



Published in final edited form as:

Structure. 2011 March 9; 19(3): 324–336. doi:10.1016/j.str.2010.11.017.

Structural Characterization of the Boca/Mesd Maturation Factors for LDL-Receptor-Type β -Propeller Domains

Mark N. Collins¹ and Wayne A. Hendrickson^{1,2,*}

¹Department of Biochemistry and Molecular Biophysics Columbia University, New York, NY 10032, USA

²Howard Hughes Medical Institute, Columbia University, New York, NY 10032, USA

SUMMARY

Folding and trafficking of low density lipoprotein receptor (LDLR) family members, which play essential roles in development and homeostasis, is mediated by specific chaperones. The Boca/Mesd chaperone family specifically promotes folding and trafficking of the YWTD β -propeller-EGF domain pair found in the ectodomain of all LDLR members. Limited proteolysis, NMR spectroscopy, analytical ultracentrifugation and x-ray crystallography were used to define a conserved core comprised of a structured domain that is preceded by a disordered N-terminal region. High-resolution structures of the ordered domain were determined for homologous proteins from three metazoans. Seven independent protomers reveal a novel ferredoxin-like superfamily fold with two distinct β -sheet topologies. A conserved hydrophobic surface forms a dimer interface in each crystal, but these differ substantially at the atomic level, indicative of non-specific hydrophobic interactions that may play a role in the chaperone activity of Boca/Mesd family.

Keywords

boca; crystal structure; mesd; molecular chaperone; protein folding; YWTD propeller; LDLR; LRP

INTRODUCTION

Molecular chaperones assist in cellular folding and maturation of many proteins. The relevant folding partners in many instances are quite promiscuous with activity toward many substrate proteins. Such chaperones include the ubiquitous Hsp70 chaperones and the Hsp60/TriC chaperonins. Other chaperones have a more restricted clientele, such as with Hsp90 involvement in the maturation of protein kinases and nuclear hormone receptors (Hartl and Hayer-Hartl, 2002; Anelli and Sitia, 2008). In yet other cases, private chaperones play a highly specialized role (Anelli and Sitia, 2008). For the low-density lipoprotein receptor (LDLR) and lipoprotein receptor-related proteins (LRPs), at least two private chaperones are dedicated to the folding and maturation of these multi-domain proteins as they move from the endoplasmic reticulum (ER) for presentation at the cell surface: (i) the

© 2011 Elsevier Inc. All rights reserved

*Correspondence: wayne@convex.hhmi.columbia.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

LRP-receptor associated protein (RAP) and (ii) the Boca/Mesd chaperone (Culi and Mann, 2003; Hsieh et al., 2003). RAP binds the cysteine-rich ligand binding modules of LDLR family members, preventing premature association with their respective ligands. RAP escorts fully folded LDLR family members from the ER to the Golgi, where the lower pH triggers RAP to dissociate and, recycle back to the ER via its retention signal (Herz, 2006). The molecular mechanism of the Boca/Mesd chaperone family is less clear. In *Drosophila*, Boca was shown to promote maturation and surface expression of several LDLR family members, through a direct, but transient, interaction (Culi and Mann, 2003; Culi et al., 2004).

The LDLR gene family represents a functionally diverse set of transmembrane proteins. The seven closely related members of this family participate in a wide range of biologically important, and often essential, roles, ranging from lipid metabolism (for LDLR itself), to embryonic development (for LRP5 and LRP6) and formation of the neuromuscular junction (LRP4) (Kim et al., 2008; MacDonald et al., 2009; May et al., 2005; Zhang et al., 2008). Other family members are multifunctional: both VLDLR (very low-density lipoprotein receptor) and APOER2 (apolipoprotein E receptor 2), like LDLR, function in cholesterol uptake but also regulate neuronal shape, migration, and synaptic plasticity (Herz, 2009). Mutations in the genes coding for these proteins have been associated with a wide spectrum of diseases including familial hypercholesterolemia (for LDLR), other less common but severe hereditary diseases such as autosomal dominant coronary artery disease type 2 (LRP6), familial exudative vitreoretinopathy and osteoporosis pseudoglioma (LRP5), as well as in cancers (LRP5 and LRP6) (Björklund et al., 2009; Li et al., 2004). Several members (ApoER2 and LRP1) have also been implicated in Alzheimer disease (Marzolo and Bu, 2009).

LDLR protein family members are characterized by the presence of three distinct types of extracellular domains; (i) cysteine-rich ligand binding modules (also called complement, or LDL-A domains), (ii) epidermal growth factor-like (EGF-like) domains, and (iii) YWTD β -propeller repeats, (named for their highly conserved tyrosine-tryptophan-threonine-aspartate sequences). The number and arrangement of domains varies greatly among these receptors; and, in humans, sizes range five-fold from the smallest LDLR (95 kDa) to the largest LRP2 (522 kDa) (Lillis et al., 2008; Marzolo and Bu, 2009). The YWTD motifs are repeated to form a six-bladed YWTD β -propeller. Each blade of the propeller has four anti-parallel β -strands, which are offset from the YWTD repeats such that the 'first' strand of the 6th blade is a C-terminal strand that follows the 5th blade to complete circularization of the toroidal YWTD β -propeller domain. The YWTD β -propeller is followed by an EGF-like domain that packs against the propeller to create a continuous hydrophobic core in the mature protein (Jeon et al., 2001).

Boca/Mesd family members are localized to the ER where they aid in the folding of YWTD β -propeller domains. The quality control process within the secretory pathway oversees protein folding, heteromeric assembly, and removal of defective proteins. Unfolded and misfolded proteins are prevented from continuing through the secretory pathway, until they are either correctly folded or targeted for degradation through the ERAD pathway (Vembar and Brodsky, 2008). ER chaperone proteins provide the most prevalent mechanism for achieving quality control during protein translation and translocation (Anelli and Sitia, 2008). Both general and private chaperones are localized within the ER, often as multiprotein complexes (Meunier et al., 2002). The promiscuous general chaperones (e.g. Hsp70 BiP) bind a wide range of substrates through common unfolded features, such as exposed hydrophobic peptide sequences (Hartl and Hayer-Hartl, 2002). Private chaperones (e.g. Boca/Mesd) are dedicated to folding one particular protein or a family of proteins, with a distinct mechanism exclusive to the substrate(s) (Anelli and Sitia, 2008).

Proteolysis and NMR studies of murine Mesd have shown this protein to comprise three regions (Koduri and Blacklow, 2007); (Köhler et al., 2006). An NMR model (residues K104 to L177) was obtained for the structured core (PDB ID 2I9S), which is situated between two large proteolytically labile disordered N- and C-terminal regions (Köhler et al., 2006). N-terminal truncations, of Mesd, do not exhibit chaperone activity (Koduri and Blacklow, 2007). Similarly, the *boca¹* allele that contains a point mutation within this N-terminal region produces a nonfunctional W49R substitution in the mature Boca chaperone (Culi and Mann, 2003). Some confusion surrounds the function of the disordered C-terminal region, which is absent in Boca and other invertebrate family members. Mesd mutants lacking this C-terminal tail can rescue LRP6 surface expression in Mesd deficient cells (Koduri and Blacklow, 2007). In normal cells, an almost identical construct is unable to promote LRP6 surface expression, yet the tail alone binds fully folded LRP6 at the cell surface with high affinity (Li et al., 2005; Li et al., 2006b; Liu et al., 2009). Boca/Mesd chaperones also prevent the formation of high molecular weight inter-molecular disulphide aggregates of LDLR members within the ER (Culi and Mann, 2003; Hsieh et al., 2003). The single cysteine residue that is present in most Boca/Mesd orthologs appears to be unnecessary for function (Koduri and Blacklow, 2007); thus, Boca/Mesd chaperones are unlikely to function as redox catalysts and their molecular mechanism remains elusive.

Here we present four high-resolution crystallographic structures of the structured domains of Boca/Mesd YWTD β -propeller (BMY) chaperones from three divergent species: *D. melanogaster* Boca (*dmBoca*), *M. musculus* Mesd (*mmMesd*) and *C. elegans* (*ceBMY-1*). NMR analysis of *dmBoca* confirms that the conserved core (CC) is split into a structured domain (SD) and an N-terminal disordered region of high sequence conservation. Limited proteolysis of *mmMesd*, *dmBoca* and *ceBMY-1* readily removed highly conserved N-terminal regions that precede the structured domain (SD) within the highly conserved core (CC) region. The seven independent SD protomers from the four crystals are structurally similar, but they deviate significantly from the monomeric NMR structure (Köhler et al., 2006). The SD structures have a novel ferredoxin-like α/β fold, but there are both 4-stranded and 5-stranded variants of the β -sheet. Each structure is a dimer, mediated by a hydrophobic interface, but relative dispositions within these dimers are variable. We suggest that these non-specific interactions represent a possible interaction site for chaperone activity. Analytical ultracentrifugation shows that the disordered N-terminal region disrupts the structured domain dimer, suggesting a possible role for this region in regulating chaperone activity.

RESULTS

Structural Dissection of Boca/Mesd Family Proteins

In undertaking a structural analysis of Boca/Mesd proteins, we first performed a sequence alignment of several homologous proteins, which revealed that these all share a conserved core of ~140 residues (CC) within the mature protein (Figure 1). A short variable N-terminal sequence precedes the CC core. The mammalian homologs have highly charged and variable C-terminal tails, and all terminate in KDEL ER-retention signals. To probe the structure of these maturation factors for YWTD β -propellers, we produced the full-length proteins from three family members: *Mus musculus* (*mm*) Mesd; *Drosophila melanogaster* (*dm*) Boca; and *Caenorhabditis elegans* (*ce*) BMY-1. The former two were previously characterized as specialized chaperones, and we identified the latter from an unnamed ORF in the *C. elegans* genome (wormbase ws125) as an obvious ortholog (41 and 48% sequence identity to *mmMesd* and *dmBoca*, respectively, in the CC portion). The full ORF was extracted from a total *C. elegans* RNA extract by RT-PCR, and the cDNA matched the predicted gene, F09E5.17. The corresponding protein is now identified as BMY-1 (WormBase, www.wormbase.org, release WS126, 07.19.2004).

We used limited proteolysis to map the domain structure of the mature proteins from three family members: *Mus musculus* (*mm*) Mesd; *Drosophila melanogaster* (*dm*) Boca; and *Caenorhabditis elegans* (*ce*) BMY-1. Each of these proteins reveals a similar proteolytic profile (Figure S1). The variable sequences from the N- and C-termini of the mature protein are removed at low protease concentrations. At higher protease concentrations, the rapid removal of an additional ~50 residues from the N-terminus generates a ~10 kDa protease-resistant fragment. This implies that CC is composed of two regions, a stable structured domain (SD) that follows a poorly ordered and proteolytically labile N-terminal region. Thus, overall, the Boca/Mesd maturation factors comprise four characteristic regions between the N-terminal signal sequence and the KDEL ER retention signal (Figure 1A): a highly variable N-terminal segment, a poorly ordered region within the CC, the protease-resistant SD, and the highly charged variable region. For further structural characterizations, we purified CC and SD proteins for each of the three Boca/Mesd chaperones.

Structural Properties of the Boca Conserved Core (CC) in Solution

To further test our conclusions on the domain structure of the CC, we recorded ^{15}N - ^1H TROSY NMR spectra on $[\text{U-}^{15}\text{N}]$ labeled *dmBoca*_{CC} and $[\text{U-}^{15}\text{N}]$ *dmBoca*_{SD} (Figure 2). Nearly all the dispersed peaks within the CC spectra overlap with those of the SD spectra, implying that SD includes all of the well structured regions of the CC. The extra ~50 residues in the N-terminal region of *dmBoca*_{CC} have resonance frequencies corresponding to those of an unstructured peptide, typified by limited dispersion centered at ^1H resonance frequencies of ~8.2 ppm. A number of the dispersed resonances from the structured domain (SD) appear to be shifted in the context of the extended N-terminus of the CC, however, suggesting that the N-terminal segment makes weak interactions with the structured portion. Whether those interactions might involve an ordered state at low abundance or come from diffuse contacts with a fully flexible extension is unclear. That there are interactions is compatible with our observation that dimer affinity is reduced for extended constructs as compared with SD constructs (see below). The evident lack of substantial order in this segment is consistent with its ready removal by four different proteases with diverse sequence recognition specificities.

Dimer Interactions in Solution

In the course of our crystallographic studies, reported below, we found evidence for SD dimerization. Consequently, we performed equilibrium analytical ultracentrifugation (AUC) with the native protein for all three SD domains and for either the corresponding CC or full-length protein (Figure S2, Table 1). The SDs all have quite similar K_{D} s, which range from 77 ± 3 to 276 ± 14 μM . The addition of the disordered N-terminal residues in the native *dmBoca*_{CC} construct, or the mature *ceBMY-1* protein weakens the SD dimer approximately two-fold. The mature *mmMesd*, which includes the long highly charged variable C-terminal tail, is for all practical purposes monomeric. The weakened dimer interactions for the larger native constructs indicate that the variable tails and disordered N-terminal region interfere with intrinsic SD dimerization propensity.

Structures of the Structured Domains (SDs) from three Boca/Mesd chaperones

We determined structures for the SDs from *dmBoca*, *mmMesd* and *ceBMY-1* by x-ray crystallography (Tables 2 and 3). The SDs in each of four crystal structures are associated as apparent dimers. In all, seven independent protomers are contained in the structures that we have determined. These seven SD protomers are superposed in Figure 3A. Each of these proteins adopts a compact globular fold with a topology similar to ferredoxins. Together these proteins represent a novel superfamily within the ferredoxin-like fold (Alexey Murzin, SCOP, personal communication). Each SD protomer has a $\beta\alpha\beta\beta\alpha\beta$ core topology, wherein each $\beta\alpha\beta$ has a right-handed conformation but with successive parallel β -strands interlaced

in an anti-parallel β -sheet (Figure 3). This places the two helices on one side of the β -sheet, packed against one another in opposite orientations. In some cases, the last β -strand forms a hairpin turn to add a fifth β -strand.

Structured Domain of Boca—Two distinct crystal forms of the *dmBoca*_{SD} were solved using experimentally determined phases. Both crystals were produced from selenomethionyl (SeMet) *dmBoca* SD. Multiwavelength anomalous diffraction (MAD) at the Se K-edge was used to determine the structure of *dmBoca*_{SD} in crystal form I (Figure 3). This crystal belongs to space group P4₁22, and has two molecules per asymmetric unit. A second crystal (form II), belonging to space group P6₅22, allowed an independent solution of the *dmBoca*_{SD} structure by single wavelength anomalous diffraction (SAD) at the Se K-peak.

The P4₁22 (form I) structure, determined at 2.3 Å resolution includes ordered residues from R92 to G172 for one protomer, while the other protomer has density for three fewer residues ranging from R92 to E169. Additionally, 121 well-ordered water molecules were also modeled. The major difference between the two protomers is in the structure of the β -sheet near the C-terminus. In protomer A, the β -sheet is 5-stranded, and includes a short strand, β 5, which follows β 4 from a type-I' hairpin turn. In protomer B, this hairpin turn is absent and β 4 strand is extended without interruption (Figure 3B).

The structure determined for the P6₅22 crystals (form II), refined to 2.0 Å resolution, has a single protomer in the asymmetric unit (Figure 3). This molecule, like protomer A in crystal form I, has five β -strands. A total of 84 ordered residues (K89-G172), three sulfate ions, one sodium ion, an acetate ion and 76 water molecules were modeled. A methionine residue from the cleaved thrombin site is also observed.

Structured Domain of Mesd—The crystals of *mmMesd*_{SD} were hemihedrally twinned throughout the P4₃ lattice, with two molecules per asymmetric unit (Figure 3). Despite this difficulty, the *mmMesd*_{SD} structure was solved by molecular replacement (MR) using the four-stranded protomer of *dmBoca*_{SD} from crystal form I as the search model. A Bijvoet-difference map from SeMet *mmMesd*_{SD} diffraction measured at the Se K-peak showed features for both the true selenium sites and the twin-related sites (data not shown). The asymmetric unit contains two five-stranded *mmMesd*_{SD} protomers, which are ordered from residues K104-Y182 and residues M99-G184, respectively. A total of 163 residues, and 108 water molecules were refined to a 2.0 Å resolution. At the C-terminus of the fifth strand, residues 181 and 182 are disordered in one protomer and displaced by a crystal packing interaction in the other.

Structured Domain of BMY-1—The SeMet *ceBMY-1*_{SD} crystal belongs to space group P1 and contains two molecules per unit cell (Figure 3). Bijvoet-difference maps calculated from the Se-K peak dataset confirmed the Se positions in the MR solution, for which the search model employed was the five-stranded SD of *dmBoca* (form I). The final structure, refined with riding hydrogen atoms to 1.3 Å resolution, contains 90 ordered residues in one protomer (S84-G174) and 86 ordered residues (Q88-G174) in the second protomer. A total of 235 waters and four chloride ions were also modeled. The β -sheets possess an extended β 4 strand conformation in both *C. elegans* SD protomers. The C-terminal residues, which are disordered in the four-stranded SD of *dmBoca* (form I), here instead form a short C-terminal tail following the extended β 4 strand. Each tail reaches across to the other SD within the unit cell (Figure 3A); however, lacking defined secondary structure, the tails trace different paths to form distinct contact patterns with their partners. A chloride ion is bound in approximately the same location on both protomers. One tail completely occludes this ion, while the other tail loops around the equivalent chloride position without making any contact.

Comparison of the Crystal Structures to Prior NMR Structures

The seven independent protomers from the four crystal structures are very similar (Table 4). Pairwise RMS deviations for 65 equivalent C α positions (see Experimental Procedures) among all seven protomers are all less than 0.9 Å. The solution structure of an *mmMesd*_{SD} construct as determined by NMR (Köhler et al., 2006) has the same overall topology as for these crystal structures. In contrast, however, the NMR structure exhibits RMSDs greater than 2.2 Å with each of the seven crystallographically determined protomer structures when calculated for the same 65 C α atoms. The structural differences are highlighted by the superposition shown in Figure 3C. The program Escet (Schneider, 2004) defines the β -sheet as the most conformationally invariant region between the NMR model and *mmMesd*_{SD} crystal structure. In this superposition, the β -sheet (27 C α positions) has an RMSD deviation less than 1 Å, while the two α -helices (28 C α positions) in the NMR model are substantially displaced relative to those in the crystal structure and with an RMSD over 4.3 Å. This suggests that, although the short-range features of the NMR structure are well-determined, the NMR restraints fail to define fully the long-range interactions between the helices and the sheet. The NMR model also forms neither of the two distinct β -sheet conformations found in the crystal structures. Rather, the NMR structure has a short β 4-strand similar to the β 4 strand found in the 5-stranded conformation of SD, but it lacks β 5 (Figure 3C).

A recent NMR solution structure of a full-length *mmMesd*_{FL} construct (PDB ID 2KGL) (Chen et al., 2010a) shows major tertiary structural differences as compared to our seven crystallographically determined SDs or to the *mmMesd*_{SD} NMR structure described above. Because this new model has poor conformational geometry and unusual biophysical properties (Table S1), we demur and avoid detailed structural comparison.

Dimer Interactions of the Structured Domains

Each of the Boca/Mesd SD structures has its subunits arranged into dimers. For the *dmBoca* structure of crystal form II, the dimer is specified by a crystallographical diad; but in the other three crystals the two protomers per asymmetric unit are related by approximate diad axes. The quasi-two-fold non-crystallographic symmetry axes for these three, *mmMesd*, *ceBMY-1* and *dmBoca* (form I), differ in screw-axis parameters (Figure 4A, 4B, Table S2). Moreover, the relationship between *dmBoca* protomers in the form I dimer is rotated $\sim 160^\circ$ from that in the form II dimer. In each case, however, the same surface of SD consistently provides a major contact between protomers in the dimer. This surface, which includes contributions from each of the β -strands of the structured core, forms a dimer interface in all cases. Differences among these dimers are substantial, resulting from different interfacial contacts from a largely overlapping set of residues. These differences are not species dependent, however, because the two crystal forms of *dmBoca* have very different arrangements.

The buried surface area (Figure 4C, Table S2) for the *mmMesd*, *dmBoca*(I) and *dmBoca*(II) structures range from 600 to 900 Å². The *ceBMY-1* structure is an outlier, burying 1700 Å² at the interface. The extensive contacts made by the irregularly structured residues in the C-terminal tails of each protomer (Figure 4A), contribute almost half of this surface. By removing the non- β residues that follow the extended β 4 strand from both *ceBMY-1* protomers, we create a model for each *ce* SD that resembles the 4-stranded *dmBoca* (form I) protomer. The hydrophobic interface created by the *ceBMY-1* dimer in the absence of these tails closely matches the surface of the other three structures (Figures 4C and D). The 846 Å² buried in the *ce* model dimer lies directly between the BSA for the other two asymmetric parallel dimers, 898 Å² for *dmBoca* (form I) and 791 Å² of *mmMesd*.

The β -sheet surface that forms the dimer interface also contains the largest patch of conserved residues (Landau et al., 2005) (Figure 4E). The conserved surface also includes residues involved in the conformational change of the β -sheet, which produce either the extended β 4-strand structure or the alternative hairpin-turn / β 5-strand conformation. Thus, conservation within the SD includes the surface of the dimer, primarily residues from strands β 1 to β 3, as well as the residues that undergo the transition between the two conformational states.

The four observed dimer interfaces share ten equivalent residues that are buried from one or the other protomer. Six of these residues, two each from strands β 1, β 2 and β 3, contribute from both protomers within all four dimers. Central to all is a buried and invariant isoleucine residue (*dmBoca* Ile138, *ceBMY-1* Ile140, *mmMesd* Ile149); however, this single residue makes different atomic contacts in all four interfaces and it even forms asymmetric interactions between the SD protomers in the *ceBMY-1* and *mmMesd* structures (Figure 4F). Each of the conserved interfacial isoleucine residues contacts two to four residues in the opposing protomer. The two crystal forms of *dmBoca* have entirely discrete contact patterns for Ile138, but individual contacts in both *dmBoca* crystal forms are replicated in the *ceBMY-1* and *mmMesd* structures. Other residues alternatively contacted by the central isoleucine (Figure 4F), are also either absolutely conserved (*dmBoca* Met 95 and Phe97) or conservatively substituted (*dmBoca* Tyr130 and Leu140 vs. *mmMesd* Phe141 and Met 151; *dmBoca* Val132 vs. *ceBMY-1* Ile134), and these are also involved in pairwise contacts between one another.

DISCUSSION

Boca/Mesd family members include a conserved core (CC) and from our studies on the homologs from *Drosophila*, *C. elegans* and mouse we characterize two structurally distinct regions within this core. A disordered N-terminal region of high sequence conservation precedes a structured domain, for which we present four crystallographically determined structures. All of these structures are very similar to one another and they have essentially the same fold as in a previously reported NMR structure of the same domain from murine Mesd; however, there are distinctions in conformation that likely reflect the greater accuracy of the crystal structures. We also find that each of these proteins dimerizes and that all of the crystal structures are of molecules in a dimeric state. Comparisons of the crystal structures reveal two new features of possible relevance to the mechanism of chaperone action.

First, a small hydrophobic patch observed in the NMR solution structure, and proposed as a possible site for intermolecular interactions, is part of a larger conserved surface involved in the protein-protein contacts between protomers within each crystal lattice. Each interface is composed almost entirely of residues from the β 1, β 2 and β 3 strands of the β -sheet, but the relative orientations of protomers differ in the four structures. The residues from these three β -strands are able to form multiple, and distinct atomic contact patterns in the different protein homologues and crystal forms. The small surface areas buried by the β -sheets at each SD interface are approximately equivalent, ranging from 600 to 900 Å² in total, and close to the minimal size required for a protein interface (Jones and Thornton, 1996). Taken together, these observations suggest that the dimer interaction is fairly weak, consistent with the affinities measured by analytical ultracentrifugation. The propensity for protein-protein interactions through this conserved surface is, however, suggestive of the slippery hydrophobic surfaces of some promiscuous general chaperones (Hartl and Hayer-Hartl, 2002). Such conserved hydrophobic surfaces in chaperone proteins usually demarcate the substrate interaction site (Saibil, 2008). It remains unclear, however, how this surface might participate in the private chaperone action of Boca/Mesd with LDLR β -propellers.

Constructs containing the highly flexible and proteolytically labile, but highly conserved N-terminal region were resistant to crystallization and weakened the dimer interaction of the structured domain (SD). If disruption of the protein-protein interface were caused by competitive inhibition in *cis*, then a second interaction between the disordered N-terminal region and the conserved, but slippery and non-specific hydrophobic surface of the SD must exist. An unpublished *mmMesd* NMR structure (PDB ID 2RQK), provides one plausible model for this interaction. In the NMR ensemble, an α -helical segment from the N-terminal region, which is connected to the SD by a large unmodeled loop, binds the same conserved β -sheet surface that is buried at the interface of each crystal structure. The invariant tryptophan residue associated with the lethal *boca*¹ allele is present in this α -helix, and contacts several surface residues, including three of the six residues buried by all seven protomers. Like the intermolecular β -sheet to β -sheet SD dimer, the presumably intramolecular α -helix to β -sheet interaction, primarily contacts the first three β -strands and not the portion of the β -sheet that undergoes conformational changes.

The second novel feature found in the crystal structures is conformational variability in the β -sheets, which appears in three independent protomers as a 4-stranded sheet, and in four protomers as a 5-stranded sheet. We attribute this variability to the flexibility of residues that can form either a rare type I' β -turn (Hutchinson and Thornton, 1994) preceding the β 5 strand or an extension of the β 4 strand without a β -hairpin. One *dmBoca* crystal structure contains both 4- and 5-stranded protomers, indicating that this conformational change is not species dependent. We considered the possibility that crystal packing might have dictated the alternative conformations, and we do see that for four of the seven protomers (Table 1) lattice interactions are only consistent with one or the other alternative; however, we also find for the 5-stranded protomer in the *dmBoca* (I) lattice and for both in the *mmMesd* lattice, a 4-stranded model can be accommodated without conflict and without apparent compromise of lattice integrity. Thus, while we cannot rule out lattice artifacts in all cases, it does appear that intrinsic energetics are defining for three of the four 5-stranded protomers.

The two crystallographically determined *mmMesd* SDs are 5-stranded but neither the fifth strand nor the extended fourth strand conformation is directly observed in the *mmMesd* (89–184) NMR model (Köhler et al., 2006). The central core of structured residues (104–175) in the NMR ensemble is contained within each of the crystal structures, but each of the crystal structures additionally contains well-ordered residues that extend beyond this core of the NMR ensemble (Figure 3C). The NMR ensemble, which shows significant differences from the crystal structures in the region of variable β -sheet architecture may reflect sampling of intermediate-like states in solution, and suggests why this region has been reported to be unstructured and proteolytically labile (Koduri and Blacklow, 2007; Köhler et al., 2006).

Residues that distinguish the two distinct β -sheet conformations are among the most conserved in the β -sheet region of Boca/Mesd proteins. On the other hand, only a few of these residues, those from the C-terminus of β 4 and the turn into β 5, participate in the dimer interface. These observations suggest the possibility of a separate function that provides evolutionary pressure for this region. An almost invariant valine-threonine residue pair (*mmMesd* Val174-Thr175, *dmBoca* Val163-Thr164, *ceBMY-1* Val165-Thr166) anchors the β 4 strand to the adjacent β 1 strand and hydrophobic core. The four residues following the threonine are also highly conserved, potentially because they must retain flexibility to form the discrete secondary structural states seen in the two conformations. The two central residues of type I' β -turns, named *i*+1 and *i*+2, are located in the α L energy minima of the Ramachandran plot whereby type I' β -turns are intrinsically less stable (Yan et al., 1993; Yang et al., 1996). Only the *i*+1 and *i*+2 residues undergo a drastic ϕ and ψ change between the two secondary structure states seen in the crystals. The ϕ and ψ angles of the *i* and *i*+3 β -turn residues remain the same between the two conformational states, allowing similar

backbone hydrogen bonds to form even though the *i*+3 residues moves ~13Å. Both these factors may reduce the energy required to undergo the transition between the β -turn and β -sheet conformations. The threonine residue preceding the β -turn may act as a conformational switch in this region, because the **p** and **m** side chain conformations (Lovell et al., 2000) correlate to the 4-strand and 5-strand β -sheet conformations.

The poorly structured N-terminal region within the CC is known to have an important role in the function of the Boca/Mesd family. Removal of the conserved N-terminal sequence from *mmMesd*, leaving the SD, prevents LRP6 from reaching the cell surface (Koduri and Blacklow, 2007; Li et al., 2005). The *boca*¹ allele, a point mutation of an invariant tryptophan (dm W49) residue within this region, also disrupts chaperone function (Culi and Mann, 2003). The reverse experiment, removal of the SD has yet to be tested: however, the SD is also probably essential for chaperone function, owing to the highly conserved amino acid sequence found throughout the CC. Both regions of the CC are likely to function together in Boca/Mesd activity as a chaperone for the YWTD β -propeller-EGF domain pair, because both regions are well conserved throughout evolution and AUC experiments suggest that interactions between these regions prevents dimer formation. How they might interact with one another and with LDLR family substrates remain as open questions.

The highly charged variable tail that follows on from the fifth strand is essentially absent in invertebrates, and it is not well conserved when compared to the preceding CC. In vertebrates, the C-terminal tail is approximately 30 to 45 residues long, ~50% of these residues are charged and a further ~38% are polar or glycine. The tail together with the last five residues of the SD binds Lrp6 on the cell surface (Li et al., 2005). Like RAP and other LDLR ligands, this is most likely achieved through the electrostatic interactions of the charged residues (Blacklow, 2007). An *mmMesd* with the C-terminal tail removed fails to bind to the Lrp6, likely because the receptor is fully folded and the CC is expected to interact with a non-native state to carry out its chaperone function (Li et al., 2005). In a separate functional assay, full-length *mmMesd* rescues LRP6 surface expression in HEK293T cells treated with *hsMesd* RNAi (Koduri and Blacklow, 2007). Another construct, containing most of the CC and a segment of the variable tail, is able to partially recover LRP6 surface expression however chaperone or trafficking activity of this protein is abolished by removing the last five residues of the SD along with the variable tails (Koduri and Blacklow, 2007). Thus, the residues that can form the fifth strand appear to be essential to the function of this chaperone.

Based on current understanding of chaperone function, two distinct binding mechanisms for the structurally discrete regions of the CC can be proposed. Disorder in a protein, as in the N-terminal region of Boca/Mesd, can provide high-specificity, low-affinity binding through a mechanism of entropy-enthalpy compensation. The entropic cost to stabilize an unstructured protein region counteracts the generally favorable enthalpic driving force of substrate binding (Dyson and Wright, 2005). In protein chaperones ~15% of residues fall within disordered regions greater than 30 residues (Tompa and Csermely, 2004). These unstructured regions uncouple affinity from specificity, providing promiscuous substrate recognition, as well as solubilizing exposed hydrophobic regions within the substrate, through association of the hydrophilic disordered regions, and by the exclusion effect, which blocks substrate molecules from approaching one another. Enthalpy transfer has also been proposed as a chaperone mechanism, in which the free energy cost of local unfolding of the substrate may be “covered” by the enthalpic cost of ordering the disordered chaperone upon binding, allowing multiple rounds of binding and release (Tompa and Csermely, 2004).

The rigid β -sheet of the SD, meanwhile, in effect provides a relatively smooth and consistent surface (Chothia and Janin, 1981), one that is the same for all of the structures including the

dissimilar *mmMesd* crystal and NMR (PDB ID 2I9S) structures. Both β -sheet structures, such as other SDs in the dimeric states, and α -helical structures, such as the N-terminal segment of the CC, interact with the relatively small surface of this stable structural element. Non-specific binding occurs through the conserved surface residues, primarily located on strands β 1, β 2 and β 3. The energetic cost in binding these distinct secondary structural elements is low, because only side chain rearrangements are required for the β -sheet to accommodate different binding partners (Chothia and Janin, 1981).

In conclusion, the surface expression of functional LDL receptors and lipoprotein receptor-related proteins (LRPs) depends on the action Boca/Mesd proteins in the ER. These private chaperones specifically assist in the proper folding of the YWTD propeller domains that characterize LRPs. Our structural studies on diverse members of the Boca/Mesd family limit options for the mechanism of chaperone action by these proteins and pave the way for future mutational studies.

EXPERIMENTAL PROCEDURES

Cloning, Protein Expression and Purification

Mature chaperone protein (MP: *dmBoca* 19–176, *mmMesd* 30–220, *ceBMY-1* 19–182), conserved core (CC: *dm* 31–172, *mm* 43–183, *ce* 24–174) and structured domain (SD: *dm* 88–172, *mm* 99–183, *ce* 84–174) for each species were PCR-amplified from full-length cDNAs of *mmMesd*, *dmBoca* (both kindly provided by J. Culi) and *ceBMY1* (Boca/Mesd chaperone for YWTD β -propeller-EGF), using appropriate 5' and 3' primers. The full *ceBMY-1* ORF was produced, using an RT-PCR kit (promega), from a worm total RNA extract (a gift from J. Etchberg), the resulting cDNA, ligated between the NdeI and BamHI restriction sites of pET22b+ (Novagen), perfectly matched the predicted gene, F09E5.17 (wormbase, release ws125).

The MP, CC and SD PCR-amplified products were all engineered with an NdeI restriction site at the 5' end, and a stop codon preceding either a BamHI or HindIII restriction site at the 3' end. These DNA fragments were ligated into the equivalent restriction sites of the vector pET28b+ (Novagen). The nine resulting fusion protein constructs contain, at the N-terminus, a hexa-histidine tag and thrombin protease recognition sequence provided by the vector. Upon thrombin cleavage, an additional GSHM sequence remains at the N-terminal end of each protein. The final *dmBoca*_{SD} product appeared to have additional residues, GSQKHM, based upon mass spectrometry, and N-term sequencing.

Each fusion protein was expressed in *E. coli* BL21(DE3) cells (Novagen), grown in rich defined auto-induction media; PA5052, or PASM5052 for SeMet incorporation, at 37°C for 6 hours followed by overnight growth at 30°C, after being inoculated 1:1000 from an overnight culture grown in PA0.5G media (Studier, 2005). Induced cells were harvested by centrifugation, resuspended in 50mM Tris pH 8.0 and 500mM NaCl supplemented with complete EDTA-free protease inhibitor (Roche) and frozen. After thawing, cells were lysed by sonication and cleared by centrifugation and filtration (0.22 μ m), before initial affinity purification using a chelating HiTrap column (GE Healthcare) charged with Ni²⁺. Each protein was eluted with a linear imidazole gradient and dialyzed, in the presence of thrombin (GE Healthcare), against 20mM Tris pH 8.0 and either 200mM or 500mM NaCl overnight at 4°C. Samples were concentrated (Amicon) and further purified by gel filtration on a Superdex 75 26/60 (GE Healthcare) column, previously equilibrated in the equivalent dialysis buffer. All protein samples appeared to be dimeric by size-exclusion chromatography and were homogeneous when analyzed by SDS-PAGE.

Crystallization

Two crystallization conditions were optimized for SeMet *dmBoca*_{SD} by hanging-drop vapor-diffusion, with a 1 μ L:1 μ L ratio of reservoir to protein solution (25mg/mL in 10mM Tris pH 8.0, 100mM NaCl). Form I crystals were grown at 4°C against a reservoir solution containing 18–21% (w/v) polyethylene glycol (PEG) 3350, 100–200mM Na citrate, 100mM 2-(cyclohexylamino)ethanesulfonic acid (CHES) pH9.0, and 10mM dithiothreitol (DTT) by macro-seeding from an equivalent conditions, which in turn were micro-seeded from a showers of small native *dmBoca*_{SD} (22mg/mL) form I crystals. The roughly cubic crystals were serially soaked in reservoir solution supplemented with gradually increasing glycerol concentrations 6.6%, 13.3% and 20%, prior to freezing in liquid nitrogen.

The second SeMet *dmBoca*_{SD} crystals (form II), was one of four crystal morphologies that grew against a reservoir solution containing 48–52% saturated AmSO₄, 100mM sodium acetate pH 5.6–5.8 at 20°C. The thin hexagonal pyramid shaped crystals were soaked in cryo-protection buffer consisting of 80% saturated AmSO₄, 100mM sodium acetate pH 5.7 and 10% GSX-goop (40% glycerol, 32% sucrose and 16% xylitol) before freezing in liquid nitrogen. The other crystal morphologies found in this condition did not produce sufficient quality diffraction data for structure determination.

Crystals of *mmMesd*_{SD} were grown by hanging drop vapor diffusion against a mother liquor containing 0.7–0.9M AmPO₄ and 100mM Hepes pH 7.5 at 4°C which were nucleated by streak-seeding from micro-crystal showers found in conditions with higher AmPO₄ precipitant concentration. Native protein solution (27 mg/mL in 10mM Tris pH 8.0, 10mM NaCl) to reservoir ratio of 1:1 (1 μ L each) was used. Stout square rod-shaped crystal grew over several weeks, and were frozen in liquid nitrogen after gradually increasing, in three steps, to a final solution that consisted of the crystallization buffer supplemented with 15% ethylene glycol and 5% glycerol.

SeMet *ceBMY-1*_{SD} crystals were grown by hanging drop vapor diffusion at 20°C against a crystallization buffer containing 15–18% PEG 3350, 300–500mM LiCl₂, 100mM Mes pH 5.5. A protein solution (38 mg/mL 10mM Tris pH 8.0, 10mM NaCl, 1mM DDT) was used with a protein to buffer ratio of 1:1 (1 μ L each). Thick square plate crystals typically appeared overnight and grew to an optimal size within seven days. Crystals were cryo-protected in one-third steps to a final solution that consisted of the crystallization buffer supplemented with either 20% glycerol or 5% glycerol and 15% ethylene glycol.

Structure Determination

***dmBoca*_{SD}(I)**—A three-wavelength MAD experiment was collected at the Se K-edge from a single frozen SeMet crystal at X25 beamline of NSLS, at BNL. Diffraction data to Bragg spacings of 2.5 Å (peak, edge), and 2.8 Å (remote) with a 440mm detector distance were collected using 6 second exposure times and 3° oscillations at each of the three wavelengths. Accurate Bijvoet differences were measured by collecting the inverse beam. A second peak-wavelength dataset, with no inverse beam, was collected through a different portion of the same crystal, to a Bragg spacings of 2.3 Å, using 2 second exposure times, 1° oscillations, and a detector distance of 320mm. Datasets were indexed and merged using Denzo and Scalepack of the HKL program package (Otwinowski and Minor, 1997). Phases were calculated using Solve (v2.08) (Terwilliger and Berendzen, 1999), which found 4 of the 6 potential Se sites, for the P4₁22 space group. An initial composite model derived from both Resolve (v2.08) (Terwilliger, 2000) and ArpWarp (v6.0) (Perrakis et al., 1999) autobuilding procedures was refined at 2.3Å resolution using CNS (v1.1) (Brünger et al., 1998) and Phenix (v1.3) (Adams et al., 2002), with iterative rounds of manual rebuilding in Coot (v0.5) (Emsley and Cowtan, 2004).

dmBoca_{SD}(II)—An f⁰ peak SAD experiment was collected at the Se K-edge on the NSLS X4C beamline from a single frozen form (II) crystal of SeMet *dmBoca_{SD}*. Diffraction data to Bragg spacings of 2.0 Å with a 100mm detector distance were recorded using 20 second exposure times, 1° oscillations, and inverse beam for accurate measurements of Bijvoet differences. Data were indexed, merged, and processed using the HKL2000 package and refined with Phenix in the space group P6₅22. Solve (v2.13) located all three Se sites, and after solvent flattening with DM (v6.1) (Cowtan, 1994), ArpWarp (v7.0.1) was able to build a complete model, that was refined in Phenix (1.3), with iterative rounds of manual rebuilding in Coot.

mmMesd_{SD}—A frozen crystal of *mmMesd_{SD}* was collected at the NSLS beamline X4C, to Bragg spacings of 2.0 Å, with 9 second exposure times, 0.5° oscillations and a detector distance of 150mm. HKL2000 was used for data processing, and the native *mmMesd_{SD}* structure was determined, in the P4₃ space group, by molecular replacement using Phaser (v2.0) (McCoy et al., 2007) with the 4 stranded *dmBoca_{SD}* (I) protomer as a search model. The initial model generated by rebuilding the MR solution using ArpWarp(v6.0) was refined using Phenix with the twin operator [h,-k,-l] and two TLS groups consisting of each peptide chain in the ASU, with iterative rounds of manual rebuilding in Coot. A Se K-edge SeMet *mmMesd_{SD}* experiment (data not shown), that was unable to produce a phase solution, confirmed the position of the Se (and twin Se) sites by calculating an anomalous map with phases from a partially refined *mmMesd_{SD}* model.

ceBMY-1_{SD}—Data from a single SeMet *ceBMY-1_{SD}* crystal were collected in two parts at the X4C beamline of NSLS. The first sweep using a detector distance of 88mm, and 1° oscillations of 8 second exposure times, diffracted x-rays to 1.37 Å resolution. The second sweep at a detector distance of 135mm, and 1° oscillations of 1 second exposure times, due to low angle reflections overloading the detector during the longer exposure times required for the first, higher resolution, sweep. HKL2000 and Phenix was used to process and merge the data, which were only compatible with a primitive triclinic space group. The molecular replacement solution found by Phaser using the 5 stranded *dmBoca_{SD}* (I) protomer as a search model, was feed to ArpWarp for phase improvement and automated model building. The resulting model with addition of riding hydrogen atoms was refined using Phenix, with iterative rounds of manual rebuilding in Coot.

Structural Analysis

All superpositions were performed with LSQMAN(v9.7.9) from the Uppsala software package (Kleywegt, 1996). The C α positions of each SD protomer were superposed, and manually adjusted to maximize the number of structurally aligned residues in both pairwise and multiple model superposition's, such that 3 or more contiguous C α positions were within 2.5Å of equivalent positions. The rotation–translation matrices of the molecular axes of symmetry were determined by TOSS (Hendrickson, 1979) using the core 65 C α position defined during the multiple model superposition. Escet (v0.7h) (Schneider, 2004) error-scaled difference distance matrices allowed for comparisons of SDs to each other, as well as for both *mmMesd_{SD}* protomers to the 20 models in the NMR ensemble (PDB ID 2I9S). CCP4 programs contact and areaimol were used to define the atoms participating in the interface and assess their change in BSA during dimer formation. An alignment of 48 orthologous Boca/Mesd sequences grouped by TreeFam (www.treefam.org) (Li et al., 2006a) was in conjunction with Consurf (consurf.tau.ac.il) (Landau et al., 2005) to map the level of evolutionary conservation at each residues position onto the SD structures. Figures were prepared with PyMOL (DeLano, 2002).

Limited Proteolytic Degradation

Subtilisin, trypsin, chymotrypsin (Sigma), and elastase (Worthington) were resuspended in cold protease buffer (10mM NaCl, 10mM Tris pH 8.0) at 1mg/mL, and the serial dilutions of these stock solutions were stored on ice. Proteolytic reactions, set up on ice, contained 9 μ L of a purified MP diluted to 1mg/mL with cold protease buffer and 1 μ L of protease at a final concentration of 10, 1, 0.1 and 0.01 μ g/ μ L, or 1 μ L PBS for the control. Reactions were moved to 20°C and terminated after 30 min by removing 2 μ L into 2 μ L of TFA, for analysis by MALDI-TOF mass spectrometry, while the remaining 8 μ L were stopped by the addition of 3 μ L 4 \times SDS-PAGE loading buffer. After boiling for 3 minutes, proteolytically digested samples were analyzed by SDS-PAGE, and both visualized with coomassie blue, and transferred to polyvinylidene difluoride (PVDF) for N-terminal sequencing by Edman degradation.

Analytical Ultracentrifugation

Sedimentation equilibrium experiments were performed with a Beckman-Coulter XLI analytical ultracentrifuge using both UV 280nm and IR 660nm optics. Native purified *dmBoca*_{SD}, *mmMesd*_{SD} and *ceBMY*-1_{SD} were dialyzed at 4°C and then diluted to approximately 0.70, 0.46 and 0.24 mg/mL with 2 \times PBS. Native purified *dmBoca*_{CC}, *mmMesd*_{MP} and *ceBMY*-1_{MP} were treated in a similar fashion, and diluted to 0.70, 0.46 and 0.24 mg/mL. Protein samples, each at the three concentrations, were loaded into three channels of a six-channel cell with sapphire windows and a path length of 12 mm: the other three channels contained buffer blank controls. The SD protein samples were sedimented to equilibrium in a Ti50An rotor at three to five rotor speeds ranging from 14,000 to 26,000 rpm at 25 °C (See Figure S2). UV absorbance and IR interference scans were taken at 1hr intervals. The program HeteroAnalysis (v1.0.114 written by J. Cole and J. Lary, University of Connecticut) enabled group fitting of the equilibrium data at all three speeds and concentrations for each protein sample.

NMR Spectroscopy

Recombinant *dmBoca*_{CC} and *dmBoca*_{SD} constructs were expressed in M9 minimal medium (Sambrook and Russell, 2001) with 99%-¹⁵NH₄Cl as the sole nitrogen source (Cambridge Isotopes) in the M9 salts. Protein purification followed the protocol described above. The NMR buffer consisted of 20mM MES pH 6.0 and 50mM NaCl. The *dmBoca*_{CC} was dialyzed overnight at 4°C against NMR buffer and concentrated to 0.66mM, the *dmBoca*_{SD} was exchanged directly into NMR buffer during concentration to 0.42mM. NMR buffers also contained 0.02% NaN₃, and 10% D₂O.

NMR ¹⁵N-¹H TROSY spectra for both ¹⁵N-labeled 0.36mM *dmBoca*_{SD} and 0.46mM *dmBoca*_{CC} were acquired on a Bruker DRX600 with a triple-resonance z-axis gradient cryogenic probe at 299.5K. The sample temperature was calibrated with 99.8% d₄-methanol and post-acquisition processing of the spectra was performed with NMRPipe (Delaglio et al., 1995) and Sparky 3 (T. D. Goddard and D. G. Kneller, University of California, San Francisco).

ACCESSION NUMBERS

Protein sequences NP_724578(dm), NP_495003(ce), Q9ERE7(Mm), Q14696(Hs), Q5ZKK4(Gg), A1L243(Dr), and AAH97859(Xl) from Entrez Protein (<http://www.ncbi.nlm.nih.gov/protein>) were used in the alignment of the Boca/Mesd family. Atomic coordinates and structure factors have been deposited at the Protein Data Bank (<http://www.rcsb.org/pdb>) with PDB ID codes 3OFE, 3OFF, 3OFG and 3OFH. See Table 3 for identifications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank L. Shapiro and P. Katsamba for critical reading of this manuscript. G. Ahlsen for help performing analytical ultracentrifugation experiments, V. Miloushev and A. Palmer for NMR studies, and J. Cheung, E. Martinez-Hackert and other Hendrickson laboratory members past and present for help with data collection, processing and refinement. We also thank staff at X4A, X4C and X25, of the NSLS at Brookhaven National Laboratory, a DOE facility. Both X4 beamlines are supported by the New York Structural Biology Center and X25 is part of the PXR initiative. This work was supported in part by NIH Biophysics Training Grant 5 T32 GM008281-19 and by NIH grant GM34102.

REFERENCES

- Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. *Acta crystallographica. Section D, Biological crystallography*. 2002; 58:1948–1954.
- Anelli T, Sitia R. Protein quality control in the early secretory pathway. *The EMBO journal*. 2008; 27:315–327. [PubMed: 18216874]
- Björklund P, Svedlund J, Olsson A-K, Akerström G, Westin G. The internally truncated LRP5 receptor presents a therapeutic target in breast cancer. *PLoS one*. 2009; 4:e4243. [PubMed: 19158955]
- Blacklow SC. Versatility in ligand recognition by LDL receptor family proteins: advances and frontiers. *Current opinion in structural biology*. 2007; 17:419–426. [PubMed: 17870468]
- Brünger, a.T.; Adams, PD.; Clore, GM.; DeLano, WL.; Gros, P.; Grosse-Kunstleve, RW.; Jiang, JS.; Kuszewski, J.; Nilges, M.; Pannu, NS., et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta crystallographica. Section D, Biological crystallography*. 1998; 54:905–921.
- Chen J, Li Q, Liu C-C, Zhou B, Bu G, Wang J. NMR structure note: solution structure of the core domain of MESD that is essential for proper folding of LRP5/6. *Journal of biomolecular NMR*. 2010a; 47:283–288. [PubMed: 20506034]
- Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography*. 2010b; 66:12–21.
- Chothia C, Janin J. Relative orientation of close-packed beta-pleated sheets in proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 1981; 78:4146–4150. [PubMed: 16593054]
- Clamp M, Cuff J, Searle SM, Barton GJ. The Jalview Java alignment editor. *Bioinformatics (Oxford, England)*. 2004; 20:426–427.
- Cowtan K. 'dm': An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*. 1994; 31:p34–38.
- Culi J, Mann RS. Boca, an endoplasmic reticulum protein required for wingless signaling and trafficking of LDL receptor family members in *Drosophila*. *Cell*. 2003; 112:343–354. [PubMed: 12581524]
- Culi J, Springer TA, Mann RS. Boca-dependent maturation of beta-propeller/EGF modules in low-density lipoprotein receptor proteins. *The EMBO journal*. 2004; 23:1372–1380. [PubMed: 15014448]
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax a. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR*. 1995; 6:277–293. [PubMed: 8520220]
- Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nature reviews. Molecular cell biology*. 2005; 6:197–208. [PubMed: 15738986]

- Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols*. 2007; 2:953–971.
- Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography*. 2004; 60:2126–2132.
- Gouet P. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Research*. 2003; 31:3320–3323. [PubMed: 12824317]
- Hartl FU, Hayer-Hartl M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science (New York, N.Y.)*. 2002; 295:1852–1858.
- Hendrickson WA. Transformations to optimize the superposition of similar structures. *Acta Crystallographica Section A*. 1979; 35:158–163.
- Herz J. The switch on the RApper's necklace.... *Molecular cell*. 2006; 23:451–455. [PubMed: 16916634]
- Herz J. Apolipoprotein E receptors in the nervous system. *Current opinion in lipidology*. 2009; 20:190–196. [PubMed: 19433918]
- Hsieh JC, Lee L, Zhang L, Wefer S, Brown K, DeRossi C, Wines ME, Rosenquist T, Holdener BC. Mesd encodes an LRP5/6 chaperone essential for specification of mouse embryonic polarity. *Cell*. 2003; 112:355–367. [PubMed: 12581525]
- Hutchinson EG, Thornton JM. A revised set of potentials for beta-turn formation in proteins. *Protein science : a publication of the Protein Society*. 1994; 3:2207–2216. [PubMed: 7756980]
- Jeon H, Meng W, Takagi J, Eck MJ, Springer TA, Blacklow SC. Implications for familial hypercholesterolemia from the structure of the LDL receptor YWTD-EGF domain pair. *Nature structural biology*. 2001; 8:499–504.
- Jones S, Thornton JM. Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America*. 1996; 93:13–20. [PubMed: 8552589]
- Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. *Briefings in bioinformatics*. 2008; 9:286–298. [PubMed: 18372315]
- Kim N, Stiegler AL, Cameron TO, Hallock PT, Gomez AM, Huang JH, Hubbard SR, Dustin ML, Burden SJ. Lrp4 is a receptor for Agrin and forms a complex with MuSK. *Cell*. 2008; 135:334–342. [PubMed: 18848351]
- Kleywegt GJ. Use of non-crystallographic symmetry in protein structure refinement. *Acta crystallographica. Section D, Biological crystallography*. 1996; 52:842–857.
- Koduri V, Blacklow SC. Requirement for natively unstructured regions of mesoderm development candidate 2 in promoting low-density lipoprotein receptor-related protein 6 maturation. *Biochemistry*. 2007; 46:6570–6577. [PubMed: 17488095]
- Köhler C, Andersen OM, Diehl A, Krause G, Schmieder P, Oschkinat H. The solution structure of the core of mesoderm development (MESD), a chaperone for members of the LDLR-family. *Journal of structural and functional genomics*. 2006; 7:131–138. [PubMed: 17342452]
- Kyte J, Doolittle RF. A simple method for displaying the hydrophobic character of a protein. *Journal of molecular biology*. 1982; 157:105–132. [PubMed: 7108955]
- Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic acids research*. 2005; 33:W299–302. [PubMed: 15980475]
- Li H, Coghlan A, Ruan J, Coin LJ, Hériché J-K, Osmotherly L, Li R, Liu T, Zhang Z, Bolund L, et al. TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic acids research*. 2006a; 34:D572–580. [PubMed: 16381935]
- Li Y, Chen J, Lu W, McCormick LM, Wang J, Bu G. Mesd binds to mature LDL-receptor-related protein-6 and antagonizes ligand binding. *Journal of cell science*. 2005; 118:5305–5314. [PubMed: 16263759]
- Li Y, Lu W, He X, Bu G. Modulation of LRP6-mediated Wnt signaling by molecular chaperone Mesd. *FEBS letters*. 2006b; 580:5423–5428. [PubMed: 16989816]
- Li Y, Lu W, He X, Schwartz AL, Bu G. LRP6 expression promotes cancer cell proliferation and tumorigenesis by altering beta-catenin subcellular distribution. *Oncogene*. 2004; 23:9129–9135. [PubMed: 15516984]

- Lillis AP, Van Duyn LB, Murphy-Ullrich JE, Strickland DK. LDL receptor-related protein 1: unique tissue-specific functions revealed by selective gene knockout studies. *Physiological reviews*. 2008; 88:887–918. [PubMed: 18626063]
- Liu C-C, Pearson C, Bu G. Cooperative folding and ligand-binding properties of LRP6 beta-propeller domains. *The Journal of biological chemistry*. 2009; 284:15299–15307. [PubMed: 19339249]
- Lovell SC, Word JM, Richardson JS, Richardson DC. The penultimate rotamer library. *Proteins*. 2000; 40:389–408. [PubMed: 10861930]
- MacDonald BT, Tamai K, He X. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Developmental cell*. 2009; 17:9–26. [PubMed: 19619488]
- Marzolo M-P, Bu G. Lipoprotein receptors and cholesterol in APP trafficking and proteolytic processing, implications for Alzheimer's disease. *Seminars in cell & developmental biology*. 2009; 20:191–200. [PubMed: 19041409]
- May P, Herz J, Bock HH. Molecular mechanisms of lipoprotein receptor signalling. *Cellular and molecular life sciences : CMLS*. 2005; 62:2325–2338. [PubMed: 16158188]
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *Journal of applied crystallography*. 2007; 40:658–674. [PubMed: 19461840]
- Meunier L, Usherwood Y-K, Chung KT, Hendershot LM. A subset of chaperones and folding enzymes form multiprotein complexes in endoplasmic reticulum to bind nascent proteins. *Molecular biology of the cell*. 2002; 13:4456–4469. [PubMed: 12475965]
- Otwinowski, Z.; Minor, W. *Macromolecular Crystallography Part A*. In: Carter, Charles W., editor. *Macromolecular Crystallography Part A*, J. Elsevier; 1997. p. 307-326.
- Perrakis, a.; Morris, R.; Lamzin, VS. Automated protein model building combined with iterative structure refinement. *Nature structural biology*. 1999; 6:458–463.
- Saibil HR. Chaperone machines in action. *Current opinion in structural biology*. 2008; 18:35–42. [PubMed: 18242075]
- Sambrook, J.; Russell, DW. *Molecular Cloning: A Laboratory Manual*. Third edn. Cold Spring Harbor Laboratory Press; 2001.
- Schneider TR. Domain identification by iterative analysis of error-scaled difference distance matrices. *Acta crystallographica. Section D, Biological crystallography*. 2004; 60:2269–2275.
- Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein expression and purification*. 2005; 41:207–234. [PubMed: 15915565]
- Terwilliger TC. Maximum-likelihood density modification. *Acta crystallographica. Section D, Biological crystallography*. 2000; 56:965–972.
- Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. *Acta crystallographica. Section D, Biological crystallography*. 1999; 55:849–861.
- Tompa P, Csermely P. The role of structural disorder in the function of RNA and protein chaperones. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2004; 18:1169–1175. [PubMed: 15284216]
- Vembar SS, Brodsky JL. One step at a time: endoplasmic reticulum-associated degradation. *Nature reviews. Molecular cell biology*. 2008; 9:944–957. [PubMed: 19002207]
- Yan Y, Tropsha A, Hermans J, Erickson BW. Free energies for refolding of the common beta turn into the inverse-common beta turn: simulation of the role of D/L chirality. *Proceedings of the National Academy of Sciences of the United States of America*. 1993; 90:7898–7902. [PubMed: 8356099]
- Yang, a.S.; Hitz, B.; Honig, B. Free energy determinants of secondary structure formation: III. beta-turns and their role in protein folding. *Journal of molecular biology*. 1996; 259:873–882. [PubMed: 8683589]
- Zhang B, Luo S, Wang Q, Suzuki T, Xiong WC, Mei L. LRP4 serves as a coreceptor of agrin. *Neuron*. 2008; 60:285–297. [PubMed: 18957220]

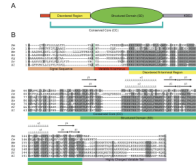


Figure 1. The Domain Organization of the Boca/Mesd Family

A) Schematic representation of the Boca/Mesd family structure. The short variable N-terminal region and highly charged variable tail, that includes the ER-retention signal labeled here as KDEL, are represented by small red and purple boxes respectively. The conserved core, CC, has high sequence identity and is divided into two regions; the disordered N-terminal region (yellow box), and the Structured domain, SD (green oval).

B) Structure-based alignment for the Boca/Mesd chaperone family. The *dmBoca* (Dm), *ceBMY-1* (Ce) and *mmMesd* (Mm) sequences for the proteins in this work along with Boca/Mesd orthologues from Human(Hs), chicken(Gg), fish (Dr), and frog (Xl). The mafft (Katoh and Toh, 2008) alignment was adjusted based on our knowledge of the coordinates for SD. Colored residues represent a 60% identity threshold between species in this alignment (Jalview) (Clamp et al., 2004). The vertical green bar shows the predicted signal peptidase cleavage site (signalP) (Emanuelsson et al., 2007), and thus the start of the mature protein. The secondary structural elements of *dmBoca*_{SD} (form I), both 4 and 5 stranded β -sheet conformations, are above the sequence (Espript) (Gouet, 2003). The colored boxes below the alignment represent the regions described (also see Figure S1).

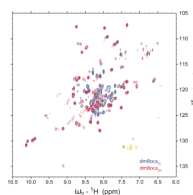


Figure 2. TROSY spectra of *dmBoca_{CC}* and *dmBoca_{SD}*

Two structurally distinct regions exist within the conserved core of *dmBoca*. The overlay of $[N^{15}-H^1]$ TROSY NMR spectra from N^{15} labeled *dmBoca_{CC}* (red contours) and *dmBoca_{SD}* (blue contours) at pH 6.0, 299.5K. The dispersed peaks of the *dmBoca_{SD}* account for almost all those of *dmBoca_{CC}* while the additional residues in N-terminal region of *dmBoca_{CC}* corresponding to those of an unstructured peptide. Green and yellow contours are negative signals for the folded resonances of Arginine Ne spins.

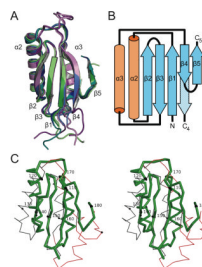


Figure 3. Similarities and Differences of the Boca/Mesd family Structured Domain

A) Structural superposition of all seven Boca/Mesd family SD protomers. The seven crystallographically determined protomers, in 4 different crystal lattices, are from 3 divergent species; mouse (green, and light green), fly (blue, light blue, and teal), and worm (purple and light purple). All seven structures are within 0.9Å RMSD. At the C-terminus of the SD the number of β -strands present in the β -sheet, dependent on which conformation is present in the crystal.

B) Boca/Mesd SD topology, both the 5-strand and extended β 4 conformations of the β -sheet are shown. The α -helices (orange) and β -strands (blue) are numbered to be consistent with NMR structure (PDB ID 2I9S). The 5-stranded topology ends at the C₅-terminus and the β 4 extension (checked blue) at the C₄-terminus.

C) Stereopicture of the C α trace for the crystal structure of *mmMesd*_{5D}, same orientation and color green as (A), superimposed with the NMR model (PDB ID 2I9S), via the Escet define conformationally invariant β -sheet (Schneider, 2004), and reveals the nearly 5Å RMSD offset for the helices in the NMR structure. Every tenth residue is presented as a sphere and numbered. The thin grey C α trace of the NMR model is highlighted in red for C α positions greater than three standard deviations from the internal RMSD of the NMR core.

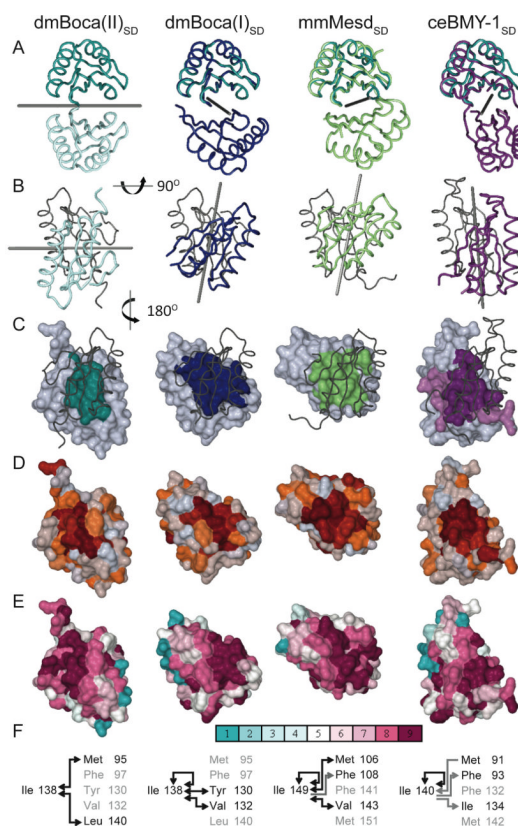


Figure 4. Properties of the Distinct Dimer Interfaces of *dmBoca_{SD}*, *mmMesd_{SD}* and *ceBMY-1_{SD}*
 Interfaces are shown for each of the four crystal structures of the Boca/Mesd family members, one in each of the four columns, with different properties shown in the rows.
 A) A backbone worm representation of each dimer pair from the four crystal structures, with each distinct molecular axes of symmetry from Table S2 (grey tubes) (Hendrickson, 1979). The superposition of each non-crystallographic dimer; *dmBoca(I)_{SD}* (blue), *mmMesd_{SD}* (green), and *ceBMY-1_{SD}* (purple) to the *dmBoca(II)_{SD}* protomer (upper, teal), further highlights the different positions of the second (lower) protomer within each dimer. The second (lower) *dmBoca(II)_{SD}* protomer is light teal to distinguish it as a crystallographically related dimer.
 B) The worm and axis representations as in (A) but rotated by 90° about the horizontal axis. The superimposed protomers are colored grey for clarity. The *dmBoca(II)_{SD}* forms an antiparallel dimer, in comparison to the parallel dimers in the other three crystal structures.
 C) The molecular surface of one protomer in each the dimer, relative orientation to the superposed protomers (grey worms) are maintained. The atoms buried by the largely overlapping set of residues at this interface are colored according to (A), the light purple surface of *ceBMY-1_{SD}* represents the additional contacts made by the unstructured residues that follow the extended β4 strand. The orientations here are as in (B) but rotated by 180° about the vertical axis. Further characterization of these surfaces can be found in Table S2.
 D) Hydrophobicity (Kyte and Doolittle, 1982) at each residues position, scaled from non-polar (red) to polar (white), in the same orientation as (C).
 E) Evolutionary conservation (Landau et al., 2005) at each residues position in the same orientation as (C), with 9 being the most highly conserved positions.
 F) An invariant isoleucine buried by all seven protomers in the four interfaces. This representative residue is typical of the overlapping but distinct contact patterns found at the atomic level in each crystal interface. Black double-headed arrows, represent an interactions

present in both protomers of the dimer, grey directional arrows signify an interaction present within the dimer from one protomer to the other, but that is not reciprocated. The residues names in grey are not contacted in that particular interface but are shown for clarity.

Table 1

Dimer Dissociation Constants

K_d (μM)		
<i>dmBoca</i> _{SD}	77	± 3.1
<i>dmBoca</i> _{CC}	155	± 4.6
<i>ceBMY1</i> _{SD}	106	± 4.2
<i>ceBMY1</i> _{FL}	220	± 4.4
<i>mmMesd</i> _{SD}	276	± 13.8
<i>mmMesd</i> _{FL}	~7000	

Table 2

Crystallographic Diffraction Data

Data Set	SeMet <i>dmBocasp</i> (I) ^d	SeMet <i>dmBocasp</i> (I) ^d	SeMet <i>dmBocasp</i> (I) ^d	SeMet <i>dmBocasp</i> (I) ^d	SeMet <i>dmBocasp</i> (II) ^d	SeMet ceBMY-I _{SD} ^c	<i>mmMesd</i> _{SD}
<i>d</i> min (Å)	2.3	2.5	2.5	2.5	2.8	1.37	2.01
Beamline	X25	X25	X25	X25	X4C	X4C	X4C
Space group	P4 ₁ 22	P4 ₁ 22	P4 ₁ 22	P4 ₁ 22	P6 ₅ 22	P1	P4 ₃ (twinned)
Wavelength (Å)	0.9794 (peak)	0.9795 (peak)	0.9795 (peak)	0.9802 (edge)	0.9790 (peak)	0.9790 (peak)	0.9790 (native)
Number of Reflections	15757	10635	10653	10653	13036	31382	12634
Average Redundancy	6.5	14.6	14.6	14.6	11.1	2.6	3.8
$\langle I \rangle / \langle \delta \rangle$ ^a	27.9 (7.5)	42.3 (9.0)	42.3 (9.0)	42.5 (9.0)	20.6 (5.3)	24.5 (9.2)	30.8 (7.8)
Completeness (%) ^a	99.3 (99.3)	85.8 (47.3)	85.9 (47.1)	85.9 (47.1)	99.0 (100)	94.1 (92.3)	99.0 (96.4)
Rmerge (%) ^b	0.69 (0.282)	0.067 (0.282)	0.067 (0.280)	0.067 (0.280)	0.129 (0.43)	0.036 (0.087)	0.038 (0.148)

^a Values in the outermost shell are given in parentheses.

^b $R_{\text{merge}} = (\sum |I_i - \langle I_i \rangle|) / \sum |I_i|$, where I_i is the integrated intensity of a given reflection.

^c Bijvoet mates are considered equivalent for scale and B-factor calculations.

^d Bijvoet mates are kept separate for scale and B-factor calculations.

Table 3

Crystallographic Refinement Statistics

Parameter	SeMet <i>dmBoca</i> _{SD} (I)	SeMet <i>dmBoca</i> _{SD} (II)	<i>mmMesd</i> _{SD}	SeMet <i>ceBMY1</i> _{SD}
Bragg spacings (Å)	2.288	2.002	2.013	1.367
Space group	P4 ₁ 22	P6 ₅ 22	P4 ₃	P1
Unit cell parameters a, b, c (Å)	34.106, 34.106, 311.238	62.284, 62.284, 89.675	71.651, 71.651, 37.102	29.522, 38.148, 40.706
α, β, γ (°)	90, 90, 90	90, 90, 120	90, 90, 90	63.39, 85.74, 84.69
Z _a ^a	2	1	2	2
β -Strands in each protomer	1 4-stranded 1 5-stranded	1 5-stranded	2 5-stranded	2 4-stranded
Solvent content (%)	46.7	52.0	49.4	35.4
R ^b	0.1892	0.1900	0.1474	0.1456
R _{free}	0.2279 ^c	0.2257 ^d	0.1744 ^d	0.1678 ^c
Number of reflections	15471	12039	12606	31382
Number of total atoms	1421	784	1401	3372
Number of total atoms (non-hydrogen)	1421	784	1401	1832
Number of protein atoms (non-hydrogen)	1300	688	1293	1593
Number of heteroatoms	0	20	0	4
Number of waters	121	76	108	235
Average B factor (Å ²)	26.75	23.58	31.32 (+ 2 TLS chains)	13.63
RMS bond ideality (Å)	0.003	0.003	0.003	0.007
RMS angle ideality (°)	0.601	0.667	0.563	1.022
Rotamer outliers ^e	0.74	1.47	1.50 %	0.00
Ramachandran (favored%/outlier%) ^e	99.35/0.00	98.80 /0.00	100.0/0.00	99.43/0.00
PDB code	3OFF	3OFE	3OFH	3OFG

^aZ_a stands for number of molecules per asymmetric unit

^bR = (||Fo| - |Fc||)/|Fo|, where Fo and Fc denote observed and calculated structure factors respectively.

^cR_{free} was calculated using 5% of data excluded from refinement.

^dR_{free} was calculated using 10% of data excluded from refinement.

^eMolprobrity analysis (<http://kinemge.biochem.duke.edu>) (Chen et al., 2010b)

Table 4

RMSD for seven crystallographically determined SDs and *mmMesd* NMR (2I9S) model.

RMSD (\AA^2) ^a	<i>dmBocasp</i> (IA)	<i>dmBocasp</i> (IB)	<i>dmBocasp</i> (II)	<i>ceBMY</i> _{1SD} A	<i>ceBMY</i> _{1SD} B	<i>mmMesdsp</i> A	<i>mmMesdsp</i> B	<i>mmMesd</i> NMR (2I9S)
<i>dmBocasp</i> (IA)		0.657 (70)	0.329 (81)	0.748 (70)	0.961 (71)	0.738 (80)	0.696 (78)	1.581 (46)
<i>dmBocasp</i> (IB)	0.591		0.569 (70)	0.683 (70)	0.909 (73)	0.850 (72)	0.841 (72)	1.424 (48)
<i>dmBocasp</i> (II)	0.290	0.508		0.670 (70)	0.887 (71)	0.690 (80)	0.693 (78)	1.501 (46)
<i>ceBMY</i> _{1SD} A	0.708	0.708	0.641		0.622 (82)	0.744 (70)	0.774 (71)	1.506 (51)
<i>ceBMY</i> _{1SD} B	0.899	0.764	0.838	0.577		0.865 (71)	0.777 (70)	1.576 (52)
<i>mmMesdsp</i> A	0.660	0.678	0.574	0.672	0.759		0.545 (79)	1.520 (50)
<i>mmMesdsp</i> B	0.594	0.652	0.540	0.621	0.714	0.402		1.443 (48)
<i>mmMesd</i> NMR (2I9S)	2.432	2.329	2.393	2.331	2.242	2.212	2.353	

^aRMSD for pairwise alignments of the all seven crystallographically determined SD protomers and the *mmMesd* NMR model (2I9S). Each pair are superimposed such that each superimposed regions contain at least three contiguous C α positions within 2.5 \AA . The upper half of the table shows the RMSD for each pair, with the total number of C α positions used in each optimized superimposition shown in brackets. The RMSD for 65 core C α positions in all crystallographically determined SD protomers and the *mmMesd* NMR model (2I9S) are shown in the lower half of the table. The 65 core C α positions are within 2.5 \AA in all seven SD and each superimposed regions contain more than 3 consecutive residues (also see Table S1).