Tseng, C.-K., Lin, Y.-C., Huang, Y.-H. & Cheng, S.-C. (2005). *Genes Dev.* **19**, 2991–3003.
Tsai, R.-T., Tseng, C.-K., Lee, P.-J., Chen, H.-C., Fu, R.-H., Chang, K., Yeh, F.-L. & Cheng, S.-C. (2007). *Mol. Cell. Biol.* **27**, 8027–8037.
Walbott, H., Mouffok, S., Capeyrou, R., Lebaron, S., Humbert, O., van Tilbeurgh, H., Henry, Y. & Leulliot, N. (2010). *EMBO J.* **29**, 2194–2204.