A Novel SNARE N-terminal Domain Revealed by the Crystal Structure of Sec22b

Received for publication, February 20, 2001, and in revised form, April 17, 2001
Published, JBC Papers in Press, April 17, 2001, DOI 10.1074/jbc.M101584200

Lino C. Gonzalez, Jr.†‡, William I. Weis§¶, and Richard H. Scheller‡∥
From the †Department of Molecular and Cellular Physiology, Howard Hughes Medical Institute and the §Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305

Intra-cellular membrane fusion is facilitated by the association of SNAREs from opposite membranes into stable α-helical bundles. Many SNAREs, in addition to their α-helical regions, contain N-terminal domains that likely have essential regulatory functions. To better understand this regulation, we have determined the 2.4-Å crystal structure of the 130-amino acid N-terminal domain of mouse Sec22b (mSec22b), a SNARE involved in endoplasmic reticulum/Golgi membrane trafficking. The domain consists of a mixed α-helical/β-sheet fold that resembles a circular permutation of the actin/polyproline binding protein, profilin, and the GAF/PAS family of regulatory modules. The structure is distinct from the previously characterized N-terminal domain of syntaxin 1A, and, unlike syntaxin 1A, the N-terminal domain of mSec22b has no effect on the rate of SNARE assembly in vitro. An analysis of surface conserved residues reveals a potential protein interaction site. Key residues in this site are distinct in two mammalian Sec22 variants that lack SNARE domains. Finally, sequence analysis indicates that a similar domain is likely present in the endosomal/lysosomal SNARE VAMP7.

Eukaryotic cells maintain highly organized internal membrane organelles that mediate transport, synthesis, and degradation of membrane and secreted proteins (1). Transport between compartments is accomplished by vesicles that accumulate specific cargo and bud from donor membranes. These vesicles are then transported to specific target membranes where they fuse, resulting in the transfer of contents. The machinery that underlies this critical membrane fusion event consists of compartment-specific proteins, which include the soluble NSF attachment protein receptors or SNAREs1 (2).

As part of the fusion process, cytoplasmic “SNARE domains” from opposed membranes assemble into an elongated, parallel four-helical bundle, termed the core complex (3, 4). Assembly of the core complex brings the membranes into close apposition, which is believed to lower the energy barrier to fusion (5–7). SNAREs can be divided into subgroups that are homologous to syntaxin 1, VAMP1 and 2 (also termed synaptobrevin), and SNAP-25 (8, 9). Syntaxin and VAMP each contribute a single helix to the complex and largely localize, via C-terminal membrane anchors, to acceptor and donor membranes, respectively. SNAP-25 contributes two helices to the core complex and is anchored to the membrane by lipid-modified cysteine residues.

A large set of regulatory factors tightly controls SNARE assembly and ultimately serves to govern the fidelity and timing of membrane trafficking in the cell (10). Examples of regulatory factors include the Rab and Sec1 proteins (2). At least 60 Rab proteins have been identified in the human genome, many localizing to specific compartments (9). The neuronal nSec1 may regulate fusion, at least in part, by binding syntaxin 1A on the plasma membrane (11). One hypothesis proposes that a Rab-effector complex signals to the nSec1-syntaxin 1A complex that the vesicle is properly docked and ready for fusion (12). Homologs of Sec1 have been identified that likely interact specifically with other syntaxin family members to regulate different transport steps (2, 13). The nSec1-syntaxin 1A interaction is mediated through both the SNARE domain and an N-terminal three-helical bundle domain of syntaxin (12, 14). These observations imply that sequences outside the SNARE domain are required for proper function in the cell and may serve critical regulatory functions. In support of this idea, the N-terminal domain of the syntaxin ortholog yeast Sso1p was recently shown to be necessary for viability (15).

Many non-syntaxin SNARE proteins also contain discrete N-terminal domains. The sequences of these domains, however, are distinct from syntaxin, perhaps because they regulate processes unique to their particular transport step. For example, the yeast SNAREs Sec9p and Spo20p function with the same syntaxin (Sso1p and 2p) and VAMP (Snc1p and 2p) partners. However, although Sec9p normally operates in vegetative cells, Spo20p is active only during sporulation. Analysis of chimeric Sec9p/Spo20p proteins demonstrated that the unique N terminus of Spo20p was required for activity during sporulation, whereas this domain was found to be inhibitory in vegetative cells (16).

In the VAMP subfamily, N-terminal sequences vary substantially and are either completely missing (VAMP5 and 8) or are less than 50 amino acids in length (VAMP1, 2, 3, and 4). The two exceptions are VAMP7 (also called T1-VAMP) and Sec22, which contain larger N-terminal domains (~130 amino acids). Interestingly, deletion of the VAMP7 N-terminal domain was found to strongly stimulate neurite outgrowth from PC12 cells (17), and expression of the deleted sequence alone inhibited neurite outgrowth. These data further support the hypothesis that sequences outside the conserved SNARE domains provide important regulatory functions.

---

‡ To whom correspondence should be addressed: Dept. of Structural Biology, Fairchild Bldg., Stanford University School of Medicine, Stanford, CA 94305-5428. Tel.: 650-725-4623; Fax: 650-723-8464; E-mail: bill.weis@stanford.edu.

¶ A Howard Hughes Medical Institute investigator.

∥ A Howard Hughes Medical Institute investigator.

The abbreviations used are: SNARE, soluble NSF attachment protein receptor; SNAP, soluble NSF attachment protein; BME, β-mercaptoethanol; CNS, Crystallography & NMR System; MAD, multilamellar membrane anomalous dispersion; r.m.s.d., root mean square difference; CD, circular dichroism; ER, endoplasmic reticulum; PBS, phosphate-buffered saline; PIP2, phosphatidylinositol 4,5-bisphosphate.
Sec22 localizes to membranes of the endoplasmic reticulum (ER), the intermediate compartment, and the cis-Golgi and has been proposed to play a role in both anterograde and retrograde trafficking (18–21). Rat Sec22b participates in a quaternary complex with three other SNAREs, rBet1, mSNAP, and syntaxin 5 (22, 23). Of these four SNAREs, all except rBet1 contain significant N-terminal domains. To begin to understand the functions of these possible regulatory domains, we have undertaken a biophysical and structural characterization of the mouse Sec22b (mSec22b) protein. The N-terminal 127 amino acids of mSec22b form an independently stable domain that consists of a mixed α-helical/β-sheet fold. Unlike syntaxin 1A, this N-terminal domain does not appear to interact with its own SNARE domain and has no effect on the rate of SNARE complex assembly. The structure, although circularly permuted, is similar to the well-characterized protein profilin and the GAF/PAS regulatory modules. Analysis of conserved surface residues within the Sec22 family, suggests that this domain serves as a protein recognition module.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—The full cytoplasmic domain of mouse Sec22b (amino acids 2–196) containing a N-terminal hexahistidine tag was expressed from a pQE-9 vector (Qiagen) in the JM109 strain of Escherichia coli. Cells were grown at 37 °C in M9 minimal media, supplemented with 10 ml/liter Kao Vitamins (Sigma Chemical Co.) and 0.1 mg/ml ampicillin. When the optical density at 600 nm reached 0.5, methionine synthesis was inhibited by the addition of 100 mM leucine and D-valine; and 60 mg/liter D/L-selenomethionine (Sigma) (24).

Mammalian cell cultures were grown in 2xYT media and induced at an optical density (600 nm) of 0.7. Protein expression and purification, using the glutathione S-transferase system, of syntaxin 1A (H3) (amino acids 191–266), mSec22b-SNARE domain (amino acids 126–195), and full-length SNAP-25 (amino acids 1–206, with cysteines 85, 88, 90, and 92 all mutated to alanine) have been described elsewhere (25). For circular dichroism studies, purification of full-length and N-terminal domains of mSec22b was carried out as described above, omitting steps for selenomethionine incorporation. Cells were grown in 2xYT media and induced at an optical density (600 nm) of 0.7. Size-exclusion chromatography was carried out in PBS with 5 mM BME. Proteins were concentrated as described above and quantitated using the Bio-Rad protein assay with bovine serum albumin as a standard.

Crystallization and Data Collection—Initial crystallization conditions were identified from hanging drop vapor diffusion at 20 °C, using the sparse matrix screen (Hampton) (26). Following optimization, the best crystals were obtained at 4 °C by mixing 1:1 ratios of −15 mg/ml protein with reservoir (−27% polyethylene glycol 1000, 100 mM ammonium sulfate, 67 mM sodium Citrate, pH 5.25, 20 mM BME, and 10% glycerol) on siliconized glass coverslips (Hampton). Crystals generally appeared after 1 week and reached full size, on average 0.2 × 0.2 × 0.1 mm, after 2 weeks. These conditions were sufficient for cryoprotection of the crystals, which were flash-frozen and stored in liquid nitrogen. The crystals belong to space group P2₁2₁2₁, with unit cell dimensions of a = 50.2, b = 57.0, and c = 99.0 Å. After optimization, the crystals diffracted to a minimum Bragg spacing of 2.4 Å. The asymmetric unit contains two molecules, corresponding to a solvent content of 51% as determined by the Matthews coefficient (27).

Diffraction data were measured at Stanford Synchrotron Radiation Laboratory beamline 9-2, using a Quantum 4 charge-coupled device detector (Area Detector Systems Corp.). Data were systematically

| Table I | Crystallographic data collection statistics |
|---------|-------------------------------------------|
| Wavelength | Resolution | % Complete | R_sym | % > 3σ(I) | Redundancy |
| λ₁ Remote | 0.92537 | 50–2.40 | 98.6 (90.9) | 0.036 (0.219) | 80.5 (51.8) | 3.63 (3.57) |
| λ₂ Peak | 0.97929 | 50–2.40 | 98.5 (90.7) | 0.035 (0.222) | 80.4 (51.2) | 3.64 (3.57) |
| λ₃ Edge | 0.89945 | 50–2.40 | 98.6 (90.7) | 0.035 (0.223) | 80.4 (50.5) | 3.64 (3.57) |

λ_sym = Σλ|I(h) − ⟨I(h)⟩|/ΣI(h), where I(h) is the Ith measurement of reflection h, and ⟨I(h)⟩ is the weighted mean of all measurements of h. Bijvoet measurements were treated as independent reflections.

Values in parentheses are for the highest resolution shell: 2.49–2.40 Å.

| Table II | Crystallographic phasing statistics |
|---------|-----------------------------------|
| Anomalous diffraction ratios | |
| λ₁ Remote | 0.0497 | 0.0444 | 0.0607 |
| λ₂ Peak | 0.0684 | 0.0353 |
| λ₃ Edge | 0.0482 |

Phasing power

(+) Friedel mate

| λ₁ Remote | Reference | 1.64 |
| λ₂ Peak | 1.46 | 2.68 |
| λ₃ Edge | 2.08 | 2.82 |

Resolution (Å) 50.0–4.70 4.70–3.79 3.79–3.31 3.31–3.01 3.01–2.80 2.80–2.63 2.63–2.50 2.50–2.40 50.0–2.40 Figure-of-merit 0.87 0.80 0.75 0.67 0.56 0.41 0.33 0.24 0.19 |

Anomalous diffraction ratios = ⟨|ΔF|/|F|⟩, where (|ΔF|) is the r.m.s. Bijvoet difference at a single wavelength (diagonal elements) or the r.m.s. dispersive difference between two wavelengths (off diagonal elements), computed between 50- and 2.40-Å resolution.

Phasing power = ⟨|F|/E⟩, where (|F|) is the r.m.s. structure factor amplitude for anomalous scatterers and E is the estimated lack-of-closure error. Phasing power is listed for each lack-of-closure expression between the reference data set (+Friedel mate at λ₁) and the + or − Friedel set at each wavelength. Phasing powers were calculated using all data between 50 and 2.40 Å.
Sec22b Crystal Structure

A

B

FIG. 1. Experimental MAD phased electron density maps at 2.4-Å resolution, contoured at 1σ. A, representative view showing a pair of anti-parallel α-helices. The refined model is also shown. Most side chains were clearly interpretable in the experimental map. B, view of an anti-parallel β-sheet. In red is the experimental map, and in blue is a refined 2Fo − Fc map, contoured at 1σ.

measured at 100 °K, at three wavelengths, in 22° wedges, using inverse-beam geometry for a total of 84° of data. Measured reflections were integrated, processed, and scaled using Denzo/Scalepack (28). Data collection statistics are presented in Table I.

Structure Determination—All six of the expected selenium sites were located using an automated Patterson heavy-atom search algorithm implemented in the Crystallography & NMR System (CNS) (29, 30). All map and phasing calculations were carried out using CNS. Experimental multiwavelength anomalous dispersion (MAD) phasing statistics are presented in Table II. The phases were improved by using solvent flipping and histogram matching implemented in CNS. Representative electron density maps at 2.4-Å resolution are shown in Fig. 1. The main-chain trace was readily identifiable with the exception of one mobile region (amino acids 21–29). Most of the side-chain density present in the refined 2Fo − Fc density maps was clearly identifiable in the initial MAD phased/solvent-flipped map.

Model Building and Refinement—All modeling was done using the program O (31). The initial model comprised ~84% of the final model, extending from amino acids 3 to 20 and 30 to 126 (the electron density for 21–29 was initially absent or weak). Strict non-crystallographic symmetry was imposed to generate the second molecule in the asymmetric unit. Refinement was performed using CNS against all data with |Fo| > 0 and was monitored by cross-validation (32), using a test set composed of randomly selected data (~10%). The initial model was refined by simulated annealing, using the MLHL target in CNS against the remote wavelength. Additional rounds of model building following by minimization were performed. Subsequent models were built using σ-related 2Fo − Fc maps and refined using the MLF target. Individual thermal factors were refined, followed by relaxation of the non-crystallographic symmetry restraints. Further cycles of model building and positional minimization, including 64 water molecules and 2 glycerol molecules, produced the final structure. The model statistics are shown in Table III. All residues from 2 to 127 are included in the model. Amino acids 23 to 27 were poorly represented in the electron-density maps and consequently have relatively high thermal factors. A number of side chains appear to be disordered in both molecules and were modeled as alanine, including Asp23, Arg28, Gln31, Gln47, Lys82, Lys101, Arg108, and Lys121. Coordinates have been deposited with the Protein Data Bank (accession code 1IFQ).

Circular Dichroism Spectroscopy—Circular dichroism (CD) data were recorded on an Aviv 62DS CD spectrophotometer equipped with a thermoelectric temperature controller. Measurements were made in PBS buffer, using either a 0.1- or 1-cm path-length quartz cuvette. Temperature measurements were recorded at 18 °C, using 15 μM protein or 60 μM for the mSec22 SNARE domain. Thermal unfolding experiments were performed at the same protein concentration by measuring the CD signal at 222 nm, allowing 1 min of equilibration per 1 °C temperature increment, averaging 30 s per measurement. Under these conditions, the unfolding transitions were irreversible. The melting temperatures are therefore apparent. Assembly reactions were performed by adding equimolar concentrations (0.8 or 8.0 μM) of the appropriate protein components to the cuvette and mixing thoroughly. Kinetic measurements were recorded at 18 °C by monitoring the CD signal at 222 nm every 10 s with a 5-s averaging time.

RESULTS

The N-terminal Domain of mSec22b (Mouse Sec22b) Is Folded and Stable—The full cytoplasmic sequence (amino acids 2–196) of mSec22b (mSec22b-FL) was expressed as an N-terminally His6-tagged fusion construct and purified on Ni2+-agarose beads. During the purification process, mSec22b (22 kDa) was proteolyzed into a smaller ~14-kDa fragment (mSec22b-NT), corresponding to approximately the first 125 amino acids (confirmed by N-terminal amino acid sequencing).

We used circular dichroism (CD) spectroscopy to determine if mSec22b possesses significant amounts of secondary structure.

| Table III | Refinement statistics |
|-----------|----------------------|
| Number of reflections | Working set: 20,615 |
| Test set: 1,906 |
| Rcryst | 0.231 |
| Rfree | 0.274 |
| Average B (Å²) | 57.1 |
| Main-chain bond-related B r.m.s.d. | 1.5 |
| Main-chain angle-related B r.m.s.d. | 2.6 |
| Side-chain bond-related B r.m.s.d. | 2.3 |
| Side-chain angle-related B r.m.s.d. | 3.4 |

| Model geometry |
|----------------|
| Bond length r.m.s.d. from ideal (Å) | 0.007 |
| Bond angle r.m.s.d. from ideal (°) | 1.30 |
| Ramachandran plot: |
| % in most favored regions | 83.1 |
| % in additionally allowed regions | 16.9 |
| % in generously allowed regions | 0.0 |

Note: *R = Σ||Fo|(|Fc|)/||Fo|| |Fo|/||Fo||. Rcryst and Rfree were calculated from the working and test reflection sets, respectively.
The data showed that the full-length and proteolytic fragment of mSec22b were structured (Fig. 2A). The measured mean molar ellipticities for both proteins were similar, but also lower than would be expected for a mostly α-helical protein. This suggests that either a fraction of the polypeptide chain was unfolded and/or there was a significant amount of β-sheet structure, which contributes less CD signal than α-helical residues. The lack of clearly defined minima at 208 and 222 nm is consistent with significant β-sheet content. Spectra measured on the isolated SNARE domain of mSec22b (residues 126–195) indicated that the polypeptide chain was predominantly unstructured (Fig. 2A). In addition, the thermal unfolding of mSec22b-FL and mSec22b-NT were virtually indistinguishable (−66 °C) (Fig. 2B), suggesting that the C-terminal domain does not significantly stabilize the N-terminal domain. Taken together, these results suggest that mSec22b consists of a possibly unstructured C-terminal SNARE domain and an independently folded and thermally stable, β-sheet containing N-terminal domain.

Overall Topology—Both the proteolytic fragment and full-length mSec22b were screened in crystallization trials. The full-length protein failed to crystallize, whereas the mSec22b-NT fragment crystallized readily from polyethylene glycol and ammonium sulfate. Selenomethionine was incorporated into the protein, and MAD phasing was used to deter-
mine the structure at 2.4-Å resolution (see “Experimental Procedures”). The experimental electron density map was of excellent quality (Fig. 1, A and B), clearly showing the presence of α-helices and β-strands.

The overall structure consists of a five-stranded, anti-parallel β-sheet and three flanking α-helices (Fig. 3, A and B, i). The domain is roughly globular with dimensions of ~30 × 30 × 30 Å. The fold begins with a β-hairpin (β1–β2), followed by a single helix (α1) that spans one face of the β-sheet, then leads into a three-stranded β-meander that completes the β-sheet. The remainder of the chain folds into a pair of anti-parallel α-helices that packs on one side of the β-sheet platform. The β-sheet gently curves around the straight α1 helix (Fig. 3, B, ii), and the α2 and α3 helices accommodate the β-sheet curvature by matching their relative helix-helix packing angle with the slope of the β-sheet.

The two monomers in the asymmetric unit are related by a ~90° rotation and bury 1195 Å² of surface area, which represents 10% of the total surface area of the crystallographic dimer. However, size-exclusion chromatography and native gel analysis indicate that mSec22b-NT does not dimerize in solution (data not shown). The interface consists of an acceptor pocket composed of residues between β3 and α1, and a donor loop (residues 108–113) and the outside surface of α3. Each mSec22b-NT monomer acts as an acceptor and a donor, generating a continuous fiber in the crystal lattice.

Structural Homology—Apart from the Sec22 or VAMP7 orthologs and variants, there are no sequences significantly homologous to mSec22b-NT in the current genome databases. Nonetheless, the three-dimensional structure might display similarity to a known structure in the Protein Data Bank, which could suggest a function. Therefore, we performed a search for structural homologs using the DALI server (33). A significant match was detected to the protein profilin (Z score = 6.4). The structural similarity between the two proteins spans about 66% of the mSec22b-NT sequence, with a root mean square deviation (r.m.s.d.) for Cα atoms of 2.6 Å (Fig. 4, A and B). The majority of the structural similarity lies in the five-stranded anti-parallel β-sheet and the pair of anti-parallel helices. The two folds, although topologically similar, are circular permutations of each other. Profilin is involved in regulating actin filament assembly and binds monomeric actin (34).

Profilin also binds to proline-rich regions of specific scaffolding proteins and, in this way, localizes to areas of actin filament assembly (35). Interaction of profilin with phosphatidylinositol 4,5-bisphosphate (PIP2) has also been reported (36).

Based on the structural similarity, we asked if mSec22b-NT displays functional similarities to profilin. First, could mSec22b-NT bind to PIP2? The biophysical details of the PIP2/profilin interaction have remained obscure. Nevertheless, we would expect that if mSec22b-NT harbors a PIP2 interaction site, a conserved electrostatic binding pocket, which is apparent in other PIP2 or phosphatidylinositol 1,4,5-trisphosphate binding domains, would be identifiable. Analysis of the surface features revealed no evidence for such a binding site; on the contrary, the electrostatic surface potential is mainly acidic (see below). Therefore, we do not consider interaction with PIP2 likely.

Poly-L-proline binds profilin in a groove between the two anti-parallel helices, interacting with highly conserved aromatic residues (Fig. 4B) (37). If mSec22b-NT binds poly-L-proline, the protein chains are color-ramped from blue to red (N to C termini). In this representation, the profilin/actin binding site is located on the right side of the molecule (green portion between β2 and β3), as indicated. The poly-proline binding groove of profilin, represented by the red line, is located between the N- and C-terminal helices.
proline in a similar manner, we would expect to find aromatic residues at these positions. There are indeed four phenylalanine residues and one tyrosine residue between the anti-parallel helices of mSec22b-NT that are conserved within the Sec22 family. In profilin, however, the two anti-parallel helices are not closely packed with each other, but instead pack independently on the surface of the β-sheet, providing a clear binding groove. In mSec22b-NT, the aromatic residues are partly involved in close helix-helix packing interactions. Thus, it is unclear whether mSec22b-NT could accommodate a poly-L-proline peptide in a manner similar to profilin. To test this possibility, we incubated mSec22b-NT with a resin coupled to poly-L-proline. Under conditions in which profilin readily binds to the resin, no detectable interaction was found with mSec22b-NT (data not shown). This result suggests that either mSec22b-NT does not interact with poly-L-proline or that the interaction is selective and requires a specific proline-rich sequence.

Actin binding is the third possible functional similarity to profilin. However, the crystal structure of profilin-β-actin complex shows that the primary actin interaction site on profilin is located in precisely a region that is absent in mSec22b-NT (Fig. 4, A and B) (38). Because of these structural differences, it is unlikely that mSec22b-NT interacts with actin in a similar manner.

Recently, the crystal structure of a GAF domain (Fig. 4C) (PDB ID: 1f5m), a regulatory module found in a variety of proteins, was reported and found to be structurally similar to profilin (3nul), a penicillin binding domain (1pmd), D-Ala, D-Ala transpeptidase (3pte), the PAS domain of photoactive yellow protein (3pyp), and an HERG potassium channel fragment (PDB ID: 1byw) (39). Each of these structural homologs were also identified by DALI to be structurally similar to mSec22b-NT. The central sheet and anti-parallel helices of mSec22b-NT and GAF are superimposable with a Ca r.m.s.d. of 2.8 Å and a DALI Z score of 4.3.

Structure-based Sequence Alignments—A structure-based sequence alignment of mSec22b-NT with two distinctly related orthologs is presented in Fig. 5. The domain is highly conserved, with 36% sequence identity between human and yeast. The most conserved region corresponds to a loop between β1 and β2 that contains an invariant pentapeptide sequence. The helix α2 is also highly conserved, as well as the second half of the loop connecting α2 to α3. The least conserved area encompasses a stretch of 30 amino acids that forms the α1 helix and the connecting loops to β-strands 2 and 3. The first connecting loop (residues 21–27) is highly variable in length and sequence and appears to be highly flexible in the crystal structure, as indicated by weak electron density and high thermal factors.

A BLAST search using hVAMP7 identified homology with yeast Sec22b in the region between amino acids 45 and 110 (E value = 0.003) (40). We therefore asked whether the remainder of the VAMP7 sequence would be compatible with the mSec22b-NT structure. In the first two β-strands and the first α-helix, the VAMP7 sequence contains a similar pattern of hydrophobic and polar residues (Fig. 5). The conserved β1-β2 loop in Sec22, however, is not present in VAMP7. The pattern of hydrophobicity matches that of mSec22b through the BLAST-detected region of homology to approximately residue 100, where the first half of the α2-α3 loop differs. The remainder of the VAMP7 sequence is sparsely but recognizably conserved with Sec22. The identity between mSec22b and hVAMP7 based on this alignment is 21%, which is in the range where structural similarity would be expected.

We also used the sequence analysis program GCG to obtain a quantitative measure of the similarity (41). Like BLAST, GCG was unable to align the first ~45 amino acids of mSec22b and hVAMP7 in a manner suggested by our structure-based alignment. However, removing the α1 region (residues 23 to 48) in mSec22b and the corresponding region in hVAMP7 from the input sequences allowed GCG to find the same alignment proposed here. In this case, based on a quality score obtained from 200 sequence randomizations and alignments, a Z score of 18.7 was obtained, where a Z score above 10 is considered significant. Therefore, it is likely that the overall topology and structure of the N-terminal domain of VAMP7, outside the α1 region, will be similar to the mSec22b-NT structure.
A recent comprehensive analysis of SNARE sequences in the human genome has suggested that two previously presumed mammalian Sec22b isoforms, Sec22a and Sec22c (20, 21, 23, 42), are not SNARE proteins (9). These variants are present in mammals but not lower organisms such as *Drosophila*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae*. Sec22c is expressed ubiquitously, whereas Sec22a is found mainly in the lung and liver. It was originally proposed that these two proteins were SNAREs, because of the homology to yeast Sec22p and Sec22b. The three proteins all contain N-terminal domains, an intervening stretch of amino acids, and a predicted transmembrane region and localize to the ER and Golgi. However, upon closer inspection, the majority of the sequence similarity lies in the N-terminal region, which corresponds to mSec22b-NT. The previously presumed SNARE domain regions of Sec22a and Sec22c in fact contain a significant number of glycine and proline residues, which are unfavorable in α-helices. Analysis with the coiled-coil prediction algorithm COILS (43), using the MTIDK matrix, (a, d)-position weighting and a 21-amino acid window, showed that the average predicted coiled-coil probability for Sec22b (residues 128–195) was 34.8%, whereas the probability for Sec22a (residues 131–219) and Sec22c (residues 131–215) were 0.02 and 0.2%, respectively. Therefore, we consider Sec22a and Sec22c non-SNARE protein variants of Sec22b.

Here we compare mouse and human Sec22b-NT, which have identical sequences, to the N-terminal domains of Sec22a and Sec22c (Fig. 5). Of these proteins, the N-terminal domains of hSec22a and hSec22c are closest in sequence (43% identity), whereas hSec22b-NT is most similar to hSec22a (39% identity). A number of surface residues that are conserved within the Sec22b family differ between the variants (Fig. 5). Such variant-specific differences suggest that these proteins may have altered binding specificities.

**Surface Features**—To identify potential sites of interactions with other molecules, we analyzed the electrostatic profile and conservation of residues on the surface of mSec22b-NT. The molecule is predominantly acidic on one side (Fig. 6A, i) and mainly uncharged on the opposite side (Fig. 6A, ii). On the neutral surface, two prominent grooves run along helix α1 (Fig. 6, A, ii and C, ii). The first groove lies between the α1 helix and β-strand β3 (named groove-β3) and is lined with a number of hydrophobic residues (Leu45, Leu54, Phe61, and Tyr68) (Fig. 6B, ii). These residues form a pocket that accommodates Ile63 from a second monomer in the asymmetric unit. Although the domain does not dimerize in solution, this crystal-mediated interaction is suggestive of a potential protein-binding site.

Mapping the conservation among the Sec22b orthologs onto the molecular surface shows that the hydrophobic pocket in groove-β3 is largely conserved (Fig. 6B, ii). The second groove, running between α1 and β2 (i.e. groove-β2), also has a conserved patch that is formed by the juxtaposition of two highly conserved loops (loop β1-β2 and loop α2-α3) (Fig. 6, B, ii, and C, respectively). This pattern of conservation and hydrophobicity is also present in Sec22a, with a hydrophobic patch formed by the juxtaposition of loops α2-α3 and β1-β2.

**Electrostatic and conserved surface features.** A, the electrostatic surface potential of mSec22b-NT was mapped onto the molecular surface and contoured at ±10kT/e, using GRASP (53). Blue patches represent positive potential, and red represents negative potential. The white surface is neutral. The views in (i) and (ii) show opposite sides of the molecule. The β2- and β3-grooves are indicated by yellow and green arrows, respectively. B, conserved residues (see Fig. 5) were mapped onto the molecular surface and color-coded orange for identical residues and magenta for conserved residues. White or gray shading represents variable side chains. All conserved surface residues are labeled except for His62 (located below Ser61) and Lys123 (located behind Thr122). (A) White patches represent positive potential, and red represents negative potential. The white surface is neutral. The views in (i) and (ii) show opposite sides of the molecule. The β2- and β3-grooves are indicated by yellow and green arrows, respectively. B, conserved residues (see Fig. 5) were mapped onto the molecular surface and color-coded orange for identical residues and magenta for conserved residues. White or gray shading represents variable side chains. All conserved surface residues are labeled except for His62 (located below Ser61) and Lys123 (located behind Thr122). C, the backbone worm representation (i, ii), showing the orientation of the molecular surface with respect to the polypeptide fold. Secondary structural elements are labeled and loops α2-α3 and β1-β2 (see text) are indicated. To facilitate comparison, A, B, and C are in identical orientations. This figure was prepared with GRASP.

**Fig. 6.** mSec22b-NT does not inhibit the rate of SNARE complex assembly. Assembly reactions were performed with syntaxin 1A (H3) and SNAP-25 as described under “Experimental Procedures,” and the CD signal at 222 nm was followed over time. The increase in CD signal represents α-helix formation as the SNARE complex is formed. The data are represented as the fraction folded (i.e. fraction of SNARE complexes assembled). The red data points were measured using mSec22b-FL in the assembly reaction, and black data points were measured using the mSec22b-SNARE domain. The data for experiments conducted at two concentrations are shown, as indicated.

**Fig. 7.** Mapping the conservation among the Sec22b orthologs onto the molecular surface shows that the hydrophobic pocket in groove-β3 is largely conserved (Fig. 6B, ii). The second groove, running between α1 and β2 (i.e. groove-β2), also has a conserved patch that is formed by the juxtaposition of two highly conserved loops (loop β1-β2 and loop α2-α3) (Fig. 6, B, ii, and C, respectively). This pattern of conservation and hydrophobicity is also present in Sec22a, with a hydrophobic patch formed by the juxtaposition of loops α2-α3 and β1-β2.

**Fig. 7.** Electrostatic and conserved surface features. A, the electrostatic surface potential of mSec22b-NT was mapped onto the molecular surface and contoured at ±10kT/e, using GRASP (53). Blue patches represent positive potential, and red represents negative potential. The white surface is neutral. The views in (i) and (ii) show opposite sides of the molecule. The β2- and β3-grooves are indicated by yellow and green arrows, respectively. B, conserved residues (see Fig. 5) were mapped onto the molecular surface and color-coded orange for identical residues and magenta for conserved residues. White or gray shading represents variable side chains. All conserved surface residues are labeled except for His62 (located below Ser61) and Lys123 (located behind Thr122). (A) White patches represent positive potential, and red represents negative potential. The white surface is neutral. The views in (i) and (ii) show opposite sides of the molecule. The β2- and β3-grooves are indicated by yellow and green arrows, respectively. B, conserved residues (see Fig. 5) were mapped onto the molecular surface and color-coded orange for identical residues and magenta for conserved residues. White or gray shading represents variable side chains. All conserved surface residues are labeled except for His62 (located below Ser61) and Lys123 (located behind Thr122). C, the backbone worm representation (i, ii), showing the orientation of the molecular surface with respect to the polypeptide fold. Secondary structural elements are labeled and loops α2-α3 and β1-β2 (see text) are indicated. To facilitate comparison, A, B, and C are in identical orientations. This figure was prepared with GRASP.
Lys124 is either a Gln in hSec22a or a Trp in hSec22c. Similar b b localizes to the hydrophobic pocket in groove-fore contribute to a binding surface. Ile113, which has been evo-
mSec22-FL through the SNARE domain, resulting in release of
SNARE domain of mSec22b to assemble into a SNARE complex
results, because we are simply probing the availability of the
SNARE complexes with mSec22b (H3) and SNAP-25, which have been shown to form stable
experiment. Instead we used the syntaxin 1A SNARE domain
complex formation. The insolubility of the syntaxin 5 SNARE
mammalian variants. One of these resides, Leu54,
localizes to the hydrophobic pocket in groove-β3 (Fig. 6B, ii). This residue, which is either a Leu, Ile, or Val in Sec22b, is an
Ala in hSec22c. Such an alanine substitution might enlarge the
pocket surface and influence the specificity of an interaction.

On the opposite side of the molecule, three variant-specific residue differences (residues 76, 79, and 83) come together in one area of the three-dimensional structure, corresponding to the conserved track in the Sec22b family (Fig. 6B, i). Other variant-specific differences are found for Lys81 and Lys124. Lys124 is either a Gln in hSec22a or a Trp in hSec22c. Similar to Ile113, a tryptophan at this solvent-exposed position could be important in mediating a protein interaction. Finally, there are differences in residues Asp90 and Asp96, which make up a significant part of the electronegative portion of the conserved track. The residue differences discussed here would likely change the surface and binding characteristics of the mamma-
lian Sec22 variants and could contribute to the specificity of a protein interaction.

The Rate of SNARE Complex Assembly in the Presence or Absence of the N-terminal Domain—The identification of a potential protein binding site on mSec22b-NT raises the possi-
bility that the C-terminal SNARE domain might fold back and interact with its own N-terminal domain, thereby inhibiting SNARE complex assembly. Such a mechanism has been con-

tirmed in vitro for syntaxin 1A and Sso1p (15, 44). Because SNARE complex formation is accompanied by a large increase in α-helical content, CD spectroscopy can be used to follow the SNARE assembly reaction over time. We used this method to assess whether the presence of the mSec22b N-terminal do-
main limits the availability of its own SNARE domain for complex formation. The insolubility of the syntaxin 5 SNARE
domain prevented us from using cognate SNAREs in this ex-
periment. Instead we used the syntaxin 1A SNARE domain
(h3) and SNAP-25, which have been shown to form stable
SNARE complexes with mSec22b in vitro (25). The use of non-cognate SNAREs should not significantly affect our results, because we are simply probing the availability of the
SNARE domain of mSec22b to assemble into a SNARE complex in the presence or absence of a potential regulatory domain. The mSec22b-FL protein is monomeric by size-exclusion chromatography above the concentrations used in these experiments (data not shown) Thus, self-oligomerization of mSec22-FL through the SNARE domain, resulting in release of the inhibitory N-terminal domain, can be excluded.

The rates of SNARE association, shown in Fig. 7, in the presence or absence of the N-terminal domain of mSec22b, are identical. The N-terminal domain of Sec22b therefore appears not to inhibit the C-terminal SNARE domain from being incor-
porated into core complexes in vitro. Thus, if the C-terminal
SNARE domain of mSec22b interacts at all with the N-termi-
nal domain, the interaction is too weak or transient to affect the overall rate of SNARE core complex formation.

DISCUSSION
A Possible Regulatory Module—The crystal structure of
mSec22b reveals a novel SNARE N-terminal domain, consist-
ing of a central β-sheet layer flanked by α-helices. The domain is structurally homologous to the actin/poly-proline binding protein profilin. Although the structural similarity is clear, the fold of mSec22b-NT is a circular permutation of the profilin fold. This fact suggests that these two proteins are not evolution-
arily related but instead have converged upon a common structural motif. As we have shown, the functional similarities between these two proteins may be limited. Perhaps the more interesting observation is that the profilin fold belongs to a larger class of proteins that includes the PAS and GAF do-

mains (39, 45). These modules are found in a number of pro-
teins that are implicated in signaling or sensory pathways and appear to bind a diverse set of regulatory small molecules (39). Also, to exert their regulatory functions, these domains must participate in protein-protein interactions. Examples include

GAFP, which regulate cGMP-specific phosphodiesterases by binding to cGMP (46), and the phototactic yellow protein, which along with a covalently bound cofactor absorbs a photon and initiates a signal transduction pathway (45). The N-termi-
nal PAS domain of the human erg potassium channel interacts with and modifies channel properties (47). Due to their diverse regulatory functions, systematic comparison of these protein classes with mSec22b-NT is not sufficient to pinpoint a com-
mon binding site. However, these protein families do contain a structural commonality in their central anti-parallel β-sheet. This β-sheet may serve as a general structural platform that is decorated with various loops and α-helices that encode specific binding sites and impart functionality. Therefore, we speculate that the N-terminal domain of mSec22b may have evolved separately and converged upon a similar structural platform that functions as a regulatory domain.

Isoform Variations—The Drosophila, C. elegans, and S. cer-
evisiae genomes each contain a single ortholog of the Sec22b protein. Mammals, however, have two additional variants that do not appear to contain SNARE domains (Sec22a and c) but still harbor a strongly conserved N-terminal domain. In our analysis of the conserved surface residues, we find significant differences between the variants, suggesting that they interact with distinct partners. Why might mammals have evolved Sec22 variants? Genomic analysis suggests that the repertoire of proteins that regulate membrane fusion has increased sub-
stantially through evolution, whereas the number of SNARE proteins has only increased modestly (9). Perhaps, therefore, mammals have evolved Sec22b variants as additional regulators of membrane trafficking between the ER and Golgi. It is likely that the Sec22b-NT-like domains of these variants retain their regulatory functions, whereas the SNARE domain in the ancestral gene diverged from its function in membrane fusion. These observations raise interesting questions, but more studies will be needed to delineate the precise roles of these vari-
ants in higher organisms.

Functional Models—Could the N-terminal domain of Sec22b regulate SNARE function in a manner analogous to the N terminus of syntaxin? The C-terminal SNARE domain of syn-
taxins is known to fold back and interact with the N-terminal,
three-helical bundle domain (15, 44). In this closed conformation, the rate at which syntaxin assembles into SNARE core complexes is dramatically reduced (48). Despite the identification of two potential binding grooves on mSec22b-NT that could accommodate a portion of the C-terminal SNARE domain, our results indicate that the N-terminal domain has no effect on the rate of core complex formation. Thus, in vitro the N-terminal domains of Sec22b and syntaxin behave differently.

These results do not preclude a model where an unidentified factor recognizes both the N- and C-terminal domains of Sec22b to limit access to partner SNAREs. The conserved, acidic track present on the surface of mSec22b-NT is suggestive of an interaction site for a larger protein. Alternatively, the similarity to the GAF/PAS domain, suggests a model in which a post-translational modification, a signaling peptide, or a small molecule ligand might bind to the N-terminal domain, resulting in a conformation capable of interacting with the SNARE domain to form a closed conformation. The N terminus may otherwise be involved in functions such as proper subcellular localization or packaging into transport vesicles, or perhaps interactions with Rab GTPases or their effectors.

Conclusion—Regulation of membrane fusion through control of SNARE complex formation is a key event governing vesicle trafficking. In vivo it is abundantly clear that SNAREs associate into very specific complexes characteristic of particular trafficking steps. However, several studies suggest that particular SNAREs participate in multiple trafficking pathways through the formation of combinatorial sets of complexes (49). For example, both syntaxin 5 and syntaxin 17 are thought to form complexes with mSec22b in mammals. In addition, both syntaxin 5 and Sed5p have been proposed to interact with overlapping sets of SNAREs (23, 50). Many of the SNAREs involved in these multiple interactions have N-terminal domains that could play critical roles in governing the combinatorial associations required for a single SNARE to participate in multiple trafficking steps. The structure provided in this work will help in the design of more rational approaches to understanding these issues.

Acknowledgments—We thank K. Ervin, S. Patel, and B. Yang for technical support, J. Bock for help with sequence analysis; S. Almo for discussions and the gift of poly-proline resin; and H. Bellamy and A. Cohen of Stanford Synchrotron Radiation Laboratory (SSRL) for beamline support. Portions of this research were carried out at the Stanford SSRl, a national user facility operated by Stanford University on behalf of the United States Department of Energy, Office of Basic Energy Sciences. The SSRl Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research and by the NCCR, Biomedical Technology Program, and NIGMS, National Institutes of Health.

REFERENCES

1. Palade, G. (1975) Science 189, 347–358
2. Jahn, R., and Sudhof, T. C. (1999) Annu. Rev. Biochem. 68, 883–911
3. Sutton, R. B., Popp, D., Jahn, R., and Brünger, A. T. (1998) Nature 395, 347–353
4. Poitr, M. A., Xiao, W., Macosko, J. C., Chan, C., Shin, Y. K., and Bennett, M. K. (1998) Nat. Struct. Biol. 5, 785–789
5. Weber, T., Zemelman, B. V., McNew, J. A., Westermarck, B., Gmachl, M., Parlati, F., Sollner, T. H., and Rothman, J. E. (1998) Cell 92, 759–772
6. Hanson, P. I., Roth, R., Morisaki, H., Jahn, R., and Heuser, J. E. (1997) Cell 90, 523–535
7. Lin, R. C., and Scheller, R. H. (1997) Neuron 19, 1087–1094
8. Weimbs, T., Mostov, K., Low, S. H., and Hofmann, K. (1998) Trends Cell Biol. 8, 260–262
9. Bock, J. B., Matern, H. T., Peden, A. A., and Scheller, R. H. (2001) Nature 409, 839–841
10. Pfeffer, S. R. (1999) Nat. Cell Biol. 1, E17–E22
11. Pevesar, J., Hsu, S. C., and Scheller, R. H. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 1445–1449
12. Misura, K. M., Scheller, R. H., and Weis, W. I. (2000) J. Appl. Crystallogr. 33, 841–845
13. Munson, M., Chen, X., Cocina, A. E., Schultz, S. M., and Hughson, F. M. (2000) Nat. Struct. Biol. 7, 894–902
14. Martinis-Arca, S., Alberts, P., Zahraoui, A., Louvard, D., and Galli, T. (2007) J. Cell Biol. 179, 889–900
15. Hay, J. C., Kumpf, A., Ososvych, V., Steegmaier, M., Kuo, C. S., and Scheller, R. H. (1998) J. Biol. Chem. 273, 1489–1502
16. Xue, D., Joglekar, A. P., Williams, A. L., and Hay, J. C. (2000) J. Biol. Chem. 275, 39831–39839
17. Hay, J. C., Chao, D. S., Kuo, C. S., and Scheller, R. H. (1997) Cell 89, 149–158
18. Yang, B., Gonzalez, L., Jr., Prekeris, R., Steegmaier, M., Advani, R. J., and Scheller, R. H. (1999) J. Biol. Chem. 274, 5649–5653
19. Jancarik, J., and Kim, S. H. (1991) J. Appl. Crystallogr. 24, 409–411
20. Matthews, B. W. (1968) J. Mol. Biol. 33, 491–497
21. Altmann, J., and Minor, W. (1997) Methods Enzymol. 276, 307–326
22. Grosse-Kunstleve, R. W., and Brunger, A. T. (1999) Acta Crystallogr. Sect. D Biol. Crystallogr. 55, 1568–1577
23. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., and Pannu, N. S. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921
24. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. O. (1991) Acta Crystallogr. Sect. A 47, 19–21
25. Brünger, A. T. (1992) Nature 355, 472–475
26. Holm, L., and Sander, C. (1993) J. Mol. Biol. 233, 123–138
27. Wohlb, A. K., Belmont, L. D., Mahoney, N. M., Almo, S. C., and Drubin, D. G. (2000) J. Cell Biol. 150, 895–904
28. Gang, F., Laine, R. O., Bubb, M. R., Southwick, F. S., and Purich, D. L. (1997) Biochemistry 36, 8384–8392
29. Söhn, R. H., Chen, J., Kohan, K. S., Bray, P. F., and Goldschmidt-Clermont, P. J. (1995) J. Biol. Chem. 270, 21114–21120
30. Mahoney, N. M., Rutwara, D. A., Fedorov, E., Fedorov, A. A., and Almo, S. C. (1999) Nat. Struct. Biol. 6, 666–671
31. Schutt, C. E., Myslik, J. C., Rozycki, M. D., Gonesekere, N. C. W., and Lindberg, U. (1993) Nature 365, 810–816
32. Ho, T. J., Burden, I. M., and Hurley, J. H. (2000) EMBO J. 19, 5288–5299
33. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 3389–3402
34. Wobble, D. D. (2000) Methods Mol. Biol. 132, 3–22
35. Tragg, B. L., Lew, D. Y., and Hong, W. (1998) Biochem. Biophys. Res. Commun. 243, 885–891
36. Lupas, A. (1996) Methods Enzymol. 266, 513–525
37. Dumbkov, I., Sugita, S., Hill, S., Hassaka, M., Fernandez, I., Sudhof, T. C., and Jörg, J. (1999) J. Biochem. 126, 4372–4382
38. Bergdahl, G. E., Williams, D. R., and Getzoff, E. D. (1995) Biochemistry 34, 6278–6287
39. Arvind, L., and Ponting, C. P. (1997) Trends Biochem. Sci. 22, 458–459
40. Morais-Cabral, J. H., Lee, A., Cohen, S. L., Chait, B. T., Li, M., and Mackinnon, R. (1998) Cell 95, 645–669
41. Scheller, R. H., Morisaki, H., Jahn, R., and Heuser, J. E. (1997) J. Biol. Chem. 272, 666–671
42. Borgstahl, G. E., Williams, D. R., and Getzoff, E. D. (1995) Biochemistry 34, 6278–6287