Cell PRESS

Structure
Article

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

Mark N. Collins[1,3] and Wayne A. Hendrickson[1,2,*]
[1]Department of Biochemistry and Molecular Biophysics
[2]Howard Hughes Medical Institute
Columbia University, New York, NY 10032, USA
[3]Present address: Institute for Cellular and Molecular Biology, University of Texas at Austin, 2500 Speedway, Austin, TX 78712, USA
*Correspondence: wayne@convex.hhmi.columbia.edu
DOI 10.1016/j.str.2010.11.017

SUMMARY

Folding and trafficking of low-density lipoprotein receptor (LDLR) family members, which play essential roles in development and homeostasis, are 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 composed 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 ferrodoxin-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 nonspecific hydrophobic interactions that may play a role in the chaperone activity of the Boca/Mesd family.

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 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 multidomain proteins as they move from the endoplasmic reticulum (ER) for presentation at the cell surface: the LRP-receptor-associated protein (RAP) and 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 cancers (LRP5 and LRP6) (Björklund et al., 2009; Li et al., 2004). Several members (ApoER2 and LRP1) have also been implicated in Alzheimer's disease (Marzolo and Bu, 2009).

LDLR protein family members are characterized by the presence of three distinct types of extracellular domains: (1) cysteine-rich ligand-binding modules (also called complement, or LDL-A domains), (2) epidermal growth factor-like (EGF-like) domains, and (3) 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; in humans, sizes range 5-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 antiparallel β strands which

are offset from the YWTD repeats such that the "first" strand of the sixth blade is a C-terminal strand that follows the fifth 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 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–L177) was obtained for the structured core (Protein Data Bank [PDB] ID code 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 *boca*^{1} 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, 2006b; Liu et al., 2009). Boca/Mesd chaperones also prevent the formation of high-molecular-weight intermolecular disulfide 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: *Drosophila melanogaster* Boca (*dm*Boca), *Mus musculus* Mesd (*mm*Mesd), and *Caenorhabditis elegans* (*ce*BMY-1). NMR analysis of *dm*Boca 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 *mm*Mesd, *dm*Boca, and *ce*BMY-1 readily removed highly conserved N-terminal regions that precede the structured domain within the highly conserved core 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 ferrodoxin-like α/β fold, but there are both four-stranded and five-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 nonspecific 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. 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: *M. musculus* (*mm*) Mesd; *D. melanogaster* (*dm*) Boca; and *C. 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 *mm*Mesd and *dm*Boca, 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, http://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: *mm*Mesd, *dm*Boca, and *ce*BMY-1. Each of these proteins reveals a similar proteolytic profile (see Figure S1 available online). 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 an ~10 kDa protease-resistant fragment. This implies that the CC is composed of two regions: a stable 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.

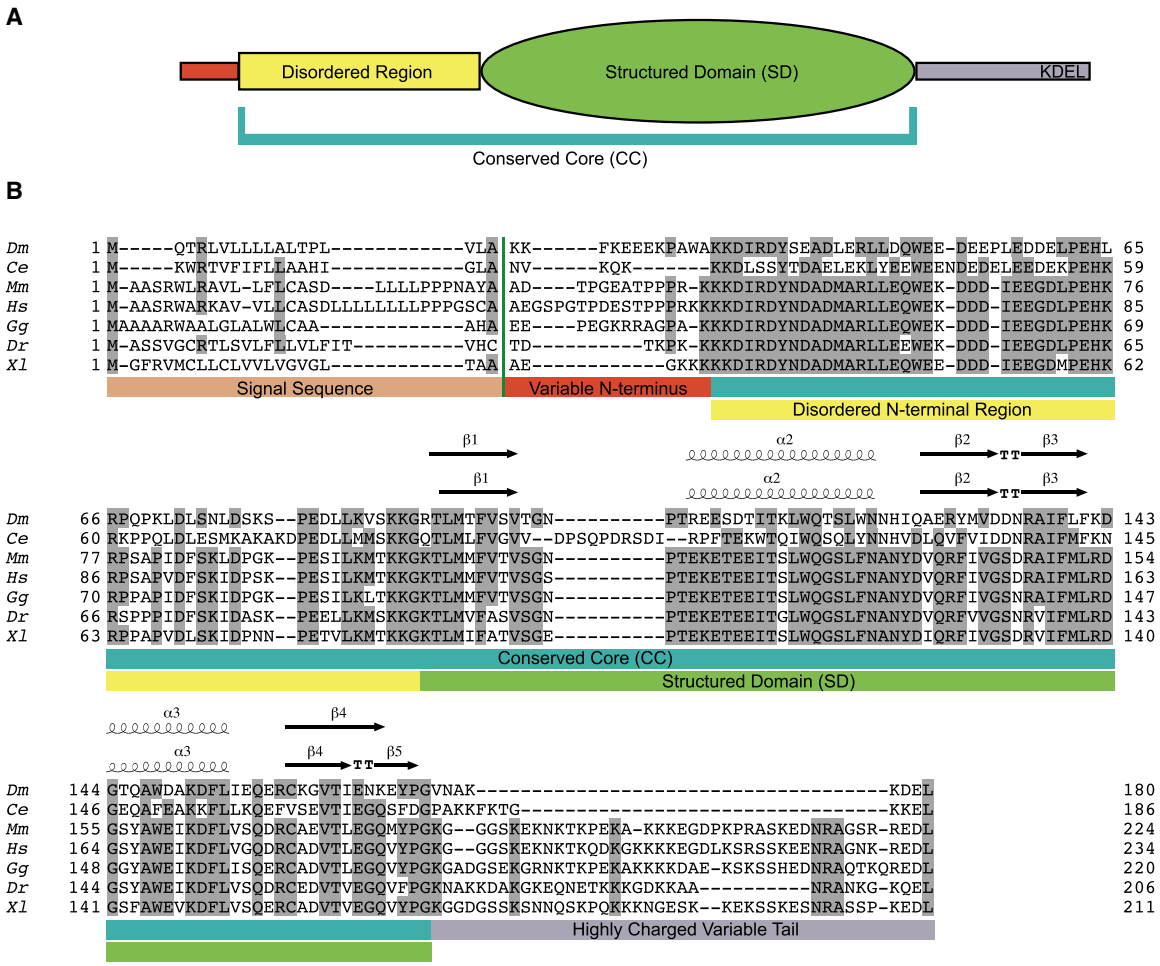

Figure 1. The Domain Organization of the Boca/Mesd Family

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

(B) Structure-based alignment for the Boca/Mesd chaperone family. The *dm*Boca (Dm), *ce*BMY-1 (Ce), and *mm*Mesd (Mm) sequences for the proteins in this work along with Boca/Mesd orthologs 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 the 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 *dm*Boca$_{SD}$ (form I), both four- and five-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).

Structural Properties of the Boca Conserved Core in Solution

To further test our conclusions on the domain structure of the CC, we recorded ^{15}N-^{1}H TROSY NMR spectra on ^{15}N-labeled *dm*Boca$_{CC}$ and ^{15}N-labeled *dm*Boca$_{SD}$ (Figure 2). Nearly all the dispersed peaks within the CC spectra overlap those of the SD spectra, implying that the SD includes all of the well-structured regions of the CC. The extra ~50 residues in the N-terminal region of *dm*Boca$_{CC}$ have resonance frequencies corresponding to those of an unstructured peptide, typified by limited dispersion centered at ^{1}H resonance frequencies of ~8.2 ppm. A number of the dispersed resonances from the 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 with the native

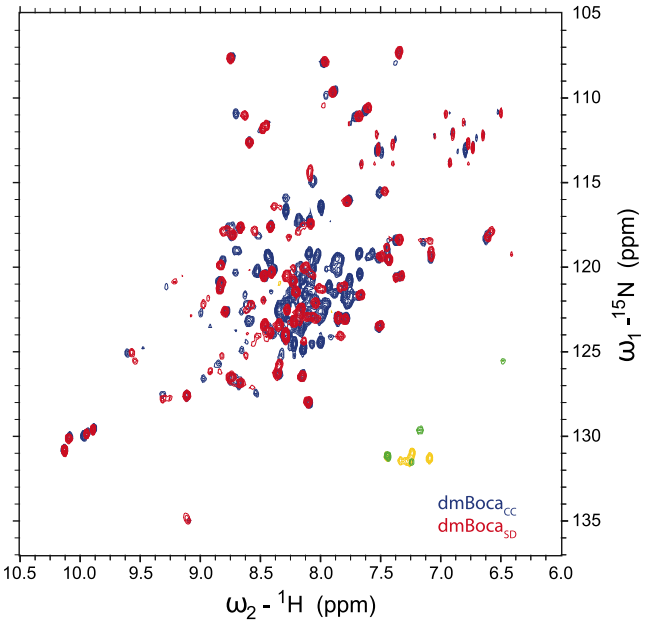

Figure 2. TROSY Spectra of *dm*Boca$_{CC}$ and *dm*Boca$_{SD}$

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

protein for all three structured domains and for either the corresponding CC or full-length protein (Figure S2; Table 1). The SDs all have quite similar K_Ds, which range from 77 ± 3 to 276 ± 14 μM. The addition of the disordered N-terminal residues in the native *dm*Boca$_{CC}$ construct or the mature *ce*BMY-1 protein weakens the SD dimer approximately 2-fold. The mature *mm*Mesd, 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 from Three Boca/Mesd Chaperones

We determined structures for the SDs from *dm*Boca, *mm*Mesd, and *ce*BMY-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 ferrodoxins. Together, these proteins represent a novel SCOP superfamily within the ferrodoxin-like fold (A. Murzin, personal communication). Each SD protomer has a βαββαβ core topology, wherein each βαβ has a right-handed conformation but with successive parallel β strands interlaced in an antiparallel 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.

Table 1. Dimer Dissociation Constants

	K_D (μM)
*dm*Boca$_{SD}$	77 ± 3.1
*dm*Boca$_{CC}$	155 ± 4.6
*ce*BMY-1$_{SD}$	106 ± 4.2
*ce*BMY-1$_{FL}$	220 ± 4.4
*mm*Mesd$_{SD}$	276 ± 13.8
*mm*Mesd$_{FL}$	~7000

Structured Domain of Boca

Two distinct crystal forms of *dm*Boca$_{SD}$ were solved using experimentally determined phases. Both crystals were produced from a selenomethionyl (SeMet) *dm*Boca SD. Multi-wavelength anomalous diffraction (MAD) at the Se K-edge was used to determine the structure of *dm*Boca$_{SD}$ in crystal form I (Figure 4). This crystal belongs to space group P4$_1$22, and has two molecules per asymmetric unit. A second crystal (form II), belonging to space group P6$_5$22, allowed an independent solution of the *dm*Boca$_{SD}$ structure by single-wavelength anomalous diffraction (SAD) at the Se K-peak.

The P4$_1$22 (form I) structure, determined at 2.3 Å resolution, includes ordered residues from R92 to G172 for one protomer, whereas 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 five stranded, and includes a short strand, β5, which follows β4 from a type I′ hairpin turn. In protomer B, this hairpin turn is absent and the β4 strand is extended without interruption (Figure 3B).

The structure determined for the P6$_5$22 crystals (form II), refined to 2.0 Å resolution, has a single protomer in the asymmetric unit (Figure 4). This molecule, like protomer A in crystal form I, has five β strands. A total of 84 ordered residues (K89–G172), 3 sulfate ions, 1 sodium ion, 1 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 *mm*Mesd$_{SD}$ were hemihedrally twinned throughout the P4$_3$ lattice, with two molecules per asymmetric unit (Figure 4). Despite this difficulty, the *mm*Mesd$_{SD}$ structure was solved by molecular replacement (MR) using the four-stranded protomer of *dm*Boca$_{SD}$ from crystal form I as the search model. A Bijvoet-difference map from SeMet *mm*Mesd$_{SD}$ 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 *mm*Mesd$_{SD}$ protomers, which are ordered from residues K104 to Y182 and M99 to G184. 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.

Table 2. Crystallographic Diffraction Data

Data Set	SeMet refine *dm*Boca$_{SD}$ (I)[a]	SeMet λ2 *dm*Boca$_{SD}$ (I)[a]	SeMet λ1 *dm*Boca$_{SD}$ (I)[a]	SeMet λ3 *dm*Boca$_{SD}$ (I)[a]	SeMet *dm*Boca$_{SD}$ (II)[a]	SeMet *ce*BMY-1$_{SD}$[b]	*mm*Mesd$_{SD}$
dmin (Å)	2.3	2.5	2.5	2.8	2.00	1.37	2.01
Beamline	X25	X25	X25	X25	X4C	X4C	X4C
Space group	$P4_122$	$P4_122$	$P4_122$	$P4_122$	$P6_522$	P1	$P4_3$ (twinned)
Wavelength (Å)	0.9794 (peak)	0.9795 (peak)	0.9802 (edge)	0.9645 (remote)	0.9790 (peak)	0.9790 (peak)	0.9790 (native)
Number of reflections	15,757	10,635	10,653	8,709	13,036	31,382	12,634
Average redundancy	6.5	14.6	14.6	15.2	11.1	2.6	3.8
$< I >/(\delta)$[c]	27.9 (7.5)	42.3 (9.0)	42.5 (9.0)	32.8 (9.6)	20.6 (5.3)	24.5 (9.2)	30.8 (7.8)
Completeness (%)[c]	99.3 (99.3)	85.8 (47.3)	85.9 (47.1)	98.5 (95.7)	99.0 (100)	94.1 (92.3)	99.0 (96.4)
R_{merge} (%)[d]	0.69 (0.282)	0.067 (0.282)	0.067 (0.280)	0.095 (0.313)	0.129 (0.43)	0.036 (0.087)	0.038 (0.148)

dmin, minimal Bragg spacing.
[a] Bijvoet mates are kept separate for scale and B factor calculations.
[b] Bijvoet mates are considered equivalent for scale and B factor calculations.
[c] Values in the outermost shell are given in parentheses.
[d] $R_{merge} = (\sum|Ii - < Ii >|)/\sum |Ii|$, where Ii is the integrated intensity of a given reflection.

Structured Domain of BMY-1

The SeMet *ce*BMY-1$_{SD}$ crystal belongs to space group P1 and contains two molecules per unit cell (Figure 4). Bijvoet-difference maps calculated from the Se K-peak data set confirmed the Se positions in the MR solution, for which the search model employed was the five-stranded SD of *dm*Boca (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

Table 3. Crystallographic Refinement Statistics

Parameter	SeMet *dm*Boca$_{SD}$ (I)	SeMet *dm*Boca$_{SD}$ (II)	*mm*Mesd$_{SD}$	SeMet *ce*BMY-1$_{SD}$
Bragg spacings (Å)	2.288	2.002	2.013	1.367
Space group	$P4_122$	$P6_522$	$P4_3$	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
Za[a]	2	1	2	2
β strands in each protomer	1 four-stranded 1 five-stranded	1 five-stranded	2 five-stranded	2 four-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	15,471	12,039	12,606	31,382
Number of total atoms	1,421	784	1,401	3,372
Number of total atoms (non-hydrogen)	1,421	784	1,401	1,832
Number of protein atoms (non-hydrogen)	1,300	688	1,293	1,593
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 ID code	3OFE	3OFF	3OFH	3OFG

[a] Za, number of molecules per asymmetric unit.
[b] $R = (||F_o| - |F_c||)/|F_o|$, where F_o and F_c denote observed and calculated structure factors, respectively.
[c] R_{free} was calculated using 5% of data excluded from refinement.
[d] R_{free} was calculated using 10% of data excluded from refinement.
[e] MolProbity analysis (http://kinemage.biochem.duke.edu) (Chen et al., 2010b).

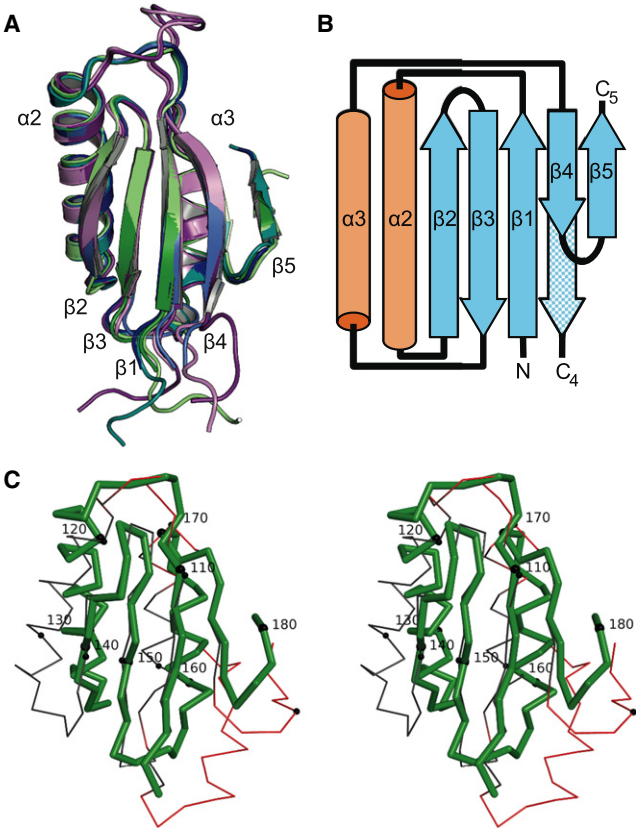

Figure 3. Similarities and Differences of the Boca/Mesd Family Structured Domain

(A) Structural superposition of all seven Boca/Mesd family SD protomers. The seven crystallographically determined protomers, in four different crystal lattices, are from three divergent species: mouse (green and light green), fly (blue, light blue, and teal), and worm (purple and light purple). All seven structures are within 0.9 Å rmsd. At the C terminus of the SD is 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 five-stranded and extended β4 conformations of the β sheet are shown. The α helices (orange) and β strands (blue) are numbered to be consistent with the NMR structure (PDB ID code 2I9S). The five-stranded topology ends at the C_5 terminus and the β4 extension (checked blue) at the C_4 terminus.

(C) Stereo picture of the Cα trace for the crystal structure of *mm*$Mesd_{SD}$, with the same orientation and color green as (A), superimposed with the NMR model (PDB ID code 2I9S) via the Escet-defined conformationally invariant β sheet (Schneider, 2004), reveals the nearly 5 Å rmsd offset for the helices in the NMR structure. Every tenth residue is presented as a sphere and numbered. The thin gray Cα trace of the NMR model is highlighted in red for Cα positions greater than three standard deviations from the internal rmsd of the NMR core.

β sheets possess an extended β4 strand conformation in both *C. elegans* SD protomers. The C-terminal residues, which are disordered in the four-stranded SD of *dm*Boca (form I), here instead form a short C-terminal tail following the extended β4 strand. Each tail reaches across to the other SD within the unit cell (Figure 4A); 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, whereas 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 root-mean-square deviations (rmsds) for 65 equivalent Cα positions (see Experimental Procedures) among all seven protomers are all less than 0.9 Å. The solution structure of an *mm*$Mesd_{SD}$ 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 Cα 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 *mm*$Mesd_{SD}$ crystal structure. In this superposition, the β sheet (27 Cα positions) has an rmsd of less than 1 Å, whereas the two α helices (28 Cα positions) in the NMR model are substantially displaced relative to those in the crystal structure and have 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 β4 strand similar to the β4 strand found in the five-stranded conformation of SD, but it lacks β5 (Figure 3C).

A recent NMR solution structure of a full-length *mm*$Mesd_{FL}$ construct (PDB ID code 2KMI) (Chen et al., 2010a) shows major tertiary structural differences as compared to our seven crystallographically determined SDs or to the *mm*$Mesd_{SD}$ 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 *dm*Boca structure of crystal form II, the dimer is specified by a crystallographical diad; in the other three crystals, however, the two protomers per asymmetric unit are related by approximate diad axes. The quasi-two-fold noncrystallographic symmetry axes for these three, *mm*Mesd, *ce*BMY-1, and *dm*Boca (form I), differ in screw-axis parameters (Figures 4A and 4B; Table S2). Moreover, the relationship between the *dm*Boca protomers in the form I dimer is rotated ~160° from that in the form II dimer. In each case, however, the same surface of the 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 *dm*Boca have very different arrangements.

The buried surface area (Figure 4C; Table S2) for the *mm*Mesd, *dm*Boca(I), and *dm*Boca(II) structures ranges from 600 to 900 $Å^2$.

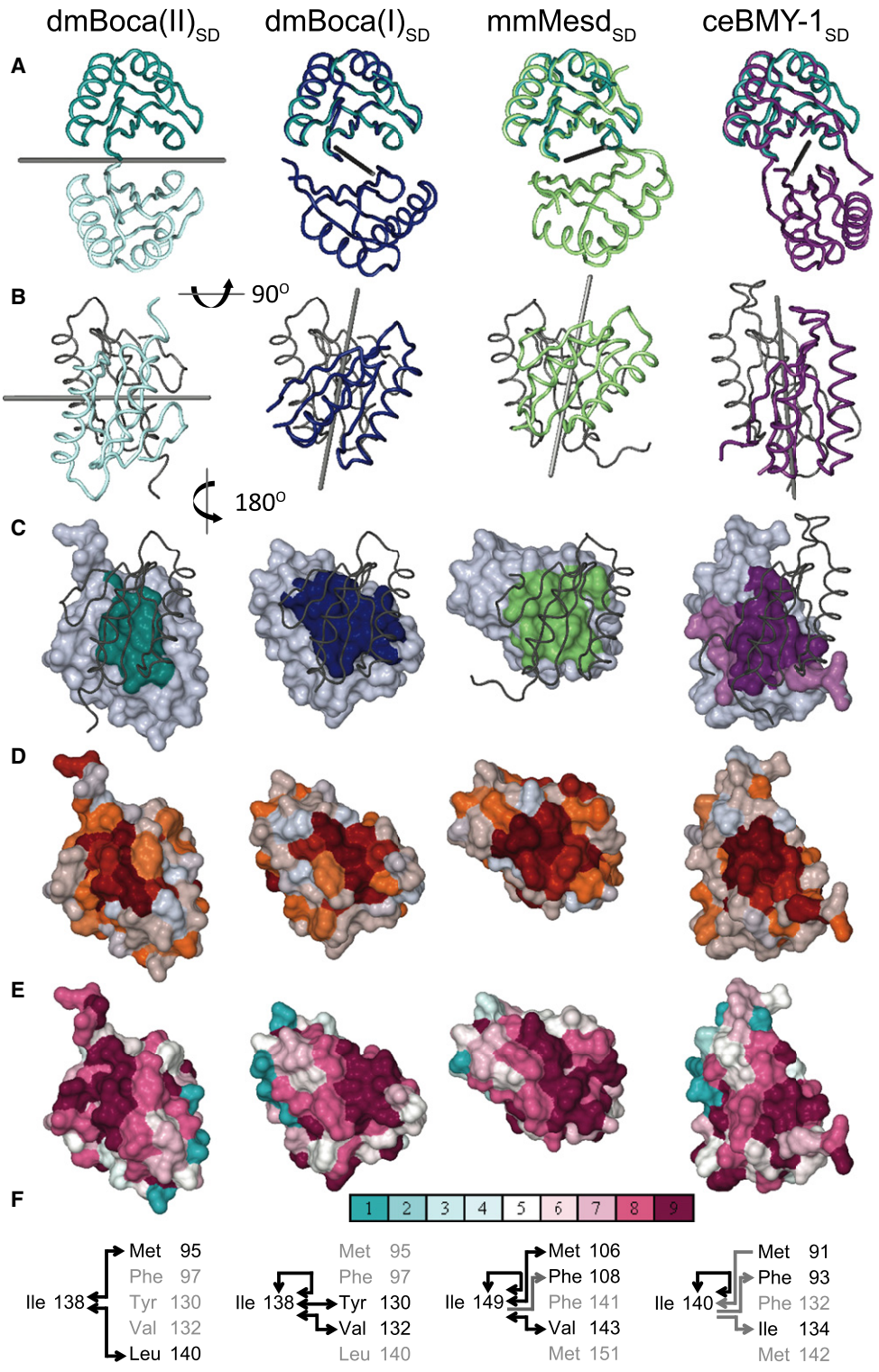

Figure 4. Properties of the Distinct Dimer Interfaces of *dm*Boca$_{SD}$, *mm*Mesd$_{SD}$, and *ce*BMY-1$_{SD}$

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 (gray tubes) (Hendrickson, 1979). The superposition of each noncrystallographic dimer, *dm*Boca(I)$_{SD}$ (blue), *mm*Mesd$_{SD}$ (green), and *ce*BMY-1$_{SD}$ (purple), to the *dm*Boca(II)$_{SD}$ protomer (upper, teal) further highlights the different positions of the second (lower) protomer within each dimer. The second (lower) *dm*Boca(II)$_{SD}$ protomer is light teal to distinguish it as a crystallographically related dimer.

Table 4. Root-Mean-Square Deviations for Seven Crystallographically Determined SDs and *mm*Mesd NMR Model

Rmsd (Å)	*dm*Boca$_{SD}$ (I)A	*dm*Boca$_{SD}$ (I)B	*dm*Boca$_{SD}$ (II)	*ce*BMY-1$_{SD}$ A	*ce*BMY-1$_{SD}$ B	*mm*Mesd$_{SD}$ A	*mm*Mesd$_{SD}$ B	*mm*Mesd NMR (2I9S)
*dm*Boca$_{SD}$ (I)A		0.657 (70)	0.329 (81)	0.748 (70)	0.961 (71)	0.738 (80)	0.696 (78)	1.581 (46)
*dm*Boca$_{SD}$ (I)B	0.591		0.569 (70)	0.683 (70)	0.909 (73)	0.850 (72)	0.841 (72)	1.424 (48)
*dm*Boca$_{SD}$ (II)	0.290	0.508		0.670 (70)	0.887 (71)	0.690 (80)	0.693 (78)	1.501 (46)
*ce*BMY-1$_{SD}$ A	0.708	0.708	0.641		0.622 (82)	0.744 (70)	0.774 (71)	1.506 (51)
*ce*BMY-1$_{SD}$ B	0.899	0.764	0.838	0.577		0.865 (71)	0.777 (70)	1.576 (52)
*mm*Mesd$_{SD}$ A	0.660	0.678	0.574	0.672	0.759		0.545 (79)	1.520 (50)
*mm*Mesd$_{SD}$ B	0.594	0.652	0.540	0.621	0.714	0.402		1.443 (48)
*mm*Mesd NMR (2I9S)	2.432	2.329	2.393	2.331	2.242	2.212	2.353	

Rmsds for pairwise alignments of all seven crystallographically determined SD protomers and the *mm*Mesd NMR model (PDB ID code 2I9S). Each pair is superimposed such that each superimposed region contains at least three contiguous Cα positions within 2.5 Å. The upper half of the table shows the rmsd for each pair, with the total number of Cα positions used in each optimized superimposition shown in parentheses. The rmsds for 65 core Cα positions in all crystallographically determined SD protomers and the *mm*Mesd NMR model (2I9S) are shown in the lower half of the table. The 65 core Cα positions are within 2.5 Å in all seven SDs, and each superimposed region contains more than three consecutive residues (also see Table S1).

The *ce*BMY-1 structure is an outlier, burying 1700 Å^2 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 β4 strand from both *ce*BMY-1 protomers, we create a model for each *ce* SD that resembles the four-stranded *dm*Boca (form I) protomer. The hydrophobic interface created by the *ce*BMY-1 dimer in the absence of these tails closely matches the surface of the other three structures (Figures 4C and 4D). The 846 Å^2 buried in the *ce* model dimer lies directly between the buried surface areas for the other two asymmetric parallel dimers, 898 Å^2 for *dm*Boca (form I) and 791 Å^2 for *mm*Mesd.

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 β4 strand structure or the alternative hairpin turn/β5 strand conformation. Thus, conservation within the SD includes the surface of the dimer, primarily residues from strands β1–β3, 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 β1, β2, and β3, contribute from both protomers within all four dimers. Central to all is a buried and invariant isoleucine residue (*dm*Boca Ile138, *ce*BMY-1 Ile140, *mm*Mesd 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 *ce*BMY-1 and *mm*Mesd structures (Figure 4F). Each of the conserved interfacial isoleucine residues contacts two to four residues in the opposing protomer. The two crystal forms of *dm*Boca have entirely discrete contact patterns for Ile138, but individual contacts in both *dm*Boca crystal forms are replicated in the *ce*BMY-1 and *mm*Mesd structures. Other residues alternatively contacted by the central isoleucine (Figure 4F) are also either absolutely conserved (*dm*Boca Met95 and Phe97) or conservatively substituted (*dm*Boca Tyr130 and Leu140 versus *mm*Mesd Phe141 and Met151; *dm*Boca Val132 versus

(B) The worm and axis representations as in (A) but rotated by 90° about the horizontal axis. The superimposed protomers are colored gray for clarity. The *dm*Boca(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 dimer; relative orientations to the superposed protomers (gray 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 *ce*BMY-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 residue's position, scaled from nonpolar (red) to polar (white), in the same orientation as (C).
(E) Evolutionary conservation (Landau et al., 2005) at each residue's position in the same orientation as (C), with 9 being the most highly conserved position.
(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 indicate an interaction present in both protomers of the dimer; gray directional arrows signify an interaction present within the dimer from one protomer to the other but that is not reciprocated. The equivalent residues names in gray are not contacted in that particular interface but are shown for clarity.

*ce*BMY-1 Ile134), and these are also involved in pairwise contacts between one another.

DISCUSSION

Boca/Mesd family members include a conserved core, 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 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 β1, β2, and β3 strands of the β sheet, but the relative orientations of the 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 homologs and crystal forms. The small surface areas buried by the β sheets at each SD interface are approximately equivalent, ranging from 600 to 900 $Å^2$ 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. 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 nonspecific hydrophobic surface of the SD must exist. An unpublished *mm*Mesd NMR structure (PDB ID code 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 four-stranded sheet and in four protomers as a five-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 β5 strand or an extension of the β4 strand without a β hairpin. One *dm*Boca crystal structure contains both four- and five-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 five-stranded protomer in the *dm*Boca(I) lattice and for both in the *mm*Mesd lattice that a four-stranded model can be accommodated without conflict and without apparent compromise of lattice integrity. Thus, although we cannot rule out lattice artifacts in all cases, it does appear that intrinsic energetics are defining for three of the four five-stranded protomers.

The two crystallographically determined *mm*Mesd SDs are five stranded, but neither the fifth-stranded nor the extended fourth-stranded conformation is directly observed in the *mm*Mesd (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 β4 and the turn into β5, 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 (*mm*Mesd Val174-Thr175, *dm*Boca Val163-Thr164, *ce*BMY-1 Val165-Thr166) anchors the β4 strand to the adjacent β1 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 φ change between the two secondary structural states seen in the crystals. The ϕ and φ 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

move ~13 Å. 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 four-stranded and five-stranded 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 *mm*Mesd, 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 analytical ultracentrifugation experiments suggest that interactions between these regions prevent dimer formation. How they might interact with one another and with LDLR family substrates remain 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 30–45 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 *mm*Mesd with the C-terminal tail removed fails to bind to LRP6, likely because the receptor is fully folded and the CC is expected to interact with a nonnative state to carry out its chaperone function (Li et al., 2005). In a separate functional assay, full-length *mm*Mesd rescues LRP6 surface expression in HEK293T cells treated with human Mesd 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 *mm*Mesd crystal and NMR (PDB ID code 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. Nonspecific binding occurs through the conserved surface residues, primarily located on strands β1, β2, and β3. The energetic cost of 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 LRPs depends on the action of 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; *dm*Boca 19–176, *mm*Mesd 30–220, *ce*BMY-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 *mesd* and *boca* (both kindly provided by J. Culi), and *BMY-1* (Boca/Mesd chaperone for YWTD β propeller-EGF) using appropriate 5′ and 3′ primers. The full *BMY-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 hexahistidine 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 *dm*Boca$_{SD}$ product appeared to have additional residues, GSQKHM, based upon mass spectrometry and N-terminal sequencing.

Each fusion protein was expressed in *Escherichia coli* BL21(DE3) cells (Novagen), grown in rich defined autoinduction media; PA5052 or PASM5052 media were used for SeMet incorporation at 37°C for 6 hr 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 50 mM Tris (pH 8.0) and 500 mM 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 μm) before initial affinity purification using a chelating HiTrap column (GE Healthcare) charged with Ni^{2+}. Each protein was eluted with a linear imidazole gradient and dialyzed in the presence of thrombin (GE

Healthcare) against 20 mM Tris (pH 8.0) and either 200 mM or 500 mM NaCl overnight at 4°C. Samples were concentrated (Amicon) and further purified by gel filtration on a Superdex 75 26/60 column (GE Healthcare) 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 *dm*Boca$_{SD}$ by hanging-drop vapor diffusion with a 1 μl:1 μl ratio of reservoir to protein solution (25 mg/ml in 10 mM Tris [pH 8.0], 100 mM NaCl). Form I crystals were grown at 4°C against a reservoir solution containing 18%–21% (w/v) polyethylene glycol (PEG) 3350, 100–200 mM Na citrate, 100 mM 2-(cyclohexylamino) ethanesulfonic acid (CHES) (pH 9.0), and 10 mM dithiothreitol by macroseeding from an equivalent condition, which in turn were microseeded from showers of small native *dm*Boca$_{SD}$ (22 mg/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 *dm*Boca$_{SD}$ crystals (form II) were one of four crystal morphologies that grew against a reservoir solution containing 48%–52% saturated $AmSO_4$, 100 mM sodium acetate (pH 5.6–5.8) at 20°C. The thin hexagonal pyramid-shaped crystals were soaked in cryoprotection buffer consisting of 80% saturated $AmSO_4$, 100 mM 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 under this condition did not produce sufficient quality diffraction data for structure determination.

Crystals of *mm*Mesd$_{SD}$ were grown by hanging-drop vapor diffusion against a mother liquor containing 0.7–0.9 M $AmPO_4$ and 100 mM HEPES (pH 7.5) at 4°C which were nucleated by streak seeding from microcrystal showers found under conditions with higher $AmPO_4$ precipitant concentration. A native protein solution (27 mg/ml in 10 mM Tris [pH 8.0], 10 mM NaCl) to reservoir ratio of 1:1 (1 μl each) was used. Stout square rod-shaped crystals 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 *ce*BMY-1$_{SD}$ crystals were grown by hanging-drop vapor diffusion at 20°C against a crystallization buffer containing 15%–18% PEG 3350, 300–500 mM $LiCl_2$, and 100 mM MES (pH 5.5). A protein solution (38 mg/ml 10 mM Tris [pH 8.0], 10 mM NaCl, 1 mM DDT) was used with a protein to buffer ratio of 1:1 (1 μl each). Thick square plate crystals typically appeared overnight and grew to an optimal size within 7 days. Crystals were cryoprotected 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 the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. Diffraction data to Bragg spacings of 2.5 Å (peak, edge) and 2.8 Å (remote) with a 440 mm detector distance were collected using 6 s 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 data set, with no inverse beam, was collected through a different portion of the same crystal, to a Bragg spacing of 2.3 Å, using 2 s exposure times, 1° oscillations, and a detector distance of 320 mm. Data sets 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 four of the six potential Se sites for the $P4_122$ 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 2.3 Å 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 f″ peak SAD experiment was collected at the Se K-edge on the NSLS X4C beamline from a single frozen form II crystal of SeMet *dm*Boca$_{SD}$. Diffraction data to a Bragg spacing of 2.0 Å with a 100 mm detector distance were recorded using 20 s exposure times, 1° oscillations, and an 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 $P6_522$. 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 (v1.3) with iterative rounds of manual rebuilding in Coot.

mmMesd$_{SD}$

Data from a frozen crystal of *mm*Mesd$_{SD}$ were collected at the NSLS beamline X4C to a Bragg spacing of 2.0 Å with 9 s exposure times, 0.5° oscillations, and a detector distance of 150 mm. HKL2000 was used for data processing, and the native *mm*Mesd$_{SD}$ structure was determined, in the $P4_3$ space group, by molecular replacement using Phaser (v2.0) (McCoy et al., 2007) with the four-stranded *dm*Boca$_{SD}$ (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 asymmetric unit, with iterative rounds of manual rebuilding in Coot. An Se K-edge SeMet *mm*Mesd$_{SD}$ experiment (data not shown), which 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 *mm*Mesd$_{SD}$ model.

ceBMY-1$_{SD}$

Data from a single SeMet *ce*BMY-1$_{SD}$ crystal were collected in two parts at the X4C beamline of NSLS. The first sweep, using a detector distance of 88 mm and 1° oscillations of 8 s exposure times, diffracted X-rays to 1.37 Å resolution. The second sweep used a detector distance of 135 mm and 1° oscillations of 1 s 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 were 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 five-stranded *dm*Boca$_{SD}$ (I) protomer as a search model was fed to ArpWarp for phase improvement and automated model building. The resulting model with an 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 Cα positions of each SD protomer were superposed and manually adjusted to maximize the number of structurally aligned residues in both pairwise and multiple model superpositions, such that three or more contiguous Cα positions were within 2.5 Å of equivalent positions. The rotation-translation matrices of the molecular axes of symmetry were determined by TOSS (Hendrickson, 1979) using the core 65 Cα positions 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 *mm*Mesd$_{SD}$ protomers to the 20 models in the NMR ensemble (PDB ID code 2I9S). CCP4 programs contact and areaimol (CCP4, 1994) were used to define the atoms participating in the interface and assess their change in buried surface area during dimer formation. An alignment of 48 orthologous Boca/Mesd sequences grouped by TreeFam (http://www.treefam.org) (Li et al., 2006a) was in conjunction with ConSurf (http://consurf.tau.ac.il) (Landau et al., 2005) to map the level of evolutionary conservation at each residue's 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 (10 mM NaCl, 10 mM Tris [pH 8.0]) at 1 mg/ml, and the serial dilutions of these stock solutions were stored on ice. Proteolytic reactions, set up on ice, contained 9 μl of a purified MP diluted to 1 mg/ml with cold protease buffer and 1 μl of protease at a final concentration of 10, 1, 0.1, and 0.01 μg/ml, or 1 μl PBS for the control. Reactions were moved to 20°C and terminated after 30 min by removing 2 μl into 2 μl of TFA for analysis by MALDI-TOF mass spectrometry, while the remaining 8 μl was stopped by the addition of 3 μl 4× SDS-PAGE loading buffer. After boiling for 3 min, proteolytically digested samples were analyzed by SDS-PAGE,

and both were visualized with Coomassie blue and transferred to polyvinyldifluoride for N-terminal sequencing by Edman degradation.

Analytical Ultracentrifugation

Sedimentation equilibrium experiments were performed with a Beckman-Coulter XLI analytical ultracentrifuge using both UV 280 nm and IR 660 nm optics. Native purified *dm*Boca$_{SD}$, *mm*Mesd$_{SD}$, and *ce*BMY-1$_{SD}$ were dialyzed at 4°C and then diluted to approximately 0.70, 0.46, and 0.24 mg/ml with 2× PBS. Native purified *dm*Boca$_{cc}$, *mm*Mesd$_{MP}$, and *ce*BMY-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 1 hr intervals. The program HeteroAnalysis (v1.0.114) (http://www.biotech.uconn.edu/auf/ha-help/HA-Help.htm) enabled group fitting of the equilibrium data at all three speeds and concentrations for each protein sample.

NMR Spectroscopy

Recombinant *dm*Boca$_{CC}$ and *dm*Boca$_{SD}$ constructs were expressed in M9 minimal medium (Sambrook and Russell, 2001) with 99%-$^{15}NH_4Cl$ as the sole nitrogen source (Cambridge Isotopes) in the M9 salts. Protein purification followed the protocol described above. The NMR buffer consisted of 20 mM MES (pH 6.0) and 50 mM NaCl. *dm*Boca$_{CC}$ was dialyzed overnight at 4°C against NMR buffer and concentrated to 0.66 mM; the *dm*Boca$_{SD}$ was exchanged directly into NMR buffer during concentration to 0.42 mM. NMR buffers also contained 0.02% NaN_3 and 10% D_2O.

NMR ^{15}N-^{1}H TROSY spectra for both ^{15}N-labeled 0.36 mM *dm*Boca$_{SD}$ and 0.46 mM *dm*Boca$_{CC}$ were acquired on a Bruker DRX-600 with a triple-resonance z axis gradient cryogenic probe at 299.5K. The sample temperature was calibrated with 99.8% d_4-methanol, and postacquisition processing of the spectra was performed with NMRPipe (Delaglio et al., 1995) and Sparky 3 (http://www.cgl.ucsf.edu/home/sparky).

Protein Alignment

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.

ACCESSION NUMBERS

Atomic coordinates and structure factors have been deposited in the Protein Data Bank (http://www.rcsb.org/pdb) under PDB ID codes 3OFE, 3OFF, 3OFG, and 3OFH. See Table 3 for identifications.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures and four tables and can be found with this article online at doi:10.1016/j.str.2010.11.017.

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 beamlines 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 PXRR Initiative. This work was supported in part by NIH Biophysics Training Grant 5 T32 GM008281-19 and by NIH grant GM34102.

Received: August 17, 2010
Revised: November 9, 2010
Accepted: November 17, 2010
Published: March 8, 2011

REFERENCES

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr. *58*, 1948–1954.

Anelli, T., and Sitia, R. (2008). Protein quality control in the early secretory pathway. EMBO J. *27*, 315–327.

Björklund, P., Svedlund, J., Olsson, A.-K., Akerström, G., and Westin, G. (2009). The internally truncated LRP5 receptor presents a therapeutic target in breast cancer. PLoS One *4*, e4243.

Blacklow, S.C. (2007). Versatility in ligand recognition by LDL receptor family proteins: advances and frontiers. Curr. Opin. Struct. Biol. *17*, 419–426.

Brünger, 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., et al. (1998). Crystallography & NMR System: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

CCP4 (Collaborative Computational Project, Number 4). (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Chen, J., Li, Q., Liu, C.-C., Zhou, B., Bu, G., and Wang, J. (2010a). NMR structure note: solution structure of the core domain of MESD that is essential for proper folding of LRP5/6. J. Biomol. NMR *47*, 283–288.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010b). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

Chen, J., Liu, C.-C., Li, Q., Nowak, C., Bu, G., and Wang, J. (2011). Two Structural and Functional Domains of MESD Required for Proper Folding and Trafficking of LRP5/6. Structure *19*, this issue, 313–324.

Chothia, C., and Janin, J. (1981). Relative orientation of close-packed β-pleated sheets in proteins. Proc. Natl. Acad. Sci. USA *78*, 4146–4150.

Clamp, M., Cuff, J., Searle, S.M., and Barton, G.J. (2004). The Jalview Java alignment editor. Bioinformatics *20*, 426–427.

Cowtan, K. (1994). 'dm': an automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography *31*, 34–38.

Culi, J., and Mann, R.S. (2003). Boca, an endoplasmic reticulum protein required for wingless signaling and trafficking of LDL receptor family members in *Drosophila*. Cell *112*, 343–354.

Culi, J., Springer, T.A., and Mann, R.S. (2004). Boca-dependent maturation of β-propeller/EGF modules in low-density lipoprotein receptor proteins. EMBO J. *23*, 1372–1380.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System (San Carlos, CA: DeLano Scientific).

Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. *6*, 197–208.

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. *2*, 953–971.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Gouet, P. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res. *31*, 3320–3323.

Hartl, F.U., and Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. Science *295*, 1852–1858.

Hendrickson, W.A. (1979). Transformations to optimize the superposition of similar structures. Acta Crystallogr. A *35*, 158–163.

Herz, J. (2006). The switch on the RAPper's necklace. Mol. Cell *23*, 451–455.

Herz, J. (2009). Apolipoprotein E receptors in the nervous system. Curr. Opin. Lipidol. *20*, 190–196.

Hsieh, J.C., Lee, L., Zhang, L., Wefer, S., Brown, K., DeRossi, C., Wines, M.E., Rosenquist, T., and Holdener, B.C. (2003). Mesd encodes an LRP5/6 chaperone essential for specification of mouse embryonic polarity. Cell *112*, 355–367.

Hutchinson, E.G., and Thornton, J.M. (1994). A revised set of potentials for β-turn formation in proteins. Protein Sci. *3*, 2207–2216.

Jeon, H., Meng, W., Takagi, J., Eck, M.J., Springer, T.A., and Blacklow, S.C. (2001). Implications for familial hypercholesterolemia from the structure of the LDL receptor YWTD-EGF domain pair. Nat. Struct. Biol. *8*, 499–504.

Jones, S., and Thornton, J.M. (1996). Principles of protein-protein interactions. Proc. Natl. Acad. Sci. USA *93*, 13–20.

Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. *9*, 286–298.

Kim, N., Stiegler, A.L., Cameron, T.O., Hallock, P.T., Gomez, A.M., Huang, J.H., Hubbard, S.R., Dustin, M.L., and Burden, S.J. (2008). Lrp4 is a receptor for Agrin and forms a complex with MuSK. Cell *135*, 334–342.

Kleywegt, G.J. (1996). Use of non-crystallographic symmetry in protein structure refinement. Acta Crystallogr. D Biol. Crystallogr. *52*, 842–857.

Koduri, V., and Blacklow, S.C. (2007). Requirement for natively unstructured regions of mesoderm development candidate 2 in promoting low-density lipoprotein receptor-related protein 6 maturation. Biochemistry *46*, 6570–6577.

Köhler, C., Andersen, O.M., Diehl, A., Krause, G., Schmieder, P., and Oschkinat, H. (2006). The solution structure of the core of mesoderm development (MESD), a chaperone for members of the LDLR-family. J. Struct. Funct. Genomics *7*, 131–138.

Köhler, C., Lighthouse, J.K., Werther, T., Andersen, O.M., Diehl, A., Schmieder, P., Du, J., Holdener, B.C., and Oschkinat, H. (2011). The Structure of MESD45–184 Brings Light into the Mechanism of LDLR Folding. Structure *19*, this issue, 338–348.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. *157*, 105–132.

Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., and Ben-Tal, N. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res. *33*, W299–W302.

Li, Y., Lu, W., He, X., Schwartz, A.L., and Bu, G. (2004). LRP6 expression promotes cancer cell proliferation and tumorigenesis by altering β-catenin subcellular distribution. Oncogene *23*, 9129–9135.

Li, Y., Chen, J., Lu, W., McCormick, L.M., Wang, J., and Bu, G. (2005). Mesd binds to mature LDL-receptor-related protein-6 and antagonizes ligand binding. J. Cell Sci. *118*, 5305–5314.

Li, H., Coghlan, A., Ruan, J., Coin, L.J., Hériché, J.-K., Osmotherly, L., Li, R., Liu, T., Zhang, Z., Bolund, L., et al. (2006a). TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic Acids Res. *34*, D572–D580.

Li, Y., Lu, W., He, X., and Bu, G. (2006b). Modulation of LRP6-mediated Wnt signaling by molecular chaperone Mesd. FEBS Lett. *580*, 5423–5428.

Lillis, A.P., Van Duyn, L.B., Murphy-Ullrich, J.E., and Strickland, D.K. (2008). LDL receptor-related protein 1: unique tissue-specific functions revealed by selective gene knockout studies. Physiol. Rev. *88*, 887–918.

Liu, C.-C., Pearson, C., and Bu, G. (2009). Cooperative folding and ligand-binding properties of LRP6 β-propeller domains. J. Biol. Chem. *284*, 15299–15307.

Lovell, S.C., Word, J.M., Richardson, J.S., and Richardson, D.C. (2000). The penultimate rotamer library. Proteins *40*, 389–408.

MacDonald, B.T., Tamai, K., and He, X. (2009). Wnt/β-catenin signaling: components, mechanisms, and diseases. Dev. Cell *17*, 9–26.

Marzolo, M.-P., and Bu, G. (2009). Lipoprotein receptors and cholesterol in APP trafficking and proteolytic processing, implications for Alzheimer's disease. Semin. Cell Dev. Biol. *20*, 191–200.

May, P., Herz, J., and Bock, H.H. (2005). Molecular mechanisms of lipoprotein receptor signalling. Cell. Mol. Life Sci. *62*, 2325–2338.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Meunier, L., Usherwood, Y.-K., Chung, K.T., and Hendershot, L.M. (2002). A subset of chaperones and folding enzymes form multiprotein complexes in endoplasmic reticulum to bind nascent proteins. Mol. Biol. Cell *13*, 4456–4469.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.

Saibil, H.R. (2008). Chaperone machines in action. Curr. Opin. Struct. Biol. *18*, 35–42.

Sambrook, J., and Russell, D.W. (2001). Molecular Cloning: A Laboratory Manual, Third Edition (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Schneider, T.R. (2004). Domain identification by iterative analysis of error-scaled difference distance matrices. Acta Crystallogr. D Biol. Crystallogr. *60*, 2269–2275.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Terwilliger, T.C. (2000). Maximum-likelihood density modification. Acta Crystallogr. D Biol. Crystallogr. *56*, 965–972.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. *55*, 849–861.

Tompa, P., and Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. FASEB J. *18*, 1169–1175.

Vembar, S.S., and Brodsky, J.L. (2008). One step at a time: endoplasmic reticulum-associated degradation. Nat. Rev. Mol. Cell Biol. *9*, 944–957.

Yan, Y., Tropsha, A., Hermans, J., and Erickson, B.W. (1993). Free energies for refolding of the common β turn into the inverse-common β turn: simulation of the role of D/L chirality. Proc. Natl. Acad. Sci. USA *90*, 7898–7902.

Yang, S., Hitz, B., and Honig, B. (1996). Free energy determinants of secondary structure formation: III. β-turns and their role in protein folding. J. Mol. Biol. *259*, 873–882.

Zhang, B., Luo, S., Wang, Q., Suzuki, T., Xiong, W.C., and Mei, L. (2008). LRP4 serves as a coreceptor of agrin. Neuron *60*, 285–297.

Note Added in Proof

We had previously compared our crystal structures to published NMR structures of Mesd (Table 4; Table S1). Now, through an exchange of accepted papers facilitated by *Structure*, we have also studied the additional NMR structures described in two other papers in this issue. We present these updated analyses in Tables S3 and S4. The core domains from both the "open" and "closed" states of the longer mouse Mesd structures, noted in the Discussion and reported here by Köhler et al. (2011) are appreciably closer in rmsd (1.51–1.84 Å versus 2.21–2.35 Å) to our X-ray structures of *mm*$Mesd_{SD}$ than was their previous shorter structure (Köhler et al., 2006), and validation parameters from our MolProbity and cavity analyses are also markedly improved. The conformation of Mesd in the new model reported here by Chen et al. (2011) is unchanged from that previously reported (Chen et al., 2010a). As this conformation is irreconcilable with that of our crystal structures, we remain unable to make a meaningful comparison. We find the validation parameters, cavity features and β sheet twist characteristics in this model for full-length Mesd to be unusual (Table S4).