

Mistic: Cellular localization, solution behavior, polymerization, and fibril formation

Hay Dvir, Matthew E. Lundberg, Samir K. Maji, Roland Riek, and Senyon Choe*

Structural Biology Laboratory, The Salk Institute for Biological Studies, San Diego, California

Received 4 February 2009; Revised 8 April 2009; Accepted 10 April 2009

DOI: 10.1002/pro.148

Published online 29 April 2009 proteinscience.org

Abstract: Mistic represents a family of unique membrane-associating proteins originally found in *Bacillus subtilis* (M110). As a fusion partner, it has been shown to assist overexpression of foreign integral membrane proteins in *E. coli*. We have expressed shorter Mistic homologs from other *Bacillus* species and surprisingly, unlike M110, found them abundant in the cytoplasm. These Mistic homologs including the corresponding shorter sequence (amino acids 27 through 110 of M110) exist as multimeric assemblies in solution in the absence of detergent. Crystals of Mistic from *B. leicheniformis* (M2) diffracted to 3.2 Å resolution, indicating that it exists as a multimer in the crystalline state as well. Moreover, we show that although M2 is mostly α -helical, it tends to polymerize and form fibrils. Such oligomerization could potentially mask the charged surface of the monomeric Mistic to assist membrane integration.

Keywords: Mistic structure; membrane proteins; fibril; domain swapping; polymerization

Introduction

Integral membrane proteins (IMPs) constitute up to 30% of all genes encoded by sequenced genomes.^{1,2} They mediate various activities such as nerve impulses, hormone action or allow pathogens to sense their environment, and elicit a virulence response, all of which makes them the largest groups of drug target proteins.³ Unfortunately, structural determination of IMPs remains an immense challenge with only ~1.5% of the protein structures in the Protein Data Bank

(PDB) are of IMPs, thus limiting our understanding of their function at the molecular level.

Most membrane proteins do not exist in abundance naturally, and therefore heterologous overexpression of IMPs is a prerequisite challenge for structural studies. Heterologous expressions of eukaryotic IMPs in bacteria has limited success because of the toxicity associated with saturating the secretory system, incompatibility in translation and traffic to the membrane, rapid degradation of the foreign protein, lack of post-translational modifications, and other misfolding problems directing the protein to inclusion bodies.

Recently, a novel membrane-associating protein discovered in *Bacillus subtilis*,⁴ Mistic, has been shown to enhance the expression levels of a large number of foreign IMPs in the inner membrane of *E. coli*, when used as a fusion partner linked to their N-termini.^{4–8} Interestingly, Mistic lacks any identified signal sequence and its overexpression, whether alone

Conflict of Interest: A patent on the use of Mistic is filed for membrane protein production in *E. coli*. Roland Riek is one of the inventors.

Grant sponsor: NIH; Grant numbers: GM74821, GM74929.

*Correspondence to: Senyon Choe, Structural Biology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037.
E-mail: choe@salk.edu

or in fusion to other IMP, lacks the toxicity issues associated with overloading the protein translation machinery and thus is likely to bypass the *E. coli* translocon complex. The solution-NMR structure of Mystic, in detergent micelle of lauryl-dimethylamine-*N*-oxide (LDAO), shows a compact four-helix-bundle fold.⁴ Structurally, Mystic differs from most of the membrane proteins as its surface area is not very hydrophobic but rather made up of many acidic residues, and none of its helices are predicted to form hydrophobic transmembrane segments. Therefore, the resident orientation of Mystic within the membrane has not been established, and the relationship of the four-helix-bundle structure to its function is still to be investigated by other techniques.

Mistic requires detergent for solubility, and it has been so far unsuccessful to crystallize. To bypass this challenge, we have studied Mistic homologs found in other *Bacillus* species.⁵ These Mistic orthologs from *B. licheniformis* (M2), *B. mojavensis* (M3), and *B. atrophaeus* (M4) all lack the N-terminal 26 amino acids of Mystic from *B. subtilis*, corresponding to the first α -helix of its solution-NMR structure. For nomenclature, we refer to the originally identified Mistic from *B. subtilis* (110 amino acids long) as M110, and to its shorter sequence⁵ starting at Met27 as M1 (Fig. 1). We have expressed these Mistic proteins, M1, M2, M3, and M4, in *E. coli* independently, and found that, unlike M110, the shorter forms are significantly more soluble in the absence of detergent. In light of the crystals of M2 we report here and our data on the formation of fibrils by Mistic proteins, we discuss the correlation between its crystalline, soluble, and aggregated states and the possible implications for membrane insertion.

Results and Discussion

Mistic orthologs express in the cytoplasm

The facts that M110 expresses primarily in the bacterial membrane,⁴ and that the shorter Mistic proteins have comparable capacity to enhance expression of IMP fused to their C-termini,⁵ stimulated us to determine whether they too express at the inner membrane of *E. coli*. Figure 2 shows that M110 is essentially insoluble and associates tightly with the membrane. Surprisingly, M1, which corresponds to the amino acid sequence 27–110 of M110, is mainly soluble with some level of association with the membrane. M2, the most distant sequence from M1, also exists mainly in the soluble fraction but shows an increased association with the membrane. Interestingly, M3 and M4, which are more closely related to the sequence of M1, interact significantly with the membrane, with M4 showing the highest existence in the pellet, comparable with that of M110.

The mutation M75A had been shown to dramatically reduce the membrane association of M110 leading to cytoplasmic protein expression and lack of enhanced IMP expression as a fusion partner.⁴ Simi-

larly, we find that the mutation corresponding to this position on M1, M49A, yields a highly soluble protein product with practically no existence at the membrane (see Fig. 2). M2 has Ile instead of Met at position 49; however, it has a Met at position 48. In an attempt to increase its solubility, we generated two M2 mutants, M48A and I49A. Our data shows that M48A, but not I49A, reduces membrane association and increased soluble expression levels of the wt M2.

The shorter Mistic proteins are soluble multimers

As seen earlier, the shorter Mistic forms, M1 through M4, are all abundant in the bacterial cytoplasm and thus do not require detergent for their purification. Moreover, expression can be readily upscaled to obtain 30–40 mg/L of pure protein from either form, which makes them potentially less challenging for crystallization than M110 that requires detergent for solubility.

All the Mistic proteins migrate as monomers on an SDS gel, however, it can be seen in Figure 2 (Soluble gel) that there are noticeable differences in their migration rates that cannot be accounted for by the differences in their mass ($\sim 10,500$ Da for all of them). Such differences in the migration rates are indicative of variation in the monomeric form, even in the presence of SDS, particularly exemplified by the mutation M48A which slows the migration of M2.

Interestingly, in the absence of detergent, all the short Mistic forms exist as multimers in solution (including M1, M2, M3, and M4), as typified by the migration of M2 on a size-exclusion chromatography (SEC) column shown in Figure 3. Static-light scattering analysis shows that there is a relatively stable oligomeric form with ~ 173 kDa, corresponding to about 16 M2 subunits (each monomer of 10,516 Da). However, higher order aggregates exist and their proportion increases daily when the proteins are stored at 4°C or -20°C . In contrast, M1 M49A exists as a discrete and stable tetramer with a measured molar mass of ~ 43 kDa (see Fig. 3). Interestingly, although the mutation M48A increases the solubility of M2, it neither changes the aggregation of M2 dramatically nor gives rise to discrete tetramerization of M2.

Crystals of M2

Data from protein crystals were collected to a resolution of 3.2 Å. Initial indexing of the intensities indicated the possibility of a primitive orthorhombic Bravais lattice (since the β angle is close to 90° , see later); however, the merging statistics were only reasonable for a monoclinic lattice. Inspection of the systematic absences clearly revealed $(0,k,0)$ with $k = 2n$ as the reflection condition, consistent with the P21 space group. The reflection data statistics for the M2 crystals are shown in Table I.

Solvent-content analysis⁹ indicates eight molecules in the asymmetric unit (ASU) for crystals

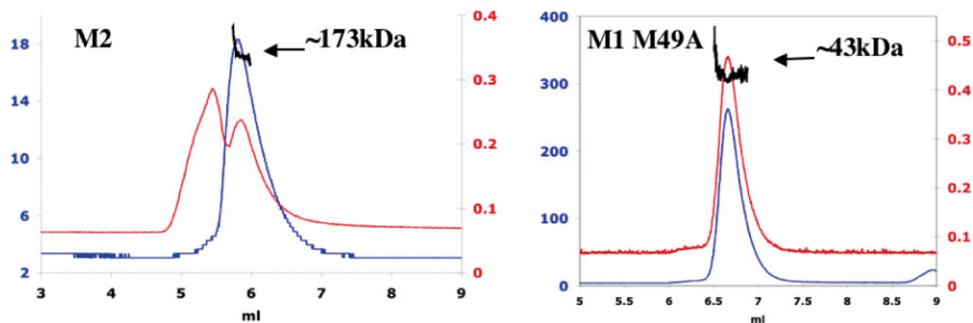


Figure 3. Oligomeric state measured by light scattering. Purified M2 (left) and M1 M49A (right) were injected on the SEC column G3000PWxl in line with a miniDAWN light scattering device. The blue trace shows the UV absorbance in mOD, the red trace shows the light scattering (Rayleigh ratio at 90°), and the black trace shows the calculated (in the ASTRA software using the Zimm model with a protein dn/dc value of 0.186) molar mass distribution around the main peaks. The chromatogram for M2 is typical for all the soluble Mystic proteins M1 through M4 with various degrees of high-order aggregation.

well. It is likely that there are variations in the structures among the eight putative molecules of M2 of ASU. Regardless, it is evident that the large number of molecules in ASU prohibits structure determination by molecular replacement. Translational element does not contribute to NCS averaging, thus provides no help in phase improvement. For the purpose of crystal structure determination, we are searching for heavy-atom derivative crystals and new crystal forms with smaller number of ASU and with better diffracting power.

Fibril formation by M2

Interestingly, M2 turns into a gel-like state once concentrated down to ~20 mg/mL at 25°C, a process that is inhibited at 4°C or on ice. M1, although more soluble than M2, can still turn into gel-like entity at around 20 mg/mL, whereas the mutant M1 M49A is highly soluble and does not form a gel up to a concentration of ~100 mg/mL. This form of polymerization is common among proteins that form amyloid fibrils. We have therefore focused on the analysis of concentrated (>20 mg/mL) samples of M2, M1, and M1 M49A by electron microscopy (EM). The EM data show that M2 forms fibrils, M1 forms short and bended protofibrils, which is a form of premature fibril, while M1 M49A does not aggregate at all (Fig. 5). Moreover, it can be seen that M2 exists primarily in the fibrillar state and no globule-like structures were observed, indicating the practical absence of low-order assemblies. In contrast, M1 M49A shows only globule-like structures typified by low-order assemblies.

Unlike amyloid fibrils, which are formed by stacking of β -sheets, the fibrils of M2 and M1 are likely composed of α -helical structure as indicated by CD [Fig. 6(A)] and by the solution-NMR structure of M110.⁴ Additional support for this notion comes from the failure of M2 fibrils to bind the traditional amyloid-specific dye, Thioflavin-T,¹² suggesting that these

fibrils are not amyloid fibrils, thus probably not made up of beta sheets. An alternate way to form highly organized aggregates has been proposed by the “runaway domain swapping” model.^{13,14} It is different from the commonly observed “close-ended” domain swapping,¹⁵ in which monomers swap small domains to form symmetric oligomers (such as dimers, trimers, tetramers, etc.). In the case of runaway domain swapping for the polymerization to continue to a higher order, it requires “open-ended” domain swapping, in which each monomer swaps a domain with the complementary domain of an adjacent monomer along the fibril allowing it to grow on both ends.

If the polymerization observed in the solution is similar to that in the multimeric crystal of M2 we herein report, then the three-dimensional packing of such a structure may shed light into the formation of runaway domain swapping. In addition, since the mutation M49A of M1 appears to inhibit fibril formation and eliminates higher-order aggregates, it is possible that the tetramer is the minimal domain-swapping unit in this case. Studying such a structure in comparison with its *wt* would be highly valuable to

Table I. X-ray Data Collection Statistics for M2 Crystals

| | |
|--|------------------------------------|
| X-ray source | ALS bl8.2.1 |
| Wavelength (Å) | 1.0781 |
| Space group | P21 |
| Unit cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å) and β (°) | 80.00, 48.23, 114.41, and 90.52 |
| Resolution range (Å) | 114.7–3.2 (3.37–3.2) |
| R_{sym}^a | 0.079 (0.410) |
| Unique reflections | 14,685 (2089) |
| Completeness (%) | 99.5 (99.2) |
| $I/\sigma(I)$ | 3.9 (1.9) |
| Multiplicity | 4.0 (3.7) |

^a $\sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$, where $I_i(\text{hkl})$ is the *i*th measurement and $\langle I(\text{hkl}) \rangle$ is the mean of all measurements of $I(\text{hkl})$ for Miller indices hkl.

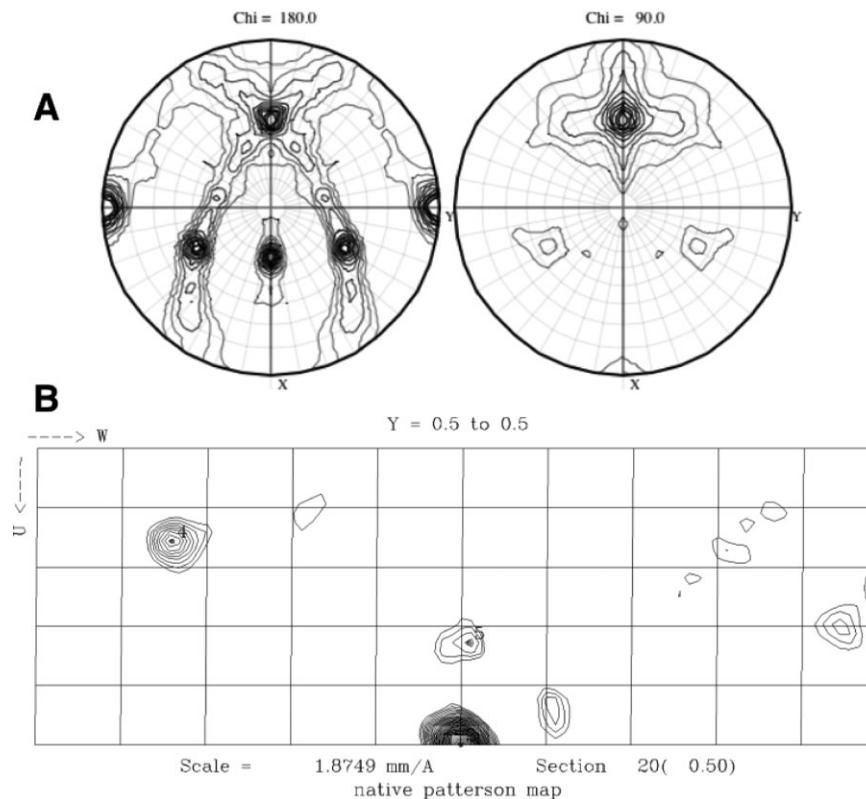


Figure 4. (A) Stereographic projection of the self-rotation function. The values of the self-rotation function (calculated from data in the resolution range 40–3.8 Å) at $\chi = 180^\circ$ (left) or $\chi = 90^\circ$ (right) are plotted with the polar angles (θ, ϕ, χ) convention looking down the z-axis. These projection have their $\tau = 0^\circ$ or 180° at the center and $\tau = 90^\circ$ at the edges, whereas the phi angle goes from 0° to 360° around the sphere starting from x in the direction of y. The $\chi = 180^\circ$ projection has its twofold crystallographic symmetry peaks along the y-axis at the edges of the sphere ($\theta = 90^\circ, \phi = 90^\circ$ or -90°). The other off-edge peaks indicates the presence of twofold NCS, whereas the $\chi = 90^\circ$ projection has a single peak that indicates the presence of fourfold NCS in the ASU; calculated using MOLREP.¹⁰ (B) Native Patterson map. The Patterson map was calculated using data from the native M2 crystal in the 40–3.8 Å resolution range excluding the reflections $F_{\text{obs}} < 3\sigma(F_{\text{obs}})$. The Harker section at $u = 0.5$ shows a strong peak (about 39% the intensity of the origin peak) at $u = 0.5, v = 0.5, w = 0.5$, indicating a pseudo-centering symmetry. The calculation was done using the Patterson routine from the CCP4.¹¹

our understanding of α -helical fibril formation. We hope that the data we present here will facilitate such analysis in the near future.

Finally, both the highly charged surface of the solution-NMR structure of Mystic and the high solubility of the short Mystic orthologs make it difficult to imagine how it inserts into the hydrophobic lipid-bilayer membrane. Polymerization, as observed for all the

soluble Mystic proteins, may help to mask the charged surface of the monomers and thereby assist membrane insertion. Intriguingly, it has been shown that fibrils tend to associate with lipid membranes,^{16–19} therefore structural insight into the polymerization of Mystic may provide insight into the unknown biophysical process of membrane insertion independent of the translocon machinery.

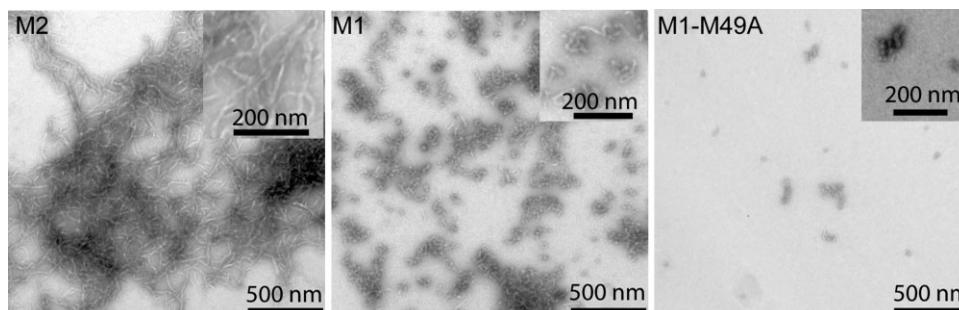


Figure 5. EM micrographs showing fibril structures of M2.

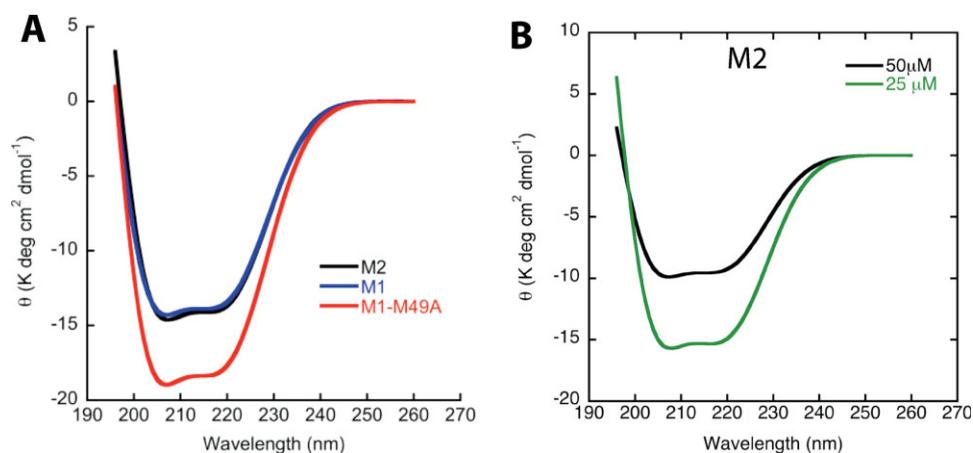


Figure 6. CD spectra of Mystic. In Panel A, the CD spectra of M2, M1, and M1M49A shows helical structure in solution. The negative ellipticity (θ) of M1M49A at 222 and 208 nm is greater than that of M2 and M1, suggesting more helical content of this protein in solution. In Panel B, the CD spectra of M2 shows decreased helicity at higher protein concentration; however, the presence of helical secondary structure is still evident.

Materials and Methods

Cloning and purification

All Mystic homologs were cloned into the plasmid pET21b (Novagen, Madison, WI) under the control of a T7 promoter, with a C-terminal thrombin cleavage site followed by a hexa-His tag. Freshly transformed BL21(DE3) *E. coli* cells (Invitrogen, Carlsbad, CA) were grown overnight in Terrific Broth (TB) supplemented with 200 $\mu\text{g}/\text{mL}$ ampicillin. The cultures (20 mL) were inoculated with 1 mL of the aforementioned starter cultures, and then the growth proceeded to an approximate optical density of 1.0. Protein production was induced with 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) for 5 h at 37°C. The cultures were harvested by centrifugation at 3000g for 15 min, and then the cell pellets were resuspended in 10 mL of lysis buffer (25 mM Tris pH 8.0, 20 mM imidazole, 200 mM NaCl, and 1 mM phenylmethylsulfonyl fluoride [PMSF]). Cells were lysed with a microfluidizer (Microfluidics, Newton, MA), and the membranes and inclusion bodies were spun down by centrifugation at 100,000g for 2 h. The supernatant was incubated with 200 μL buffer-equilibrated NiNTA beads (Qiagen, Hilden, Germany) for 2 h at 4°C. The mixture was columned and washed five times by gravity with 1 column volume of wash buffer (lysis buffer without PMSF). Bound protein was eluted with 1 mL elution buffer (wash buffer with 0.5M imidazole). Protein extraction from the membranes was done by resuspending the pellets in 1 mL lysis buffer containing 0.5% (17.5 mM) SDS and shaking for 2 h at 25°C. The SDS-insoluble fraction (inclusion bodies) was spun down by centrifugation at 20,000g for 15 min.

Large-scale expressions of soluble Mystic proteins were carried out similarly from cultures of 1-L TB, using the soluble fraction only for purification. The His-tags were cleaved off from the proteins with

bovine thrombin (Sigma-Aldrich, St. Louis, MO) digestion during dialysis against 50 mM NaCl for O/N at 4°C. The protein samples were analyzed for cleavage with NuPAGE 4–12% Bis-Tris gel with MES running buffer (Invitrogen, Carlsbad, CA). The completely cleaved proteins were concentrated using 50K MWCO Vivaspin concentrators (Vivascience AG, Hannover, Germany) and further purified by SEC with Superdex-200 (16/60) column (Amersham Pharmacia Biotech, Piscataway, NJ) pre-equilibrated with 10 mM Tris pH 8.0, 100 mM NaCl.

Light scattering

Static light-scattering measurements coupled with SEC were performed using Wyatt miniDAWN (Wyatt Technology, Santa Barbara, CA) and G3000PWXL column (Tosoh Bioscience, Montgomeryville, PA). Molar masses were derived by the Zimm fit method in ASTRA V (Wyatt Technology, Santa Barbara, CA) using concentrations obtained from UV at 280 nm and protein-specific refractive index increment (dn/dc) of 0.186.

Crystallization

Before crystallization, M2 was concentrated to 10–15 mg/mL in the SEC buffer. The crystals of M2 grow (up to ~ 0.35 mm in length) via both the hanging and the sitting-drop vapor-diffusion methods, by mixing equal volumes of protein stock solution and the reservoir (38% MPD, 0.1M Tris pH 8.8 and 0.2M MgAc). The protein crystals grow in 2–4 days at 15°C. Before X-ray data collection, protein crystals were transferred to a fresh mother liquor solution containing 40% of MPD.

X-ray data analysis

Diffraction data from the native protein crystal were collected on an ADSC Q210 CCD detector at 100 K on

beamline 8.2.1 (ALS, Berkeley California). All X-ray data were indexed and integrated using the program MOSFLM v.6.2.4²⁰ and merged and scaled with SCALA²¹ from the CCP4 program package.¹¹

Electron microscopy

Protein samples were concentrated to at least 20 mg/mL and kept at 4°C. The concentrated samples were diluted in Buffer A (10 mM Tris, 1 mM EDTA, 25 mM dithiothreitol, 7.5) to a final concentration of 50 μM and spotted on a glow-discharged, carbon-coated Formvar grid (Electron Microscopy Sciences), incubated for 5 min, washed with distilled water, and then stained with 1% (w/v) aqueous uranyl formate solution. Uranyl formate solutions were filtered through 0.2-μm sterile syringe filters (Corning, Corning, NY) before use. EM analysis was performed using a JEOL JEM-100CXII electron microscope at 80 kV with nominal magnifications between 36,000 and 72,000. The images were recorded digitally by using the SIS Megaview III imaging system. At least two independent experiments were carried out for each sample.

Thioflavin T binding

The concentrated samples were diluted with Buffer A to a final concentration of 50 μM, and then the solution was mixed with 2 μL of 1 mM Thio T prepared in the same buffer. Fluorescence was measured immediately after the addition of Thio T. The experiment was measured on a spectrofluorimeter (Photon Technology International, Lawrenceville, NJ) with excitation at 450 nm and emission at 482 nm. A rectangular 10-mm quartz microcuvette was used. Three independent experiments were performed for each sample.

Circular dichroism spectroscopy

Protein samples were concentrated to at least 20 mg/mL and kept at 4°C. The concentrated samples were diluted into Buffer A to a final concentration of 25 μM. The solution was then placed into a 0.1-cm path-length quartz cell (Hellma, Forest Hills, NY). Spectra were acquired using an instrument of BioLogic MOS-450 (Molecular Kinetics, Pullman, WA). All measurements were done at 23°C. Spectra were generally recorded over the wavelength range of 196–260 nm. Three independent experiments were performed with each sample. Raw data were manipulated by smoothing and subtraction of buffer spectra, according to the manufacturer's instructions.

References

1. Wallin E, von Heijne G (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci* 7:1029–1038.

2. Boyd D, Schierle C, Beckwith J (1998) How many membrane proteins are there? *Protein Sci* 7:201–205.
3. Speers AE, Wu CC (2007) Proteomics of integral membrane proteins—theory and application. *Chem Rev* 107:3687–3714.
4. Roosild TP, Greenwald J, Vega M, Castronovo S, Riek R, Choe S (2005) NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. *Science* 307:1317–1321.
5. Roosild TP, Vega M, Castronovo S, Choe S (2006) Characterization of the family of Mistic homologues. *BMC Struct Biol* 6:10.
6. Kefala G, Kwiatkowski W, Esquivies L, Maslennikov I, Choe S (2007) Application of Mistic to improving the expression and membrane integration of histidine kinase receptors from *Escherichia coli*. *J Struct Funct Genomics* 8:167–172.
7. Liu X, Wu L, Deng G, Li N, Chu X, Guo F, Li D (2008) Characterization of mitochondrial trifunctional protein and its inactivation study for medicine development. *Biochim Biophys Acta* 1784:1742–1749.
8. Lee KE, Kim HM, Lee JO, Jeon H, Han SS (2008) Regulation of CD40 reconstitution into a liposome using different ratios of solubilized LDAO to lipids. *Colloids Surf B: Biointerfaces* 62:51–57.
9. Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* 33:491–497.
10. Vagin A, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *J Appl Cryst* 30:1022–1025.
11. CCP4 (1994) Collaborative Computational Project, Number 4. *Acta Cryst D Biol Cryst* 50:760–763.
12. Naiki H, Higuchi K, Hosokawa M, Takeda T (1989) Fluorometric determination of amyloid fibrils in vitro using the fluorescent dye, thioflavin T1. *Anal Biochem* 177:244–249.
13. Bennett MJ, Schlunegger MP, Eisenberg D (1995) 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci* 4:2455–2468.
14. Guo Z, Eisenberg D (2006) Runaway domain swapping in amyloid-like fibrils of T7 endonuclease I. *Proc Natl Acad Sci USA* 103:8042–8047.
15. Bennett MJ, Choe S, Eisenberg D (1994) Refined structure of dimeric diphtheria toxin at 2.0 Å resolution. *Protein Sci* 3:1444–1463.
16. McLean LR, Balasubramaniam A (1992) Promotion of beta-structure by interaction of diabetes associated polypeptide (amylin) with phosphatidylcholine. *Biochim Biophys Acta* 1122:317–320.
17. Mirzabekov TA, Lin MC, Kagan BL (1996) Pore formation by the cytotoxic islet amyloid peptide amylin. *J Biol Chem* 271:1988–1992.
18. Anderlueh G, Gutierrez-Aguirre I, Rabzelj S, Ceru S, Kopitar-Jerala N, Macek P, Turk V, Zerovnik E (2005) Interaction of human stefin B in the prefibrillar oligomeric form with membranes. Correlation with cellular toxicity. *FEBS J* 272:3042–3051.
19. Kagan BL, Hirakura Y, Azimov R, Azimova R, Lin MC (2002) The channel hypothesis of Alzheimer's disease: current status. *Peptides* 23:1311–1315.
20. Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, No. 26.
21. Evans PR (1997) Scala. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* 33:22–24.