



NMR Structure of Mystic, a Membrane-Integrating Protein for Membrane Protein Expression

Tarmo P. Roosild, *et al.*

Science **307**, 1317 (2005);

DOI: 10.1126/science.1106392

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 23, 2009):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/307/5713/1317>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/307/5713/1317/DC1>

This article **cites 22 articles**, 7 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/307/5713/1317#otherarticles>

This article has been **cited by** 50 article(s) on the ISI Web of Science.

This article has been **cited by** 3 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/307/5713/1317#otherarticles>

This article appears in the following **subject collections**:

Biochemistry

<http://www.sciencemag.org/cgi/collection/biochem>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

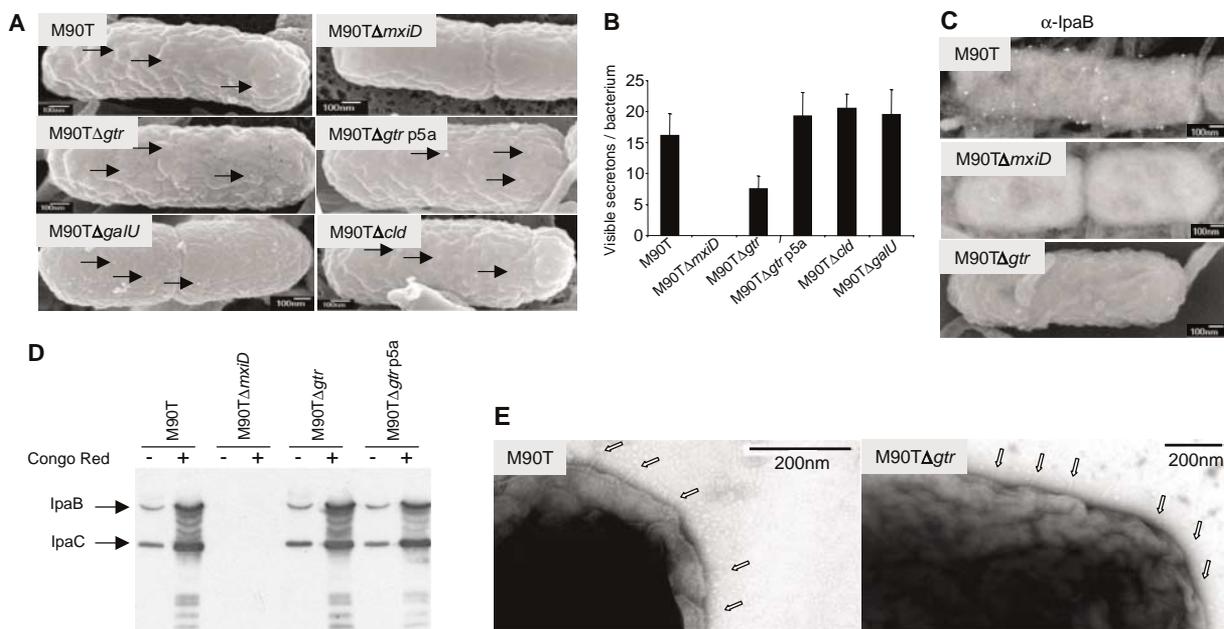
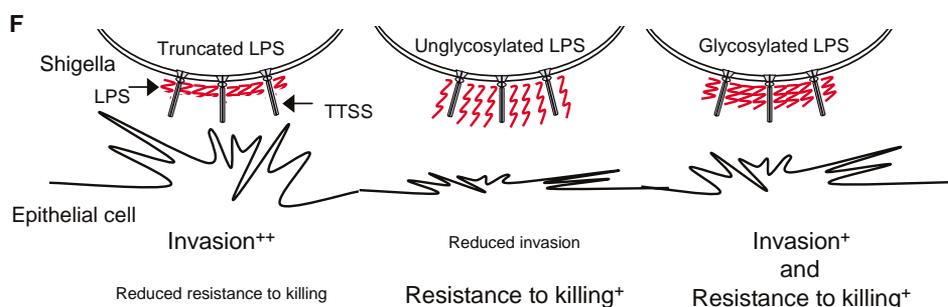


Fig. 4. (A) Scanning electron microscopy of bacteria demonstrating the abundance of TTSS secretions (arrowed) on the cell surface; scale bars and strains are shown. (B) Number of visible secretions per bacterium (error bars show SD). (C) Immunogold labeling of IpaB at the bacterial surface, and (D) secretion of IpaB and IpaC into culture supernatants detected by Coomassie are not affected by LPS glucosylation; secretion was induced by the addition of Congo Red (+). (E) Negative-staining electron microscopy showed that the number of TTSS emerging from the bacterial outer membrane was not altered in the strain lacking the *gtr* operon. (F) The interaction between LPS and TTSS. Hyperinvasive strains that have truncated LPS molecules are susceptible to being killed by the innate immune response in vivo. Bacteria



expressing unglucosylated LPS are compromised for invasion and therefore attenuated. Glucosylation of the O antigen halves the length of the LPS molecule, which allows efficient function of the TTSS while it retains resistance to antimicrobial factors in vivo.

NMR Structure of Mystic, a Membrane-Integrating Protein for Membrane Protein Expression

Tarmo P. Roosild, Jason Greenwald, Mark Vega, Samantha Castronovo, Roland Riek,* Senyon Choe*

Although structure determination of soluble proteins has become routine, our understanding of membrane proteins has been limited by experimental bottlenecks in obtaining both sufficient yields of protein and ordered crystals. Mystic is an unusual *Bacillus subtilis* integral membrane protein that folds autonomously into the membrane, bypassing the cellular translocon machinery. Using paramagnetic probes, we determined by nuclear magnetic resonance (NMR) spectroscopy that the protein forms a helical bundle with a surprisingly polar lipid-facing surface. Additional experiments suggest that Mystic can be used for high-level production of other membrane proteins in their native conformations, including many eukaryotic proteins that have previously been intractable to bacterial expression.

Integral membrane (IM) proteins, constituting nearly 30% of eukaryotic genomes, play central roles in cellular transport processes, inter-

cellular signaling, and growth regulation. However, of the more than 28,000 high-resolution protein structures known, only some

25 unique families of IM proteins are represented. This disparity is accounted for by two bottlenecks in membrane protein structural analysis: high-yield protein production and crystallization. Recombinant expression of IM proteins in *Escherichia coli*, the primary protein source for biophysical studies, has met with limited success (1). Two complications likely account for this difficulty. First, IM proteins must be trafficked to the membrane, requiring targeting signals that may not be recognized by the bacterial host. Second, high-level expression of membrane proteins that can use *E. coli* translocon machinery will competitively exclude production of other vital host membrane proteins, leading to toxicity. Most successful attempts at expression of IM proteins in bacteria have used low-copy-number plasmids with weak promoters to produce low levels of protein, compensated by large culture volumes (2).

Structural Biology Laboratory, Salk Institute, San Diego, CA 92037, USA.

*To whom correspondence should be addressed. E-mail: choe@salk.edu, riek@salk.edu

Alternatively, one can target IM proteins to inclusion bodies (3), but this requires subsequent renaturation of the desired protein from these insoluble deposits, a process with limited success rates.

The established procedure of using fusion partner proteins to aid production of recombinant proteins has also had limited utility in the production of eukaryotic IM proteins (4), because the fusion proteins currently available do not target the construct to the membrane or facilitate membrane insertion. An ideal fusion partner for IM protein production would autonomously traffic its cargo to the membrane, bypassing the translocon and associated toxicity issues while retaining the characteristics of other successful fusion partner proteins, including relatively small size, *in vivo* folding, and high stability. Several proteins (particularly bacterial toxins) and some synthetic peptides (5) have many of these characteristics, which suggests that an ideal fusion partner specialized for recombinant IM protein production in *E. coli* is likely to exist.

Crystallization is another obstacle in the determination of IM protein structures, because such proteins must be solubilized in detergent micelles that are inherently resistant to forming ordered crystal lattices. Nuclear magnetic resonance (NMR) spectroscopy offers an alternative method for determining atomic resolution structures of proteins (6, 7). To date, however, protocols for NMR structure determination of IM proteins have been established only for very small, structurally simplistic IM proteins (8, 9) and for outer membrane bacterial porins (10–12), whose β -barrel fold allows collection of ample interstrand long-range backbone-backbone nuclear Overhauser effects (NOEs) that are sufficient to determine the fold of the protein. The development of new techniques to specifically address the inherent characteristics of α -helical IM proteins is necessary to bring the powerful tools of NMR to bear on this class of molecules.

We have isolated a 110-amino acid (13 kD) *B. subtilis* protein called Mistic (an acronym for “membrane-integrating sequence for translation of IM protein constructs”). Mistic associates tightly with the bacterial membrane when expressed recombinantly in *E. coli* (Fig. 1A) (13). Surprisingly, however, Mistic is highly hydrophilic, lacking a recognizable signal sequence. Detergent-solubilized Mistic binds tightly to micelles and aggregates rapidly when stripped of surfactant. Mistic solubilized in lauryl dimethylamine oxide (LDAO) was found to be monomeric by static light scattering analysis in combination with detection of ultraviolet absorption (fig. S1), forming a protein-detergent complex (PDC) of ~25 kD containing ~50 molecules of LDAO (relative molecular mass = 229.4) per molecule of Mistic.

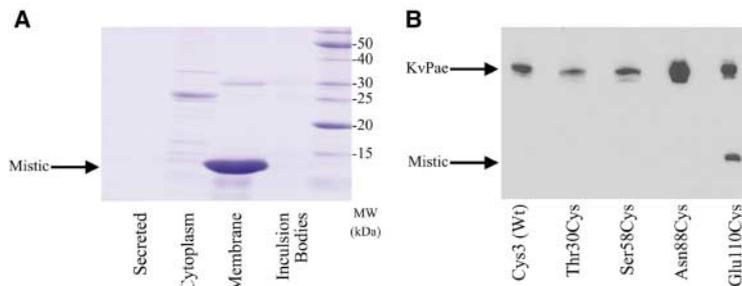


Fig. 1. Mistic characterization. (A) SDS–polyacrylamide gel electrophoresis (PAGE) results for Ni–nitrilotriacetic acid (NTA) elutions from fractionation of a culture recombinantly expressing octahistidine-tagged Mistic. Mistic is found abundantly only in the bacterial membrane. (B) Topology analysis of Mistic as assessed by biotinylation of monocysteine variations of Mistic by the membrane-impermeable, thiol-reactive probe MPB. Only Glu¹¹⁰ at the C terminus is well exposed periplasmically. Cys³ at the N terminus of the protein and the centrally located Ser⁵⁸, both also putatively on the extracellular side of the membrane, are nonreactive with MPB in right-side-out (RSO) membrane vesicle preparations, consistent with these side chains being embedded in the membrane. In support of this hypothesis, Cys³ mutation to Ser is functionally disruptive, whereas mutation to hydrophobic Val, Leu, or Ile is well tolerated. Mistic constructs were expressed as a fusion to a bacterial potassium channel (KvPae) and subsequently separated by cleavage with thrombin. The channel, identical in all constructs, serves as an internal control for calibrating expression, extraction, biotinylation, and detection efficiency among the samples.

The *in vivo* topology of this protein in *E. coli* was analyzed by evaluating the accessibility of an array of monocysteine mutants to the membrane-impermeable thiol biotinylating reagent 3-(*N*-maleimido-propinyl) biocytin (MPB) (14). In addition to the single naturally occurring cysteine (residue 3), cysteine mutations were introduced individually at the C terminus (residue 110) and in predicted loop regions at positions 30, 58, and 88 (Fig. 2A), with the naturally occurring cysteine mutated to valine. This experiment revealed a well-exposed periplasmic C terminus (Fig. 1B) (fig. S2). The lack of reactivity of the other locations indicates that they are either intracellular or membrane-embedded in Mistic’s native conformation.

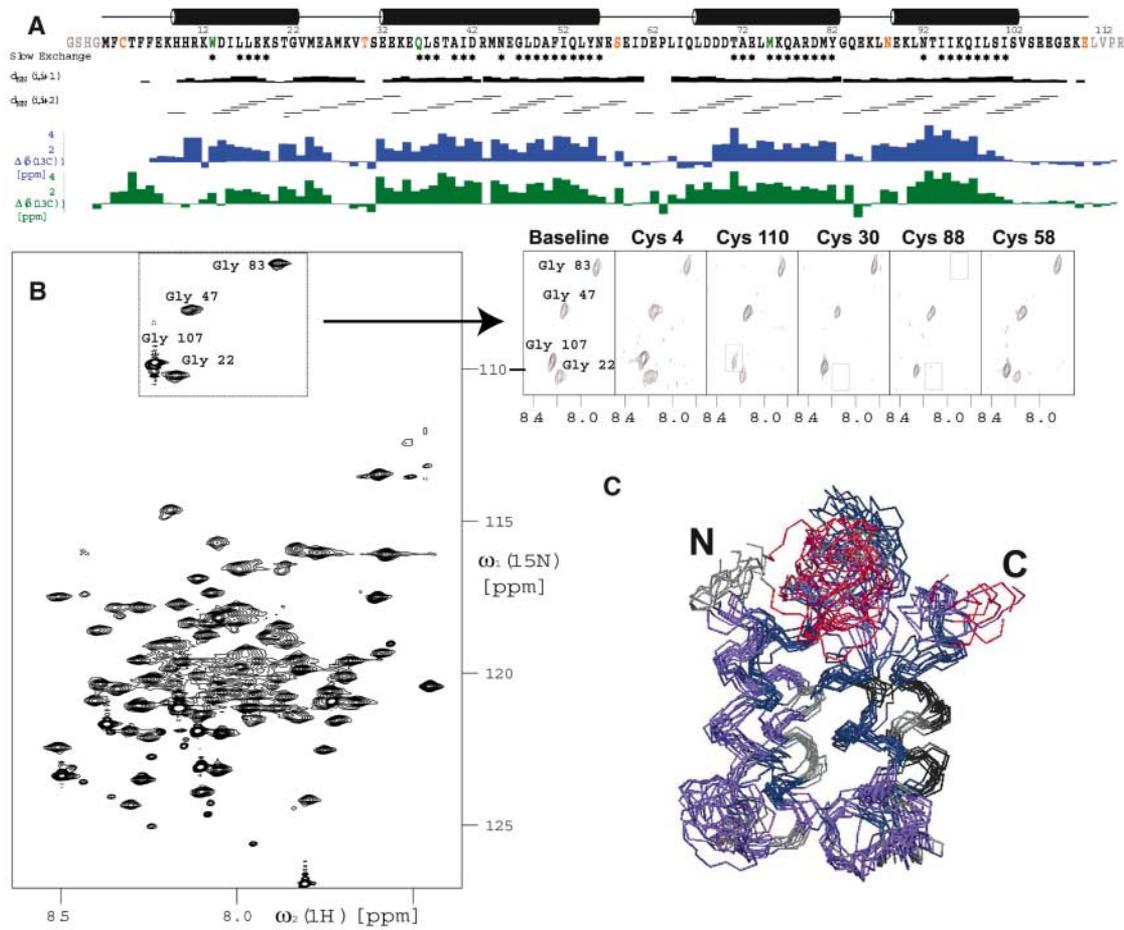
NMR *de novo* structure determination began with sequential backbone assignment, including the use of transverse relaxation optimized spectroscopy (TROSY)–HNCA (15, 16), TROSY–HNCA^{coiled}CO (17), and TROSY-based ¹⁵N-resolved [¹H,¹H]–nuclear Overhauser effect spectroscopy (NOESY) (mixing time 200 ms) of a ²H, ¹⁵N, and ¹³C-labeled sample (fig. S3). The ¹³C α chemical shift deviation from “random coil” values, the observed NOE pattern, and slow ¹H_N exchange with solvent strongly indicate the presence of four helices comprising residues 8 to 22, 32 to 55, 67 to 81, and 89 to 102 (Fig. 2A). Although intrasite, sequential, and medium-range NOEs and angle restraints enabled the assignment of secondary structure, without long-range restraints the fold of the protein could not be determined. We thus used the monocysteine mutant library described in the topology assay (see above) to incorporate site-directed spin labels within Mistic that produce distance-dependent line-broadening perturbations in the NMR spectra (18) that could be translated into distances for structure determination (19). [¹⁵N,¹H]–TROSY

experiments were measured on Mistic samples modified with the thiol-reactive nitroxide label (1-oxyl-2,2,5,5-tetramethyl- Δ^3 -pyrroline-3-methyl) methanethiosulfonate (MTSL) (Fig. 2B). The signal changes observed for the five spin-labeled samples were transformed into 197 long-range upper-distance and 290 lower-distance restraints (fig. S4).

Initial structure calculation was performed with CYANA (20) using the collected NOE data, chemical shift–derived angle restraints, and restraints derived from spin labeling. In addition, α -helical hydrogen bond restraints were implemented for residues that show all of the three following properties: slow HN exchange, a helical ¹³C chemical shift, and helical backbone NOEs (Fig. 2A). In an iterative process, the derived scaffold was used to collect long-range and medium-range NOEs and to refine calibration of the spin-label restraints. In the end, 29 long-range NOEs between methyl or aromatic protons and amide protons were identified. Because these distances are intrinsically large in a helical bundle and concomitantly result in weak NOEs, the use of a cryoprobe and long mixing times of 200 ms were essential.

The final structure calculation was performed with 573 NOE distance restraints, 346 angle restraints from chemical shifts and NOEs, and 478 distance restraints from the spin-label experiments (table S1). A total of 100 conformers were initially generated by CYANA; in Fig. 2C, the bundle of 10 conformers with the lowest target function is used to represent the three-dimensional NMR structure. The resulting structure is a four-helix bundle (Fig. 3A). Although all helices except α 2 are slightly shorter (~14 amino acids) than expected for a bilayer-traversing helix, this is likely due to partial unraveling of the ends of the helices in the detergent micelle environment, especially at the N and C termini (α 1 and α 4). Helix α 2 has a kink,

Fig. 2. Secondary structure and long-range interactions of Mystic. (A) Primary sequence of Mystic displaying location of monocysteine probing residues (orange), structural disruption mutants (green), and cloning artifact residues (gray) with secondary structural boundaries above the sequence. (First line) $^1\text{H}_\text{N}$ protection from solvent exchange indicative for hydrogen bond formation (stars). The solvent protection is determined by the absence of a cross-peak between the chemical shifts of $^1\text{H}_\text{N}$ and water in the ^{15}N -resolved TROSY- ^{15}N -NOESY spectrum. (Second and third lines) NOEs observed in the ^{15}N -resolved TROSY- ^{15}N -NOESY. Thin, medium, and thick bars represent weak (4.5 to 5.5 Å), medium (3 to 4.5 Å), and strong (< 3 Å) sequential NOEs [$d_{\text{NN}}(i, i + 1)$]. The medium-range NOEs [$d_{\text{NN}}(i, i + 2)$] are shown by lines starting and ending at the positions of the residues related by the NOE. (Fourth and fifth lines) Deviation of the $^{13}\text{C}^\alpha$



chemical shifts from corresponding "random coil" chemical shifts in 0 mM K^+ (blue) and 100 mM K^+ (green), as independently assigned. Values larger than 1.5 ppm are indicative of an α -helical secondary structure; values smaller than -1.5 ppm are indicative of β -sheet secondary structure. (B) The 2D ^{15}N , ^1H -TROSY spectrum of Mystic is shown along with parts of the 2D ^{15}N , ^1H -TROSY spectra in the presence of paramagnetic spin labels at positions Cys³, Thr³⁰Cys, Ser⁵⁸Cys, Asn⁸⁸Cys, and Glu¹¹⁰Cys. Comparison of peaks' heights between perturbed spectra and multiple reference spectra was used to obtain long-range distance restraints. (C) Superposition of 10 conformers

representing the final NMR structure. The bundle is obtained by superimposing the backbone C^α carbons of residues 13 to 62 and 67 to 102. The bundle is colored by $^{15}\text{N}\{^1\text{H}\}$ NOE data by the following color code: black, 1 to 0.8; navy, 0.8 to 0.6; blue, 0.6 to 0.4; red, 0.4 to 0.2. $^{15}\text{N}\{^1\text{H}\}$ NOE as well as T1 (^{15}N) and T2 (^{15}N) relaxation data indicate that the dynamics of the structure is generally reflected in the variance of the conformers. In particular, the loop connecting α_2 and α_3 , as well as the C terminus of Mystic, are more mobile. The T1/T2 ratio of ^{15}N was used to estimate the effective global rotational correlation time at 11 ns. This value corresponds to a spherical molecule of ~ 22 kD.

centrally positioned and putatively within the membrane. Most surprising, Mystic retains an unexpectedly hydrophilic surface for an IM protein even though it is assembled internally with a typical hydrophobic core (Fig. 3, B and C).

Given the membrane-traversing topology demonstrated by the MPB labeling experiment (Fig. 1B), this unusual surface property is very intriguing. To confirm the orientation of Mystic with respect to the membrane, we measured and assigned NOEs between Mystic and its solubilizing LDAO detergent micelle. When sites with NOE signals are mapped to the surface of the Mystic structure, a concentric ring of detergent interactions around the helical bundle is observed, as expected for a membrane-integrated protein (Fig. 4, A to C). Additionally, we perturbed Mystic spectra with paramagnetic probes that selectively partition to hydrophilic or hydrophobic environments (fig. S5) (21). The results from this study are

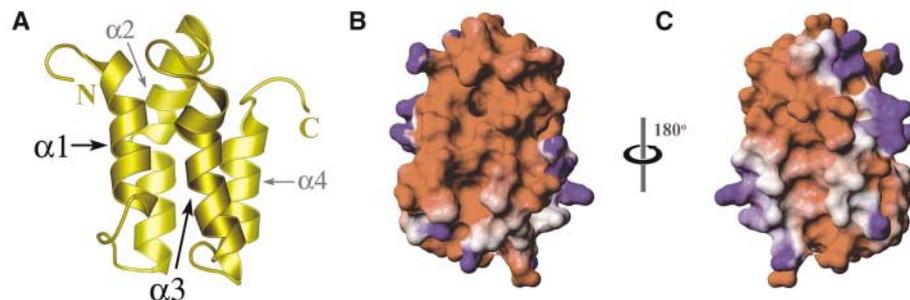


Fig. 3. Mystic structure. (A) Ribbon diagram of the lowest energy conformer highlighting the four α -helix bundle. (B) Surface representation of Mystic, oriented as in (A), mapping electrostatic potential. Color code is blue for positive charges, red for negative charges, and white for neutral surface. (C) Electrostatic potential of Mystic, viewed from the opposite face from that shown in (B).

also consistent with Mystic being embedded within the LDAO micelle.

We hypothesized that Mystic might be exploited to target another protein to the bacterial membrane, when fused to Mystic's C ter-

minus, such that it too could readily fold into its native, lipid bilayer-inserted conformation. We tested the Mystic-assisted expression of three topologically and structurally distinct classes of eukaryotic IM proteins: voltage-gated K^+ chan-

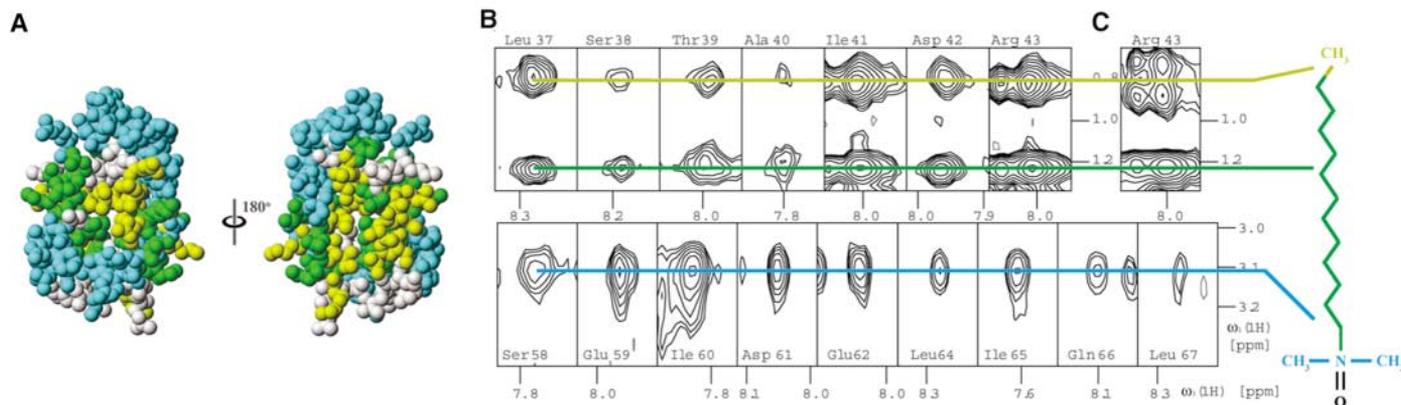


Fig. 4. Mystic-detergent interactions. (A) Surface representation of Mystic indicating observed NOE interactions between detergent molecules and the protein. Observed interactions are coded blue between the head methyl (CH₃) groups of LDAO and backbone amides (¹H_N) of the protein, yellow between the hydrophobic CH₃ end of LDAO and ¹H_N, and green between the LDAO chain (CH₂) and ¹H_N. NOEs were never observed from the same residue to both the head-group methyl and the aliphatic chain tail methyl of LDAO for the same residue. (B) A selection of intermolecular NOEs between LDAO and residues 37 to 43 and 58 to 67 of Mystic; [¹⁵N,¹H]

strips from the ¹⁵N-resolved TROSY [¹H,¹H]-NOESY are shown. The detergent-protein NOEs are marked by a bar colored as in (A), pointing to the appropriate portion of the chemical structure of LDAO. (C) For the differentiation between intramolecular and intermolecular NOEs, a second NOESY experiment was measured without decoupling on ¹³C during ¹H evolution, yielding doublets for protein-protein NOEs but single peaks for detergent-protein NOEs. Arg⁴³ for this measurement is shown in comparison with Arg⁴³ in (B), showing the presence of a protein-protein NOE at 0.8 ppm and the presence of a detergent-protein NOE at 1.2 ppm.

nels, receptor serine kinases of the transforming growth factor-β (TGF-β) superfamily, and G protein-coupled receptors (GPCRs) (Fig. 5A). Although expression success varied according to induction conditions, proteolytic susceptibility of the target gene, and the length of the amino acid linker from Mystic to the fusion protein, in most cases (15 of 22 tested constructs, table S2) the desired product could be isolated from the membrane fraction of recombinant bacteria at yields exceeding 1 mg per liter of culture (Fig. 5B). The *Aplysia* potassium channel, aKv1.1, was extracted and purified in LDAO to verify that the expressed proteins resemble their native conformations; size exclusion chromatography (Fig. 5C) showed that it retains a tetrameric assembly. Additionally, several TGF-β receptors were found to retain native ligand-binding affinity and specificity (22). Taken in combination with the fact that all of these proteins partition to the membrane fraction of cell extracts, we conclude that there exists a high propensity for this system to produce IM proteins fully folded in their native conformations.

To validate Mystic's direct role in assisting in the production of these recombinant IM proteins, we introduced mutations at three potentially structurally disruptive sites within the core of the protein (Figs. 2A and 6A). Expression tests of these Mystic variants, alone and fused to aKv1.1, indicate that the integrity of Mystic's structure is essential to its ability to chaperone cargo proteins to the bacterial lipid bilayer (Fig. 6B). The single mutation of a core methionine (Met⁷⁵) to alanine, in particular, sufficiently destabilized Mystic's structure such that it partitioned between the membrane and the cytoplasm. This same mutant

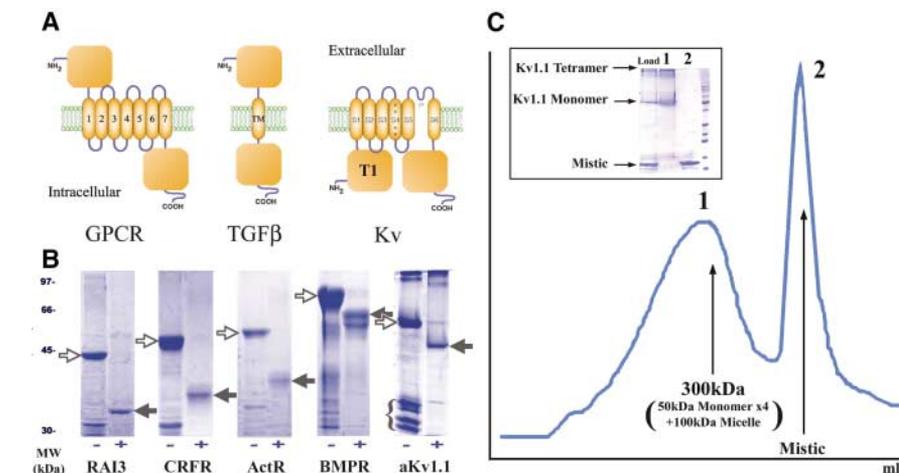


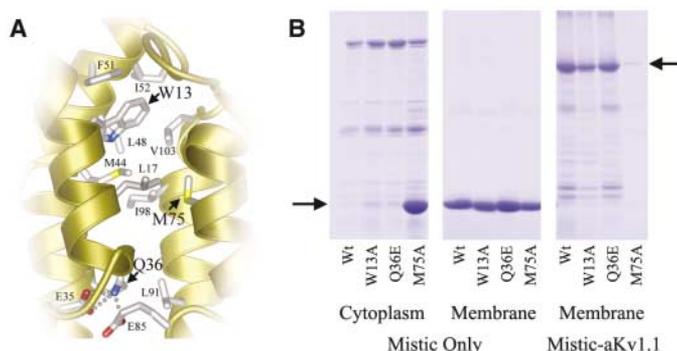
Fig. 5. Mystic-assisted eukaryotic IM protein expression. (A) Topological depictions of the three protein classes studied in this report: GPCRs, TGF-β family receptors, and voltage-gated K⁺ channels (Kv). (B) SDS-PAGE results for various eukaryotic IM proteins. Lane pairs reveal expression of the desired protein from LDAO-solubilized membrane fractions after purification by Ni-NTA affinity chromatography. The Mystic-fused protein is shown on the left (open arrow); the final product after removal of Mystic by thrombin digestion is on the right (solid arrow). Protein identities were verified for select samples [including retinoic acid-induced protein 3 (RAI3), bone morphogenetic protein receptor type II (BMPR II), and aKv1.1] by N-terminal Edman degradation sequencing of at least 14 residues of the target protein after separation from Mystic. The additional bands in the sample of aKv1.1 before digestion (bracket) were determined to be truncated products containing fragments of the N-terminal T1 domain of this channel. The region between T1 and the membrane-spanning domains of this channel is known to be flexible and proteolytically susceptible. (C) Gel filtration profile of thrombin-digested aKv1.1 run in 3 mM LDAO on a Superose-6 column. aKv1.1 elutes as a detergent solubilized tetramer subsequent to Mystic removal. (Inset) Baseline separation between aKv1.1 (lane 1) and Mystic (lane 2) allows two-step purification of aKv1.1 to near-homogeneity.

yielded no protein expression when fused to aKv1.1, confirming that Mystic's structure and resulting membrane affinity are critical for its ability to facilitate the production of target IM proteins.

Given the highly acidic surface of Mystic (Fig. 3, B and C), it is still conceivable that the conformation of Mystic in the cell membrane

differs from the structure observed in the Mystic-detergent complex. Recently, charged transmembrane helices have been shown to play dynamic roles within the lipid bilayer in ion channels and transporters (23, 24). Conformational flexibility, such as rotation of the four helices about their helical axes or even partial unraveling of the helical bundle, may allow Mystic to adapt

Fig. 6. Mutational disruption of Mystic's structure and function. (A) Residues forming the core of Mystic, with those mutated in structural disruption studies highlighted with arrows. **(B)** Mystic mutated singly at three core residues displays varying structural stability and functionality. Mutation of Trp¹³ to Ala (W13A) reduces the overall yield of fused aKv1.1 by a factor of 2 to 3. More important, mutation of Met⁷⁵ to Ala (M75A) destabilizes the structure of Mystic sufficiently such that, when expressed by itself, it partitions substantially into the cytoplasm (fourth lane from left), in stark contrast to wild-type Mystic or any of the other mutants analyzed. This results in a functionally disabled protein; thus, when M75A is fused to aKv1.1, there is no detectable yield of this protein (rightmost lane).



to the lipid environment in a fashion analogous to the mechanisms of membrane integration for the chloride channel CLIC1 (25) or diphtheria toxin (26), both of which exist alternately in soluble and membrane-integrated forms. Molecular interplay between lipid composition and membrane insertion of IM protein structures is another intriguing possibility (27). Although complete understanding of the integration dynamics of Mystic requires further study, all available data suggest that it must autonomously associate with the bacterial membrane and that this property alone accounts for its high efficiency in chaperoning the production and integration of downstream cargo proteins (fig. S6). Taken together with the NMR techniques and protocols developed and used for Mystic structure determination, Mystic's unique ability to assist in the production of IM proteins opens new avenues around traditional obstacles in the study of IM proteins, particularly those of eukaryotic origin.

References and Notes

1. C. G. Tate, *FEBS Lett.* **504**, 94 (2001).
2. R. Laage, D. Langosch, *Traffic* **2**, 99 (2001).
3. H. Kiefer, R. Vogel, K. Maier, *Receptors Channels* **7**, 109 (2000).
4. J. Tucker, R. Grishammer, *Biochem. J.* **317**, 891 (1996).
5. W. C. Wimley, S. H. White, *Biochemistry* **39**, 4432 (2000).
6. K. Wuthrich, *Nature Struct. Biol.* **5**, 492 (1998).
7. G. M. Clore, A. M. Gronenborn, *Nature Struct. Biol.* **4**, 849 (1997).
8. V. K. Rastogi, M. E. Girvin, *Nature* **402**, 263 (1999).
9. K. R. MacKenzie, J. H. Prestegard, D. M. Engelman, *Science* **276**, 131 (1997).
10. C. Fernandez, C. Hilty, G. Wider, P. Guntert, K. Wuthrich, *J. Mol. Biol.* **336**, 1211 (2004).
11. P. M. Hwang et al., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13560 (2002).
12. A. Arora, F. Abildgaard, J. H. Bushweller, L. K. Tamm, *Nature Struct. Biol.* **8**, 334 (2001).
13. See supporting data on Science Online.
14. V. Ramamurthy, D. Oliver, *J. Biol. Chem.* **272**, 23239 (1997).
15. K. Pervushin, R. Riek, G. Wider, K. Wuthrich, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12366 (1997).
16. M. Salzmann, G. Wider, K. Pervushin, K. Wuthrich, *J. Biomol. NMR* **15**, 181 (1999).
17. C. Ritter, T. Luhrs, W. Kwiatkowski, R. Riek, *J. Biomol. NMR* **28**, 289 (2004).
18. P. A. Kosen, *Methods Enzymol.* **177**, 86 (1989).

19. J. L. Battiste, G. Wagner, *Biochemistry* **39**, 5355 (2000).
20. P. Guntert, *Methods Mol. Biol.* **278**, 353 (2004).

21. C. Hilty, G. Wider, C. Fernandez, K. Wuthrich, *Chembiochem* **5**, 467 (2004).
22. T. P. Roosild et al., data not shown.
23. R. B. Bass, P. Strop, M. Barclay, D. C. Rees, *Science* **298**, 1582 (2002).
24. Y. Jiang et al., *Nature* **423**, 33 (2003).
25. D. R. Littler et al., *J. Biol. Chem.* **279**, 9298 (2004).
26. S. Choe et al., *Nature* **357**, 216 (1992).
27. W. Zhang, M. Bogdanov, J. Pi, J. Pittard, W. Dowhan, *J. Biol. Chem.* **278**, 50128 (2003).
28. We thank E. Wiater for help in TGF-β receptor binding studies and C. Park for Edman degradation sequencing of proteins. Supported by NIH grant GM056653. R.R. is a Pew scholar. The bundle of 10 conformers representing the NMR structure is deposited in the PDB database with accession code 1YGM. The coding sequence of Mystic has been deposited in GenBank with accession code AY874162.

Supporting Online Material
www.sciencemag.org/cgi/content/full/307/5713/1317/DC1
 Materials and Methods
 Figs. S1 to S6
 Tables S1 and S2
 References

14 October 2004; accepted 6 January 2005
 10.1126/science.1106392

The Genome of the Basidiomycetous Yeast and Human Pathogen *Cryptococcus neoformans*

Brendan J. Loftus,^{1*} Eula Fung,² Paola Roncaglia,³ Don Rowley,² Paolo Amedeo,¹ Dan Bruno,² Jessica Vamathevan,¹ Molly Miranda,² Iain J. Anderson,¹ James A. Fraser,⁴ Jonathan E. Allen,¹ Ian E. Bosdet,⁵ Michael R. Brent,⁶ Readman Chiu,⁵ Tamara L. Doering,⁷ Maureen J. Donlin,⁸ Cletus A. D'Souza,⁹ Deborah S. Fox,^{4,10} Viktoriya Grinberg,¹ Jianmin Fu,¹¹ Marilyn Fukushima,² Brian J. Haas,¹ James C. Huang,⁴ Guilhem Janbon,¹² Steven J. M. Jones,⁵ Hean L. Koo,¹ Martin I. Krzywinski,⁵ June K. Kwon-Chung,¹³ Klaus B. Lengeler,^{4,14} Rama Maiti,¹ Marco A. Marra,⁵ Robert E. Marra,^{4,15} Carrie A. Mathewson,⁵ Thomas G. Mitchell,⁴ Mihaela Perlea,¹ Florenta R. Riggs,¹ Steven L. Salzberg,¹ Jacqueline E. Schein,⁵ Alla Shvartsbeyn,¹ Heesun Shin,⁵ Martin Shumway,¹ Charles A. Specht,¹⁶ Bernard B. Suh,¹⁷ Aaron Tenney,⁶ Terry R. Utterback,¹⁸ Brian L. Wickes,¹¹ Jennifer R. Wortman,¹ Natasja H. Wye,⁵ James W. Kronstad,⁹ Jennifer K. Lodge,⁸ Joseph Heitman,⁴ Ronald W. Davis,² Claire M. Fraser,¹ Richard W. Hyman²

Cryptococcus neoformans is a basidiomycetous yeast ubiquitous in the environment, a model for fungal pathogenesis, and an opportunistic human pathogen of global importance. We have sequenced its ~20-megabase genome, which contains ~6500 intron-rich gene structures and encodes a transcriptome abundant in alternatively spliced and antisense messages. The genome is rich in transposons, many of which cluster at candidate centromeric regions. The presence of these transposons may drive karyotype instability and phenotypic variation. *C. neoformans* encodes unique genes that may contribute to its unusual virulence properties, and comparison of two phenotypically distinct strains reveals variation in gene content in addition to sequence polymorphisms between the genomes.

With an increased immunocompromised population as a result of AIDS and widespread immunosuppressive therapy, *Cryptococcus neoformans* has emerged as a major pathogenic microbe in patients with impaired immunity (1). *C. neoformans* elaborates two

coccus neoformans has emerged as a major pathogenic microbe in patients with impaired immunity (1). *C. neoformans* elaborates two