Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Structure of mouse muskelin discoidin domain and biochemical characterization of its self-association

Kook-Han Kim,[a]‡ Seung Kon Hong,[a,b]‡ Kwang Yeon Hwang[b] and Eunice EunKyeong Kim[a]*

[a]Biomedical Research Institute, Korea Institute of Science and Technology, Hwarang-ro 14-gil 5, Seongbuk-gu, Seoul 136-791, Republic of Korea, and [b]Division of Biotechnology, College of Life Sciences and Biotechnology, Korea University, 145 Anam-ro, Seoul 136-701, Republic of Korea

‡ These authors contributed equally to this work.

Correspondence e-mail: eunice@kist.re.kr

Received 7 April 2014
Accepted 21 August 2014

PDB reference: discoidin domain from muskelin, 4pqq

Muskelin is an intracellular kelch-repeat protein comprised of discoidin, LisH, CTLH and kelch-repeat domains. It is involved in cell adhesion and the regulation of cytoskeleton dynamics as well as being a component of a putative E3 ligase complex. Here, the first crystal structure of mouse muskelin discoidin domain (MK-DD) is reported at 1.55 Å resolution, which reveals a distorted eight-stranded β-barrel with two short α-helices at one end of the barrel. Interestingly, the N- and C-termini are not linked by the disulfide bonds found in other eukaryotic discoidin structures. A highly conserved MIND motif appears to be the determinant for MK-DD specific interaction together with the spike loops. Analysis of interdomain interaction shows that MK-DD binds the kelch-repeat domain directly and that this interaction depends on the presence of the LisH domain.

1. Introduction

Muskelin (MKLN1 or TWA2) was first identified as an intracellular protein that acts as a mediator of cell adhesion to the C-terminal domain of thrombospondin-1. When over-expressed, it has also been shown to be an effector of the formation and dynamics of focal adhesions in cells grown on fibronectin (Adams *et al.*, 1998). Although the exact mechanisms and details of the interactions are not fully understood, muskelin has been reported to interact with the carboxyl-terminal region of prostaglandin receptor EP3 isoform α (Hasegawa *et al.*, 2000) and the Cdk5-activating protein p39 for actin binding and cytoskeleton organization (Ledee *et al.*, 2005). It has also been reported to bind to the T-box DNA-binding domain of the TBX20b isoform [T-box (TBX) transcription factor] directly using its kelch-repeat domain in the cytoplasm (Debenedittis *et al.*, 2011). Muskelin interacts with $GABA_A$ receptor α1 ($GABA_AR\ \alpha$), which is involved in neuronal co-transport, and plays a central role at the subcellular level by acting as a trafficking protein regulating the transport of $GABA_A$ receptors (Heisler *et al.*, 2011). In addition, muskelin has been reported to associate with RanBPM (Ran-binding protein in the microtubule-organizing centre; also known as RanBP9) and TWA1 (two hybrid-associated protein 1 with RanBPM) in transfected COS-7 cells (Umeda *et al.*, 2003). These three proteins in turn form a 670 kDa complex together with MAEA, Rmnd5 and Armc8. This 670 kDa complex resembles the 600 kDa *Saccharomyces cerevisiae* glucose-induced degradation-deficient (GID) complex that mediates the polyubiquitination of fructose-1,6-bisphosphatase *via* E3 ubiquitin ligase activity (Francis *et al.*, 2013).

© 2014 International Union of Crystallography

Muskelin is expressed in various tissues including the brain, eyes, heart, skeletal muscles and kidneys (Adams *et al.*, 1998; Ledee *et al.*, 2005; Tagnaouti *et al.*, 2007). This 85 kDa protein is comprised of an N-terminal discoidin domain (DD) followed by a lissencephaly 1 homology (LisH) domain, a C-terminal to LisH (CTLH) domain, a kelch-repeat domain of six kelch motifs and an unannotated C-terminal region, as shown in Fig. 1(*a*). Although it is considered to be part of the kelch-repeat superfamily, which has a set of five to seven kelch repeats that form a β-propeller tertiary structure, muskelin is unique in its domain structure. The homologue with regard to domain structure is fungal galactose oxidase, which lacks the LisH and CTLH domains (Firbank *et al.*, 2001). Most of the interacting proteins mentioned above interact with the C-terminal region of muskelin, which includes the kelch-repeat domain (Prag *et al.*, 2007), while the binding and regulation of $GABA_AR\,\alpha1$ is mediated by the region including the discoidin and LisH domains (Heisler *et al.*, 2011). Interestingly, muskelin is reported to self-associate through its discoidin and kelch-repeat domains in a head-to-tail fashion in addition to the aforementioned interactions with other proteins (Prag *et al.*, 2004).

The discoidin and discoidin-like domains (also known as C2-like, F5/8C domain and galactose-binding domain; IPR008979) are functional and structural units that are found in both eukaryotic and prokaryotic proteins. They bind diverse ligand molecules such as phospholipids, carbohydrates, collagens and other partner proteins (Kiedzierska *et al.*, 2007). They include blood coagulation factors such as factor V and VIII, lectins such as discoidin I and II, secreted enzymes such as galactose oxidase and sialidase, cell-surface tyrosine kinase receptors such as discoidin domain receptors, sperm–egg adhesion protein/milk fat globule-EGF factor and proteins involved in neural development such as neurexin IV and neuropilins (Wu & Molday, 2003; Baumgartner *et al.*, 1998; Vogel, 1999; Franco-Pons *et al.*, 2006). Most of these proteins are extracellular or membrane-associated and are involved in the cellular adhesion, migration or aggregation events associated with organogenesis and other developmental processes. They all have a β-barrel structure with the N- and C-termini tied by a disulfide bond and the loops protruding from the barrel. The loops on the opposite side of the N- and C-termini serve as the recognition platform for ligands, and three loops in particular are named spike1, spike2 and spike3 (Kiedzierska *et al.*, 2007).

Although it is known that discoidin and kelch-repeat domains form β-structures, namely a β-barrel and β-propeller, respectively, and LisH and CTLH form α-helical domains, to date no structural information has been available. Since these domains function as a structural module, detailed information on the nature of the protein surface is crucial in understanding its interactions. In this study, we report the crystal structure of the discoidin domain of mouse muskelin for the first time. We also carried out an analysis of the interdomain interaction using isothermal titration calorimetry (ITC) to understand the mechanism of self-association.

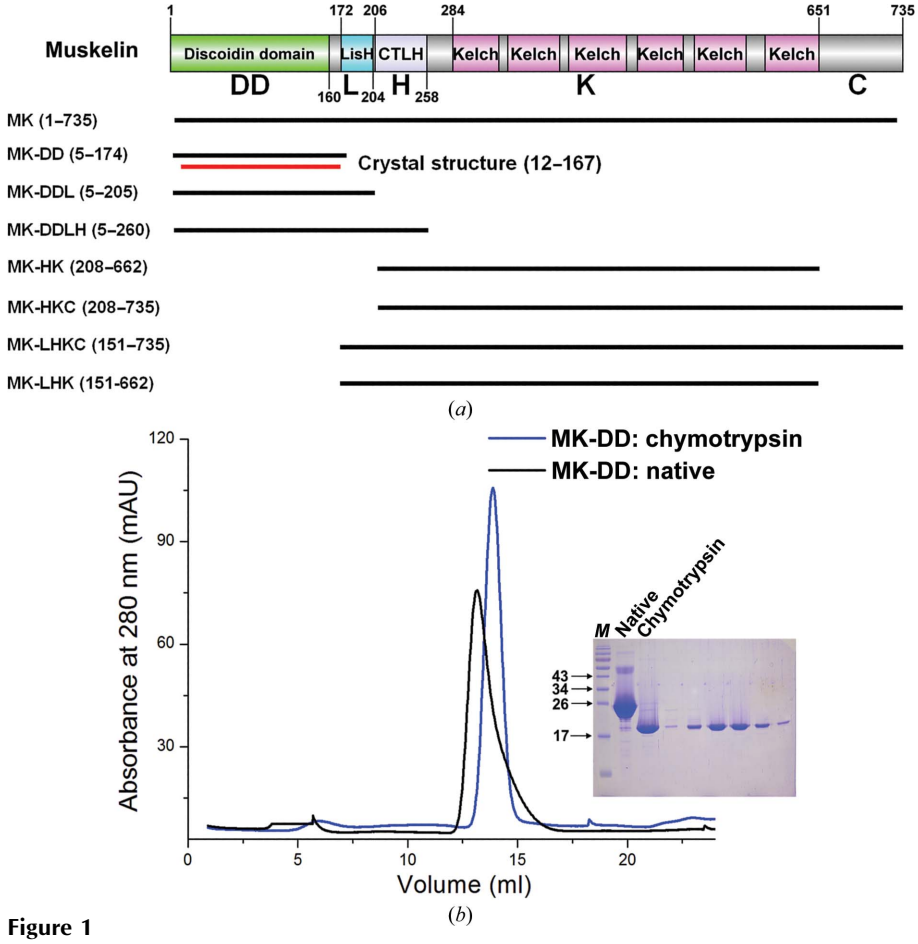


Figure 1
Schematic representation and purification of muskelin. (*a*) Domain organization and the constructs used. The discoidin domain is followed by LisH, CTLH and six kelch repeats and an unannotated C-terminal region. Constructs used in this study: MK, 1–735; MK-DD, 5–174; MK-DDL, 5–205; MK-DDLH, 5–260; MK-HK, 208–662; MK-HKC, 208–735; MK-LHKC, 151–735; MK-LHK, 151–662. The region in the crystal structure is indicated in red. (*b*) SEC profile and SDS–PAGE of MK-DD. Purification of MK-DD with and without chymotrypsin treatment is shown in blue and black, respectively. Lane *M* of the SDS–PAGE contains molecular-mass markers (labelled in kDa).

2. Materials and methods

2.1. Protein expression and purification

The gene encoding the discoidin-like domain of mouse muskelin (SWISS-PROT entry O89050; residues 1–174) was amplified by polymerase chain reaction and ligated into pET-32a vector (Novagen) with a thioredoxin-His_6 tag at the N-terminus. The cloned vector

was transformed into *Escherichia coli* Rosetta (DE3) strain (Novagen). The cells were grown at 310 K in Luria–Bertani medium containing ampicillin (100 µg ml^{-1}) and expression was induced by 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside at an optical density of about 0.6 at 600 nm. The cells were allowed to grow at 291 K overnight after induction and were harvested by centrifugation and resuspended in buffer consisting of 50 m*M* Tris–HCl pH 8.0, 300 m*M* NaCl, 2 m*M* β-mercaptoethanol, 1 m*M* phenylmethylsulfonyl fluoride. The cells were disrupted by sonication and the crude lysate was centrifuged at 18 000 rev min^{-1} (Hanil Supra 22K) for 40 min at 277 K and the cell debris was discarded. The supernatant was loaded onto a nickel-chelated HiTrap chelating column (GE Healthcare) and was eluted with a linear gradient of 25–500 m*M* imidazole in 50 m*M* Tris–HCl pH 8.0, 300 m*M* NaCl, 2 m*M* β-mercaptoethanol. Fractions were pooled based on SDS–PAGE analysis and were subjected to thrombin treatment overnight at room temperature. Next, the protein was further purified using gel-filtration chromatography on HiLoad 26/60 Superdex 75 (GE Healthcare) which was pre-equilibrated with 50 m*M* Tris–HCl pH 7.4, 100 m*M* NaCl, 5 m*M* DTT. The purified protein was concentrated using a Vivaspin20 concentrator (Sartorius) and was stored at 203 K. Since no crystals were obtained from this preparation, we tested limited proteolysis using trypsin and chymotrypsin, and the 19 kDa fragment from chymotrypsin treatment eventually gave usable crystals (Fig. 1*b*). The protein cleaved with chymotrypsin was subjected to gel-filtration chromatography in 50 m*M* Tris–HCl pH 7.4, 100 m*M* NaCl, 5 m*M* DTT. The resulting protein was concentrated to 22 mg ml^{-1} for crystallization trials. The selenomethionine (SeMet)-substituted discoidin domain was produced in the methionine auxotroph *E. coli* B834 (DE3) (Novagen) and purified using the same procedure as described as for the native protein. The full length as well as the six truncated forms of the protein, namely DD-LisH (MK-DDL; residues 5–205), DD-LisH-CTLH (MK-DDLH; residues 5–260), CTLH-Kelch (MK-HK; residues 208–662), CTLH-Kelch-C-terminal (MK-HKC; residues 208–735), LisH-CTLH-Kelch (MK-LHK; residues 151–662) and LisH-CTLH-Kelch-C-terminal (MK-LHKC; residues 151–735), shown in Fig. 1(*a*) were cloned. The boundaries of the various domain constructs are slightly different from those described earlier by Valiyaveettil *et al.* (2008). Four muskelin variants, namely MK-HK, MK-HKC, MK-LHK and MK-LHKC, were expressed and purified using affinity and gel-filtration chromatography and were dialyzed against 20 m*M* HEPES pH 7.5, 100 m*M* NaCl. Constructs with kelch repeats alone or the C-terminal region alone did not yield soluble protein.

2.2. Crystallization and data collection

All purified proteins were subjected to crystallization screening by the sitting-drop vapour-diffusion method using commercially available kits from Hampton Research, Molecular Dimensions and Emerald BioSystems, 96-well IntelliPlates (Hampton Research) and a Hydra II Plus One (Matrix Technology) at 295 K. The initial screening gave microcrystals in Wizard IV condition No. 17 (Emerald BioSystems) and PEG/Ion 2 condition No. 25 (Hampton Research) for the MK-DD and MK-LHKC constructs, respectively, and the conditions were optimized using hanging-drop vapour-diffusion experiments. However, the MK-LHKC crystals only diffracted to low resolution. Diffraction-quality crystals of MK-DD were obtained by mixing equal volumes of the protein at 22 mg ml^{-1} in 50 m*M* Tris–HCl pH 7.4, 100 m*M* NaCl, 5 m*M* DTT and reservoir solution consisting of 25% polyethylene glycol (PEG) 1500, 100 m*M* succinate–phosphate–glycine (SPG) buffer pH 5.5. Rhombohedral crystals of MK-DD appeared in a few days. They were cryoprotected using reservoir solution supplemented with an additional 30%(*v*/*v*) sucrose and were flash-cooled in liquid nitrogen. Diffraction data were collected at 100 K on beamline 4A equipped with an ADSC Quantum 315r CCD detector at Pohang Light Source, Pohang, Republic of Korea. The native crystal of the discoidin domain diffracted to 1.55 Å resolution and belonged to space group *R*3, with unit-cell parameters $a = b = 98.3$, $c = 57.3$ Å, $\alpha = \beta = 90.0$, $\gamma = 120.0^\circ$. SeMet-substituted crystals were obtained by seeding the native crystal into a drop containing SeMet-substituted protein and the reservoir solution that gave the native crystal. The cross-seeded crystal melted as soon as it was transferred, but crystals with similar morphology appeared after a few days and were harvested and cryoprotected in the same manner as for the native crystals. The

Table 1
Statistics of X-ray data collection and refinement.

Values in parentheses are for the highest resolution shell.

Data set	SeMet MK-DD	Native MK-DD
Data-collection statistics		
Beamline	PAL-5C	PAL-5C
X-ray wavelength (Å)	0.97918	0.97951
Resolution range (Å)	50.0–1.80 (1.86–1.80)	50.0–1.55 (1.61–1.55)
Space group	*R*3	*R*3
Unit-cell parameters (Å)	$a = b = 98.411$, $c = 57.269$, $\alpha = \beta = 90$, $\gamma = 120$	$a = b = 98.303$, $c = 57.317$, $\alpha = \beta = 90$, $\gamma = 120$
Total/unique reflections	1349625/19251	408202/30011
Completeness (%)	95.3 (88.1)	98.2 (90.8)
Multiplicity	5.1 (2.5)	4.1 (2.5)
Mean $I/\sigma(I)$	13.2 (2.6)	24.4 (5.7)
$R_{\rm merge}$† (%)	10.6 (31.4)	6.9 (18.4)
Refinement statistics		
Resolution range (Å)		50.0–1.55
$R/R_{\rm free}$‡ (%)		15.3/17.6
No. of protein atoms		1288
No. of water molecules		203
No. of ligand molecules		29
Average *B* factor (Å^2)		19.3
R.m.s.d., bond lengths (Å)		0.007
R.m.s.d., bond angles (°)		1.142
Ramachandran analysis (%)		
Most favoured		97.0
Additionally allowed		3.0
Outliers		0

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all *i* measurements. ‡ $R_{\rm free}$ was calculated from a randomly selected 5% set of reflections that were not included in calculation of the *R* value.

SeMet-substituted crystal diffracted to 1.8 Å resolution and data were collected in the same way as for the native crystal. The presence of one molecule of MK-DD in the asymmetric unit gives a crystal volume per protein weight (V_M) of 2.92 Å^3 Da^{-1}, with a corresponding solvent content of 57.8% (Matthews, 1968). All data were processed and scaled with *DENZO* and *SCALEPACK* from the *HKL*-2000 suite (Otwinowski & Minor, 1997). The statistics of data collection are summarized in Table 1.

2.3. Structure solution and refinement

All attempts at molecular replacement with *Phaser* (McCoy *et al.*, 2007) using previously reported DD structures such as bovine lactadherin C2 domain (PDB entry 2pqs; Lin *et al.*, 2007) and galactose oxidase (PDB entry 2eib; Rogers *et al.*, 2007) as search models failed to give a reasonable solution. Also, attempts to obtain the phase using heavy-atom derivatives such as mercury and platinum all failed. However, single-wavelength anomalous dispersion (SAD) phasing using an SeMet-substituted protein crystal yielded the structure of MK-DD. A reasonable phase was obtained with a mean FOM of 0.6 to a resolution of 2.0 Å and all three potential Se sites were found and refined using *SOLVE* (Terwilliger & Berendzen, 1999). Electron-density modification and automated model building were carried out using *RESOLVE* (Terwilliger, 2003). The resulting electron-density map with a partial model revealed clear main-chain density with substantial side-chain details. Manual building was performed using *Coot* (Emsley & Cowtan, 2004) and refinement was carried out using *CNS* (Brünger *et al.*, 1998). All atoms of MK-DD except for residues 1–11 and 168–174 were well defined in the electron-density maps. In addition, two PEG molecules, one phosphate and 203 water molecules were placed. The final model refined using *PHENIX* (Adams *et al.*, 2010) had an *R* factor of 15.3% and an R_{free} of 17.6%, and the refinement statistics are summarized in Table 1. The final structure was validated using *PROCHECK* (Laskowski *et al.*, 1993). The Ramachandran plot produced by *PROCHECK* (Laskowski *et al.*, 1993) shows that 97% of the residues are within the most favoured regions and 3% are in additional allowed regions. The *DALI* server (Holm & Rosenström, 2010) was used to search for proteins with similar folds. Solvent-accessible and interaction areas were calculated using *PISA* (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). Sequence alignment was performed using *ClustalW* and an image was produced using *ESPript* 3.0 (http://espript.ibcp.fr). All figures for the structure were generated using *PyMOL* (Schrödinger).

2.4. Isothermal titration calorimetry (ITC) measurements

ITC measurements were performed on an ITC200 calorimeter (GE Healthcare) at 298 K. ITC experiments were carried out to determine the self-association of muskelin using MK-HK, MK-HKC, MK-LHK, MK-LHKC and MK-DD. All proteins were dialyzed in buffer consisting of 20 m*M* HEPES

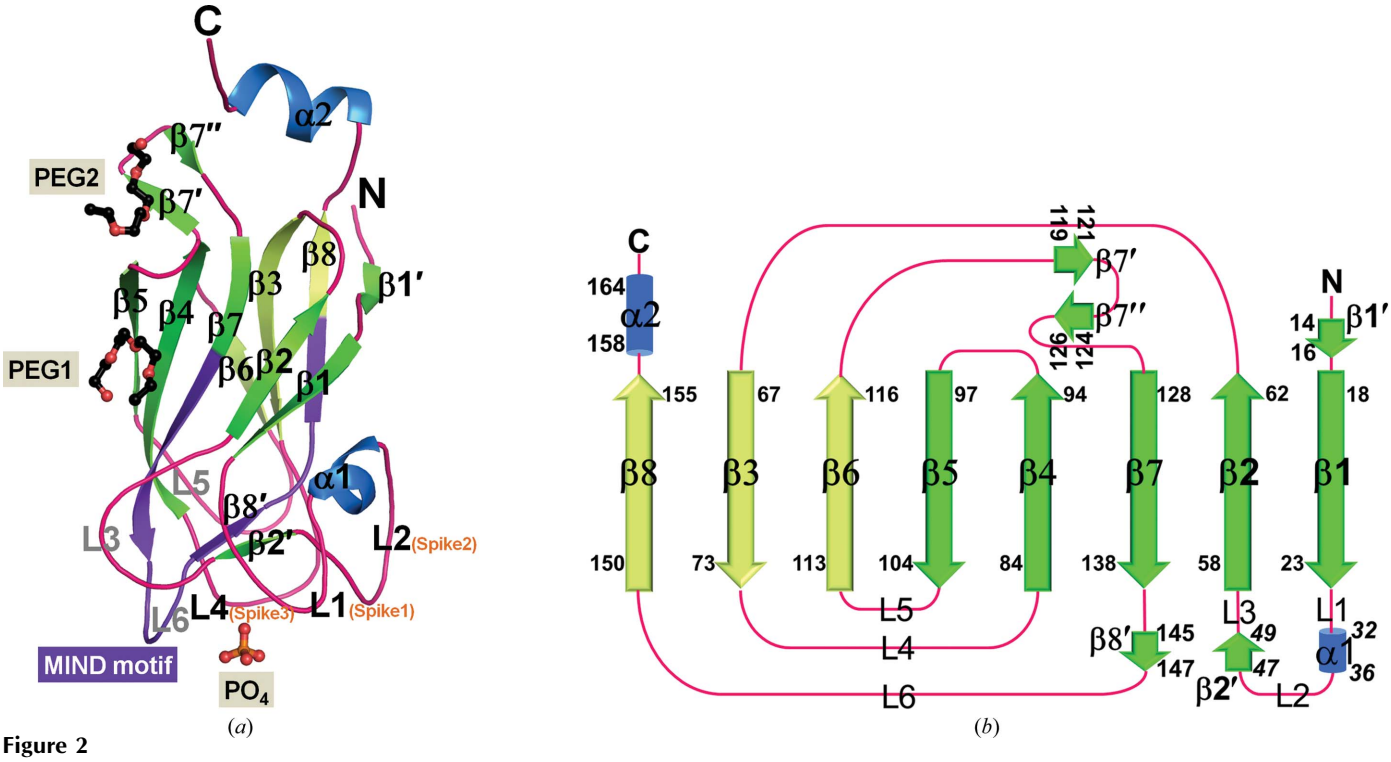


Figure 2
Overall structure of MK-DD. (*a*) Ribbon diagram of MK-DD. Sheet A composed of five β-strands and sheet B composed of three core β-strands are coloured green and lime, respectively. α-Helices are coloured marine. PEGs and phosphate ion are shown in ball-and-stick representation and are labelled. The conserved MIND motif and spike loops of MK-DD are coloured purple and labelled in orange, respectively. (*b*) Topology of the overall structure of MK-DD. α-Helices and β-strands are represented by blue cylinders and green arrows, respectively.

pH 7.5, 100 m*M* NaCl prior to use. Experiments were carried out by titrating 400 µ*M* of the discoidin domain (MK-DD) into a cell containing protein at 20 µ*M* in identical buffer to the injectant. Data were analysed using the *Origin* software (MicroCal) and the stoichiometry (*n*), association constant (K_a) and change in enthalpy (*H*) were obtained by fitting the isotherm to a one-site binding model.

2.5. Protein Data Bank accession code

The atomic coordinates and structure factors for the mouse muskelin discoidin domain have been deposited in the Protein Data Bank (http://www.pdb.org/) and are accessible under the code 4pqq.

3. Results and discussion

MK-DD of mouse muskelin (residues 1–174), which corresponds to the 19 kDa fragment obtained from limited proteolysis using chymotrypsin (Fig. 1*b*), gave rhombohedral-shaped crystals that diffracted to 1.55 Å resolution. The structure was solved by SAD phasing using crystals grown from SeMet-substituted protein seeded by a native crystal. There is one molecule of MK-DD in the asymmetric unit with no significant contacts from neighbouring molecules, and this agrees well with the result from the gel-filtration experiment, *i.e.* with a monomeric form being the major peak (Fig. 1*b*). The final model includes protein atoms corresponding to residues 12–167, two PEG molecules, one phosphate and 203 water molecules refined to an *R* factor of 15.3% and an R_{free} of 17.6%.

3.1. Overall structure

The overall structure and the topology of MK-DD are shown in Figs. 2(*a*) and 2(*b*), respectively. As seen in the figure, the overall structure of MK-DD consists of eight β-strands in a jelly-roll motif with a five-stranded β-sheet facing a three-stranded β-sheet. The first β-sheet is composed of five anti-parallel β-strands (sheet A: strands β1, 18–23; β2, 58–62; β7, 128–138; β4, 84–94; β5, 97–104) and the second β-sheet is composed of three β-strands (sheet B: strands β3, 67–73; β6, 113–116; β8, 150–155). The two sheets are facing each other with hydrophobic residues interdigitating, *i.e.* residues Leu60, Leu62, Phe87, Val89, Leu102, Ile131 and Ile133 from sheet A and residues Val68, Ile71, Phe73, Phe115, Val150 and Leu152 from sheet B make interactions with the side chains of Phe87 in sheet A and Phe73 and Phe115 in sheet B, which form an aromatic cluster. In addition to this core structure, there are two short antiparallel β-structures formed by strands β2′–β8′ and β7′–β7″ and one short α-helix near the flank of β2′. The β2′ strand is on the flanking loop region between β1 and β2, and the β8′ strand is on the loop connecting strands β7 and β8. The residues in β2′ and β8′ are absolutely conserved and they serve as a lid to the β-barrel, with the side chains of Trp47 and Ile147 facing the centre of the barrel (Figs. 2 and 3). The β-structure of β7′–β7″ is positioned between strands β6 and β7. In addition, there are two β-hairpin structures: the first hairpin in the β4 and β5 strands involves residues Glu95/Glu96 and the second is formed by residues Asp122 and Glu123. The carboxyl-terminus of MK-DD forms an α-helix near the N-terminus. An almost identical structural arrangement has been observed in other members of the DD subfamily. Since the LisH domain follows this helix and LisH is predicted to have a highly helical secondary structure, it is possible that this region may be part of the same helix.

The loops connecting the strands at both the top and the bottom of the β-barrel protrude, as can be seen in Fig. 2(*a*). Both the N- and C-termini lie at one end of the β-barrel, similar to as observed in other DD structures, and there are additionally three loops. On the opposite side of the barrel there are six loops, namely

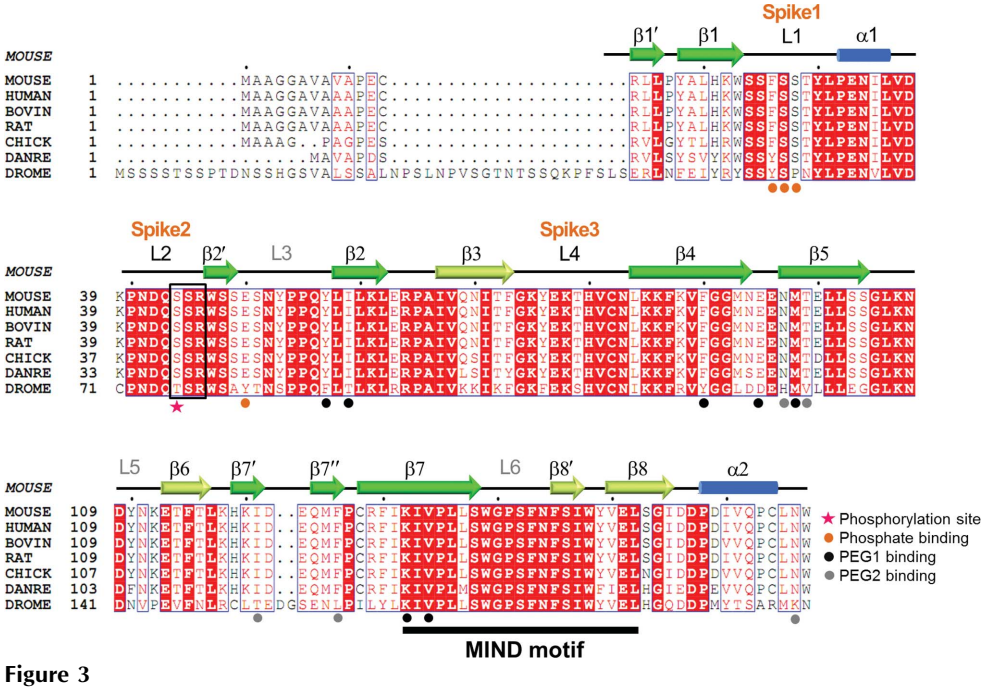


Figure 3
Sequence alignments of MK-DD. Discoidin domain sequences from *Mus musculus* (O89050), *Homo sapiens* (Q9UL63), *Bos taurus* (A5D7A3), *Rattus norvegicus* (Q99PV3), *Gallus gallus* (Q5ZI84), *Danio rerio* (Q8AYJ5) and *Drosophila melanogaster* (Q0E9A5) were aligned using *ClustalW*. SWISS-PROT entry numbers are given in parentheses. The residues in red boxes are strictly conserved and the phosphorylation site is highlighted by an asterisk. The secondary structures of MK-DD are depicted above the sequence; α-helices and β-strands are indicated by cylinders and arrows, respectively. Residues within 4 Å of the bound PEG1, PEG2 and phosphate are indicated by circles and the MIND motif is indicated by a black bar.

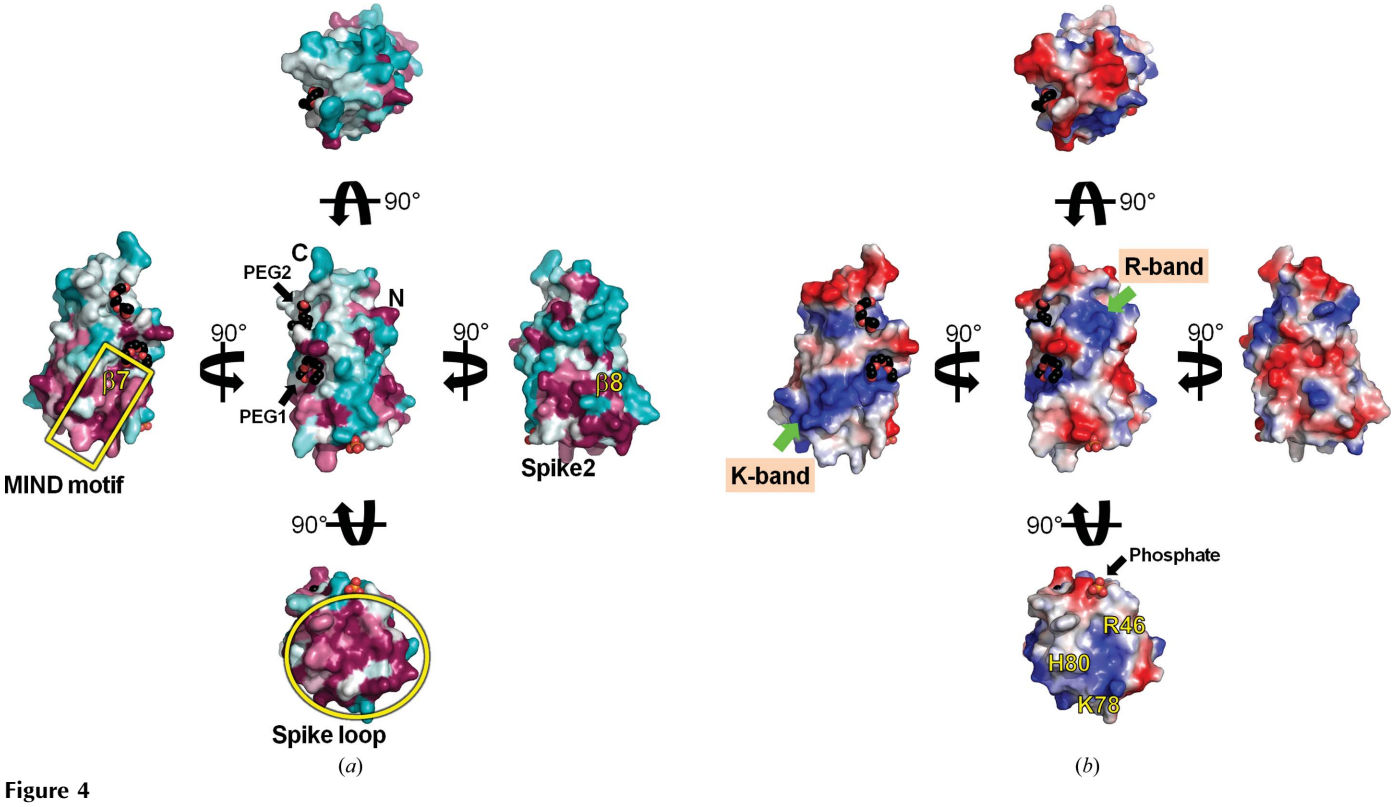


Figure 4
Surface properties of MK-DD. (*a*) Sequence conservation of the muskelin discoidin domain. Residues are colour-coded based on sequence conservation from maroon to blue to white as the degree of conservation decreases. The highly conserved MIND motif and spike loops are indicated in yellow. (*b*) The electrostatic potential of MK-DD calculated by *PyMOL* is shown in blue for positively charged areas and red for negatively charged areas. The K-band and R-band are marked by green arrows.

L1–L6, and it is on this surface that ligands are reported to bind in other DD proteins (Fig. 2*a*). The three spike loops that are highly conserved within the family correspond to loops L1, L2 and L4 and they are also reasonably well conserved in muskelin, as indicated in Fig. 3. Since MK-DD of muskelin is highly conserved, *e.g.* mouse MK-DD shares a sequence identity of about 99% with that from human and above 86% identity with MK-DD from other mammalian species, the structure of these will be the same. On the other hand, *Drosophila melanogaster* DD shares only 53% sequence identity with human MK-DD and the N-terminus is longer by 32 residues in this case.

3.2. Structural features that may be relevant to function

There are two features of MK-DD that should be pointed out. Firstly, of the seven phosphorylation sites in animal muskelins, five by protein kinase C and two by casein kinase 2, MK-DD includes one site, namely Ser44 (Adams, 2002). As seen in Fig. 3, the highly conserved Ser44 is located on spike2 and is therefore is readily available for phosphorylation. Secondly, the highly conserved 21-residue stretch (residues 132–152) called the muskelin identity in N-terminal domain (MIND) motif (Fig. 3) is identified as part of strands β7 and β8 with L6 in between in the structure. These residues are highlighted in purple in Fig. 2(*a*). The sequence is strictly conserved from vertebrates to *Drosophila*, while only 13 residues are conserved in *Schizosaccharomyces pombe* (Adams, 2002; Prag *et al.*, 2004). The loop region of the MIND motif, *i.e.* L6, is somewhat exposed, but many of the residues are not readily accessible, *i.e.* the hydrophobic residues on the loop are stabilized by adjacent spike loops. For example, Trp148 of L6 makes hydrophobic contacts with Arg46 and Tyr76 of spike2 and spike3. Therefore, the MIND motif appears to play an important role in stabilizing this region of the structure.

The electrostatic nature and the conservation reveal several interesting points (Fig. 4) which might shed some light on possible protein–protein interactions. Firstly, there are two positively charged bands, referred to as the K-band and the R-band. The K-band consists of five lysine residues, namely Lys85, Lys86 and Lys88 from β4, Lys107 from L5 and Lys132 from β7, while the R-band consists of three arginines and two lysines, namely the residues Arg14 from $\beta1'$, Arg64 from the β2–β4 loop, Arg129 from β7, Lys61 from β2 and Lys118 from the β6–β7 loop. The K-band forms a narrow long positive groove along with the MIND region, as illustrated in Fig. 4, and is highly conserved. In contrast, the residues in the R-band show low conservation. There are two PEG molecules bound near these regions and these will be described later. Another noticeable feature is found around the spike loops. As mentioned earlier, this region is highly conserved and displays both a positively charged surface, contributed to by Arg46 from spike2 and His80 and Lys78 from spike3, and a hydrophobic surface.

Most of the reported eukaryotic DDs are extracellular or transmembrane proteins. The N- and C-terminal ends of the eukaryotic DDs are linked by a disulfide bridge (Supplementary Fig. S1[1]) and these cysteine residues are considered to be markers of the beginning and end of the DD domain (Kiedzierska *et al.*, 2007). This kind of architectural signature is often found in multi-domain proteins, and it facilitates domain shuffling and rearrangement. Based on this knowledge and on the results of an Ellman's assay, muskelin was suggested to have two disulfide bonds, one in the DD and another in the kelch-repeat domain (Kiedzierska *et al.*, 2008). The crystal structure of MK-DD shows no disulfide bridges, although there are four cysteine residues, namely Cys13, Cys82 and Cys128 in DD and Cys164 in the C-terminal helix α2. Except for Cys82, which is located on spike3, the cysteine residues are not solvent-accessible and they are not in close proximity (Supplementary Fig. S1). Muskelin functions as an intracellular protein in the cytosol and there has also been a report of a membrane-associated fraction that acts as a cytosolic transporter involved in cell signalling and cytoskeletal organization (Valiyaveettil *et al.*, 2008). To date, only two other DD-containing proteins have been found to localize in the intracellular fraction, namely XRCC1-NTD (X-ray cross-complementing group 1 N-terminal domain; PDB entry 1xna) and APC10 (anaphase-promoting complex subunit 10; PDB entry 1jhj), which mediate the specific binding as a component of a large multi-molecular complex (Marintchev *et al.*, 1999; Wendt *et al.*, 2001).

As mentioned earlier, one phosphate and two PEG molecules, most likely from the crystallization solution, are located in the electron-density map (Fig. 5). Phosphate is bound near spike1 and L3 (Fig. 2*a*); it forms direct hydrogen bonds to Ser28 and Glu50 and is further stabilized by the backbone carbonyl O atoms of Phe26 and Ser27 of the spike1 loop through water molecules (Fig. 5*a*). Since muskelin is reported to associate with the membrane in addition to being present in the cytoplasm (Adams *et al.*, 1998), initially this was thought to be a possible membrane-binding region based on the fact that phosphate often indicates a possible position of the phospho-head group of the eukaryotic plasma membrane (Karathanassis *et al.*, 2002). However, considering the fact that the proximity of the bound phosphate is negatively charged, as seen in Fig. 4(*b*), it would be unlikely that this surface would associate with the membrane. A possible mechanism of membrane association for MK-DD will be discussed later.

After all of the protein atoms had been placed, there still remained two contiguous extra electron densities on the side of the β-barrel with residues 96–100 lying in between them. The ethylene glycol units from PEG 1500 in the crystallization buffer accounted well for the extra density (as shown in Fig. 5*b*). The first PEG molecule makes hydrophobic interactions with the side-chain atoms of Tyr57, Ile59, Phe90, Met98 and Val134, and forms a hydrogen bond to residue Lys132 that is located in the centre of sheet A. The second PEG molecule is located between two β-hairpin structures (β4–β5 and $\beta7'$–$\beta7''$) and the C-terminal helix. In this case, it forms a hydrogen bond to the side-chain atom of Asn166 in addition to nonbonded contacts with Asn97, Thr99, Ile121 and Phe126. It is noteworthy that a PEG molecule is also found in the N-terminal DD of discoidin II and is also located on the surface of the five β-stranded β-sheet, which is somewhat positively charged (Aragão *et al.*, 2008).

3.3. Comparison with other discoidin-like domains

In order to search for structurally similar proteins, the *DALI* server was used (Holm & Rosenström, 2010). Hits included galactose-binding domain-like fold and discoidin family proteins, with *Z*-scores ranging from 17 to 13 and r.m.s.d. values between 2 and 3 Å for 124–134 aligned C^{α}

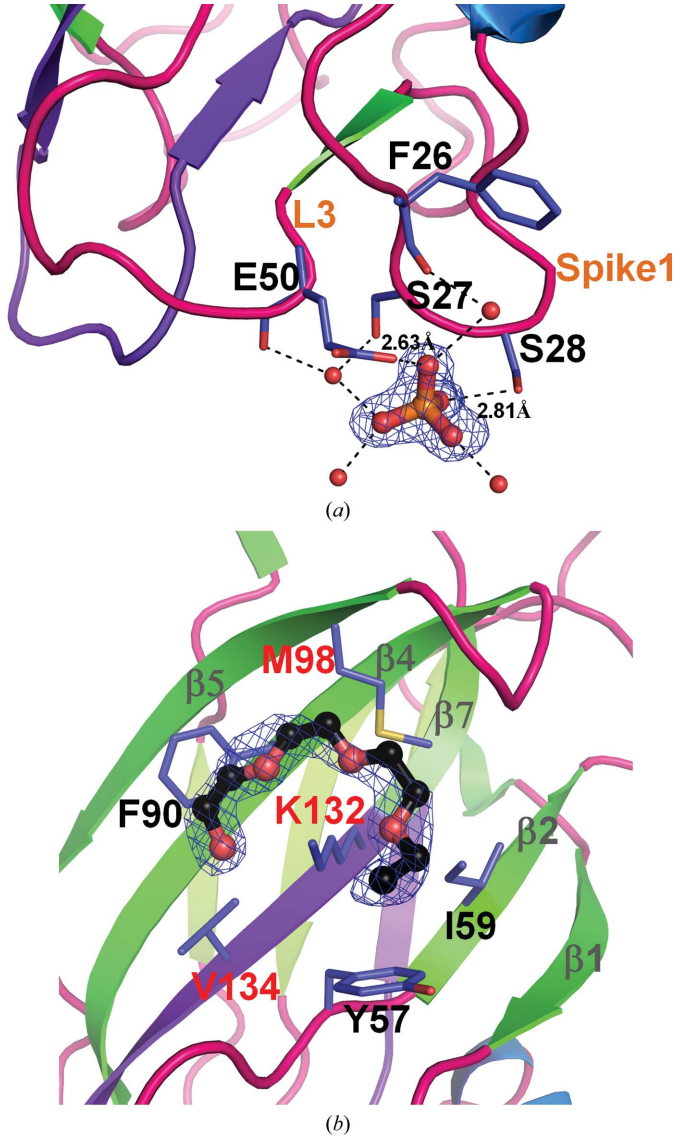


Figure 5
Phosphate and PEG binding to MK-DD. The electron-density map ($2F_o - F_c$) of (*a*) phosphate and (*b*) PEG1 is contoured at the 1σ level and the residues involved in binding are shown as stick models. Hydrogen bonds are indicated as dotted lines.

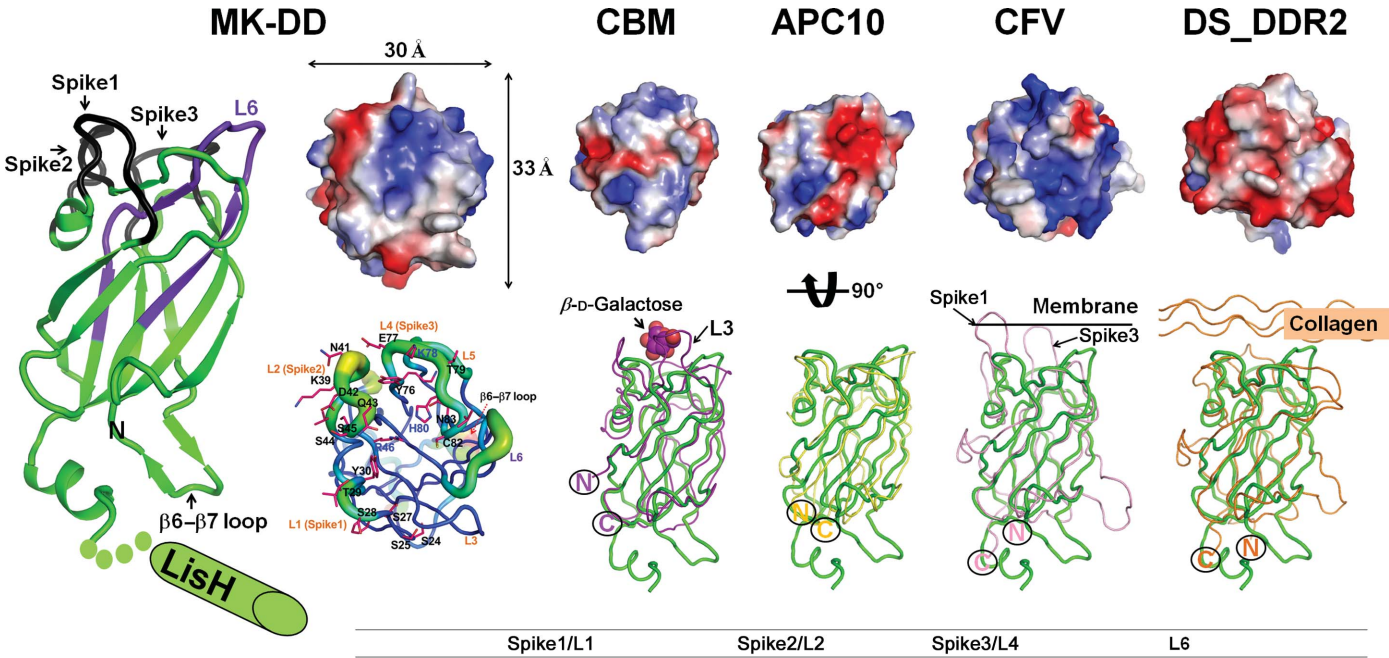
[1] Supporting information has been deposited in the IUCr electronic archive (Reference: MH5139).

Table 2
Comparison of discoidin domains.

Mv, *Micromonospora viridifaciens*; *Dd*, *Dactylium dendroides*; *Ms*, *Morone saxatilis*; *h*, human; *b*, bovine.

Protein	PDB code	R.m.s.d. (Å) (aligned C^{α} atoms)	Sequence identity (%)	Cellular localization (disulfide bond)	Ligands	Biological process
*Mv*CBM	2bzd	1.79 (122)	12	Extracellular	Carbohydrate	Cell adhesion, pathogenic factors
*Dd*GO	1gof	1.98 (124)	10	Extracellular (+)	Carbohydrate	Cell adhesion, oxidation of primary alcohols
*Ms*FBP32	3cqo	2.13 (124)	12	Secreted (+)	Carbohydrate	Cell adhesion
*h*APC10	1jhj	2.19 (125)	14	Intracellular	Protein	Cell cycle, Ubl conjugation pathway
*h*XRCC1-NTD	3k77	2.10 (117)	19	Intracellular	Protein	DNA damage, DNA repair
*h*CFV	1czt	2.13 (129)	17	Extracellular (+)	Phospholipids	Blood coagulation, haemostasis
*b*Lact-C2	3bn6	2.01 (124)	18	Extracellular (+)	Phospholipids	Angiogenesis, cell adhesion, fertilization
*h*DS_DDR1	4ag4	2.23 (125)	15	Transmembrane (+)	Collagen	Lactation, pregnancy
*h*DS_DDR2	2wuh	2.17 (124)	17	Transmembrane (+)	Collagen	Osteogenesis
*h*b1_ Npn1	1kex	2.14 (128)	18	Transmembrane (+)	Semaphorin	Angiogenesis, differentiation, neurogenesis

atoms. They can be divided into four groups depending on the ligand types. The first group consists of carbohydrate-binding modules (CBMs) of the bacterial sialidase family including *N*-acetylneuraminosyl glycohydrolase, neuraminidase (CBM; PDB entry 2bzd; Newstead *et al.*, 2005) and galactose oxidase (GO; PDB entry 1gof; Ito *et al.*, 1991), and sugar-binding protein FBP32 of the lectin family (FBP32; PDB entry 3cqo; Bianchet *et al.*, 2010). They bind various carbohydrates and their derivatives, which are attached to the cell surface. The second group consists of the APC10/Doc1 subunit of the human anaphase-promoting complex (APC10; PDB entry 1jhj; Wendt *et al.*, 2001) and the N-terminal domain of DNA-repair protein XRCC1 (XRCC1-NTD; PDB entry 3k77; Cuneo & London, 2010). These two are components of a multi-subunit complex in the cytoplasm and they bind nucleotides or proteins. The third group includes the membrane-binding C2 domain of human coagulation factors V/VIII (CFV; PDB entry 1czt; Macedo-Ribeiro *et al.*, 1999)

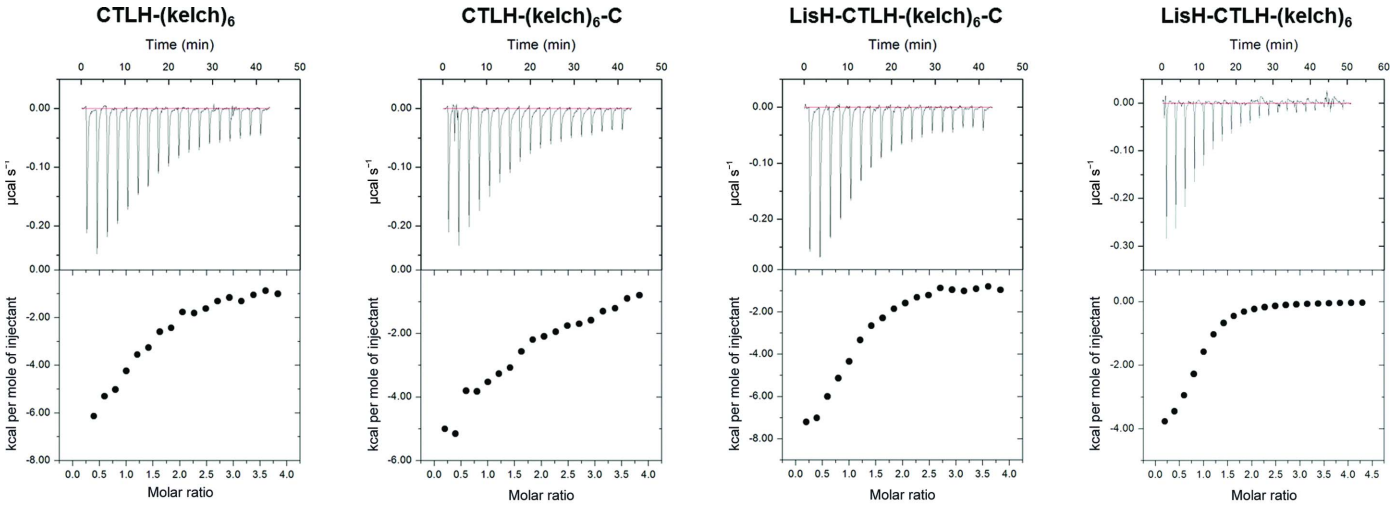


	Spike1/L1	Spike2/L2	Spike3/L4	L6
MK-DD	^{24}SSFSSTYL^{31}	^{37}VDKPNDQSSR^{46}	^{74}GKYEKTHVCN^{83}	^{139}WGPSFNFS^{146}
CBM (2bzd)	^{515}DSEETAREDGR^{525}	^{531}DGNPSTF^{537}	^{571}RRQNSANEQVA^{581}	^{630}QTGHKYAA^{637}
APC10 (1jhj)	^{34}SSCKPGFG^{41}	^{47}DDNLETY^{53}	^{81}ADYKSDESYTP^{91}	^{145}HQNGRDTH^{152}
CFV (1czt)	^{21}SFKKSWWGDYWE^{32}	^{38}RLNAQGRVNA^{46}	^{75}GCKSLSSEMY^{84}	^{144}NQSIT^{148}
DDR2 (2wuh)	^{49}SSQWSESTA^{57}	^{61}GRLDSEEGDGA^{71}	^{104}GRHAGGHGIEFAP^{116}	^{170}TDHSMNVC^{177}

Figure 6
Structural comparison of discoidin domains. Representatives of each groups superposed on MK-DD. Structures of CBM (PDB entry 2bzd; bacterial sialidase), APC10 (PDB entry 1jhj; APC10/DOC1 subunit of human anaphase-promoting complex), CFV (PDB entry 1czt; human coagulation factor V) and DS_DDR2 (PDB entry 2wuh; discoidin domain receptor 2) are superposed onto MK-DD using C^{α} atoms and are shown in magenta, yellow, pink and orange, respectively. A C^{α} drawing of MK-DD is shown on the left. The electrostatic surface and *B* factors of MK-DD are shown next. Electrostatic surface of the ligand-binding region and the C^{α} drawings with ligands for CBM, APC10, CFV and DDR2 are shown at the top and bottom, respectively. Bound β-D-galactose and collagen are indicated. The sequences of the three spike loops and L6 are listed.

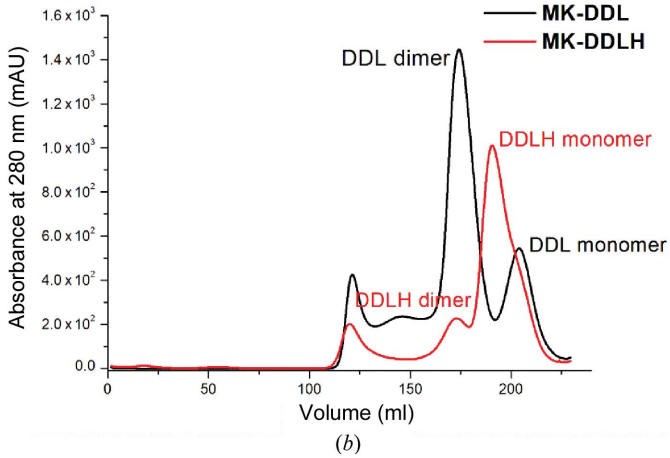
and the lactadherin C2 domain (Lact-C2; PDB entry 3bn6; Shao *et al.*, 2008). They bind phospholipids on the outside of the mammalian cell membrane. The fourth group includes discoidin domain receptor 1 (DS_DDR1; PDB entry 4ag4; Carafoli *et al.*, 2012), discoidin domain receptor 2 (DS_DDR2; PDB entry 2wuh; Carafoli *et al.*, 2009) and neuropilin-1/2 b1 domain (b1_Npn1; PDB entry 1kex; Lee *et al.*, 2003). These proteins bind collagen and semaphorin and regulate a variety of cellular and developmental processes.

Representatives of each group are shown in Fig. 6. Superposition of these on MK-DD shows r.m.s.d. values ranging from 1.79 to 2.23 Å for about 120 residues, indicating a good agreement for the core structure (Table 2). Despite the structural similarities, they share sequence identities of only 10–19%. As mentioned above, the overall structure is almost identical at the core and the structural differences mostly involve the loop regions at the top and bottom of the barrel. The end with the N- and C-termini shows less variation since the loops are short; however, MK-DD does have a few unique features. Firstly, there is an additional α-helix that would be connected to the LisH domain in MK-DD and this is tightly packed against the residues connecting β2 and β3. Secondly, the residues between strands β6 and β7 of MK-DD are much longer than those in CBM, CFV or DDR2, and these residues



	n	K_d (µ*M*)	ΔH (kcal mol^{-1})	ΔS (cal mol^{-1} K^{-1})	
MK-HK(208–662)	1.04 ± 0.15	19.3	−14.08 ± 2.64	−25.7	H K
MK-HKC(208–735)	1.39 ± 0.29	29.6	−10.96 ± 3.20	−16.0	H K C
MK-LHKC(151–735)	1.01 ± 0.08	10.4	−12.03 ± 1.28	−17.5	L H K C
MK-LHK(151–662)	0.84 ± 0.07	2.2	−4.38 ± 0.53	11.2	L H K

(*a*)



(*b*)

Figure 7
Analysis of the molecular association of muskelin. (*a*) Interaction between MK-DD and MK-HK, MK-HKC, MK-LHKC and MK-LHK. ITC measurements were carried out in 20 m*M* HEPES, 100 m*M* NaCl pH 7.5. MK-DD was injected into four constructs containing kelch-repeat domains. The binding isotherm was fitted to a one-site model with 1:1 stoichiometry. A summary of the thermodynamic properties is shown in the table. (*b*) SEC profile of MK-DDL and MK-DDLH.

form a hairpin structure with $\beta7'$ and $\beta7''$ arranged in an antiparallel manner. This region is also longer in *h*APC10 and *h*XRCC2-NTD, which are also intracellular proteins, but these proteins adopt a somewhat different conformation (Fig. 6 and Supplementary Fig. S2).

On the other hand, the structure at the opposite end of the barrel, where the ligands bind, differs. In all cases the antiparallel $\beta2'$–$\beta8'$ pair acts as a lid to the barrel with Trp from $\beta2'$ and Ile/Leu/Met from $\beta8'$ facing the centre of the barrel (Fig. 2). However, the length and the conformation of the loops vary and provide a different surface, as depicted in Fig. 6. CBM has 37 residues between $\beta1$ and $\beta2$, which includes spike1 to L3, while MK-DD has 34 residues (Fig. 3). Compared with MK-DD, spike1 and L3 of CBM are longer by three and five residues, respectively, and they protrude significantly and stabilize the bound carbohydrate, as seen in Fig. 6. The surface reflects these properties by having both hydrophobic and polar residues. In the case of the C2 domain of CFV, spike1 is four residues longer and the spike loops harbour more positively charged residues as well as more bulky hydrophobic residues. Spike1 and spike3, in particular the indoles of Trp26 and Trp27 from spike1 and Leu79 from spike3, are suggested to be immersed in the apolar membrane core, while the basic patch formed by several Lys and Arg residues on the spike loops makes favourable contacts with negatively charged membrane phosphate groups (Macedo-Ribeiro *et al.*, 1999). Similarly, the cationic β-groove formed by the spike loops in the PKCα C2 domain is reported to be associated with membrane binding (Lemmon, 2008). The DD of DDR2 displays a hydrophobic surface surrounded by a negatively charged surface, as shown in Fig. 6. Both DDR1 and DDR2 bind to collagen through the DD, and Trp52 and Thr56 in spike1 and Arg105 and Glu113 in spike3, which are strictly conserved in both, interact with collagen. The lengths of the loops of APC10 are similar to those in MK-DD, but the compositions are such that the surface has a quite different charge distribution. In the case of the lactadherin C2 domain, hydrophobic residues (Trp26, Leu28, Phe31 and Phe81) are proposed to interact with the membrane (Shao *et al.*, 2008).

As seen in Figs. 4 and 6, MK-DD has a relatively flat surface with a positively charged patch formed by Arg46, His80 and Lys78 with Val81 from spike3 and Pro141 and Phe143 from L6 protruding. Most of the residues on this surface are highly conserved, with the abovementioned residues being strictly conserved (Figs. 3 and 4*a*). Also nearby is the highly conserved MIND motif (Fig. 4). The presence of the protruding hydrophobic residues of L6 and the positive patch mimics the C2 domains somewhat; however, the surface nature of muskelin is unique and significantly different from those of other DDs, suggesting that MK-DD would bind different ligands and binding partners (Fig. 6). Muskelin is known to be a cytosolic protein, but its membrane association has been described in numerous studies (Adams *et al.*, 1998; Tagnaouti *et al.*, 2007). In the case of $GABA_AR$ $\alpha1$, the cytosolic region between TM3 and TM4 (residues 399–420) has been reported to make a direct interaction with residues 90–200 of muskelin (Heisler *et al.*, 2011), and these include the residues of L6 and the MIND region of DD as well as the LisH motif of muskelin.

3.4. LisH modulates interaction between the discoidin and kelch-repeat domains

Although the functional significance and the physiological mechanism are not fully understood, muskelin is reported to self-associate through a head-to-tail mechanism involving the DD (Prag *et al.*, 2004, 2007). In order to investigate this, further binding analysis of various fragments of muskelin, namely MK-HK, MK-HKC, MK-LHK and MK-LHKC, with MK-DD (see Fig. 1*a*) has been carried out. The binding affinities resulting from ITC experiments are summarized in Fig. 7(*a*). MK-DD binds to MK-LHKC with a dissociation constant (K_d) of 10.4 μM. However, when the C-terminal region is removed, *i.e.* MK-LHK, the K_d value becomes 2.2 μM. A similar trend is seen for MK-HKC *versus* MK-HK, *i.e.* there is an increase in the K_d value of about 1.5-fold, suggesting that the C-terminal 80 or so residues perhaps interfere in the interaction. On the other hand, the LisH motif appears to play the opposite role. There is about a tenfold and a threefold decrease in the K_d values of MK-LHK *versus* MK-HK and MK-LHKC *versus* MK-HKC, respectively.

Size-exclusion chromatography (SEC) shows that MK-DD exists as a monomer, while the full-length muskelin eluted as a size larger than a dimer. PAGE analysis under nondenaturating conditions gives a molecular weight of about 540 kDa (data not shown). Earlier, Kiedzierska *et al.* (2008) reported the full-length protein to be a hexameric unit based on gel-filtration and native PAGE analysis. Since both the LisH and CTLH domains are known to be dimerization and oligomerization motifs (Emes & Ponting, 2001), we tested whether the LisH and/or CTLH domains have any effect on the molecular status of muskelin using MK-DDL and MK-DDLH. As seen in Fig. 7(*b*), MK-DDL exists predominantly as a dimer, while MK-DDLH mostly exists as a monomer. On native PAGE both MK-DDL and MK-DDLH appeared as tetrameric and higher oligomeric units (data not shown). This suggests that the LisH domain induces dimerization and possibly further oligomerization, while the CTLH domain seems to interfere with further oligomerization, possibly by forming a stable interaction with the LisH and CTLH domains. However, the effect of these domains on dimerization and the binding affinity could also be through contact with the other domains in muskelin. A structural analysis using the full-length protein is necessary to assess the physiologically relevant mechanism.

Based on the findings that muskelin forms particles in the intact cell and both the kelch-repeat domain and the DD are necessary for particle formation, and the fact that a number of kelch-repeat proteins tend to self-associate, the two domains of muskelin have been suggested to be involved in both *cis* and *trans* interactions (Prag *et al.*, 2004). In fact, the two domains in the galactose oxidase crystal structure show extensive interactions including backbone-to-backbone hydrogen bonds between the DD and the kelch-repeat domain (Ito *et al.*, 1991). Superposition of the two structures suggests that similar

interactions may well be possible for muskelin. Furthermore, the $\beta7'$–$\beta7''$ of the loop connecting $\beta6$ and $\beta7$ which is unique to MK-DD is positioned such that it may contribute to the stabilization of such an interaction. Our ITC and SEC results above suggest that the domains of muskelin participate in both intermolecular and intramolecular interactions. The DD alone does not form a dimer, but a dimer is formed when LisH is present. Also, the interaction between MK-DD and the kelch-repeat domain is affected by the presence of LisH and the C-terminal region. Therefore, it appears that LisH may be involved in the self-association and eventual oligomerization of DD and the kelch-repeat domain. LisH and the C-terminus have been reported to regulate the subcellular localization of muskelin (Valiyaveettil *et al.*, 2008). Based on this, LisH and the C-terminal region seem to regulate both self-association and subcellular localization.

4. Conclusions

Muskelin is an intracellular protein belonging to the kelch-repeat-containing protein superfamily. It plays a pivotal role in the cytoskeletal response and acts as an intracellular mediator in receptor-mediated signalling by interacting with a large number of proteins including EP3 isoform α, p39, RanBPM, the TBX20b isoform and $GABA_AR$ $\alpha1$. The localization and domain architecture of muskelin are reported to reflect its cellular function. The crystal structure of MK-DD determined in this study shows a β-barrel structure with two short helices at the N-terminal end. The opposite end of the barrel has a somewhat positively charged surface formed by Arg46, His80 and Lys78 with Val81 from spike3, Pro141 and Phe143 from L6 protruding. This surface, together with the highly conserved MIND motif which includes the L6 loop, might be the binding site for partner proteins or intramolecular domains. Also, the residues between $\beta6$ and $\beta7$ form a hairpin structure which is unique and may participate in interaction with partner proteins or domains. Analysis of the interdomain interactions suggests the association of DD and kelch-repeat domain, with LisH playing an important role in self-association. The results from this study should serve as a basis for understanding the molecular mechanisms and eventually the cellular function of muskelin.

We thank the beamline staff at the Pohang Light Source, Republic of Korea (beamline 5C-SBII) for assistance during data collection. This work was supported financially by the Global Research Laboratory Program (grant No. 2013056409) of the Ministry of Science, ICT and Future Planning of Korea and an institutional grant from the Korea Institute of Science and Technology.

References

Adams, J. C. (2002). *Gene*, **297**, 69–78.
Adams, J. C., Seed, B. & Lawler, J. (1998). *EMBO J.* **17**, 4964–4974.
Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Aragão, K. S., Satre, M., Imberty, A. & Varrot, A. (2008). *Proteins*, **73**, 43–52.
Baumgartner, S., Hofmann, K., Chiquet-Ehrismann, R. & Bucher, P. (1998). *Protein Sci.* **7**, 1626–1631.
Bianchet, M. A., Odom, E. W., Vasta, G. R. & Amzel, L. M. (2010). *J. Mol. Biol.* **401**, 239–252.
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.
Carafoli, F., Bihan, D., Stathopoulos, S., Konitsiotis, A. D., Kvansakul, M., Farndale, R. W., Leitinger, B. & Hohenester, E. (2009). *Structure*, **17**, 1573–1581.
Carafoli, F., Mayer, M. C., Shiraishi, K., Pecheva, M. A., Chan, L. Y., Nan, R., Leitinger, B. & Hohenester, E. (2012). *Structure*, **20**, 688–697.
Cuneo, M. J. & London, R. E. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 6805–6810.
Debenedittis, P., Harmelink, C., Chen, Y., Wang, Q. & Jiao, K. (2011). *Biochem. Biophys. Res. Commun.* **409**, 338–343.
Emes, R. D. & Ponting, C. P. (2001). *Hum. Mol. Genet.* **10**, 2813–2820.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Firbank, S. J., Rogers, M. S., Wilmot, C. M., Dooley, D. M., Halcrow, M. A., Knowles, P. F., McPherson, M. J. & Phillips, S. E. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 12932–12937.
Francis, O., Han, F. & Adams, J. C. (2013). *PLoS One*, **8**, e75217.
Franco-Pons, N., Virgos, C., Vogel, W. F., Ureña, J. M., Soriano, E., del Rio, J. A. & Vilella, E. (2006). *Neuroscience*, **140**, 463–475.
Hasegawa, H., Katoh, H., Fujita, H., Mori, K. & Negishi, M. (2000). *Biochem. Biophys. Res. Commun.* **276**, 350–354.
Heisler, F. F., Loebrich, S., Pechmann, Y., Maier, N., Zivkovic, A. R., Tokito, M., Hausrat, T. J., Schweizer, M., Bähring, R., Holzbaur, E. L., Schmitz, D. & Kneussel, M. (2011). *Neuron*, **70**, 66–81.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Ito, N., Phillips, S. E., Stevens, C., Ogel, Z. B., McPherson, M. J., Keen, J. N., Yadav, K. D. & Knowles, P. F. (1991). *Nature (London)*, **350**, 87–90.
Karathanassis, D., Stahelin, R. V., Bravo, J., Perisic, O., Pacold, C. M., Cho, W. & Williams, R. L. (2002). *EMBO J.* **21**, 5057–5068.
Kiedzierska, A., Czepczynska, H., Smietana, K. & Otlewski, J. (2008). *Protein Expr. Purif.* **60**, 82–88.
Kiedzierska, A., Smietana, K., Czepczynska, H. & Otlewski, J. (2007). *Biochim. Biophys. Acta*, **1774**, 1069–1078.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Ledee, D. R., Gao, C. Y., Seth, R., Fariss, R. N., Tripathi, B. K. & Zelenka, P. S. (2005). *J. Biol. Chem.* **280**, 21376–21383.
Lee, C. C., Kreusch, A., McMullan, D., Ng, K. & Spraggon, G. (2003). *Structure*, **11**, 99–108.
Lemmon, M. A. (2008). *Nature Rev. Mol. Cell Biol.* **9**, 99–111.
Lin, L., Huai, Q., Huang, M., Furie, B. & Furie, B. C. (2007). *J. Mol. Biol.* **371**, 717–724.
Macedo-Ribeiro, S., Bode, W., Huber, R., Quinn-Allen, M. A., Kim, S. W., Ortel, T. L., Bourenkov, G. P., Bartunik, H. D., Stubbs, M. T., Kane, W. H. & Fuentes-Prior, P. (1999). *Nature (London)*, **402**, 434–439.
Marintchev, A., Mullen, M. A., Maciejewski, M. W., Pan, B., Gryk, M. R. & Mullen, G. P. (1999). *Nature Struct. Biol.* **6**, 884–893.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Newstead, S. L., Watson, J. N., Bennet, A. J. & Taylor, G. (2005). *Acta Cryst.* D**61**, 1483–1491.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Prag, S., Collett, G. D. & Adams, J. C. (2004). *Biochem. J.* **381**, 547–559.
Prag, S., De Arcangelis, A., Georges-Labouesse, E. & Adams, J. C. (2007). *Int. J. Biochem. Cell Biol.* **39**, 366–378.

Rogers, M. S., Tyler, E. M., Akyumani, N., Kurtis, C. R., Spooner, R. K., Deacon, S. E., Tamber, S., Firbank, S. J., Mahmoud, K., Knowles, P. F., Phillips, S. E., McPherson, M. J. & Dooley, D. M. (2007). *Biochemistry*, **46**, 4606–4618.
Shao, C., Novakovic, V. A., Head, J. F., Seaton, B. A. & Gilbert, G. E. (2008). *J. Biol. Chem.* **283**, 7230–7241.
Tagnaouti, N., Loebrich, S., Heisler, F., Pechmann, Y., Fehr, S., De Arcangelis, A., Georges-Labouesse, E., Adams, J. C. & Kneussel, M. (2007). *BMC Neurosci.* **8**, 28.
Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 38–44.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Umeda, M., Nishitani, H. & Nishimoto, T. (2003). *Gene*, **303**, 47–54.
Valiyaveettil, M., Bentley, A. A., Gursahaney, P., Hussien, R., Chakravarti, R., Kureishy, N., Prag, S. & Adams, J. C. (2008). *J. Cell Biol.* **182**, 727–739.
Vogel, W. (1999). *FASEB J.* **13**, S77–S82.
Wendt, K. S., Vodermaier, H. C., Jacob, U., Gieffers, C., Gmachl, M., Peters, J. M., Huber, R. & Sondermann, P. (2001). *Nature Struct. Biol.* **8**, 784–788.
Wu, W. W. H. & Molday, R. S. (2003). *J. Biol. Chem.* **278**, 28139–28146.