Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Structural insights into the novel ARM-repeat protein CTNNBL1 and its association with the hPrp19–CDC5L complex

Jae-Woo Ahn,[a] Sangwoo Kim,[b] Eun-Jung Kim,[a] Yeo-Jin Kim[a] and Kyung-Jin Kim[a]*

[a]Structural and Molecular Biology Laboratory, School of Life Sciences and Biotechnology, Kyungpook National University, Daehak-ro 80, Buk-ku, Daegu 702-701, Republic of Korea, and [b]School of Nano-Bioscience and Chemical Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, Republic of Korea

Correspondence e-mail: kkim@knu.ac.kr

Received 29 August 2013
Accepted 8 December 2013

PDB references: CTNNBL1, residues 77–563, 4mfu; residues 33–563, 4mfv

The hPrp19–CDC5L complex plays a crucial role during human pre-mRNA splicing by catalytic activation of the spliceosome. In order to elucidate the molecular architecture of the hPrp19–CDC5L complex, the crystal structure of CTNNBL1, one of the major components of this complex, was determined. Unlike canonical ARM-repeat proteins such as β-catenin and importin-α, CTNNBL1 was found to contain a twisted and extended ARM-repeat structure at the C-terminal domain and, more importantly, the protein formed a stable dimer. A highly negatively charged patch formed in the N-terminal ARM-repeat domain of CTNNBL1 provides a binding site for CDC5L, a binding partner of the protein in the hPrp19–CDC5L complex, and these two proteins form a complex with a stoichiometry of 2:2. These findings not only present the crystal structure of a novel ARM-repeat protein, CTNNBL1, but also provide insights into the detailed molecular architecture of the hPrp19–CDC5L complex.

1. Introduction

Pre-mRNA splicing is a modification of the nascent pre-mRNA transcript in which introns are removed and exons are joined, and is catalyzed by the spliceosome, a dynamic megadalton-sized complex composed of four ribonucleoprotein particles (snRNPs; U1, U2, U5 and U4/U6) and non-snRNP proteins such as the hPrp19–CDC5L complex (Wahl *et al.*, 2009; Staley & Guthrie, 1998; Makarov *et al.*, 2002). snRNPs are formed by snRNAs, along with their associated proteins, and bind to specific sequences on the pre-mRNA substrate (Guo *et al.*, 2009). snRNPs and non-snRNP proteins assemble to form the spliceosome in consecutive order and in a highly dynamic procedure (Wahl *et al.*, 2009; Krämer, 1996; Murray & Jarrell, 1999). The first recognition of pre-mRNA involves binding of the U1 snRNP to the 5′ splice site, followed by binding of the U2 snRNP to the branch-point sequence (BPS), forming the prespliceosome or complex A (Will & Lührmann, 1997). The U4/U6–U5 tri-snRNP is then recruited to the assembling spliceosome to form complex B. Following dramatic compositional and structural rearrangements of complex B, the first transesterification reaction occurs and generates complex C, which then catalyzes the second splicing reaction. After completion of catalysis, the postspliceosomal complex is disassembled and the mRNA and U2, U5 and U6 snRNPs are released for reuse in subsequent splicing reactions.

In addition to the snRNPs, various non-snRNP proteins are also involved in the formation of the spliceosome, and many of them are known to play crucial roles in the splicing reaction. One well known non-snRNP protein complex is the hPrp19–CDC5L complex in humans, which is referred to as the NTC (**n**ine**t**een **c**omplex) in yeast (Tarn *et al.*, 1993, 1994). The

hPrp19 protein, which is a major component of the hPrp19–CDC5L complex, is known to be a highly conserved splicing factor required for the activation of the spliceosome, and exhibits E3 ubiquitin ligase activity *in vitro* (Hatakeyama *et al.*, 2001; Ohi *et al.*, 2003; Chan *et al.*, 2003; Ohi & Gould, 2002). The hPrp19–CDC5L complex consists of seven proteins: hPrp19, CDC5L, PRL1, AD002, SPF27, CTNNBL1 and HSP73 (Makarova *et al.*, 2004). While homologues of four proteins of the hPrp19–CDC5L complex, hPrp19, CDC5L, PRL1 and SPF27, are found in the yeast NTC, homologues of CTNNBL1, HSP73 and AD002 are absent in the yeast NTC (Chen *et al.*, 2002; Ohi & Gould, 2002; Tarn *et al.*, 1994).

Human CTNNBL1 (catenin-β-like 1), a component of the hPrp19–CDC5L complex, has been identified as a widely expressed nuclear protein that is expressed in multiple human tissues (Jabbour *et al.*, 2003) and is conserved in a variety of vertebrates. Studies of the molecular architecture of the hPrp19–CDC5L complex have shown that CTNNBL1 interacts with CDC5L and AD002 in order to form a complex. The limited proteolysis of the purified hPrp19–CDC5L complex also suggests that the CTNNBL1–AD002 heterooligomer binds the hPrp19–CDC5L complex core *via* the N-terminal domain of CDC5L (Grote *et al.*, 2010). Recently, it was proven that the hPrp19–CDC5L complex is formed by the interaction of CTNNBL1 with CDC5L and Prp31; in particular, NLS3 located in the N-terminal domain of CDC5L is crucially involved in the binding to CTNNBL1 (Ganesh *et al.*, 2011).

Interestingly, it has been reported that CTNNBL1 also interacts with activation-induced deaminase (AID), suggesting the involvement of CTNNBL1 in antibody diversity (Conticello *et al.*, 2008). This interaction suggests that the nuclear localization signal motif directs AID to the nucleoli, where it colocalizes with CTNNBL1 and physically associates with nucleolin and nucleophosmin (Hu *et al.*, 2013). More interestingly, a deficiency of the spliceosome-associated factor CTNNBL1 results in embryonic lethality in mice (Chandra *et al.*, 2013). In addition, a genome-wide survey and a functional brain-imaging study have identified CTNNBL1 as a memory-related protein (Papassotiropoulos *et al.*, 2013).

Recently, a crystal structure of CTNNBL1 was reported, showing a monomeric structure (Huang *et al.*, 2013). Here, we determined the dimeric structure of CTNNBL1 and found that the protein exhibits a novel ARM-repeat structure compared with the canonical ARM-repeat proteins such as β-catenin and importin-α. Based on the dimeric structure of CTNNBL1, we identified the CDC5L binding site of the protein and revealed that the CTNNBL1–CDC5L complex is formed with a 2:2 stoichiometry. Our findings therefore provide new insights into the molecular architecture of the hPrp19–CDC5L complex.

2. Materials and methods

2.1. Protein production

The gene coding for human CTNNBL1 (residues 77–563) was amplified by polymerase chain reaction (PCR) from the human cDNA library, and the amplified fragment was digested with *Nde*I and *Sal*I restriction enzymes and ligated into a modified VHb fusion vector (pPosKJ) which contained a 6×His tag and bacterial haemoglobin at the N-terminus of the target protein (Kwon *et al.*, 2005). In order to obtain mutant CTNNBL1 proteins, site-directed mutagenesis was performed using *PfuTurbo* DNA polymerase (Stratagene). All oligonucleotides used in the study were purchased in a salt-free grade from Bioneer (Daejeon, Republic of Korea) and are summarized in the Supporting Information[1]. The resulting plasmid pPosKJ:*CTNNBL1* was transformed into *Escherichia coli* strain B834(DE3) and was grown to an OD_{600} of 0.6 in LB medium containing 100 mg l^{-1} ampicillin at 310 K. After induction with 1.0 m*M* IPTG, the cells were grown for a further 20 h at 295 K. The cultured cells were harvested by centrifugation at 5000*g* for 20 min at 277 K, and the cell pellet was resuspended in ice-cold buffer *A* (50 m*M* Tris–HCl pH 8.0) containing 10 m*M* imidazole and disrupted by ultrasonication. The cell debris was removed by centrifugation at 11 000*g* for 1 h and the lysate was bound to Ni–NTA agarose (Qiagen). After washing with buffer *A* containing 20 m*M* imidazole, the bound proteins were eluted with 300 m*M* imidazole in buffer *A*. The 6×His tag and VHb were then removed by treatment with TEV protease (Invitrogen) and Ni–NTA chromatography. Trace amount of contaminants were removed by applying ion-exchange and size-exclusion chromatography (Superdex 200, GE Healthcare). The purified protein was concentrated to approximately 35 mg l^{-1} in 40 m*M* Tris–HCl pH 8.0 and stored at 193 K for crystallization trials. SDS–PAGE analysis of the purified protein showed a single band corresponding to the calculated molecular weight of the target protein. SeMet-substituted CTNNBL1 and the mutant proteins were prepared similarly to the wild-type CTNNBL1. For the production of $CDC5L^{141–377}$, the gene coding for $CDC5L^{141–377}$ was cloned into pET-30a expression vector and the $CDC5L^{141–377}$ protein with a 6×His tag at the C-terminus was purified similarly to CTNNBL1 but without TEV protease treatment.

2.2. Crystallization, data collection and structure determination

The initial crystallization conditions were obtained by sparse-matrix screening. Crystals suitable for diffraction appeared using 100 m*M* CAPS pH 9.5, 8% PEG 8K, 0.2 *M* NaCl at 295 K within 2 d and reached their maximal size in about 7 d, with dimensions of approximately 0.2 × 0.1 × 0.1 mm. For data collection, 30%(*v*/*v*) glycerol was added to the crystallizing precipitant as a cryoprotectant and the crystals were immediately placed in a 100 K nitrogen-gas stream. X-ray diffraction data were collected from native crystals to a resolution of 2.7 Å on beamline 7A of the Pohang Accelerator Laboratory (PAL; Republic of Korea) using a Quantum 270 CCD detector (San Diego, California, USA). The data were then indexed, integrated and scaled using the *HKL*-2000 suite

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: WA5060).

(Otwinowski & Minor, 1997). The crystals belonged to space group $P3_121$, with unit-cell parameters $a = b = 90$, $c = 178$ Å, $\gamma = 120°$. With one CTNNBL1 molecule in the asymmetric unit, the crystal volume per unit of protein weight was 3.74 Å^3 Da^{-1}, corresponding to a solvent content of 67.1% (Matthews, 1968). SeMet-substituted protein crystals were obtained using the same crystallization condition as used for the native protein crystal. Single-wavelength anomalous dispersion (SAD) data were collected from an SeMet protein crystal on beamline 7A at PAL to a wavelength of 0.97953 Å. 14 Se atoms out of the expected 17 in the asymmetric unit were identified at 3.2 Å resolution using *SOLVE* (Terwilliger & Berendzen, 1999). The electron density was improved by density modification using *RESOLVE* (Terwilliger, 2000), resulting in 52% of the cloned residues being automatically built. Further model building was performed manually using *WinCoot* (Emsley & Cowtan, 2004), and refinement was performed with *REFMAC*5 (Murshudov *et al.*, 2011) in *CCP*4 (Winn *et al.*, 2011) and *CNS* (Brünger *et al.*, 1998) (Supplementary Fig. S3). The data statistics are summarized in Table 1. Refined models of human CTNNBL1 consisting of residues 77–563 and residues 33–563 have been deposited in the Protein Data Bank as entries 4mfu and 4mfv, respectively.

Table 1
Data-collection, phasing and refinement statics for human CTNNBL1.

Values in parentheses are for the highest resolution shell.

	CTNNBL1 (77–563)	CTNNBL1 (33–563)	SeMet CTNNBL1 (77–563)
Data collection			
Space group	$P3_121$	$P3_1$	$P3_121$
Unit-cell parameters			
a (Å)	90.15	90.04	90.03
b (Å)	90.15	90.04	90.03
c (Å)	178.56	175.75	175.74
$\alpha = \beta$ (°)	90.00	90.00	90.00
γ (°)	120.00	120.00	120.00
Resolution (Å)	50.00–2.74 (2.84–2.74)	50.00–2.92 (3.02–2.92)	50.00–3.00 (3.11–3.00)
R_{merge} (%)	9.5 (32.8)	7.8 (26.6)	9.1 (80.7)
$\langle I/\sigma(I)\rangle$	36.9 (2.2)	30.3 (2.6)	41.8 (3.2)
Completeness (%)	95.8 (71.5)	94.1 (77.6)	100.0 (100.0)
Multiplicity	13.9 (4.7)	7.2 (3.1)	10.1 (9.8)
Refinement			
Resolution (Å)	50.00–2.74	50.00–2.92	
No. of reflections	20631	30828	
R_{work}/R_{free} (%)	24.7/31.9	23.3/30.5	
No. of protein atoms	3917	7818	
Overall B factor (Å^2)	89.03	92.29	
R.m.s. deviations			
Bond lengths (Å)	0.019	0.019	
Bond angles (°)	1.426	1.481	

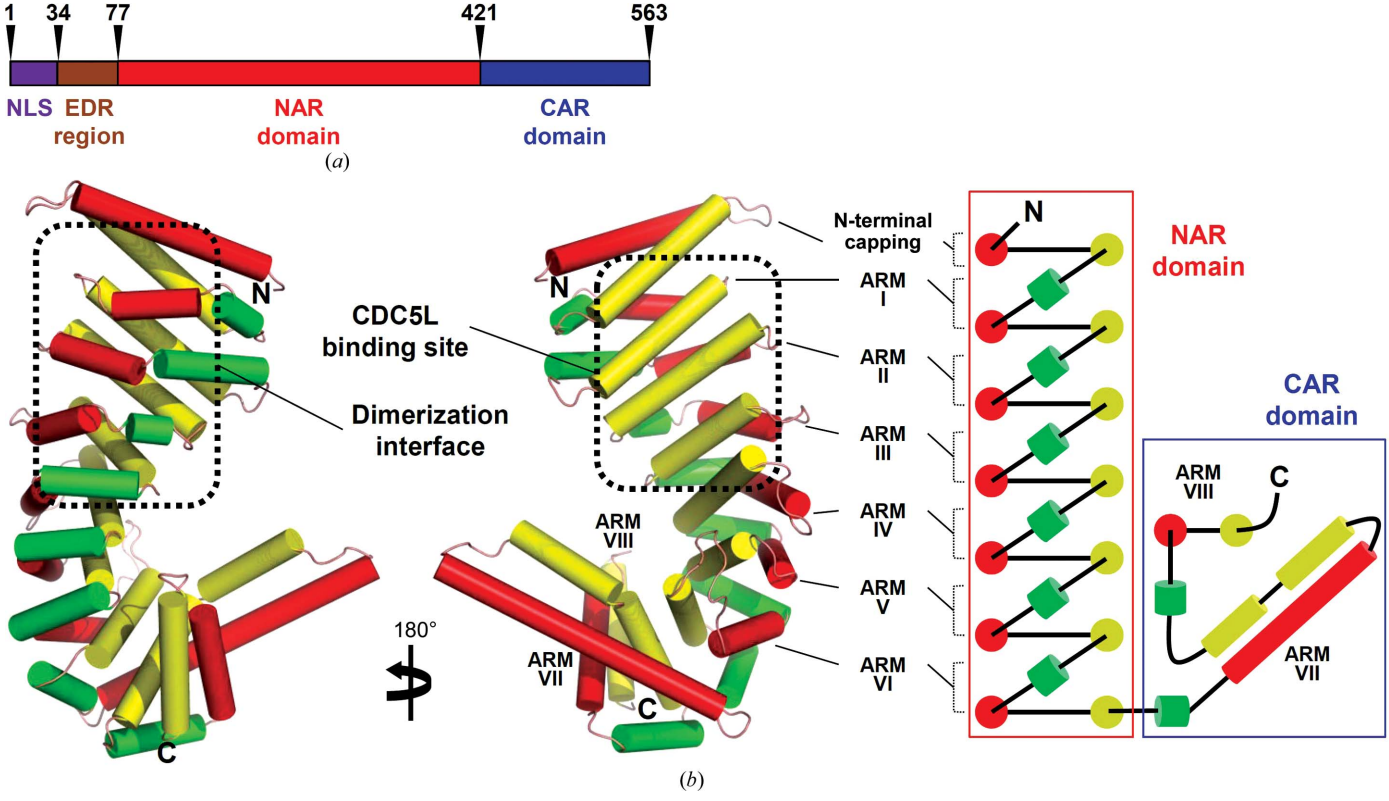


Figure 1
Crystal structure of human CTNNBL1. (*a*) Domain identification of CTNNBL1. The NLS and EDR regions represent the nuclear localization sequence and the glutamate/aspartate-rich region, respectively, and are labelled appropriately. The NAR and CAR domains represent the N-terminal and C-terminal ARM-repeat domains, respectively. (*b*) The monomeric structure of CTNNBL1. The secondary-structure elements are shown as cylindrical cartoon representations. The H1, H2 and H3 helices of each ARM repeat are distinguished by green, red and yellow colours, respectively, and the N-terminal capping region and the eight ARM repeats are labelled. The CDC5L binding site and the dimerization interface are indicated by dotted rectangles and are labelled appropriately. A schematic diagram of the ARM-repeat arrangement of CTNNBL1 is presented (right), and the NAR and CAR domains are labelled.

2.3. *In vitro* pull-down assay

In order to identify the CDC5L binding site of CTNNBL1, a pull-down assay was performed. For binding to the Ni–NTA column, the CDC5L$^{141–377}$ protein was purified with a 6×His tag at the C-terminus and CTNNBL1 was purified as the native protein. The wild type and various mutants of CTNNBL1 were mixed with the CDC5L$^{141–377}$ protein and incubated at 277 K for 2 h. The mixture was loaded onto an Ni–NTA column and washed thoroughly with 10 m*M* imidazole to avoid nonspecific binding of CTNNBL1 to the Ni–NTA resin. The sample was eluted with 300 m*M* imidazole and run on SDS–PAGE.

2.4. Size-exclusion chromatographic analysis

In order to confirm the dimerization of CTNNBL1 and to measure the molecular weight of the protein, 500 µl 2 mg l^{-1} CTNNBL1 was injected onto a Superdex 200 10/300 column (GE Healthcare) at a flow rate of 0.5 ml min^{-1} at various concentrations of NaCl: 0, 100, 150, 200, 250 and 300 m*M*. The molecular weights of the eluted samples were calculated based on the calibration curve of standard samples. The molecular weight of the CTNNBL1–CDC5L$^{141–377}$ complex was determined using a procedure similar to that used for CTNNBL1.

2.5. Multi-angle light-scattering analysis (MALS)

The molecular weight and dimerization state of CTNNBL1 were monitored by size-exclusion chromatography multi-angle light scattering (SEC-MALS). The purified CTNNBL1 was injected onto a Superdex 200 HR 10/30 gel-filtration column (GE Healthcare) equilibrated with 20 m*M* Tris pH 8.0 containing 0, 100, 200, 300 and 400 m*M* NaCl. The chromatography system was coupled to a three-angle light-scattering detector (mini-DAWN TREOS) and a refractive-index detector (Optilab DSP) (Wyatt). Data were collected at a flow rate of 0.4 ml min^{-1}. Data acquisition and processing was carried out using the *ASTRA* software (Wyatt).

3. Results and discussion

3.1. CTNNBL1 is a novel ARM-repeat protein

To study the structural features of human CTNNBL1 and the molecular architecture of the hPrp19–CDC5L complex, we determined the crystal structure of an N-terminally truncated form of human CTNNBL1 (amino-acid residues 77–563). The truncated N-terminal region consisted of the nucleus localization sequence (NLS; amino-acid residues 1–33) and the glutamate/aspartate-rich region (EDR region; amino-acid residues 34–76) (Fig. 1*a* and Supplementary Fig. S1). The crystal belonged to space group $P3_121$, and one polypeptide was present in the asymmetric unit of the structure. The crystal structure of CTNNBL1 showed that the protein consists mainly of α-helices and forms an armadillo (ARM) repeat structure (Fig. 1*b*). The first two helices located in the N-terminal region of the structure showed a HEAT-repeat-like structure and formed an N-terminal capping region. The

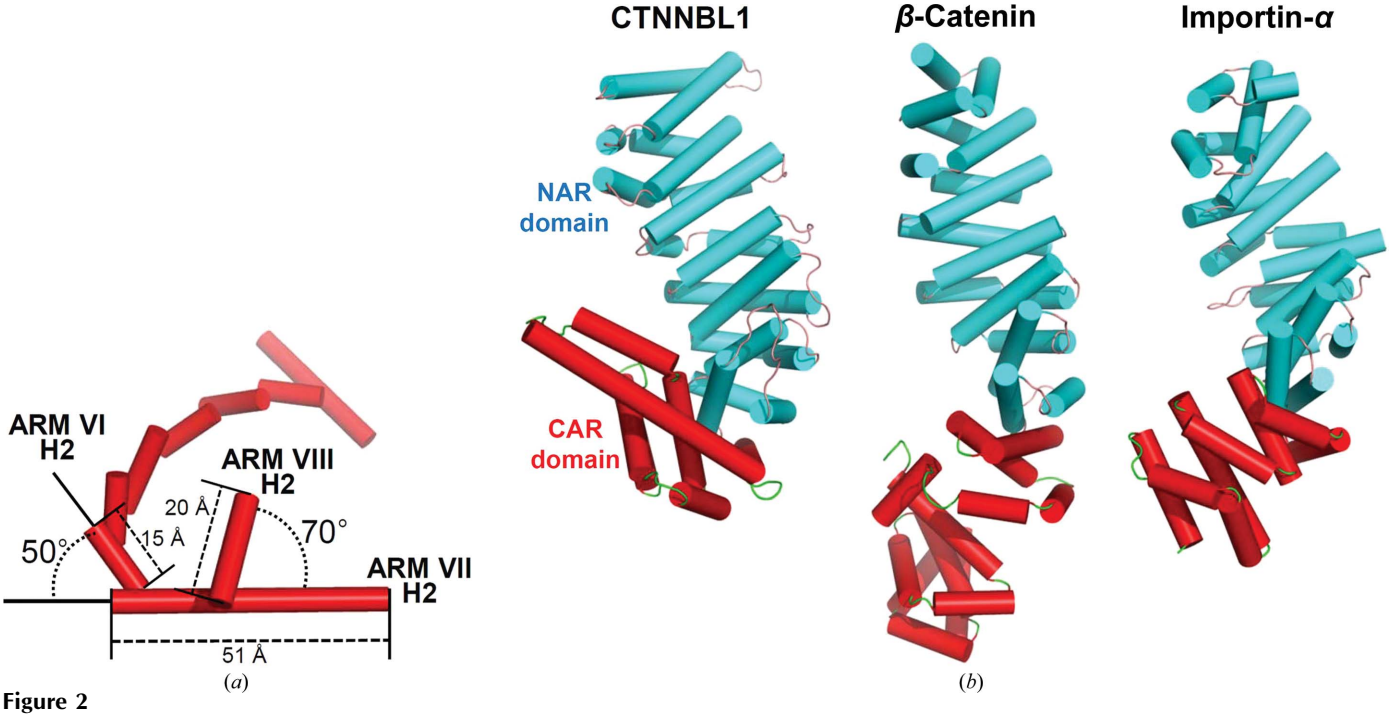


Figure 2
Extended and twisted ARM-repeat structures of the CAR domain. (*a*) Twist angles and lengths of the H2 helices of ARMs VI, VII and VIII. The H2 helices of the CTNNBL1 structure are shown as a cylindrical cartoon representation. Twist angles are measured based on the H2 helices. The twist angles and lengths of the H2 helices of ARMs VI, VII and VIII are indicated. (*b*) Comparison of the structure of CTNNBL1 with other canonical ARM-repeat proteins: β-catenin and importin-α. The structures of CTNNBL1, β-catenin and importin-α are shown as cylindrical cartoon representations in the same orientation. The NAR and CAR domains are distinguished by cyan and red colours, respectively, and are labelled appropriately. The NAR and CAR domain portions of β-catenin and importin-α are distinguished using the same colour scheme as for CTNNBL1.

rest of the helices constituted eight ARM repeats and, as observed in ARM-repeat proteins such as β-catenin and importin-α, each ARM in CTNNBL1 is composed of three helices (H1, H2 and H3). The H2 and H3 helices are packed antiparallel to each other and the H1 helix is perpendicular to them (Fig. 1).

However, the CTNNBL1 structure showed that the protein contains unique ARM-repeat structures in the C-terminal region of the protein. The N-terminal capping region and the first six ARM repeats (ARM I–ARM VI) of CTNNBL1 showed a canonical ARM-repeat structure that superposed well with the structures of other ARM-repeat proteins such as β-catenin and importin-α. Interestingly, the last two ARM repeats (ARM VII and ARM VIII) of the protein exhibited unique structural conformations (Fig. 2). In canonical ARM-repeat structures each repeat is related to its neighbour by a right-handed rotation of 30° about an axis that is approximately parallel to the superhelix axis. In CTNNBL1, the rotation angles of each ARM from the N-terminal capping region to ARM VI were similar to those in β-catenin and importin-α, that is 30° on an average. However, ARM VII and ARM VIII of CTNNBL1 are extremely twisted, with angles of 50° and 70°, respectively (Fig. 2*a*). Another interesting structural feature of CTNNBL1 is the extended H2 and H3 helices of ARM VII. The length of the ARM VII H2 helix is twice that of the other helices, and H3 is separated into two tandem helices (H3 and H3′) (Fig. 2). As a result of these unique structural features of ARM VII and ARM VIII of CTNNBL1, we divided the protein into two distinct domains: an N-terminal canonical ARM-repeat (NAR) domain and a C-terminal extended and twisted ARM-repeat (CAR) domain. In fact, the unique structural conformation of the CAR domain

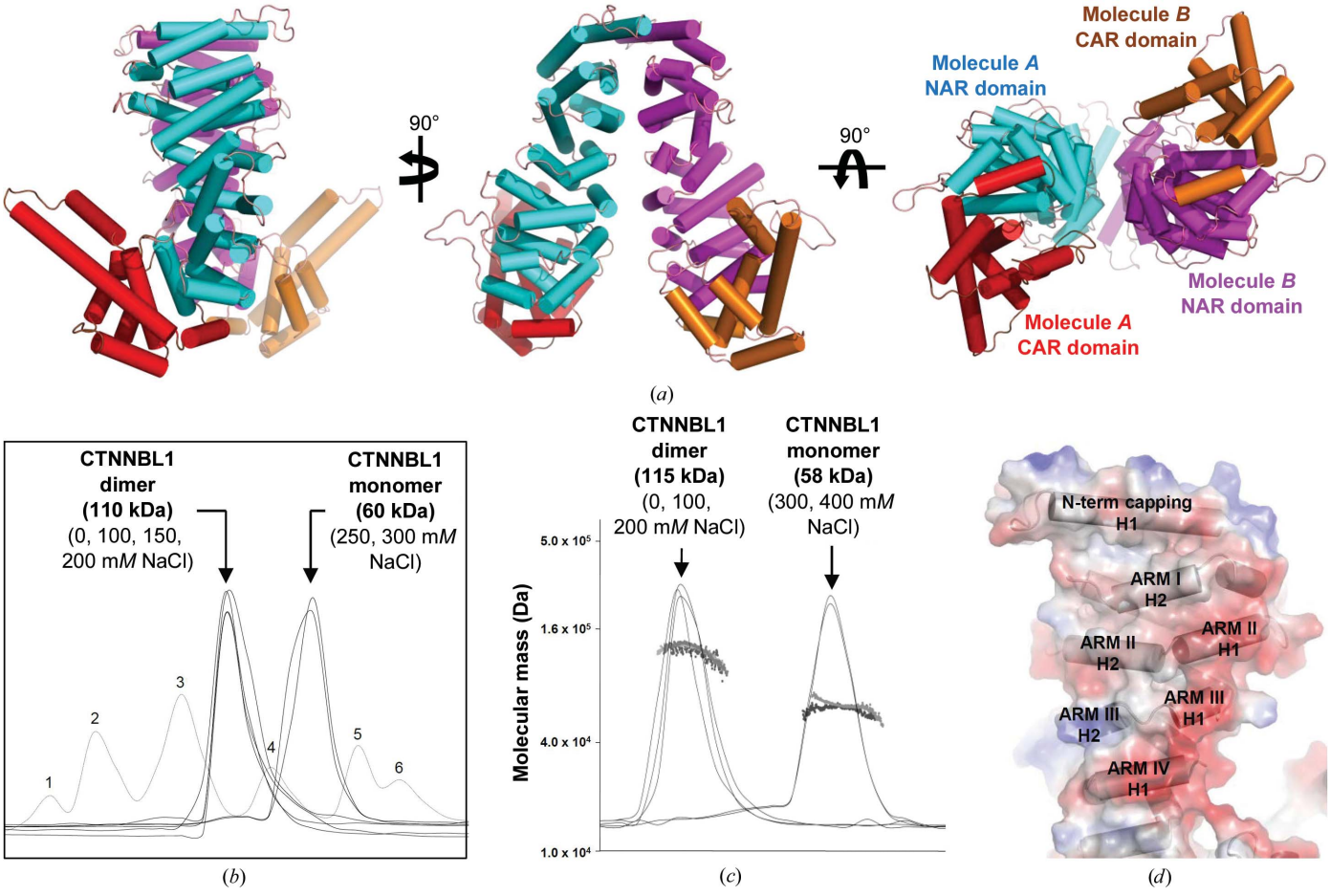


Figure 3
Dimerization of CTNNBL1. (*a*) Structure of the CTNNBL1 dimer. The structure of the CTNNBL1 dimer is shown as a cylindrical cartoon representation. The NAR and CAR domains of molecules *A* and *B* are distinguished in cyan and red and in magenta and orange, respectively, and are labelled appropriately. (*b*) Size-exclusion chromatographic analysis. Dimer formation of CTNNBL1 was analyzed by the addition of various concentrations of NaCl. The elution peaks corresponding to the dimeric and monomeric forms of CTNNBL1 are indicated by arrows, and the concentrations of NaCl added to the protein solution are shown in parentheses. Dimeric and monomeric CTNNBL1 eluted at molecular weights of ∼110 and ∼60 kDa, respectively. For precise analysis of the molecular weight, standard samples of thyroglobulin (669 kDa), ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (44 kDa) and carbonic anhydrase (29 kDa) were used for calibration of the size-exclusion chromatography and are labelled 1–6, respectively. (*c*) SEC-MALS analysis of CTNNBL1 with various concentrations of NaCl. Molecular mass and refractive index are plotted as a function of elution volume. Peaks for dimers and monomers are indicated by arrows. (*d*) The dimerization interface of CTNNBL1. The structure of CTNNBL1 is presented as an electrostatic potential model, and the secondary-structure elements involved in the dimerization of the protein are shown as a cylindrical cartoon representation and labelled appropriately.

positioned these two domains in sufficient proximity to form a hydrogen bond between Asp274 from ARM IV and Glu544 from ARM VIII.

3.2. CTNNBL1 forms a stable dimer

To date, all ARM-repeat proteins have been reported to function as monomers. However, the crystallographic $P3_121$ symmetry of CTNNBL1 unexpectedly showed that the protein forms a dimer, even though a monomeric protein was present in the asymmetric unit of our structure (Fig. 3*a*). The dimeric state of CTNNBL1 was further verified by size-exclusion chromatography, which showed a molecular weight of ~110 kDa, similar to the calculated molecular weight for the CTNNBL1 dimer (Fig. 3*b*). Interestingly, the N-terminal capping region and the first four ARM repeats (ARM I, ARM II, ARM III and ARM III) participated in the dimerization of CTNNBL1, even though, as described above, these repeats are structurally quite similar to those of canonical ARM-repeat proteins. The N-terminal capping region and the first four ARM repeats from one polypeptide contacted the corresponding repeats from another polypeptide, and because of the rotation of the ARM repeats the contact modes of each repeat were quite different from each other (Figs. 3*a* and 3*d*). The H2 helices of the N-terminal capping region and ARM I formed antiparallel helix dimers between the H2 helices of each monomer and seem to play a major role in CTNNBL1 dimerization. The connecting loops of H1 and H2 of each ARM repeat were involved in the contacts of ARM II and ARM II and of ARM III and ARM III. The contact between two ARM IV repeats was constituted by the formation of an antiparallel helix dimer between the H1 helices of each ARM IV. Detailed observation of the CTNNBL1 structure showed that dimerization of the protein was mediated by combinations of hydrophobic and hydrophilic interactions between two monomers. Several hydrophobic residues in the N-terminal capping region, ARM I, ARM II and ARM III were involved in hydrophobic interactions: Ile86, Leu87 and Ile101 from the N-terminal capping region, Leu147, Gly148 and Gly151 from ARM I, Val183, Ala187, Gly191 and Leu196 from ARM II and Gly236 from ARM III. These hydrophobic interactions do not seem to provide sufficient force to form a dimer, and many hydrophilic residues assist the dimerization of the protein by the formation of hydrogen bonds to hydrophilic residues from the partner polypeptide (Fig. 3*d*). 29 amino acids are involved in dimerization in each monomer, and 963.3 and 963.7 Å^2 of solvent-accessible surface area is buried per monomer, corresponding to 3.8% of the total surface area of each monomer.

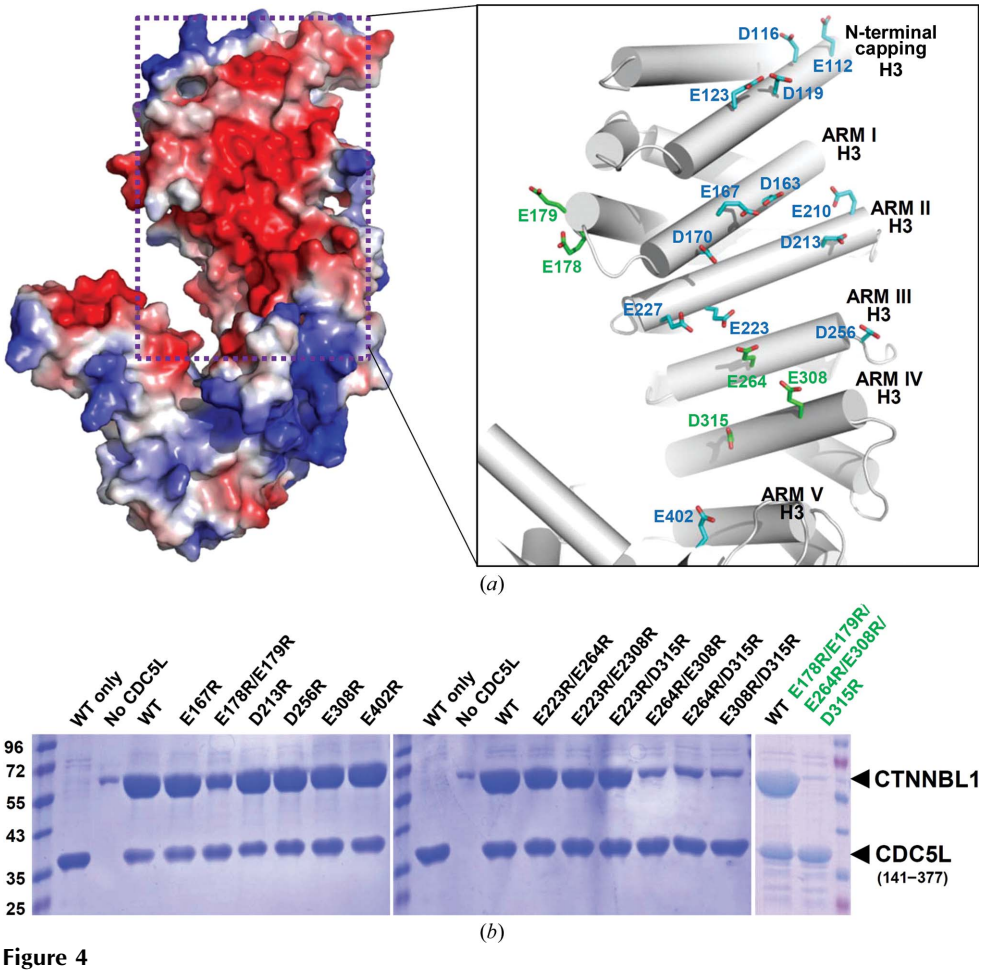


Figure 4
Interaction between CTNNBL1 and CDC5L. (*a*) CDC5L binding site of CTNNBL1. The structure of CTNNBL1 is presented as an electrostatic potential model and the CDC5L binding site is indicated by a dotted rectangle (left). The CDC5L binding site of CTNNBL1 is shown as a cylindrical cartoon representation (right). The negatively charged residues located in the H3 helices of the NAR domain are presented as stick models in cyan and are labelled. Residues whose simultaneous mutation abolishes the binding to CDC5L are distinguished in green. (*b*) CTNNBL1–CDC5L$^{141–377}$ pull-down assay. Recombinant CTNNBL1 protein without a 6×His tag was mixed with CDC5L$^{141–377}$ protein with a 6×His tag at the C-terminus and Ni–NTA chromatography was performed. The eluted proteins were run on SDS–PAGE. Each mutation of CTNNBL1 is labelled at the top of the figure. The molecular weights of the standard size markers are labelled on the left side of the figure in kDa, and the CTNNBL1 and CDC5L$^{141–377}$ proteins are indicated by arrows on the right side of the figure and are labelled.

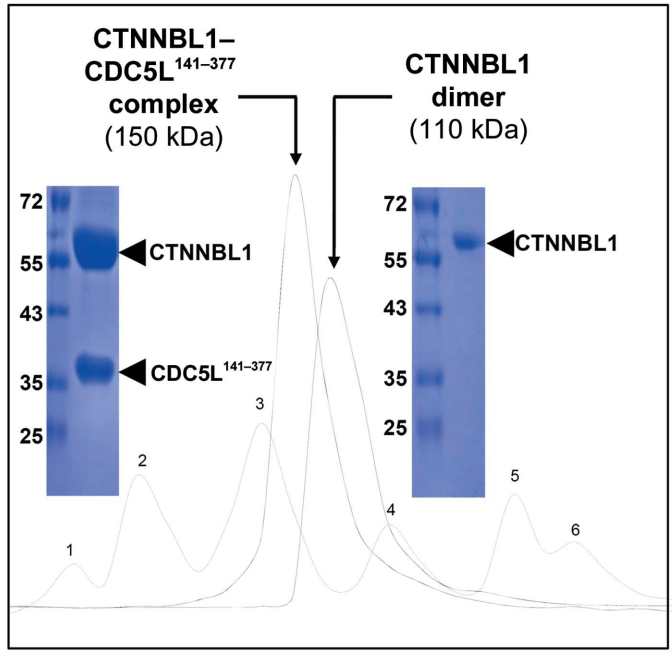
In order to confirm the dimerization of the CTNNBL1 protein and the involvement of hydrophilic interactions in the dimerization contacts, we tested how the addition of NaCl affected the dimerization state of the protein. When size-exclusion chromatography was performed using

protein solutions with various concentrations of NaCl (0, 100, 150, 200, 250 and 300 m*M*), the CTNNBL1 protein still eluted as a stable dimer at NaCl concentrations of up to 200 m*M*. The dimerization of the protein started to break up on addition of 250 m*M* NaCl, and the protein was completely separated into monomers by the addition of 300 m*M* NaCl (Fig. 3*b*). The successful crystallization of CTNNBL1 as a dimer from a crystallization condition that contained 200 m*M* NaCl further confirmed that the protein forms a stable dimer when NaCl is used at a concentration of at least 200 m*M*. The dimerization of CTNNBL1 (amino-acid residues 77–563) was further confirmed using size-exclusion chromatography multi-angle light-scattering (SEC-MALS). CTNNBL1 protein eluted with buffer containing 0, 100 or 200 m*M* NaCl showed a molecular weight of ∼115 kDa, which corresponds to the dimeric state of the protein. The protein eluted with buffer containing 300 or 400 m*M* NaCl eluted with a molecular weight of ∼58 kDa, which corresponds to the monomeric state of the protein (Fig. 3*c*). We also performed SEC-MALS analysis with NLS-deleted full-length CTNNBL1 (amino-acid residues 33–365), and the results showed that the full-length CTNNBL1 protein forms a dimer in buffer containing 0, 100 or 200 m*M* NaCl and is separated into monomers in buffer containing 300 or 400 m*M* NaCl (data not shown). Because the cellular salt concentration is known to be lower than 200 m*M*, we suggest that the CTNNBL1 protein also functions as a dimer *in vivo*. A search for related crystal structures in the Protein Data Bank showed that all ARM-repeat proteins exist as monomers, and CTNNBL1 is therefore to our knowledge the first ARM-repeat protein that functions as a dimer. Considering that CTNNBL1 is a component of the hPrp19–CDC5L complex, the dimerization of the protein might provide important structural insights into the molecular architecture of the hPrp19–CDC5L complex, which will be described later.

3.3. CDC5L binds to the negatively charged patch of CTNNBL1

CDC5L binds CTNNBL1 in order to form the hPrp19–CDC5L complex. It has also been reported that CDC5L contains four highly positively charged NLS regions that are involved in binding to CTNNBL1 (Ganesh *et al.*, 2011). Interestingly, our CTNNBL1 structure showed that the protein forms a highly negatively charged patch at a position similar to the NLS recognition site of importin-α (Fig. 4*a* and Supplementary Fig. S2). The negatively charged patch of CTNNBL1 was formed mainly by glutamate and aspartate residues located in the H3 helices of ARM I, ARM II, ARM III and ARM IV (Fig. 4). The highly negatively charged patch formed in the N-terminal canonical ARM-repeat region led us to speculate that binding between CTNNBL1 and CDC5L might be mediated by charge–charge interactions, as are generally found when importin-α binds to its NLS binding partner.

In order to identify the binding mode between CTNNBL1 and CDC5L, we generated various CTNNBL1 mutants by replacing glutamate and/or aspartate residues with arginines and studied how the charge-replaced mutants affected the formation of the CTNNBL1–CDC5L complex. Firstly, we generated six single or double CTNNBL1 mutants (E167R, E178R/E179R, D213R, D256R, E308R and E402R) and performed pull-down assays using the mutants and recombinant CDC5L protein containing four NLS regions (CDC5L$^{141–377}$; residues 141–377). Our results showed that compared with wild-type CTNNBL1, the five single-point CTNNBL1 mutations (E167R, D213R, D256R, E308R and E402R) did not noticeably affect formation of the CTNNBL1–CDC5L$^{141–377}$ complex, while the E178R/E179R mutation decreased complex formation by approximately 50%, indicating that the binding surface between CTNNBL1 and CDC5L$^{141–377}$ is extensive (Fig. 4*b*). We then selected additional negatively charged residues such as Glu223, Glu264 and Asp315 and generated double-point CTNNBL1 mutants (E223R/E264R, E223R/E308R, E223R/D315R, E264R/



(*a*)

Protein	Band intensity	Calculated MW (kDa)	Band intensity/ MW (kDa)	Ratio
CTNNBL1	12171	56	217	**1**
CDC5L$^{141–377}$	5955	27	221	**1.02**

(*b*)

Figure 5
Stoichiometry of the CTNNBL1–CDC5L$^{141–377}$ complex. (*a*) Size-exclusion chromatography of the CTNNBL1–CDC5L$^{141–377}$ complex and comparison with that of the CTNNBL1 dimer. The eluted proteins at the peak were run on SDS–PAGE. The same standard samples were used as shown in Fig. 3(*b*) and are labelled in kDa. For precise analysis of the molecular weight, the size-exclusion column was calibrated using the standard samples. (*b*) The band intensity was analyzed using *GelQuant.NET* (BiochemLabSolutions.com) and those of each protein were measured using the SDS–PAGE of the CTNNBL1–CDC5L$^{141–377}$ complex presented in (*a*).

E308R, E264/D315R and E308R/D315R). Interestingly, there was a dramatic decrease in complex formation when the E264R/E308R and E308R/D315R mutants were used. Because the three residues Glu264, Glu308 and Asp315 are located in the central regions of the H3 helices of ARM III and ARM IV, we speculated that the CDC5L binding surface of CTNNBL1 is formed by the negatively charged residues located in the H3 helices of the canonical ARM repeats (Fig. 4*a*). To further confirm the requirement of the H3 helices for binding to CDC5L, we generated a five-point CTNNBL1 mutant (E178R/E179R/E264/E308/D315) and performed a pull-down assay with the CDC5L$^{141–377}$ protein. As expected, complex formation between the CTNNBL1 mutant and CDC5L$^{141–377}$ was completely abolished, confirming that the highly negatively charged patch formed by the acidic residues in the H3 helices of the canonical ARM repeats recognizes the highly positively charged NLS of CDC5L (Fig. 4*b*). Based on these results, we conclude that although CTNNBL1 has a completely different function to importin-α, the protein mimics the NLS recognition mode of importin-α for its interaction with CDC5L by using the novel ARM-repeat structure.

3.4. CTNNBL1 binds CDC5L with a 2:2 stoichiometry

The dimerization of CTNNBL1 and the formation of its complex with CDC5L raise the question of the stoichiometry of the CTNNBL1–CDC5L complex, which might provide important information on the molecular architecture of the hPrp19–CDC5L complex. Firstly, we performed size-exclusion chromatography to analyze the molecular weight of the CTNNBL1–CDC5L$^{141–377}$ complex, and the complex eluted at a molecular weight of ∼150 kDa (Fig. 5*a*). Considering that the molecular weight of the CDC5L$^{141–377}$ monomer is 27 kDa, this result indicates that the molecular weight of the eluted CTNNBL1–CDC5L$^{141–377}$ complex is smaller than the calculated molecular weight of a complex with a 2:2 CTNNBL1:CDC5L stoichiometry (∼164 kDa), but is larger than that of a complex with a 2:1 stoichiometry (∼137 kDa). Here, we need to note that CDC5L$^{141–377}$ bound to CTNNBL1 might be located inside the groove formed between the NAR and the CAR domains of CTNNBL1 (Fig. 4*a*), which might lead to a smaller molecular weight of the eluted CTNNBL1–CDC5L$^{141–377}$ complex than the calculated molecular weight, and we suspect that CTNNBL1 and CDC5L might form a complex with 2:2 stoichiometry. We then quantified the band intensities on SDS–PAGE using the eluted complex from size-exclusion chromatography, and the stoichiometry of CTNNBL1:CDC5L$^{141–377}$ was calculated to be close to 1:1 (Fig. 5*b*). Considering that CTNNBL1 exists as a dimer, the result indicates that a CTNNBL1 dimer binds two molecules of CDC5L$^{141–377}$ with a 2:2 stoichiometry. It has been reported that CDC5L forms a homodimer *in vivo* and *in vitro* (Gräub *et al.*, 2008; Grote *et al.*, 2010), and size-exclusion chromatography using CDC5L$^{141–377}$ also supported dimerization of the protein. Taken together, we suggest that one CTNNBL1 dimer might interact with one CDC5L$^{141–377}$ dimer for formation of the CTNNBL1–CDC5L$^{141–377}$ complex.

4. Conclusion

The crystal structure of CTNNBL1 showed that the protein has a novel type of ARM-repeat structure with two distinct properties. Firstly, although the NAR domain contains a canonical ARM-repeat structure similar to those found in other reported ARM-repeat structures such as β-catenin and importin-α, the CAR domain shows an extended and twisted ARM-repeat structure. The unique structural features of the CAR domain cause positioning of the domain proximal to the NAR domain. Secondly, the protein forms a stable dimer; a similar structure has not been reported to date among known ARM-repeat proteins. These structural differences might be explained by the functional difference between CTNNBL1 and other canonical ARM-repeat proteins. As a component of the hPrp19–CDC5L complex, CTNNBL1 seems to adapt the canonical ARM-repeat structure in the N-terminal domain and to form a negatively charged patch for its interaction with CDC5L. Interestingly, the CAR domain does not have a

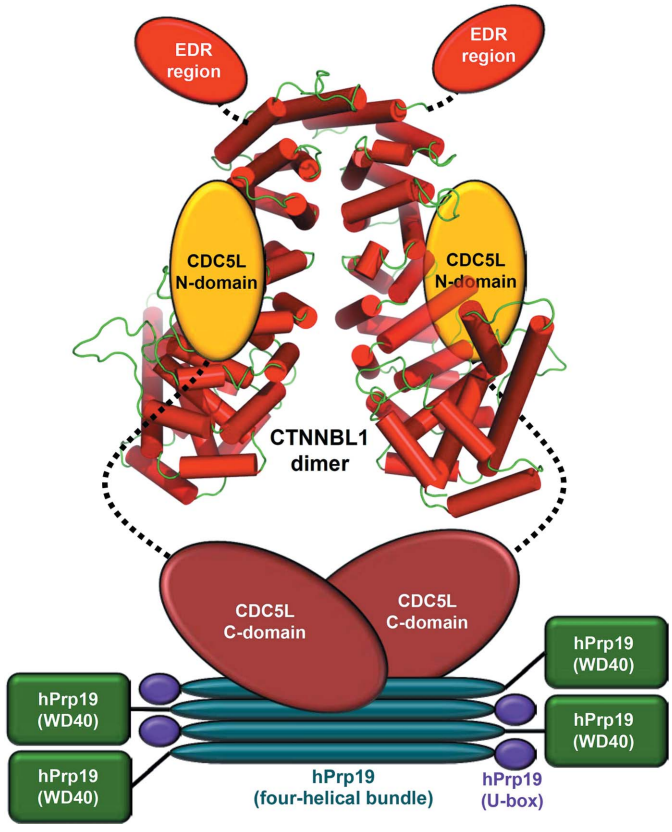


Figure 6
Schematic diagram of the molecular architecture of the hPrp19–CDC5L complex. The identified domains of each of the components of the hPrp19–CDC5L complex are distinguished with different colours and are labelled appropriately. The CTNNBL1 dimer is shown as a cylindrical cartoon representation and the other components of the complex are shown as diagrams. The diagram reflects the dimerization of CTNNBL1 and its interaction with the N-terminal domain of CDC5L with a 2:2 stoichiometry.

negatively charged patch like the NAR domain, but rather shows a broad positively charged surface, which is consistent with the fact that this domain is not involved in interaction with CDC5L. Although we were unable to define the role of the CAR domain from our study, we speculated that this domain might be involved in interaction with other components of the hPrp19–CDC5L complex. The glutamate/aspartate-rich region (EDR region; amino-acid residues 34–76) might provide another binding site for other components of the hPrp19–CDC5L complex. In order to examine the structural features of the EDR region and its interaction with the ARM-repeat structure of CTNNBL1, we also crystallized and determined the structure of NLS-deleted full-length CTNNBL1 (amino-acid residues 33–365). Our results showed no electron-density map for the EDR region (data not shown), indicating that the EDR region does not interact with the ARM-repeat structure but rather interacts with other proteins in the hPrp19–CDC5L complex such as AD002 and HSP73.

Recent studies of the molecular architecture of the hPrp19–CDC5L complex using various methods such as electron microscopy, analytical ultracentrifugation, co-immunoprecipitation and mass spectrophotometry have shown that (i) human Prp19 forms a tetramer (Grote *et al.*, 2010; Ohi *et al.*, 2005; Vander Kooi *et al.*, 2010; Gräub *et al.*, 2008), (ii) human CDC5L forms a homodimer both *in vivo* and *in vitro* (Grote *et al.*, 2010; Gräub *et al.*, 2008) and (iii) human Prp19 interacts with the C-terminal domain of CDC5L with a 4:2 stoichiometry (Smits *et al.*, 2013; Schmidt *et al.*, 2010). Our crystallographic and stoichiometric studies showed that CTNNBL1 forms a dimer and interacts with the N-terminal NLS of CDC5L with a 2:2 stoichiometry. Taken together, we propose that CTNNBL1, CDC5L and Prp19 form a complex with a 2:2:4 stoichiometry and, combined with the relative sizes of the proteins, we present a more detailed schematic diagram of the hPrp19–CDC5L complex (Fig. 6). Further structural studies of the individual components of the hPrp19–CDC5L complex and their complexes are needed to obtain a more detailed molecular architecture as well as the cellular function of the complex.

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIP) (NRF-2009-C1AAA001-2009-0093483) and by the Advanced Biomass R&D Center (ABC) of the Global Frontier Project funded by the MSIP (ABC- 2012-053895), and was also funded by a part of the project titled 'Gyeongbuk Sea Grant Program' funded by the MOF, Korea. J-WA was supported by the NRF-2013-Global PhD Fellowship Program of the Korean Government.

References

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Chan, S.-P., Kao, D.-I., Tsai, W.-Y. & Cheng, S.-C. (2003). *Science*, **302**, 279–282.

Chandra, A., van Maldegem, F., Andrews, S., Neuberger, M. S. & Rada, C. (2013). *Cell Cycle*, **12**, 732–742.

Chen, C.-H., Yu, W.-C., Tsao, T. Y., Wang, L.-Y., Chen, H.-R., Lin, J.-Y., Tsai, W.-Y. & Cheng, S.-C. (2002). *Nucleic Acids Res.* **30**, 1029–1037.

Conticello, S. G., Ganesh, K., Xue, K., Lu, M., Rada, C. & Neuberger, M. S. (2008). *Mol. Cell*, **31**, 474–484.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Ganesh, K., Adam, S., Taylor, B., Simpson, P., Rada, C. & Neuberger, M. (2011). *J. Biol. Chem.* **286**, 17091–17102.

Gräub, R., Lancero, H., Pedersen, A., Chu, M., Padmanabhan, K., Xu, X.-Q., Spitz, P., Chalkley, R., Burlingame, A. L., Stokoe, D. & Bernstein, H. S. (2008). *Cell Cycle*, **7**, 1795–1803.

Grote, M., Wolf, E., Will, C. L., Lemm, I., Agafonov, D. E., Schomburg, A., Fischle, W., Urlaub, H. & Lührmann, R. (2010). *Mol. Cell. Biol.* **30**, 2105–2119.

Guo, Z., Karunatilaka, K. S. & Rueda, D. (2009). *Nature Struct. Mol. Biol.* **16**, 1154–1159.

Hatakeyama, S., Yada, M., Matsumoto, M., Ishida, N. & Nakayama, K. I. (2001). *J. Biol. Chem.* **276**, 33111–33120.

Hu, Y., Ericsson, I., Torseth, K., Methot, S. P., Sundheim, O., Liabakk, N. B., Slupphaug, G., Di Noia, J. M., Krokan, H. E. & Kavli, B. (2013). *J. Mol. Biol.* **425**, 424–443.

Huang, X., Wang, G., Wu, Y. & Du, Z. (2013). *Acta Cryst.* D**69**, 1598–1608.

Jabbour, L., Welter, J. F., Kollar, J. & Hering, T. M. (2003). *Genomics*, **81**, 292–303.

Krämer, A. (1996). *Annu. Rev. Biochem.* **65**, 367–409.

Kwon, S.-Y., Choi, Y.-J., Kang, T.-H., Lee, K.-H., Cha, S.-S., Kim, G.-H., Lee, H.-S., Kim, K.-T. & Kim, K.-J. (2005). *Plasmid*, **53**, 274–282.

Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M. & Lührmann, R. (2002). *Science*, **298**, 2205–2208.

Makarova, O. V., Makarov, E. M., Urlaub, H., Will, C. L., Gentzel, M., Wilm, M. & Lührmann, R. (2004). *EMBO J.* **23**, 2381–2391.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Murray, H. L. & Jarrell, K. A. (1999). *Cell*, **96**, 599–602.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Ohi, M. D. & Gould, K. L. (2002). *RNA*, **8**, 798–815.

Ohi, M. D., Vander Kooi, C. W., Rosenberg, J. A., Chazin, W. J. & Gould, K. L. (2003). *Nature Struct. Biol.* **10**, 250–255.

Ohi, M. D., Vander Kooi, C. W., Rosenberg, J. A., Ren, L., Hirsch, J. P., Chazin, W. J., Walz, T. & Gould, K. L. (2005). *Mol. Cell. Biol.* **25**, 451–460.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Papassotiropoulos, A. *et al.* (2013). *Mol. Psychiatr.* **18**, 255–263.

Schmidt, C., Lenz, C., Grote, M., Lührmann, R. & Urlaub, H. (2010). *Anal. Chem.* **82**, 2784–2796.

Smits, A. H., Jansen, P. W. T. C., Poser, I., Hyman, A. A. & Vermeulen, M. (2013). *Nucleic Acids Res.* **41**, e28.

Staley, J. P. & Guthrie, C. (1998). *Cell*, **92**, 315–326.

Tarn, W.-Y., Hsu, C.-H., Huang, K.-T., Chen, H.-R., Kao, H.-Y., Lee, K.-R. & Cheng, S.-C. (1994). *EMBO J.* **13**, 2421–2431.

Tarn, W.-Y., Lee, K.-R. & Cheng, S.-C. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 10821–10825.

Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.

Vander Kooi, C. W., Ren, L., Xu, P., Ohi, M. D., Gould, K. L. & Chazin, W. J. (2010). *Structure*, **18**, 584–593.

Wahl, M. C., Will, C. L. & Luhrmann, R. (2009). *Cell*, **136**, 701–718.

Will, C. L. & Lührmann, R. (1997). *Curr. Opin. Cell Biol.* **9**, 320–328.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.