Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structure of full-length human CTNNBL1 reveals a distinct member of the armadillo-repeat protein family

Xiaolan Huang,[a] Guan Wang,[b] Yuhong Wu[b] and Zhihua Du[b]*

[a]Department of Computer Science, Southern Illinois University, 1000 Faner Drive, Carbondale, IL 62901, USA, and [b]Department of Chemistry and Biochemistry, Southern Illinois University, 1245 Lincoln Drive, Carbondale, IL 62901, USA

Correspondence e-mail: zdu@chem.siu.edu

Received 28 January 2013
Accepted 25 April 2013

PDB References: CTNNBL1, full length, 4hm9; fragment, 4hnm

Catenin-β-like protein 1 (CTNNBL1) is a highly conserved protein with multiple functions, one of which is to act as an interaction partner of the antibody-diversification enzyme activation-induced cytidine deaminase (AID) for its nuclear import and subnuclear trafficking. Here, the crystal structure of full-length human CTNNBL1 is reported. The protein contains six armadillo (ARM) repeats that pack into a superhelical ARM domain. This ARM domain is unique within the ARM protein family owing to the presence of several unusual structural features. Moreover, CTNNBL1 contains significant and novel non-ARM structures flanking both ends of the central ARM domain. A strong continuous hydrophobic core runs through the whole structure, indicating that the ARM and non-ARM structures fold together to form an integral structure. This structure defines a highly restrictive and discriminatory protein-binding groove that is not observed in other ARM proteins. The presence of a cluster of histidine residues in the groove implies a pH-sensitive histidine-mediated mechanism that may regulate protein binding activity. The many unique structural features of CTNNBL1 establish it as a distinct member of the ARM protein family. The structure provides critical insights into the molecular interactions between CTNNBL1 and its protein partners, especially AID.

1. Introduction

Human catenin-β-like protein 1 (CTNNBL1; 563 amino acids; calculated MW 65.2 kDa) was named after β-catenin, the prototypical member of the armadillo-repeat protein family, because it was predicted that CTNNBL1 contained armadillo (ARM) repeats (Jabbour *et al.*, 2003). ARM repeats were first characterized in the *Drosophila* segment polarity protein armadillo, the ortholog of mammalian β-catenin (Peifer *et al.*, 1994). A typical ARM repeat contains ~42 residues forming three α-helices (H1, H2 and H3) that are arranged in a roughly triangular shape. Tandemly arranged ARM repeats form a superhelical ARM domain, with its inner (concave) face defined by the H3 helices and its outer (convex) face defined by the shorter helices H1 and H2. Most ARM domains bind partner proteins using the concave face (Huber *et al.*, 1997; Xu & Kimelman, 2007; Tewari *et al.*, 2010).

The amino-acid sequence of CTNNBL1 is highly conserved across widely divergent species (Supplementary Fig. S1[1]), indicating that this protein carries out certain essential and fundamental biological functions in eukaryotic cells.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: MN5029). Services for accessing this material are described at the back of the journal.

In about a dozen genome-wide association studies (GWAS), single-nucleotide polymorphisms (SNPs) in the *CTNNBL1* gene have been associated with obesity (Liu *et al.*, 2008; Cho *et al.*, 2009; Andreasen *et al.*, 2009; Yin *et al.*, 2012), colorectal cancer (Huhn *et al.*, 2011), bone mineral density (BMD; Ichikawa *et al.*, 2010; van Meurs *et al.*, 2008; Richards *et al.*, 2008; Styrkarsdottir *et al.*, 2008, 2010) and episodic memory performance (Papassotiropoulos *et al.*, 2013). Because GWAS are able to identify novel genes without *a priori* knowledge of gene functions, these findings provide particularly strong evidence for the important roles of CTNNBL1 in normal biological processes as well as in complex diseases and conditions.

At the molecular level, protein-association studies have identified several binding partners of CTNNBL1. A direct interaction of CTNNBL1 with the antibody-diversifying enzyme AID (activation-induced deaminase) has been established (Conticello *et al.*, 2008; Ganesh *et al.*, 2011; Hu *et al.*, 2013). AID deaminates deoxycytidine residues in immunoglobulin genes to generate antibody diversification. Inactivation of CTNNBL1 and disruption of the CTNNBL1–AID interaction severely diminished antibody diversity (Conticello *et al.*, 2008; Ganesh *et al.*, 2011). Other CTNNBL1-binding proteins include CDC5L (cell division cycle 5-like protein), which is a component of the Prp19-containing RNA-splicing complexes, Prp31, which is another U4/U6·U5 tri-snRNP-associated splicing factor, and importin-α, which is an ARM protein in the canonical nuclear import pathway (Ganesh *et al.*, 2011).

It has been shown that the interactions of CTNNBL1 with CDC5L and Prp31 are mediated by binding of the nuclear localization signals (NLSs) of CDC5L and Prp31 to CTNNBL1 (Ganesh *et al.*, 2011). CTNNBL1 exhibited distinctive NLS-recognition specificities compared with the canonical nuclear import protein importin-α. While both CTNNBL1 and importin-α can bind the NLS of CDC5L, only CTNNBL1 can bind the NLS of Prp31 and only importin-α can bind the classic NLS of SV40 T antigen (Ganesh *et al.*, 2011). AID does not contain a linear NLS. Its nuclear import depends on a so-called conformational NLS that is defined by a number of basic (arginine/lysine) residues (Arg8, Arg9, Lys16, Arg19, Lys22, Arg24, Arg50, Arg112 and Arg171; Conticello *et al.*, 2008; Ganesh *et al.*, 2011; Hu *et al.*, 2013). It has also been found that another stretch of residues (A_{39}TSFS) is important for interaction with CTNNBL1 and antibody diversification (Ganesh *et al.*, 2011). These observations suggest that CTNNBL1 functions as a novel NLS-binding protein that recognizes selective cargo proteins.

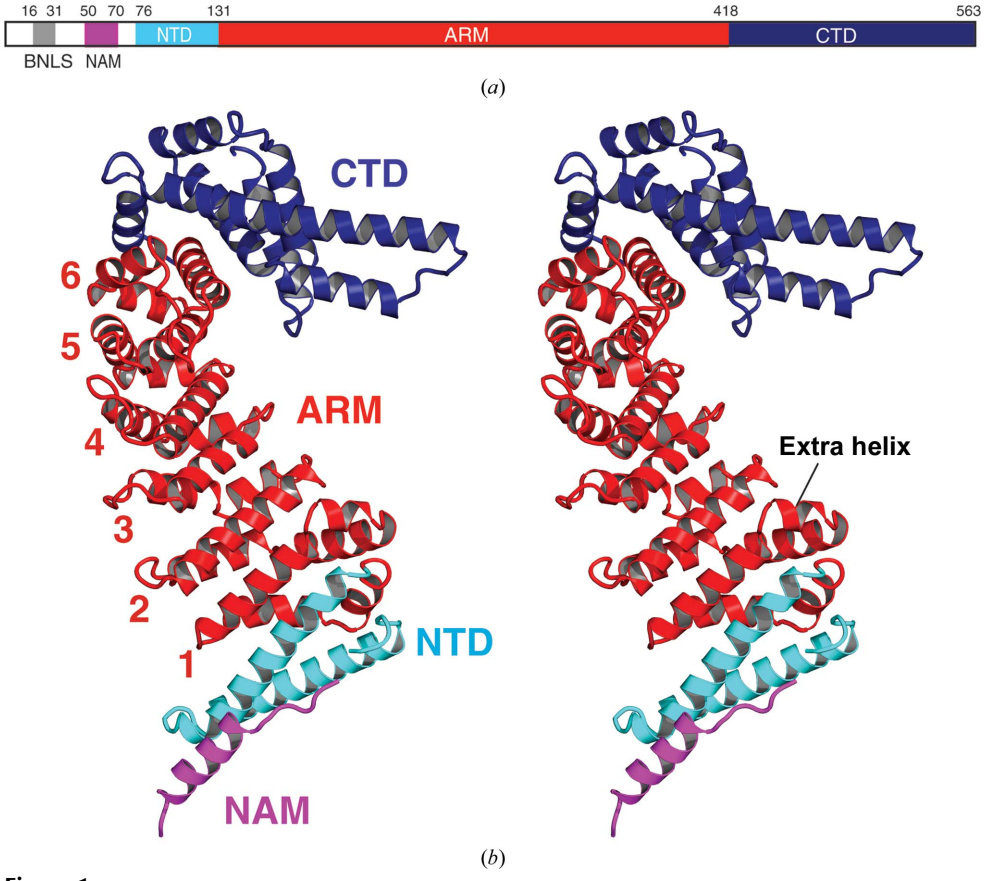

Figure 1
Overall structure of CTNNBL1. (*a*) Schematic representation of CTNNBL1. BNLS, bipartite nuclear localization signal; NAM, N-terminal anchoring motif; NTD, N-terminal domain; ARM, armadillo domain; CTD, C-terminal domain. (*b*) Stereoview of the overall structure of full-length human CTNNBL1. The NAM, NTD, ARM and CTD domains are colored magenta, cyan, red and blue, respectively. The numbers 1–6 indicate the six ARM repeats within the ARM domain.

CTNNBL1 itself contains an N-terminal bipartite nuclear localization signal (BNLS; K_{16}RPRDDEEEEQKMRRK$_{31}$; Fig. 1*a*). This BNLS is recognized by importin-α and mediates nuclear localization of CTNNBL1 and a reporter protein (Ganesh *et al.*, 2011). It is hypothesized that the interaction between CTNNBL1 and importin-α provides a means for the cargo-bound CTNNBL1 to enter the canonical nuclear-import pathway mediated by importin-α/β (Ganesh *et al.*, 2011).

Most known ARM proteins are multifunctional. This may also be the case for CTNNBL1. Using RNAi screening, the *CTNNBL1* gene was identified as a putative regulator of the canonical Wnt signaling pathway (mediated by β-catenin), acting upstream of or in parallel to β-catenin (Huhn *et al.*, 2011). In another study, CTNNBL1 was identified by a yeast two-hybrid system as a binding partner of osteopontin, which has established roles in

bone formation, cell adhesion and tumorigenesis (Rual *et al.*, 2005).

The results of sequence analysis, GWAS and protein-association studies not only indicate that CTNNBL1 has important and unique functions but also begin to reveal the molecular mechanisms of certain functions of CTNNBL1. However, no structure of CTNNBL1 is available to date. This lack of knowledge has hindered progress towards a better understanding of the functions of CTNNBL1.

Here, we report the crystal structures of full-length human CTNNBL1 and of an N-terminally truncated fragment (~488 amino acids) at resolutions of 3.1 and 2.9 Å, respectively. The structures show that CTNNBL1 harbors a central ARM domain consisting of six ARM repeats with several unusual features. Significant non-ARM domains are present flanking both the N-terminal and C-terminal ends of the ARM domain. The ~55-amino-acid N-terminal domain (NTD) consists of two antiparallel helices. It caps the otherwise open-ended putative protein-binding groove formed by the superhelical concave face of the ARM domain. The cap is further augmented by an N-terminal helical motif (NAM; N-terminal anchoring motif) that packs against the second helix of the NTD. The C-terminal domain (CTD, consisting of the last 145 amino acids) adopts a novel protein fold. About one-third of the CTD structures participate in tight packing against the ARM domain. The remaining structures lean towards the concave face of the ARM domain, forming a roughly V-shaped deep cleft. A strong and continuous hydrophobic core runs through the whole structure, indicating that the three domains (NTD, ARM and CTD) fold together to form an integral and relatively rigid structure. This structure may define a more specific and discriminatory protein-binding groove than those of other ARM proteins. The presence of four histidine residues in the groove implies a pH-sensitive histidine-mediated control mechanism that regulates the protein-binding activity of CTNNBL1. These unique structural features distinguish CTNNBL1 as a distinct member of the ARM-repeat protein family. The structure provides crucial insights into the molecular mechanisms of the functions of CTNNBL1, especially its interaction with AID.

2. Materials and methods

2.1. Protein-sample preparation and crystallization

The full-length human CTNNBL1 protein was expressed in *Escherichia coli* as a fusion protein with the HaloTag (Los *et al.*, 2008). To facilitate purification and tag removal, a His_6 tag and an eight-amino-acid recognition sequence for HRV 3C protease were engineered between the HaloTag and the target protein. Protein expression was carried out using NiCo21 (DE3) *E. coli* cells (New England Biolabs). The expressed protein was purified using NTA affinity resin. The fusion tag was then cleaved using HRV 3C protease. The cleaved tag was separated from the target protein by passing the protein solution through a small column packed with NTA resin. The target protein was further purified by SP Sepharose ion-exchange chromatography. The fractions (eluted with a 50–1000 m*M* NaCl gradient in 25 m*M* Tris buffer pH 8.0) containing pure protein were combined and concentrated to ~10 mg ml^{-1}.

To obtain SeMet-labelled protein, the bacteria were grown in M9 minimal medium until they reached an OD_{600} of 0.6–0.8, whereupon leucine, isoleucine, lysine, phenylalanine, threonine and valine were added to the culture (50–100 mg l^{-1}) to inhibit methionine biosynthesis. After 30 min, L-selenomethionine (50 mg l^{-1}) was added, followed by IPTG to induce the expression of SeMet-labelled protein at 285 K overnight.

The full-length protein degraded over a period of 3–4 months, yielding a stable fragment (referred to here as CTNNBL1^{SF}) that is about 10 kDa smaller than the full-length protein (Supplementary Fig. S2*a*). We suspected that the degradation was owing to the presence of residual protease activity in the preparation. After the full-length protein had all degraded to CTNNBL1^{SF}, SP Sepharose ion-exchange chromatography was used to repurify CTNNBL1^{SF}.

Crystals of full-length CTNNBL1 (in 25 m*M* Tris buffer pH 8.0, 50 m*M* NaCl) were obtained at 285 K by hanging-drop vapor diffusion against 0.5 ml well solution consisting of 15% PEG 2000, 50 m*M* Tris–HCl pH 8.0, 50 m*M* KCl, 10%(*v*/*v*) glycerol. Crystals of CTNNBL1^{SF} were obtained using a different well buffer consisting of 12% PEG 600, 50 m*M* Tris–HCl pH 8.0, 50 m*M* KBr, 12%(*v*/*v*) glycerol.

2.2. Data collection, data processing and structure determination

Data collection was carried out on beamlines 21ID-D, 21ID-F and 21ID-G of LS-CAT at the Advanced Photon Source, Argonne National Laboratory. Data were processed, integrated and scaled with the programs *MOSFLM* (Battye *et al.*, 2011) and *SCALA* in *CCP*4 (Winn *et al.*, 2011). The structure of CTNNBL1^{SF} was solved by SAD using a data set collected at the peak wavelength of Se (0.97872 Å) from a single crystal of SeMet-labelled protein. Location of the Se atoms and building of the initial structure were performed using the *PHENIX* package (Adams *et al.*, 2011). Interactive model building was carried out with *Coot* (Emsley *et al.*, 2010). The structure was refined using *PHENIX*. The structure of full-length CTNNBL1 was solved by molecular replacement using the program *Phaser* (Storoni *et al.*, 2004) as implemented in the *PHENIX* package. To avoid bias of the relative orientation between the ARM domain and the CTD, the structure of CTNNBL1^{SF} was split into two substructures containing the NTD-ARM and the CTD, respectively. These two structures were used as the search models for molecular replacement. All plausible space groups of the *P*222 family were tried. A single solution was found in space group $P2_12_12_1$ with TFZ = 22.8 and LLG = 4806, indicating that the solution was absolutely correct. Using *Coot*, interactive model building was carried out to fit those electron densities that were not observed in the CTNNBL1^{SF} structure. Structure refinement was carried out in *PHENIX*. In *PHENIX* refinement, the initial isotropic *B*

factors were all set to 20 Å^2 and individual B factors were refined. No restraints were used in the refinement except that weightings between X-ray and stereochemistry/ADP terms were optimized using the default values. No H atoms were included in the refinement. The input coordinates for refinement did not contain water molecules. Water molecules were automatically updated during the refinement process. However, no water molecules were added during the refinement process owing to the relatively low resolution. The refinement was carried out in real space using simulated annealing (Cartesian, 5000–300 K). Structure-determination statistics are shown in Table 1. The figures were prepared with the program *PyMOL* (v.1.5.0.4; Schrödinger).

Atomic coordinates and diffraction data for full-length CTNNBL1 and for CTNNBL1^{SF} have been deposited in the Protein Data Bank with accession codes 4hm9 and 4hnm, respectively.

Table 1
Data-collection and refinement statistics for full-length CTNNBL1 and CTNNBL1^{SF}.

Values in parentheses are for the highest resolution shell.

	Full-length CTNNBL1	CTNNBL1^{SF}
Data collection		
Space group	$P2_12_12_1$	$P4_12_12$
Unit-cell parameters (Å)	a = 48.99, b = 116.36, c = 137.4	a = b = 95.00, c = 172.63
Resolution (Å)	42.6–3.10 (3.27–3.10)	49.2–2.90 (3.06–2.90)
Completeness (%)	100 (100)	99.9 (99.8)
R_{merge} (%)	12.6 (53.0)	7.5 (51.0)
$\langle I/\sigma(I)\rangle$	11.7 (3.8)	18.3 (4.3)
Multiplicity	7.1 (7.1)	10.6 (10.4)
Refinement		
Resolution (Å)	42.6–3.10	49.2–2.90
No. of reflections	14885	18213
R_{work}/R_{free} (%)	17.42/22.53	21.53/27.12
No. of atoms		
Protein	4117	3890
Water	0	0
Protein B factor (Å^2)	29.8	35.0
R.m.s.d., bonds (Å)	0.003	0.003
R.m.s.d., angles (°)	0.597	0.690
Ramachandran plot		
Favored (%)	97.8	95.0
Outliers (%)	0.20	0.21

3. Results

3.1. Overall structure

The purified and concentrated full-length CTNNBL1 protein is very stable. It can be stored at 277 K for 2–3 months without degradation (as judged by SDS–PAGE analysis). However, degradation did occur over longer periods of time. Different preparations of the full-length protein consistently yielded the same major degradation product that is about 10 kDa smaller than the full-length protein (Supplementary Fig. S2*a*). This fragment of CTNNBL1 did not degrade further (over almost one year to date), indicating that it harbors a very stable structure. This stable fragment is referred to as CTNNBL1^{SF}. Both the full-length CTNNBL1 and CTNNBL1^{SF} proteins yielded crystals that diffracted to decent resolution.

We first solved the structure of CTNNBL1^{SF} by SAD phasing using data collected from an SeMet-labelled protein crystal. The structure of full-length CTNNBL1 was then solved by molecular replacement using the CTNNBL1^{SF} structure as the search model. The presence of full-length CTNNBL1 in the crystals was verified by SDS–PAGE analysis of washed and dissolved crystals (Supplementary Fig. S2*b*). The crystals of CTNNBL1^{SF} and of full-length CTNNBL1 belonged to different space groups. Both crystals contained one molecule in the asymmetric unit. Table 1 shows the statistics of data collection, structure determination and refinement.

The structure of CTNNBL1^{SF} contained residues 75–563, with the exception of five residues (amino acids 105–109) in a loop connecting the two helices of the NTD, for which there was no electron density. The structure of full-length CTNNBL1 contained

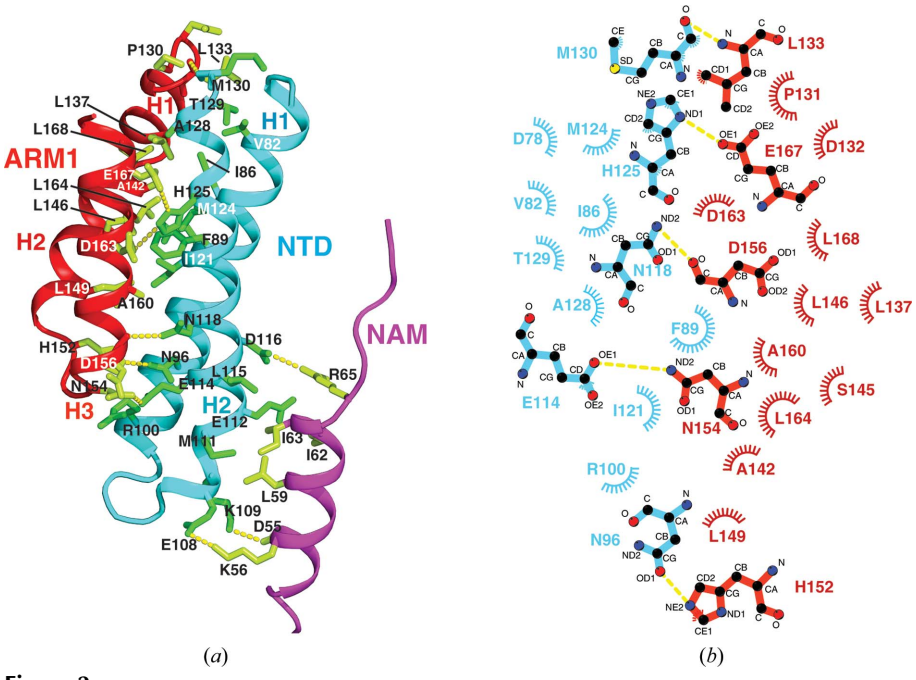

Figure 2
Structure of the N-terminal region of CTNNBL1. (*a*) Structure of the NTD–ARM1 and NTD–NAM interdomain interfaces. The NAM, the NTD and ARM1 are colored magenta, cyan and red, respectively. Residues involved in interdomain hydrophobic and/or hydrogen-bonding interactions are shown as stick representations in different shades of green for the different domains. Hydrogen bonds are represented as yellow dashed lines. (*b*) *LigPlot* presentation of the interactions between the NTD (cyan) and ARM1 (red). Hydrogen bonds are designated by yellow dashed lines. Hydrophobic interactions are represented as starbursts.

residues 50–70 and 74–563. Electron density was not observed for residues 1–49 and 71–73, indicating that these N-terminal sequences are flexible. Because the structure of CTNNBL1^{SF} is very similar to the corresponding structure within the full-length protein (r.m.s.d. of 0.917 Å for 3890 common atoms), the following presentation will mainly focus on the full-length protein. Differences between the two structures will be addressed in the appropriate context.

As expected, CTNNBL1 contains a central ARM domain (Fig. 1*b*). However, the number of ARM repeats (six) is lower than previously estimated (Jabbour *et al.*, 2003). Unusual structural features are also present within the ARM domain (see below for details). The six repeats pack together to form a superhelical structure.

Unlike any of the previously characterized ARM proteins, CTNNBL1 contains significant non-ARM structures flanking the ARM domain. The NTD (amino acids 76–131) contains two antiparallel α-helices that pack tightly against ARM1 (Figs. 1*b* and 2*a*). N-terminal to the NTD, the NAM (amino acids 52–65) forms a three-turn helix that packs against the N-terminal portion of helix H2 of the NTD (Fig. 2*a*). The acidic linker (D_{66}GEEEEEEEE) between the NAM and the NTD is highly flexible, as indicated by missing electron density and higher *B*-factor values. This linker is where the spontaneous degradation that produces CTNNBL1^{SF} occurs. The CTD (amino acids 418–563) adopts an all-helical structure consisting of seven α-helices (Fig. 3*a*). One end of the CTD packs tightly against ARM6 (Figs. 3*b* and 3*c*). The other portions of the CTD lean towards the concave face of the ARM domain. The peculiar relative structural relationship between the CTD and the ARM domain creates a deep roughly V-shaped cleft with one side defined by the CTD and the other side by the H3 helices of ARM3–ARM6 in the ARM domain (Fig. 1*b*).

The interdomain molecular interactions at the NTD–ARM1 and CTD–ARM6 interfaces are comparable to (if not stronger than) the interactions present at the ARM–ARM interfaces (see below for details). The three domains (NTD, ARM and CTD) are most likely to fold together to form an integral structure.

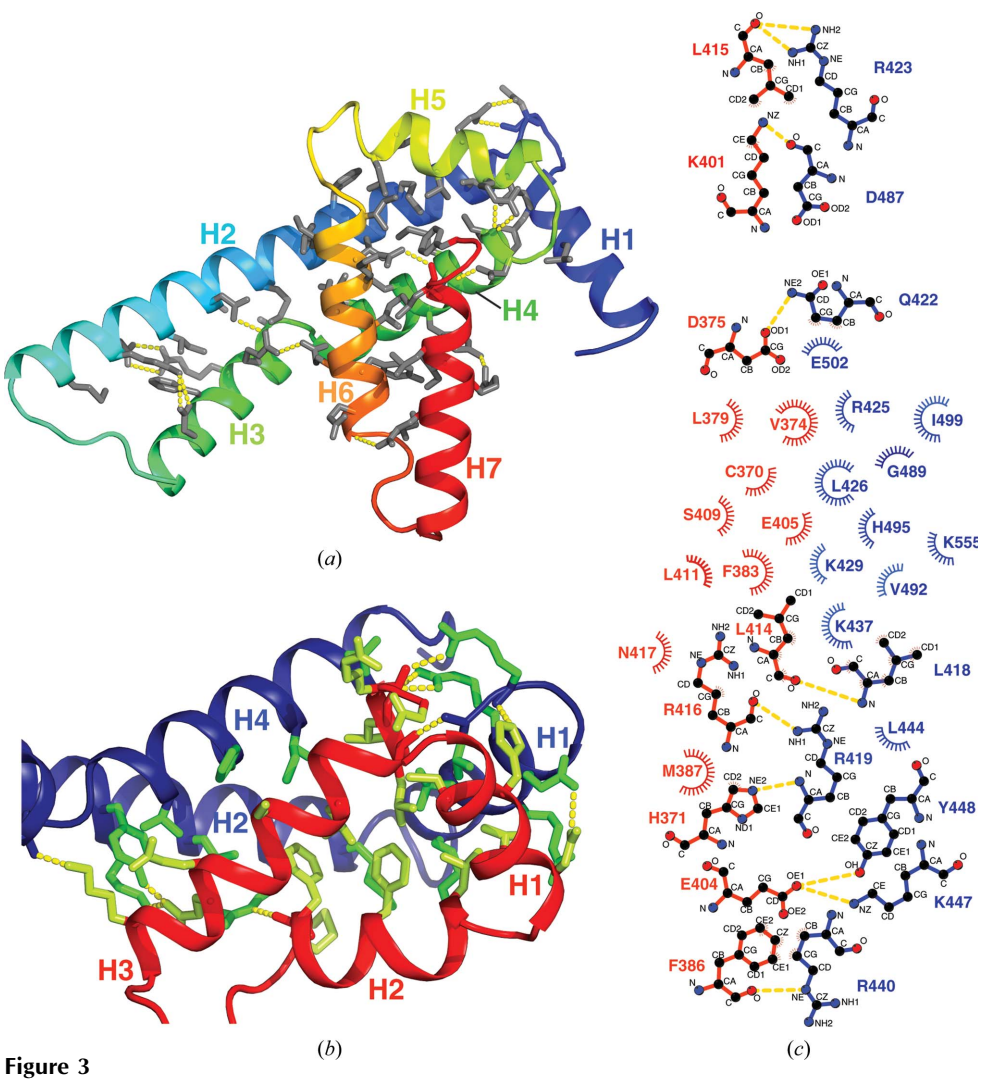

Figure 3
Structure of the C-terminal domain (CTD) and its interaction with the ARM domain. (*a*) Structure of the CTD shown in a cartoon representation color-ramped from the N-terminus in blue to the C-terminus in red. H1–H7 indicate the seven helices of the CTD. Residues that form the hydrophobic core of the domain and/or participate in inter-helix hydrogen bonds are shown as stick representations in gray. Inter-helix hydrogen bonds are represented as yellow dashed lines. (*b*) CTD–ARM6 interaction. CTD and ARM6 are colored blue and red, respectively. Residues involved in interdomain hydrophobic contacts and/or hydrogen bonds are shown as stick representations in different shades of green for the different domains. Hydrogen bonds are represented as yellow dashed lines. (*c*) *LigPlot* presentation of the interactions between CTD (blue) and ARM6 (red). Hydrogen bonds are indicated by yellow dashed lines. Hydrophobic interactions are represented as starbursts.

3.2. Structures and interactions of the terminal domains

Packing of the NTD against ARM1 is predominantly mediated by hydrophobic and hydrogen-bonding interactions. As shown in Figs. 2(*a*) and 2(*b*), residues involved in hydrophobic and/or hydrogen-bonding interactions include those from NTD H1 (Asp78, Val82, Ile86, Phe89, Asn96 and Arg100) and H2 (Glu114, Asn118, Ile121, Met124, His125, Ala128, Thr129 and Met130) and ARM1 H1 (Pro131, Asp132, Leu133 and Leu137), H2 (Ala142, Ser145, Leu146, Leu149 and His152) and H3 (Asn154, Asp156, Ala160, Asp163, Leu164, Glu167 and Leu168).

The outward-facing solvent-exposed surfaces of the H1 and the H2 helices of the NTD are dominated by residues with hydrophilic side chains. At the N-terminal end of H2, the two hydrophobic side chains of Met111 and Leu115 are outward-facing. These two side chains and the aliphatic portion of the Glu112 side chain engage in hydrophobic interactions with the side chains of Leu59, Ile62 and Ile63 from the NAM (Fig. 2*a*). Several hydrogen bonds and salt bridges further strengthen the interaction between the NAM and the NTD (Fig. 2*a*). The NAM–NTD interaction seems to stabilize the structure of the NTD. In the structure of full-length CTNNBL1 all of the residues of the NTD are well defined, including the linker residues between helices H1 and H2. In the structure of CTNNBL1^{SF} the linker region of the NTD is ill-defined and several residues of the connecting loop are missing.

The CTD contains seven helices: H1–H7. The longest nine-turn helix H2 is slightly kinked after the third turn (Fig. 3*a*). A three-dimensional structure-similarity search using the *DALI* server (Holm & Rosenström, 2010) did not identify any significant match for the CTD. This structure therefore represents a new protein fold.

Hydrophobic interactions play a crucial role in the folding of the CTD. Its hydrophobic core is formed by the packing of a large number of hydrophobic residues from helices H1 (Leu427 and Phe430), H2 (Val438, Met442, Phe446, Leu449, Met452, Ala455, Ile459 and Met466), H3 (Phe481 and Ala488), H4 (Leu493, Cys497, the aromatic ring of Tyr498, M5et00, Ala501, Ile503 and Cys504), H5 (Ile511, Val515, Ile518 and Leu519), H6 (Ile526, Ile528, Val529, Ile532, Ile533, the aromatic ring of Tyr536 and Ile540) and H7 (Ile556, Leu557, Leu560 and Phe563). The structure is further stabilized by 14 interhelix hydrogen bonds (Fig. 3*a*). All of the connecting loops except for that between H6 and H7 are short and constrained.

Superimposition of the structures of full-length CTNNBL1 and CTNNBL1^{SF} (Supplementary Fig. S3) reveals a small but noticeable difference between the two structures in the region of CTD helix H3 and the C-terminal portion of H2. In the structure of CTNNBL1^{SF} the kink in helix H2 is less pronounced. Although the difference is most likely owing to different crystal packings, its existence nonetheless indicates some degree of plasticity for the protruding parts of the CTD.

Packing of the CTD against ARM6 is mainly mediated by helices H1, H4 and the N-terminal portion of H2. Extensive interdomain hydrophobic and hydrogen-bonding interactions (ten) are present at this CTD–ARM6 interface. Residues involved in hydrophobic and hydrogen-bonding interactions include those from CTD H1 (Leu418, Arg419, Gln422, Arg423, Arg425, Leu426 and Lys429), H2 (Lys437, Arg440, Leu444, Lys447 and Tyr448) and H4 (Asp487, Gly489, Val492, His495, Ile499 and Glu502), as well as those from ARM6 H1 (Cys370, His371, Val374 and Asp375), H2 (Leu379, Phe383, Phe386 and Met387) and H3 (Lys401, Glu404, Glu405, Ser409, Leu411, Leu414, Leu415, Arg416 and Asn417) (Figs. 3*b* and 3*c*). Several residues from CTD helices H3 and H7 and the linker between H6 and H7 also participate in six hydrogen-bonding/electrostatic/hydrophobic interactions with ARM6 (not shown).

The CTD helices H3, H6 and H7, as well as the connecting loop (amino acids 541–545) between helices H6 and H7, lean towards the concave face of the ARM domain, forming a roughly V-shaped deep cleft. The residues participating in the NTD–ARM1 and CTD–ARM6 interdomain interactions are generally highly conserved (Supplementary Fig. S1). A strong and continuous hydrophobic core extends through the structure, indicating the presence of an integral structure encompassing all three domains which corresponds to the very stable CTNNBL1^{SF}.

3.3. Structure of the ARM domain

The ARM domain contains six repeats, each with three helices: H1, H2 and H3. The six repeats pack together to form a superhelical structure with its concave face defined by the H3 helices (Fig. 4*a*).

Several unusual features are observed in the ARM domain. There is an extra helix between ARM1 H3 and ARM2 H1 (Fig. 4*a*). The sequence of this extra helix is highly conserved

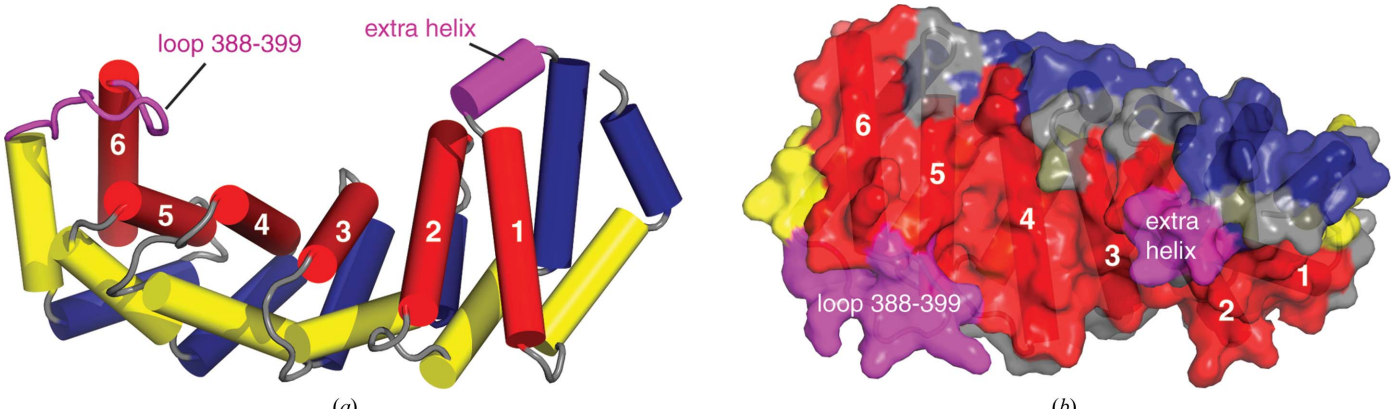

Figure 4
Structure of the ARM domain. (*a*) A cartoon representation of the ARM domain. The helices are shown as cylinders. The three helices H1, H2 and H3 in each of the six ARM repeats are colored blue, yellow and red, respectively. The H2–H3 connecting loop in ARM6 (amino acids 388–399) and the extra helix between ARM1 H3 and ARM2 H1 are colored magenta. (*b*) Molecular-surface rendering with 20% transparency. The view into the concave surface is rotated about 90° from that in (*a*). The numbers 1–6 in both figures indicate the H3 helices of the corresponding ARM repeats.

(Supplementary Fig. S1), suggesting that such a helix is a common feature in the ARM domain of CTNNBL1. Another unusual feature is exhibited by the relative orientation between ARM4 and ARM5. In a typical ARM domain, two adjacent ARM repeats show a twist of ~30°; packing of multiple twisted ARM repeats results in a right-handed superhelical structure of the ARM domain. The ARM domain of CTNNBL1 also adopts an overall right-handed superhelical structure. However, there is no twist between ARM4 and ARM5. The H3 helices of ARM4 and ARM5 are coplanar (Fig. 4*a*). Correspondingly, the groove surface area defined by these two helices is relatively level compared with other groove surface areas, which are typically slanted owing to the superhelical twist (Fig. 4*b*). More interestingly, the bottom edge of this portion of the groove (near the N-terminal ends of the H3 helices of ARM4 and ARM5) is blocked by a ridge formed by the loop connecting the H2 and H3 helices of ARM6 (amino acids 388–399). The position of this loop is very unusual. To the best of our knowledge, this is the first structure that shows a well defined H2–H3 connecting loop that is located on the concave face of the ARM domain. Except for a partially defined loop in β-catenin ARM10 (Huber *et al.*, 1997), all of the H2–H3 connecting loops are found on the convex faces of the ARM domains (for examples, see Fig. 4*a*).

It is clear from Fig. 4(*b*) that all of these unusual structural features are directly involved in the definition of the protein-binding groove. With these unusual features, the ARM domain of CTNNBL1 is unique in the ARM protein family.

3.4. Structure of the putative protein-binding groove

ARM proteins use the concave face of the ARM domain to bind their partner proteins (Huber *et al.*, 1997; Xing *et al.*, 2004). As a member of the ARM protein family, CTNNBL1 should also bind its protein partners using the concave face formed by its six ARM repeats. However, owing to the presence of significant non-ARM structures at both ends of the ARM domain and the unusual structural features within the ARM domain, the putative protein-binding groove of CTNNBL1 is unique in the family.

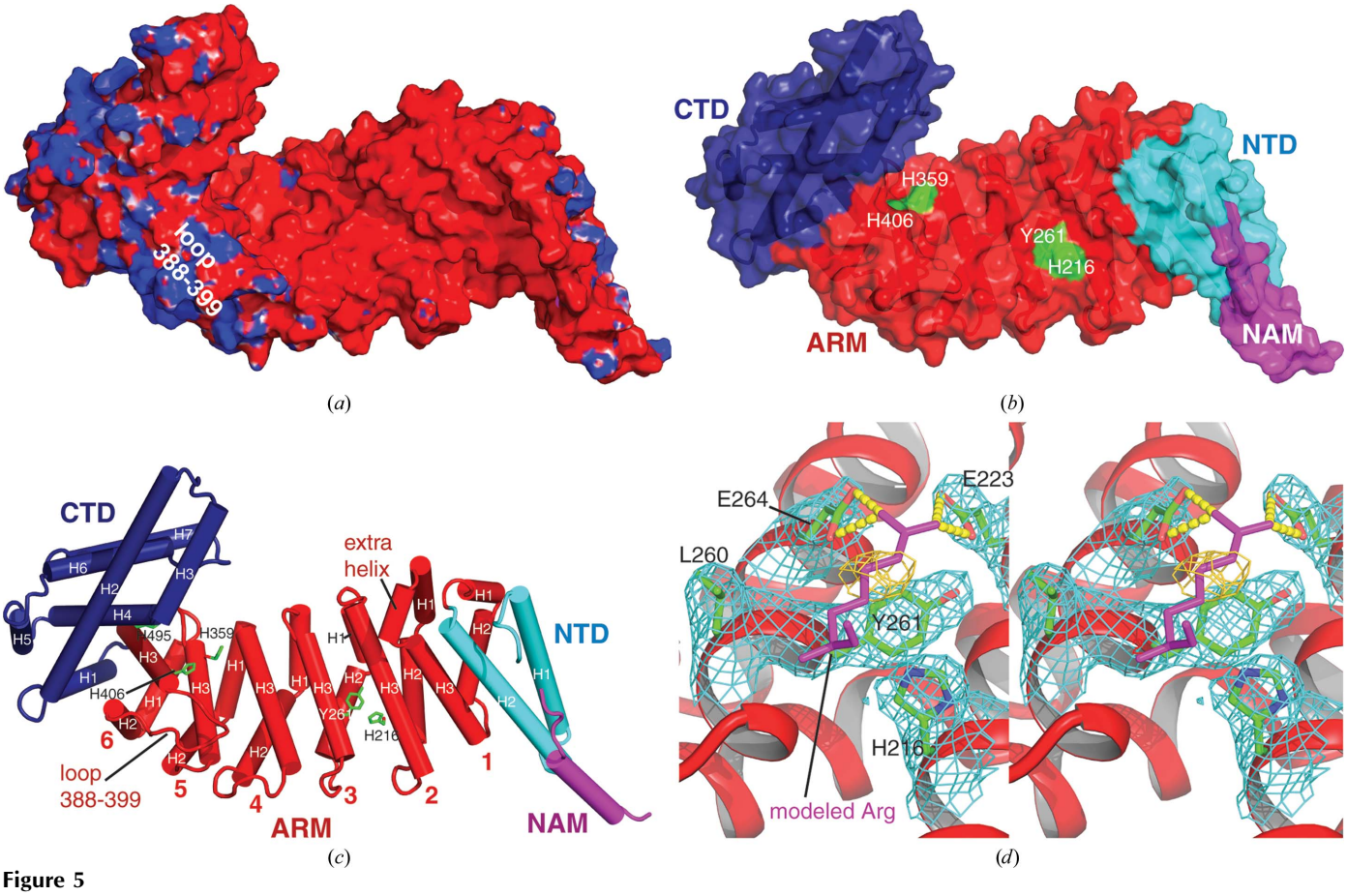

Figure 5
Structure of the putative protein-binding groove. (*a*) Electrostatic surface representation of the structure of full-length CTNNBL1, viewed into the binding groove. Blue and red represent regions of positive and negative potential, respectively. (*b*) Molecular-surface rendering with 20% transparency. The coloring scheme for the different domains is identical to that in Fig. 1. One tyrosine and three histidines that may participate in protein binding are highlighted in green. (*c*) Cartoon representation from the same view as in (*a*) and (*b*). The helices are represented as cylinders. The side chains of the four residues highlighted in (*b*) and His495 from the CTD are shown as sticks in green. (*d*) Stereoview of the His216–Tyr261 region with a modelled arginine to illustrate its possible manner of recognition. A $2F_o - F_c$ electron-density map contoured at 1σ is shown for five residues (shown as sticks and colored by element) that define the binding surface in this region. Potential hydrogen-bonding and electrostatic interactions between the arginine and glutamates are indicated by yellow dashed lines. An unmodelled density above the benzene ring of Tyr261 is also shown as an orange mesh.

At the N-terminal end, the H2 helix of the NTD runs roughly parallel to the H3 helix of ARM1, in effect forming a ridge that caps the concave groove of the ARM domain (Figs. 5*a*, 5*b* and 5*c*). The ridge is further augmented by the NAM, which packs against the N-terminal portion of the NTD H2 helix. This helix–helix interaction may provide an anchor to dock the N-terminal sequences of CTNNBL1 to the NTD–ARM–CTD core structure. The extra helix between ARM1 H3 and ARM2 H1 also seems to augment the ridge at the C-terminal end of NTD H2. These structures (the NTD, the NAM, the extra helix and the flexible linker between the NTD and the NAM) together pose steric restrictions near the N-terminal end of the putative protein-binding groove.

The binding groove is defined by surface residues of the H3 helices of the ARM repeats, including those from ARM1 (Asp156, Ala160, Asp163, Gln166 and Glu167), ARM2 (Lys209, Glu210, Asp213, His216, Asn217, Ala220, Glu223, Asn224 and Glu227), ARM3 (Ala257, Tyr261, Glu264, Ile268 and Gln271), ARM4 (Glu305, Glu308, Glu311, Asn312, Asp315 and Ser319), ARM5 (Ser351, Ser352, Lys355, Asp358 and His359) and ARM6 (Glu402, Glu405, His406, Ser409 and Ser413). Obviously, the concave surface is dominated by hydrophilic residues. Negatively charged residues (15 Asp and Glu residues) significantly outnumber positively charged residues (two Lys and three His residues; note that the histidine side chain is not charged at neutral pH). The overall concave surface of the ARM domain is therefore negatively charged (Fig. 5*a*). Most of the surface residues are highly conserved, indicating their functional importance.

At the C-terminal end, a large portion of the CTD leans towards the concave face of the ARM domain, turning the otherwise shallow and widely accessible concave face into a roughly V-shaped deep cleft. One side of the cleft is defined by the CTD, involving residues from helices H3 (Glu479, Glu480, Leu483 and Asp487), H4 (Phe491 and His495) and H6 (Tyr536), as well as the connecting loop G_{541}DGRS between helices H6 and H7. The other side of the cleft is defined by the H3 helices of ARM3–ARM6.

On one edge of the binding groove near the CTD, loop 388–399 forms a ridge that is ∼6–7 Å higher than the relatively level floor of the groove surface defined by the H3 helices of ARM4 and ARM5 (Figs. 5*a* and 5*b*).

With all the unique structural features described above, the putative protein-binding groove of CTNNBL1 appears to be much more restrictive and selective than those of other ARM proteins.

3.5. Comparison with other ARM protein structures

Crystal structures of many ARM proteins have been determined with or without their binding partners (Huber *et al.*, 1997; Xu & Kimelman, 2007; Tewari *et al.*, 2010). The number of ARM repeats varies in different proteins (from five to 17 repeats). The most extensively studied ARM domains in β-catenin and importin-α contain 12 and ten ARM repeats, respectively (Huber *et al.*, 1997; Conti *et al.*, 1998). Both proteins have a shallow elongated and open-ended binding groove that contains two peptide-binding sites. CTNNBL1 contains six ARM repeats. This number is at the low end of the range. Correspondingly, the putative protein-binding groove in CTNNBL1 is more confined.

Most previous structures contain only ARM domains. Non-ARM domains are only observed in a few structures. In the structure of a full-length β-catenin from zebrafish (Xing *et al.*, 2008), an extra helix is present at both the N- and C-terminal ends. The N-terminal helix is actually an extension of ARM1 H1. The ∼15-amino-acid C-terminal helix runs parallel to H3 of ARM12. Binding assays suggested that the C-terminal helix participates in the binding of some β-catenin partners. More substantial non-ARM structures are seen in the C-terminal sequences of mDia1 (mammalian Diaphanous 1), the regulatory subunit H of the yeast V-type ATPase and the Golgi protein p115. In mDia1, a 56-amino-acid C-terminal sequence forms a dimerization domain which is connected to the ARM domain by an interdomain helix. Contact between the dimerization and ARM domains is limited (Lammers *et al.*, 2005; Rose *et al.*, 2005). In subunit *H*, a C-terminal domain of ∼125 amino acids (which is actually a separate ARM structure with two ARM repeats) also shows very limited interaction with the central ARM domain (Sagermann *et al.*, 2001).

Substantial interaction between ARM and non-ARM domains is only observed in the Golgi protein p115 (Striegl *et al.*, 2009; An *et al.*, 2009). An ∼72-amino-acid C-terminal sequence (called the USO element) assumes a multihelical structure that packs tightly against the H3 helices of the last four ARM repeats (ARM8–ARM11) of the ARM domain. The C-terminal portion of the concave face of the ARM domain is therefore fully blocked and no longer accessible for ligand binding. It was shown that the C-terminal domain mediated dimerization of the protein.

In CTNNBL1, the central ARM domain is sandwiched between the NTD and the CTD. Moreover, the three domains fold together as an integral structure. The last three helices (H5–H7) of the CTD show a certain degree of similarity to the C-terminal domain of p115 (Supplementary Fig. S4). However, the spatial relationships between the C-terminal domain and the central ARM domain are substantially different in the two proteins. In CTNNBL1, most areas of the concave face of the ARM domain do not directly contact the CTD. Therefore, these surface areas are still present in the protein-binding groove. Accessibility to these areas just becomes more restrictive owing to the presence of the CTD (Figs. 5*a* and 5*b*). The CTD of CTNNBL1 is not involved in dimerization.

It is known that CTNNBL1 possesses distinct NLS-recognition specificities compared with importin-α (Ganesh *et al.*, 2011). Importin-α contains two NLS-binding sites (the major and minor sites; Kobe, 1999; Matsuura & Stewart, 2004; Conti *et al.*, 1998). Each site can bind a monopartite NLS sequence. Both sites are required to bind a bipartite NLS sequence, with each site recognizing one of the Lys/Arg-rich clusters of the BNLS. In comparison, the much shorter length of the groove in CTNNBL1 should only be able to accommodate one binding site. It has been shown that recognition of the BNLS

of CTNNBL1 by importin-α required both of the basic clusters (Ganesh *et al.*, 2011). Similar recognition by the ARM domain of CTNNBL1 is unlikely. Indeed, static intramolecular ARM–BNLS interaction is not observed in the structure of full-length CTNNBL1.

The residues involved in NLS recognition in importin-α are highly conserved, including the Trp-Asn array (W*XXX*N motifs in ARM repeats 2, 3, 4, 7 and 8) and negatively charged residues. The bulky and hydrophobic side chains of the tryptophan residues engage in hydrophobic contacts with the aliphatic portions of the side chains of Lys/Arg residues from the NLSs (Matsuura & Stewart, 2004; Dias *et al.*, 2009; Giesecke & Stewart, 2010). There is no tryptophan in the putative binding groove of CTNNBL1. Instead, four highly/strictly conserved histidine residues (His216, His359, His406 and His495 in ARM2, ARM5, ARM6 and CTD H2, respectively) and a tyrosine residue (Tyr261 in ARM3) are found (Figs. 5*b* and 5*c*). To obtain an idea of how these residues might mediate NLS recognition, we modeled an arginine residue in the His216/Tyr261 region. As shown in Fig. 5(*d*), in such a model His216, Leu260 and Tyr261 provide hydrophobic contacts with the aliphatic portion of the arginine side chain, while Glu223 and Glu264 mediate hydrogen-bonding and electrostatic interactions with the guanidinium group of the arginine side chain. Interestingly, a blob of electron density is present around a portion of the modeled arginine side chain (Fig. 5*d*). It is possible that a transient interaction between the N-terminal sequence and the ARM domain exists that gives rise to the unaccounted density.

At around neutral pH the imidazole side chain of a histidine residue is deprotonated, noncharged and hydrophobic, enabling hydrophobic interaction. At lower pH (5.0–5.5) the histidine side chain may become protonated, positively charged and hydrophilic, and therefore unfavorable for hydrophobic interaction with the aliphatic portions of the side chains of the Lys/Arg residues from the NLSs. The positively charged histidine side chains in the protein-binding groove of CTNNBL1 may also repel the positively charged side chains of the Lys/Arg residues in the NLSs.

Protonation/deprotonation of histidine residues is well established as a pH-dependent mechanism that regulates the ligand-binding activity of proteins (Rötzschke *et al.*, 2002; Yang *et al.*, 2011). The presence of a histidine cluster in the putative protein-binding groove of CTNNBL1 may suggest a pH-sensitive histidine-mediated mechanism for regulation of the protein-binding activity of CTNNBL1.

4. Discussion

The confirmation of CTNNBL1 as an ARM protein comes as no surprise. However, the six-repeat ARM domain is much smaller than previously predicted (ten repeats comparable to those in importin-α; Ganesh *et al.*, 2011). The ARM structure accounts for only half of the protein sequence (Fig. 1*a*). Significant non-ARM structures (NTD, CTD and NAM) are present. The well defined non-ARM structures cap the concave face of the central ARM domain at both ends. The NTD, the ARM domain and the CTD are packed together by extensive hydrophobic and hydrogen-bonding interactions to form a seamless and relatively rigid unibody core structure. Such a structure is unique among ARM proteins with known structures.

ARM proteins use the concave faces of their ARM domains as the binding grooves for target proteins. These grooves are typically elongated, shallow and open-ended at both ends owing to the absence of significant non-ARM structures. In CTNNBL1, the non-ARM structures and the unusual structural features within the ARM domain work synergistically to create a much more restrictive and discriminatory partner protein-binding groove. At the N-terminal end, the otherwise open-ended concave face is capped by a ridge formed by the NTD, the NAM, the linker between the NTD and the NAM, and the extra helix within the ARM domain. The presence of this ridge may pose steric hindrance to accessing the groove surface from the N-terminal end. At the C-terminal end, the CTD contributes to the formation of a deep cleft. A partner protein bound in this cleft may need to contact the H3 helices of the ARM domain on one side and the H3, H4, H6 and H7 helices of the CTD on the other side. For recognition of the conformational NLS of AID, AID may need to adopt a peculiar structure that is complementary to the restrictive binding groove of CTNNBL1.

To gain insights into CTNNBL1–AID interaction, we have built a structural model of the CTNNBL1–AID complex (Fig. 6). Based on the crystal structures of two other human

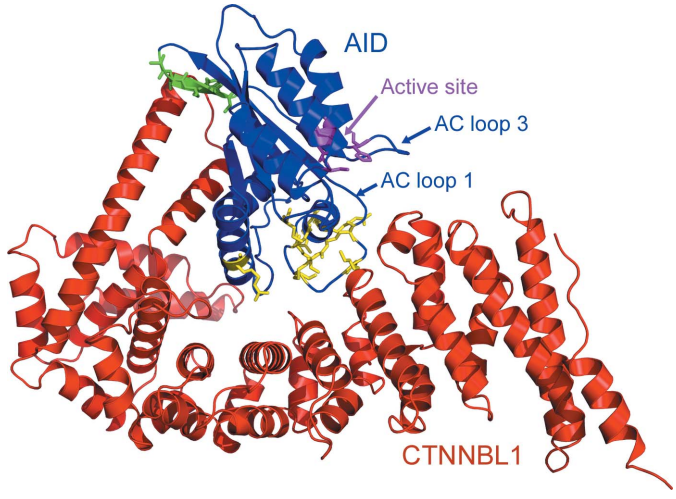

Figure 6
A structural model of CTNNBL1–AID interaction. The structural model of AID was generated by *Modeller* using the crystal structure of APOBEC3G (PDB entry 3e1u). The model of the complex was generated by manually docking the structure of AID to the protein-binding groove of CTNNBL1 using *Chimera*. CTNNBL1 and AID are colored red and blue, respectively. The putative active-site residues of AID (His56, Glu58, Cys87 and Cys90) are colored magenta with their side chains shown as sticks. The active-center (AC) loops 1 and 3 are indicated. A number of basic residues of AID (Arg8, Arg9, Lys16, Arg19, Lys22, Arg24, Arg50, Arg112 and Arg171) that are involved in interaction with CTNNBL1 are shown as sticks and colored yellow; a stretch of residues (^{39}ATSFS) that is also involved in CTNNBL1 interaction is colored green.

cytidine deaminases, APOBEC2 (PDB entry 2nyt; Prochnow *et al.*, 2007) and APOBEC3G (PDB entry 3e1u; Holden *et al.*, 2008), a homologous structure of human AID was built using the modeling program *MODELLER* (Eswar *et al.*, 2008). The modeled structure of AID was then manually docked to the structure of CTNNBL1 using *Chimera* (Pettersen *et al.*, 2004). Although this is only a rough structure model (without energy minimization and consideration of structural rearrangement upon binding), it clearly illustrates several important points about the AID–CTNNBL1 interaction. Firstly, the size and overall shape of the structure of AID neatly fit the protein-binding groove of CTNNBL1. Secondly, all of the Lys/Arg residues of AID (Arg8, Arg9, Lys16, Arg19, Lys22, R24, Arg50, Arg112 and Arg171) that are known to be involved in the interaction are located on the 'bottom surface' of the molecule, facing the negatively charged binding groove of the CTNNBL1 ARM domain. Thirdly, the stretch of residues (A_{39}TSFS) known to be important for the interaction is located near the protruding end of the CTD of CTNNBL1. Because the cluster of basic residues and the A_{39}TSFS stretch are located at opposite ends of the AID structure, simultaneous recognition of these two sites can be achieved by CTNNBL but not by importin-α. Lastly, the putative active site of AID (defined by the four active-site residues His56, Glu58, Cys87 and Cys90) and the single-stranded DNA-binding groove (between the active-center loops AC loop 1 and AC loop 3) are fully accessible. Substantial interaction between CTNNBL1 and AID may not interfere with the enzymatic activity of AID. In short, the model is in good agreement with available biochemical data regarding AID–CTNNBL1 interaction.

The putative partner protein-binding groove in CTNNBL1 is undoubtedly unique, which provides the structural basis for the profound mechanistic difference in NLS recognition between CTNNBL1 and importin-α. It has been shown that CTNNBL1 and AID colocalize in nucleoli, where they physically associate with nucleolin and nucleophosmin (Hu *et al.*, 2013); it is likely that the highly specific interaction between CTNNBL1 and AID is important not only for the nuclear import of AID but also for the subnuclear targeting of AID to the immunoglobulin gene loci.

In the importin-α-mediated nuclear import pathway, cargo-bound importin-α binds to importin-β through the highly basic N-terminal importin-β-binding domain (IBB; Cingolani *et al.*, 1999; Conti & Izaurralde, 2001). Without bound cargo, the IBB engages in an intramolecular interaction with the ARM domain. NLS binding displaces the autoinhibitory IBB so that it can interact with importin-β. It is tempting to speculate that a similar autoinhibitory mechanism exists in CTNNBL1 because CTNNBL1 binds NLSs and itself contains a functional BNLS. However, a static intramolecular BNLS–ARM interaction is not observed in the structure of full-length CTNNBL1, although the possibility of a transient interaction cannot be ruled out.

The degradation of full-length CTNNBL1 to CTNNBL1^{SF} indicates that CTNNBL1^{SF} constitutes a stable structure within CTNNBL1. Sequence alignment of CTNNBL1 from different species shows that the N-terminal sequences (amino acids 1–75 in human CTNNBL1) are much less conserved than the sequences covering the NTD–ARM–CTD domains (Supplementary Fig. S1). The N-terminal sequences are rich in hydrophilic residues, especially negatively charged residues. In human CTNNBL1, the presence of a transcription factor-like acidic domain from Asp20 to Glu79 was speculated (45% Asp and Glu; Jabbour *et al.*, 2003). However, the structure of full-length CTNNBL1 indicates that this region is largely unstructured. The only observed structure within the N-terminal sequence is the NAM, which forms a three-turn helix packed against the N-terminal portion of NTD H2. This packing places the NAM right at the N-terminal end of the otherwise open-ended groove of the ARM domain. It is noted that the the NAM–NTD interaction is weaker than the NTA–ARM1 and CTD–ARM6 interactions. It is possible that the NAM–NTD interaction somehow sequesters the N-terminal sequences, making the BNLS unavailable for binding to importin-α. Binding a partner protein may disrupt the NAM–NTD interaction, therefore releasing the BNLS for importin-α interaction.

The CTD adopts a novel protein fold. As shown in Fig. 3(*a*), the structure of this protein fold is held together by extensive hydrophobic and hydrogen-bonding interactions. The presence of these interactions suggests that the CTD may fold autonomously as a separate domain, although in the native context it folds together with the ARM domain in the integral structure of CTNNBL1. It has been shown that a protein construct containing the last 122 amino acids (residues 442–563, corresponding to the CTD lacking the first helix H1 and one turn of helix H2) induced apoptosis in transfected CHO cells (Jabbour *et al.*, 2003), suggesting not only the functional importance of the C-terminal region but also its ability to autonomously fold into a functional structure. It is possible that, besides capping the ARM domain and contributing to the formation of a unique protein-binding groove, the CTD may also have other functions.

It is interesting to note that there is an alternatively spliced product of the *CTNNBL1* gene known as testes development-related NYD-SP19 (Jabbour *et al.*, 2003). The sequence of NYD-SP19 corresponds to amino acids 189–563 of the CTNNBL1 sequence. Without the BNLS (amino acids 16–31; Fig. 1*a*) of CTNNBL1, NYD-SP19 is most likely to be a cytoplasmic protein that may assume different functions from those of CTNNBL1. In the structure of CTNNBL1, residue 189 locates at the end of the H1 helix of ARM2. Therefore, the N-terminal structures of CTNNBL1, including the NAM, the NTD, ARM1, the extra helix between ARM1 and AMR2 and the first helix of ARM1, are not present in NYD-SP19. Assuming that NYD-SP19 adopts a structure similar to the corresponding portion of the CTNNBL1 structure, this structure would have five ARM repeats (with the N-terminal repeat having only two helices). The N-terminal end of the putative protein-binding groove (defined by the H3 helices of the ARM repeats) would be open, as in other ARM domains. Owing to this structural difference, protein binding to NYD-SP19 may be less restrictive than to CTNNBL1.

In summary, the structures reported here provide much-needed structural knowledge for better understanding of the biological functions of CTNNBL1. The structure not only consolidates the available biochemical data and hypotheses about the structure and function of CTNNBL1, but also provide new insights to guide further investigation. Given the many unique structural features of human CTNNBL1, there is no doubt that CTNNBL1 is a distinct member of the ARM protein family. Previously, it has been proposed that ARM proteins should be classified into four subfamilies represented by β-catenin, p120 catenin, importin-α and the p115 head domain, respectively (Striegl *et al.*, 2010). The structure of CTNNBL1 is substantially different from any of these structures (in this sense, the name CTNNBL1 is actually somewhat misleading). CTNNBL1 may therefore define a new subfamily of ARM proteins within the structurally and functionally diverse ARM protein superfamily.

We thank Drs Spencer Anderson, Joseph Brunzelle, Elena Kondrashkina and Zdzislaw Wawrzak at LS-CAT of the Advanced Photon Source, Argonne National Laboratory for assistance with data collection. The work was supported by a start-up fund and a seed grant from Southern Illinois University Carbondale.

References

Adams, P. D. *et al.* (2011). *Methods*, **55**, 94–106.
An, Y., Chen, C. Y., Moyer, B., Rotkiewicz, P., Elsliger, M.-A., Godzik, A., Wilson, I. A. & Balch, W. E. (2009). *J. Mol. Biol.* **391**, 26–41.
Andreasen, C. H., Mogensen, M. S., Borch-Johnsen, K., Sandbaek, A., Lauritzen, T., Almind, K., Hansen, L., Jørgensen, T., Pedersen, O. & Hansen, T. (2009). *BMC Med. Genet.* **10**, 17.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Cho, Y. S. *et al.* (2009). *Nature Genet.* **41**, 527–534.
Cingolani, G., Petosa, C., Weis, K. & Müller, C. W. (1999). *Nature (London)*, **399**, 221–229.
Conti, E. & Izaurralde, E. (2001). *Curr. Opin. Cell Biol.* **13**, 310–319.
Conti, E., Uy, M., Leighton, L., Blobel, G. & Kuriyan, J. (1998). *Cell*, **94**, 193–204.
Conticello, S. G., Ganesh, K., Xue, K., Lu, M., Rada, C. & Neuberger, M. S. (2008). *Mol. Cell*, **31**, 474–484.
Dias, S. M., Wilson, K. F., Rojas, K. S., Ambrosio, A. L. & Cerione, R. A. (2009). *Nature Struct. Mol. Biol.* **16**, 930–937.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Eswar, N., Eramian, D., Webb, B., Shen, M.-Y. & Sali, A. (2008). *Methods Mol. Biol.* **426**, 145–159.
Ganesh, K., Adam, S., Taylor, B., Simpson, P., Rada, C. & Neuberger, M. (2011). *J. Biol. Chem.* **286**, 17091–17102.
Giesecke, A. & Stewart, M. (2010). *J. Biol. Chem.* **285**, 17628–17635.
Holden, L. G., Prochnow, C., Chang, Y. P., Bransteitter, R., Chelico, L., Sen, U., Stevens, R. C., Goodman, M. F. & Chen, X. S. (2008). *Nature (London)*, **456**, 121–124.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
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.
Huber, A. H., Nelson, W. J. & Weis, W. I. (1997). *Cell*, **90**, 871–882.
Huhn, S. *et al.* (2011). *Int. J. Mol. Epidemiol. Genet.* **2**, 36–50.
Ichikawa, S., Koller, D. L., Padgett, L. R., Lai, D., Hui, S. L., Peacock, M., Foroud, T. & Econs, M. J. (2010). *J. Bone Miner. Res.* **25**, 1821–1829.
Jabbour, L., Welter, J. F., Kollar, J. & Hering, T. M. (2003). *Genomics*, **81**, 292–303.
Kobe, B. (1999). *Nature Struct. Mol. Biol.* **6**, 388–397.
Lammers, M., Rose, R., Scrima, A. & Wittinghofer, A. (2005). *EMBO J.* **24**, 4176–4187.
Liu, Y.-J. *et al.* (2008). *Hum. Mol. Genet.* **17**, 1803–1813.
Los, G. V. *et al.* (2008). *ACS Chem. Biol.* **3**, 373–382.
Matsuura, Y. & Stewart, M. (2004). *Nature (London)*, **432**, 872–877.
Meurs, J. B. J. van *et al.* (2008). *JAMA*, **299**, 1277–1290.
Papassotiropoulos, A. *et al.* (2013). *Mol. Psych.* **18**, 255–263.
Peifer, M., Berg, S. & Reynolds, A. B. (1994). *Cell*, **76**, 789–791.
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.
Prochnow, C., Bransteitter, R., Klein, M. G., Goodman, M. F. & Chen, X. S. (2007). *Nature (London)*, **445**, 447–451.
Richards, J. B. *et al.* (2008). *Lancet*, **371**, 1505–1512.
Rose, R., Weyand, M., Lammers, M., Ishizaki, T., Ahmadian, M. R. & Wittinghofer, A. (2005). *Nature (London)*, **435**, 513–518.
Rötzschke, O., Lau, J. M., Hofstätter, M., Falk, K. & Strominger, J. L. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 16946–16950.
Rual, J.-F. *et al.* (2005). *Nature (London)*, **437**, 1173–1178.
Sagermann, M., Stevens, T. H. & Matthews, B. W. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 7134–7139.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Striegl, H., Andrade-Navarro, M. A. & Heinemann, U. (2010). *PLoS One*, **5**, e8991.
Striegl, H., Roske, Y., Kümmel, D. & Heinemann, U. (2009). *PLoS One*, **4**, e4656.
Styrkarsdottir, U. *et al.* (2008). *New Engl. J. Med.* **358**, 2355–2365.
Styrkarsdottir, U. *et al.* (2010). *PLoS One*, **5**, e13217.
Tewari, R., Bailes, E., Bunting, K. A. & Coates, J. C. (2010). *Trends Cell Biol.* **20**, 470–481.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Xing, Y., Clements, W. K., Le Trong, I., Hinds, T. R., Stenkamp, R., Kimelman, D. & Xu, W. (2004). *Mol. Cell*, **15**, 523–533.
Xing, Y., Takemaru, K.-I., Liu, J., Berndt, J. D., Zheng, J. J., Moon, R. T. & Xu, W. (2008). *Structure*, **16**, 478–487.
Xu, W. & Kimelman, D. (2007). *J. Cell Sci.* **120**, 3337–3344.
Yang, L., Zhang, J., Ho, B. & Ding, J. L. (2011). *PLoS One*, **6**, e19647.
Yin, Q., Yang, H.-W., Han, X.-L., Fan, B. & Liu, B. (2012). *Mol. Biol. Rep.* **39**, 4485–4490.