

## Structure of the VHS Domain of Human Tom1 (Target of Myb 1): Insights into Interactions with Proteins and Membranes<sup>‡</sup>

Saurav Misra, Bridgette M. Beach, and James H. Hurley\*

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases,  
National Institutes of Health, Bethesda, Maryland 20892-0580

Received June 14, 2000; Revised Manuscript Received July 25, 2000

**ABSTRACT:** VHS domains are found at the N-termini of select proteins involved in intracellular membrane trafficking. We have determined the crystal structure of the VHS domain of the human Tom1 (target of myb 1) protein to 1.5 Å resolution. The domain consists of eight helices arranged in a superhelix. The surface of the domain has two main features: (1) a basic patch on one side due to several conserved positively charged residues on helix 3 and (2) a negatively charged ridge on the opposite side, formed by residues on helix 2. We compare our structure to the recently obtained structure of tandem VHS–FYVE domains from Hrs [Mao, Y., Nickitenko, A., Duan, X., Lloyd, T. E., Wu, M. N., Bellen, H., and Quiocho, F. A. (2000) *Cell* 100, 447–456]. Key features of the interaction surface between the FYVE and VHS domains of Hrs, involving helices 2 and 4 of the VHS domain, are conserved in the VHS domain of Tom1, even though Tom1 does not have a FYVE domain. We also compare the structures of the VHS domains of Tom1 and Hrs to the recently obtained structure of the ENTH domain of epsin-1 [Hyman, J., Chen, H., Di Fiore, P. P., De Camilli, P., and Brünger, A. T. (2000) *J. Cell Biol.* 149, 537–546]. Comparison of the two VHS domains and the ENTH domain reveals a conserved surface, composed of helices 2 and 4, that is utilized for protein–protein interactions. In addition, VHS domain-containing proteins are often localized to membranes. We suggest that the conserved positively charged surface of helix 3 in VHS and ENTH domains plays a role in membrane binding.

The intracellular flow of lipids and proteins between organelles is central to a vast array of cellular processes. The molecular machinery that directs this membrane traffic has come into sharper focus with the determination of the three-dimensional structures of key endocytic (1, 2), exocytic (3, 4), and endosomal (5–8) proteins and domains. The VHS<sup>1</sup> domain, found in the Vps27, Hrs, and STAM proteins among others (9), has recently been implicated in membrane trafficking processes.

VHS domains are ~150-residue domains located at the N-termini of three groups of proteins. Proteins in the first group, including the yeast Vps27 protein, its mammalian homologue Hrs (hepatocyte growth factor-regulated tyrosine kinase substrate), and the related Hrs-2 protein, contain a FYVE domain C-terminal to the VHS domain. FYVE domains (found in Fab1p, YOTB, Vac1p, and EEA1) are

membrane localization domains that specifically bind phosphatidylinositol 3-phosphate (10–12). Vps27 and Hrs regulate endosome maturation and trafficking between endosomes and the yeast vacuole or mammalian lysosomes (13). Hrs-2 is an ATPase involved in Ca<sup>2+</sup>-dependent transport of synaptic vesicles to the plasma membrane and interacts with the synaptic membrane-associated protein SNAP-25 (14).

A second group of VHS domain-containing proteins is involved in signaling by growth factor receptors and in receptor endocytosis. EAST (epidermal growth factor receptor-associated protein with SH3 and TAM domains) associates with the EGF receptor and clathrin-coated pits, and becomes phosphorylated in response to EGF (15). EAST also associates with Eps15, another EGFR substrate. Eps15 is localized to clathrin-coated vesicles, binds the  $\alpha$  subunit of the endocytic adaptor complex AP-2, and is required for receptor-mediated endocytosis (16–18). EAST also colocalizes with the actin cytoskeleton and focal adhesion components and may mediate EGF effects on the cytoskeleton (15). An N-terminal portion of EAST, including its VHS domain and ~30 additional residues, appears to mediate the interaction between EAST and the actin cytoskeleton. The STAM (signal transducing adaptor molecule) protein participates in cytokine-mediated signaling and is phosphorylated by the Jak2 and Jak3 tyrosine kinases. STAM enhances the induction of c-myc by cytokine stimulation (19) and associates with Hrs under these conditions (20). The domain organization of STAM and EAST is very similar (9); in addition to VHS

<sup>‡</sup> X-ray coordinates have been deposited in the Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ, under ID code 1ELK.

\* To whom correspondence should be addressed. Phone: (301) 402-4703. Fax: (301) 496-0201. E-mail: jh8e@nih.gov.

<sup>1</sup> Abbreviations: VHS, domain found in Vps27, Hrs, and STAM; Hrs, hepatocyte growth factor receptor-regulated tyrosine kinase substrate; FYVE, domain found in Fab1p, YOTB, Vac1p, and EEA1; EGF and EGFR, epidermal growth factor and epidermal growth factor receptor, respectively; STAM, signal transducing adaptor molecule; EAST, epidermal growth factor receptor-associated protein with SH3 and TAM domains; ITAM, immunoreceptor tyrosine-based activation motif; GGA, Golgi-localized,  $\gamma$ -ear containing, ARF-binding; ARF, ADP ribosylation factor; Tom1, target of myb 1; ENTH, epsin N-terminal homology; SeMet, selenomethionine; PH, pleckstrin homology.

domains, both contain SH3 domains and ITAM motifs. ITAMs (immunoreceptor tyrosine-based activation motifs) function as SH2 domain binding sites when they are phosphorylated. An uncharacterized yeast gene product, YHA2 (*Saccharomyces cerevisiae* open reading frame YHL002W), contains VHS and SH3 domains in a similar region of its sequence as EAST and STAM but lacks an ITAM domain (9). YHA2 has been shown to interact with Vps27 in a large-scale study of yeast protein–protein interactions (21); this interaction may be analogous to the STAM–Hrs interaction mentioned above.

Recently, a third group of proteins containing VHS domains has been identified. Termed GGAs (Golgi-localized,  $\gamma$ -ear containing, ARF-binding), these proteins also contain a central region (called a GAT domain) which interacts with ARFs (ADP ribosylation factors). The C-termini of the GGAs are homologous to the ear domain of the  $\gamma$ -adaptin subunit from the AP-1 adaptor complex (22–25). The ear domains of adaptins bind to accessory proteins, helping to recruit them to the AP complexes and to the membrane surface (1, 26). The homologous regions of the GGAs bind to some of the same proteins as  $\gamma$ -adaptin (24). GGAs are localized to the trans-Golgi network and late Golgi membranes, and may function as adaptors themselves, or even as components of a Golgi-specific coat complex (23, 24).

An 85-residue subsection of the ARF-binding GAT region of the GGAs has limited homology to portions of two other VHS domain-containing proteins, Tom1 (target of myb 1) and Tom1Like (23, 27). While less is known about the function of Tom1 than other VHS domain-containing proteins, it has been shown that transcription of the Tom1 gene is induced by the oncogene v-Myb in transformed cells. v-Myb or its cellular protooncogene counterpart, c-Myb, activates Tom1 transcription in cooperation with the Ets transcription factor, another protooncogene (27). c-Myb is involved in cell differentiation processes, including hematopoiesis (28). Tom1 may help to regulate membrane trafficking in response to c-Myb activity during cellular differentiation or transformation.

Thus, VHS domains are found in a variety of proteins that span the breadth of intracellular membrane traffic. They are invariably found at the N-termini of these proteins. The function of VHS domains has been investigated through membrane localization studies. Green fluorescent protein (GFP) fusions with the VHS domain of GGA3 are primarily cytosolic (23). In contrast, the VHS domain of GGA2 localizes diffusely to intracellular membranes (but not specifically to the Golgi) when expressed by itself (25). The VHS domain of EAST associates with the plasma membrane and colocalizes with clathrin (29). These results suggest that VHS domains may partially determine the intracellular localization of the proteins in which it is found.

To further investigate the function of the VHS domain, we attempted to determine a VHS domain structure by subcloning several VHS domains and screening them for efficacious overexpression and crystallization. We were successful in crystallizing the VHS domain from the Tom1 protein, and we present its structure at 1.5 Å resolution. Mao et al. (7) recently published the structure of the VHS and FYVE domains from the Hrs protein. Hrs belongs to the first group of VHS domain-containing proteins classified above, while Tom1 is similar to the GGAs in the third group. We

compare the Tom1 VHS and Hrs VHS structures in this paper. In addition, the structure of the ENTH (epsin N-terminal homology) domain of epsin-1 was recently reported (8). ENTH domains are found at the N-termini of a group of proteins involved in endocytosis. The best-characterized ENTH-containing proteins, the epsins, bind clathrin and AP2 (30, 31); these binding interactions map to the C-terminal halves of the epsins, and do not appear to involve the ENTH domains. The structure of the epsin-1 ENTH domain exhibits remarkable similarity to that of the Tom1 VHS domain despite the lack of discernible sequence homology. Structural comparison between the two VHS structures and the ENTH structure emphasizes elements of protein sequence and structure that are conserved between the domains.

## EXPERIMENTAL PROCEDURES

**Expression and Purification.** Residues 2–153 of *Homo sapiens* Tom1 were amplified from an expressed sequence tag (IMAGE clone 667343; Soares\_NhHMPu\_S1 library) using PCR. The sequence was cloned into the *Bam*HI and *Eco*RI sites of the pHis-Parallel2 vector (32), which adds a six-His tag and TEV (tobacco etch virus) protease cleavage site to the N-terminus of the expressed protein. The fusion protein was expressed in *Escherichia coli* BL21(DE3) cells (Novagen) grown for 20 h at 20 °C after induction with 0.5 mM IPTG. Cells in lysis buffer [150 mM NaCl, 50 mM Tris-HCl (pH 8.0), 2 mM DTT, 100  $\mu$ g/mL AEBSF, and leupeptin, pepstatin, and aprotinin (1  $\mu$ g/mL each)] were lysed using a French press (SLM-Aminco). The His-tagged protein was purified on a Ni–NTA Superflow column (Qiagen) in a 50 to 400 mM imidazole gradient. The six-His tag was cleaved off by overnight incubation, at 4 °C, with 1000 units of TEV protease (Gibco-BRL). Cleaved six-His tags, uncleaved fusion protein, and TEV protease (which is also His-tagged) were removed by another pass over the Ni–NTA column in 5 mM imidazole. The cleaved protein was concentrated to 20–30 mg/mL and stored in 20% glycerol at –80 °C. The protein yield was ~5 mg/L of culture. The final expressed protein contains the N-terminal sequence GAMGS, contributed by the vector, and Tom1 residues 2–153. Selenomethionylated (SeMet) protein was expressed in the *E. coli* methionine auxotroph B834 (Novagen) grown in defined medium. SeMet protein was purified using the same protocol that was used for the native protein, but with only ~50% of the native yield.

**Protein Crystallization.** Freshly thawed protein was dialyzed into a low-salt buffer [50 mM NaCl, 20 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, and 10 mM DTT] at a concentration of 12–18 mg/mL (a range of protein concentrations was always screened). Native and SeMet crystals were grown by vapor diffusion in hanging drops consisting of 2  $\mu$ L each of protein and crystallization buffer [20% PEG 8000 and 100 mM HEPES (pH 7.5)]. Crystals grew at 20 °C or at room temperature (unregulated) in 5–10 days, to a maximum size of 0.6 mm  $\times$  0.3 mm  $\times$  0.3 mm.

**Crystallographic Data Collection.** Crystals were cryoprotected by brief soaking in 100 mM HEPES (pH 7.5), 17% PEG 8000, and 15% PEG 400. Cryoprotected native crystals were frozen by immersion in liquid propane; SeMet crystals were frozen directly in the Nitrogen vapor cryostream,

Table 1: Crystallographic Data Collection, Phasing, and Refinement Statistics

		Crystallographic Data and Phasing				
space group		$P2_1$				
cell dimensions		$a = 42.07 \text{ \AA}, b = 52.86 \text{ \AA}, c = 73.22 \text{ \AA}, \beta = 99.75^\circ$				
		$d_{\min}$ (\AA)	no. of reflections	complete- ness	$\langle I \rangle / \langle \sigma \rangle$	$R_{\text{sym}}^a$ (%)
SeMet $\lambda_1$ (0.97925 \AA) <sup>b</sup>		1.90	44630 (4245) <sup>c</sup>	93.5 (89.4)	20.6 (4.6)	3.4 (16.0)
native (0.97930 \AA)		1.50	49338 (4775)	97.0 (94.8)	22.4 (6.3)	7.6 (27.6)
		Observed Dispersive and Bijvoet Ratios <sup>d</sup>				
		$\lambda_1$	$\lambda_2$	$\lambda_3$		
		$\lambda_1$	0.060	0.023	0.036	
		$\lambda_2$		0.038	0.042	
		$\lambda_3$			0.046	
figure of merit		0.78 (0.69)				
		Refinement				
resolution range		16–1.5 \AA				
no. of reflections		48257				
$R^e$		19.3% (17.4%) <sup>c</sup>				
$R_{\text{free}}^f$		22.1% (21.4%)				
Luzzati coordinate error		0.16 \AA				
cross-validated Luzzati coordinate error		0.19 \AA				
bond-length deviation		0.019 \AA				
bond-angle deviation		1.9^\circ				
improper-angle deviation		1.38^\circ				
dihedral deviation		21.2^\circ				
average $B$ factor		13.2 \AA^2				
bonded main chain atom $B$ factor rmsd		1.16 \AA^2				
bonded side chain atom $B$ factor rmsd		2.3 \AA^2				
residues in the most favored $\phi$ - $\psi$ region		92.6%				
residues in disallowed regions		0.0%				

<sup>a</sup>  $R_{\text{sym}} = \sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)$ . <sup>b</sup> Statistics for  $\lambda_2$  (0.97972 \AA) and  $\lambda_3$  (0.95740 \AA) were very similar to those for  $\lambda_1$ . <sup>c</sup> Values in parentheses are for the highest-resolution bin. <sup>d</sup> Ratios are calculated as  $\langle \Delta|F| \rangle / \langle |F| \rangle$ , using data between 5 and 2.5 \AA. <sup>e</sup>  $R = \sum(|F_{\text{obs}}| - k|F_{\text{calc}}|) / \sum|F_{\text{obs}}|$ . <sup>f</sup>  $R_{\text{free}}$  is the  $R$  value calculated for a test set of reflections, comprising a randomly selected 10% of the data, not used during refinement.

warmed (annealed) by brief resoaking in cryoprotectant, and refrozen. In the absence of annealing, the SeMet crystals almost invariably had split diffraction spots and high mosaicity. Native crystals also benefited from annealing. Initial data were collected using mirror-focused Cu K $\alpha$  radiation from a Rigaku RU-200 rotating anode source at 100 mA and 50 kV and an RAXIS-IV image plate detector. Autoindexing using HKL (33) showed that the crystals were in space group  $P2_1$  with the following unit cell dimensions:  $a = 42.1 \text{ \AA}, b = 52.9 \text{ \AA}, c = 73.2 \text{ \AA}$ , and  $\beta = 99.8^\circ$ . Native and (three-wavelength) SeMet MAD data were collected at beamline X9B (National Synchrotron Light Source, Brookhaven National Laboratory, Upton, NY) (Table 1). No inverse beam or other collection strategy was adopted. The native and SeMet crystals diffracted to 1.5 and 1.9 \AA, respectively, at the beamline. Synchrotron data were also indexed and merged using HKL.

**Structure Determination and Refinement.** Identification and refinement of the heavy-atom sites, MAD phasing and density modification (solvent flipping), and refinement were carried out in CNS 0.9 (34). Five of eight possible selenium sites were used in phasing. Solvent flattening using a solvent fraction of 0.43 and data to 2.4 \AA yielded an interpretable density map, which was used to build an atomic model in O

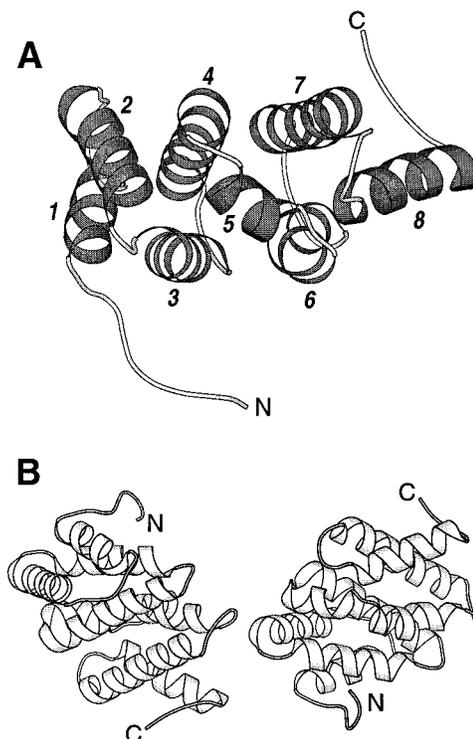


FIGURE 1: Structure of the Tom1 VHS domain. (A) Secondary structure of Tom1 VHS. (B) Tom1 VHS dimer in the crystallographic asymmetric unit. The interaction between the monomers is limited, suggesting that the dimer is nonphysiological. Structures were drawn using Molscript (58).

(35). Refinement was carried out using torsional dynamics and the maximum likelihood target function (34), the stereochemical restraints of Engh and Huber (36), and the thermal factor restraints of Tronrud (37). The free  $R$  factor (38) calculated with  $\sim 10\%$  of the observed reflections was used to monitor the refinement. The final structure includes two Tom1 VHS domains in the asymmetric unit and 385 water molecules. Tom1 residues 2–153 and a single N-terminal residue (Ser) contributed by the vector are resolved in each of the protein molecules. The stereochemical quality of the structure is as good as or better than expected for structures of this resolution, as gauged in PROCHECK (39). No non-glycine residues are in the disallowed or generously allowed regions of the Ramachandran plot.

## RESULTS

**Structure of the Tom1 VHS Domain.** The human Tom1 VHS domain, consisting of residues 2–153, is a right-handed superhelix of eight helices (Figure 1A). Helix 5 is only half as long as the other helices and runs perpendicular to them. Helix 8 points away from the others, but the cleft between helix 8 and helices 6 and 7 is filled with hydrophobic residues from all three helices. The core of the Tom1 VHS domain consists almost exclusively of hydrophobic residues, with polar, acidic, and basic residues restricted to the surface of the domain. The Tom1 VHS domain appears to be a monomer in solution from gel filtration data (not shown). The VHS domain, however, crystallizes as a dimer in crystallographic space group  $P2_1$  (Table 1). The contacts between the two molecules in the asymmetric unit are limited (Figure 1B), and the crystallographic dimer is likely nonphysiological.

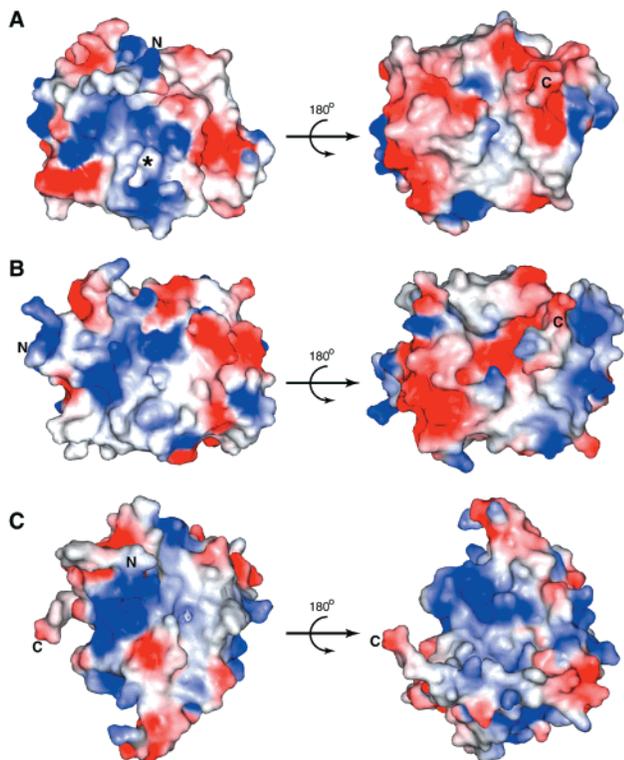


FIGURE 2: Molecular surface of Tom1 VHS, Hrs VHS, and epsin-1 ENTH using a 1.4 Å probe. The surfaces are colored according to electrostatic potential, with saturating colors at 10 kT (blue) and  $-10$  kT (red) at 298 K and a salt concentration of 150 mM. Electrostatic calculations were performed and surfaces rendered and colored using SPOCK (59). The N- and C-termini are marked. (A) View of the Tom1 VHS domain looking at helices 1, 3, 6, and 8 (left) and a view rotated 180°, looking at helices 2, 4, and 7 (right). The asterisk marks a small pocket lined with basic residues (see the text). (B) Hrs VHS domain in the same orientation as in panel A. (C) Epsin-1 ENTH domain in the same orientation as in panel A.

The N-terminal loop preceding the first helix is longer in Tom1 than most other VHS domains and is primarily hydrophobic. In the crystal, the loop prominently extends away from the rest of the protein. The extended conformation of the N-terminal loop is stabilized by packing against the C-terminus of helix 8 and the loop between helices 6 and 7 of a symmetry-related molecule. The N-terminal loop also makes limited contact with the N-terminus of helix 3 from a second symmetry-related molecule.

**Surface of Tom1 VHS.** Two features dominate the Tom1 VHS domain surface: a positively charged patch on one side of the domain and a negatively charged ridge on the opposite side (Figure 2A). In the Tom1 VHS structure, the pocket is partially shielded on one side by the extended N-terminal loop. The positive charges are contributed by four basic residues gathered together on helix 3: Arg52, Lys55, Lys56, and Arg57. At least two of the latter three residues are conserved or conservatively substituted in most VHS domains; the first is less well conserved. The positive patch surrounds a small nearby pocket, formed by a gap between the helix 3–helix 4 interhelical loop and helices 6 and 7 (Figure 2A, marked with an asterisk). Lys62, Arg101, and Lys106 cause the rim of this pocket to be positively charged, but the bottom of the pocket is lined with hydrophobic residues. Six water molecules occupy the pocket in our structure.

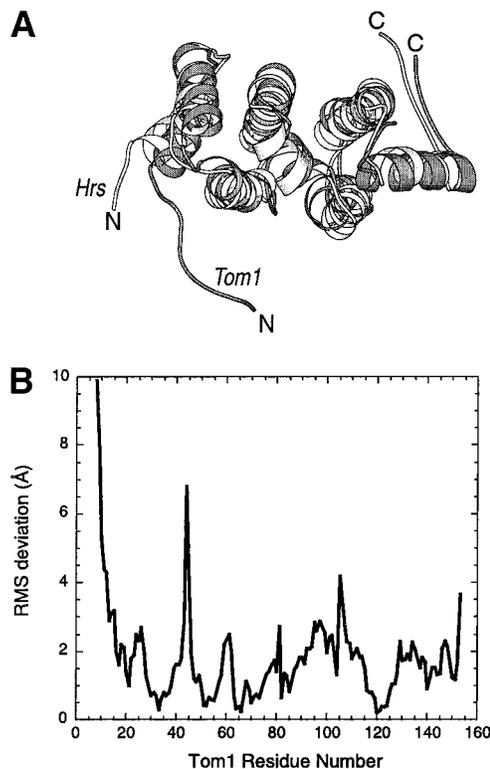


FIGURE 3: Structural comparisons between Tom1 VHS and Hrs VHS domains. (A) Superimposed structures of Tom1 VHS (dark gray) and Hrs VHS (white). Structures were drawn using Molscript (58). (B) Root-mean-square deviation between Tom1 and the corresponding Hrs residues.

A set of acidic residues that stretch across helix 2, the neighboring loops, and part of helix 3 give rise to a negatively charged surface on the opposite side of the domain. These include two residues whose counterparts in Hrs interact with FYVE domains (see below). These acidic residues include aspartic acids 22, 29, 38, and 40 and glutamic acids 28, 35, 42, 44, and 45. All of these residues are conserved in at least some of the other known VHS domains. Only one basic residue, Lys48, is near this series of acidic residues. The result is a negatively charged “ridge” on one side of the domain (Figure 2A).

**Comparison between VHS Domains from Tom1 and Hrs.** The sequences of the Tom1 and Hrs VHS domains are 33% identical. The VHS domain structures align structurally with an rms deviation of 1.54 Å over 127 residues that align closely, and 2.20 Å overall. The biggest differences are found at the N- and C-terminal loops; the N-terminal loop of Tom1 VHS is seven residues longer (Figure 3). Helices 6 and 8 align less closely than helices 1–5 and 7. Four residues on helix 2 of the Hrs VHS domain (Pro21, Trp23, Pro24, and Asp31) make contact with FYVE domains from the same molecule and from the partner molecule of a homodimer (7). They are located on the same side of helix 2 and point away from the VHS domain. Trp23 and Asp31 are conserved in Tom1 VHS. As in the Tom1 VHS domain, these residues are located on a negatively charged ridge (Figure 2B). The positively charged pocket due to basic residues on helix 3 is, however, also present on the Hrs surface (Figure 2B).

**Comparison between Tom1 VHS and the ENTH Domain of Epsin-1.** Like Tom1 VHS, the ENTH domain of epsin-1 consists of eight helices (8). The first seven helices of epsin-1

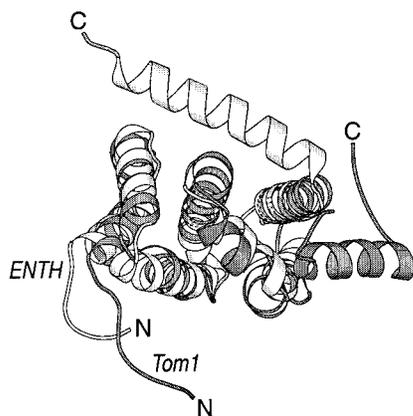


FIGURE 4: Superimposed structures of Tom1 VHS (dark gray) and epsin-1 ENTH (white). Structures were drawn using Molscript (58).

ENTH overlap with those of Tom1 VHS (rms deviation of 1.38 Å over 97 closely aligned residues, and 2.67 Å overall). The eight helices are completely different (Figure 4). The eighth helix of the epsin-1 ENTH domain lies perpendicularly across helices 2 and 4 and the helix 4–helix 5 interhelical loop. Other than helix 8, the regions of greatest divergence between the two structures are the N-terminus and the helix 6–helix 7 intrahelical loop. The latter is six residues longer in the ENTH structure and extends outward from the domain.

On the basis of structural similarity, 18 of some 33 residues conserved or conservatively substituted among ENTH domains (8) are also conserved in the VHS domains (Figure 5). Most of these are hydrophobic residues which make up the cores of the domains, but several acidic and basic residues on helices 3, 4, and 7 are also conserved. This leads to a surface charge distribution on the ENTH domain that is similar to that of the VHS domains in several respects (Figure 2C). Basic residues on helices 1 and 3 contribute to a positive surface patch similar to the positive VHS surface discussed above. Another positive surface patch unique to the ENTH domain is located between helices 6 and 7, which would be covered by helix 8 in the VHS domains. On the other side of the ENTH domain, a positive patch is present between helices 7 and 4. Acidic residues on helices 1–3 form a negatively charged face that is a smaller counterpart to the negatively charged ridge of the VHS domains. These residues include Asp45, which is the equivalent of Hrs Asp31, one of the Hrs residues that contacts the FYVE domains. The side of the ENTH domain formed by helices 2, 4, 5, and 7 is covered by helix 8 and presents a convex rather than slightly concave surface, unlike the VHS domains. The main electrostatic features of the VHS domains, i.e., a positively charged patch on one side and a negatively charged ridge on the other, are preserved in the ENTH domain. The major differences between the respective surfaces are due to the different conformations of helix 8 in the two structures.

*Similarity between Tom1 VHS and Other Superhelical Domains.* Tom1 VHS, like the VHS domain of Hrs and the ENTH domain of epsin, exhibits similarity to other, larger, superhelical proteins. For example, helices 3–7 superimpose with an rms deviation of 3.5 Å with helices in HEAT repeats 10–12 of karyopherin  $\beta$  (40). Other structures with appreciable similarity to VHS were found using DALI (41). These include the N-terminal domain of the son-of-sevenless

protein (42),  $\beta$ -catenin (43), the N-terminal domain of chondroitinase (44), and the  $\alpha$ -subunit of protein farnesyltransferase (45). In all these proteins, except karyopherin, the corresponding superhelical region engages in intra- or inter-protein contacts. In  $\beta$ -catenin and farnesyltransferase, the apposition of repeating helical motifs results in large, curved modules with concave and convex sides. In both cases, the concave side is the one that engages in heteromeric interactions. The N-terminal helical domain of son-of-sevenless, however, packs against the rest of the protein through a slightly convex surface. Although VHS domains are much smaller than the  $\beta$ -catenin and farnesyltransferase superhelices, they also demonstrate a slight curvature. It is the slightly concave side (including helix 2) that contacts other domains or proteins (7).

## DISCUSSION

*A Conserved Protein–Protein Interaction Surface.* The Tom1 VHS, Hrs VHS, and ENTH domains share a similar tertiary structure. Of particular interest is the portion of the VHS domain surface that interacts with FYVE domains in the tandem VHS–FYVE structure of Hrs (7). In Hrs, the VHS domain interacts with the FYVE domain from the same molecule as well as a linker segment between the FYVE and VHS domains. The VHS domain also interacts with the FYVE domain from a second, partner molecule in a homodimer. These interactions cluster on one side of the VHS domain so that the two FYVE domains interact with each other in an antiparallel manner (Figure 6A). The two FYVE domains bury 850 Å<sup>2</sup> of accessible surface area on the Hrs VHS domain. Residues on helices 2 and 4 and the helix 4–helix 5 interhelical loop form the primary points of interaction between the VHS domain and the two FYVE domains. Several residues that interact with the FYVE domains are conserved in Tom1 VHS, even though the full-length Tom1 protein does not contain a FYVE domain.

Trp30 in Tom1 is the counterpart of Hrs Trp23 (Figure 6B). This Trp is also conserved in the EAST and STAM proteins and the GGAs. Hrs Trp23 and Pro21 interact with residues located on the loop between  $\beta$ -sheets 3 and 4 of the intramolecular FYVE domain (7). Also on helix 2, Tom1 Asp38 is the counterpart of Hrs Asp31. All known VHS domains have an acidic residue at the corresponding position. Residues at two positions make contacts with the symmetry-related FYVE domain in the Hrs structure. The carbonyl oxygen of Hrs Pro24 (the counterpart of Tom1 Ala31) interacts with the side chain of Thr174 through a water molecule, while the Hrs Asp31 side chain interacts directly with the backbone amides of Thr174 and Phe173. Several residues on helix 4 of Tom1 are also conserved among VHS domains. These include Tom1 Lys79 and Asn80, which is conserved in all known VHS domains except the GGAs. The Hrs counterparts of both residues interact with the inter-domain loop located between the VHS and FYVE domains (Figure 6B). Specifically, the side chain and amide group of Asp153 make two hydrogen bonds to the side chain of Asn72, while the carbonyl oxygen of Phe150 makes a hydrogen bond to the carbonyl group of Lys71 via a water molecule. The side chain of Phe150 itself protrudes into a cleft between helices 4 and 7 which is lined with hydrophobic residues. The conservation of the VHS residues described

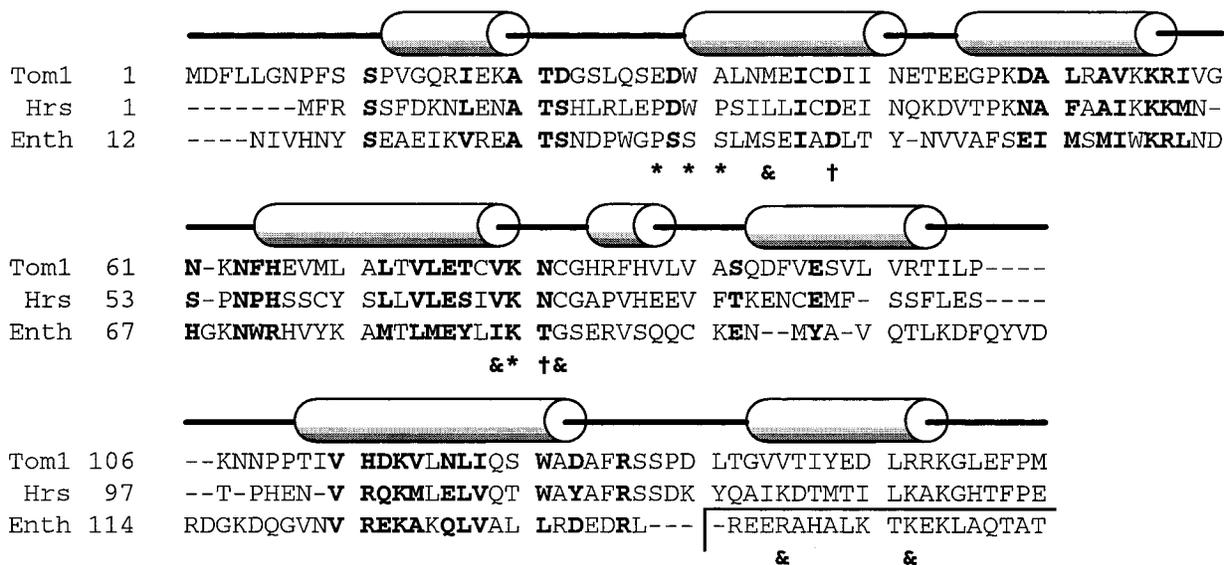


FIGURE 5: Structure-based sequence alignment between Tom1 VHS, Hrs VHS, and epsin-1 ENTH. The secondary structure that is shown is for Tom1 VHS. Conserved and conservatively substituted residues are shown in boldface. The symbols below the alignment designate residues that participate in protein-protein contacts as follows: (\*) in VHS domains, (&) in ENTH domains, and (†) in VHS and ENTH domains. The eighth helix of the ENTH domain is boxed off, to emphasize that it does not overlap spatially with the eighth helices of the VHS domains.

above suggests that they may interact with other domains or with an interdomain linker in Tom1 as well as Hrs.

Several VHS residues that contact the FYVE domains are also conserved in the epsin-1 ENTH domain, despite the limited overall sequence similarity between the two domains. From a superposition of the ENTH and Hrs VHS domains, it is evident that helix 8 of the ENTH domain is spatially equivalent to the interdomain linker of Hrs (Figure 6C). Residues on ENTH helices 2 and 4 interact with helix 8 rather than with another domain or with an interdomain linker (Figure 6A). Epsin Asp45, which is equivalent to Tom1 Asp38, makes direct hydrogen bonds with the Lys151 side chain. Asp45 and Lys151 are conserved or conservatively substituted in almost all other ENTH domains. On helix 4, Thr87 is the epsin counterpart of Tom1 Asn80. Thr87 interacts via water molecules with Lys151 and Ala147 (Figure 6C). The side chain of epsin Arg144 interacts with backbone atoms of two other residues on helix 4, Ile85 and Gly88. These interactions form a spatial (but not sequence) counterpart to the interaction involving Lys71 and Phe150 in Hrs. In fact, the interactions in both domains cluster into three regions that are roughly spatially equivalent (compare panels B and C of Figure 6). The fact that the interactions between VHS and FYVE domains have counterparts in the ENTH domain, coupled with the conservation of several of the residues that participate in these interactions, emphasizes that helices 2 and 4 of VHS and ENTH domains constitute a conserved protein-protein interaction surface.

The interactions define surface "hot spots" that may be used by VHS domains for making protein-protein contacts (Figure 7) with domains other than FYVE domains. Other domains that are found in VHS domain-containing proteins include SH3 domains, present in STAM, EAST, and *S. cerevisiae* YHL002W (9). Another such domain is the Arf-interacting GAT domain found in the GGAs (23). Tom1 includes an 85-residue sequence homologous to a part of GAT domains. The domains listed above are all relatively small ( $\leq 160$  residues). SH3 domains, especially, are pri-

marily  $\beta$ -stranded and have interstrand loops that are topologically similar to the FYVE domain loops that interact with the VHS domain (7, 46).

In contrast to the VHS domains, the surface of the ENTH domain includes helix 8, which covers the interaction hot spots shown in Figure 7. Helix 8 buries 830  $\text{\AA}^2$  of accessible surface area on helices 1-7, almost as much area as is buried by the two FYVE domains on the VHS domain of Hrs. The contrasting positions of helix 8 in VHS and ENTH domains raise the possibility that this helix can adopt more than one conformation in these domains. Even though they do not correspond spatially, the eighth helices of the ENTH and VHS domains share some sequence similarity. Two positively charged residues on ENTH helix 8 (epsin-1 residues Arg144 and Lys151) that interact with helices 2 and 4 (Figure 6C) have counterparts in helix 8 of Hrs with a similar spacing (Hrs residues Lys127 and Lys134). The latter is a basic residue in most VHS domains. Under the proper regulatory circumstances, helix 8 of Hrs could displace the interdomain linker and adopt an alternate conformation, similar to that of ENTH helix 8. This would provide a method for "decoupling" FYVE and VHS domains. It is clear that the mere absence of the FYVE domain is not sufficient to cause helix 8 of the VHS domains to adopt an ENTH-like conformation, since helix 8 has the same conformation in the Tom1 VHS domain as in the Hrs VHS-FYVE tandem structure. On the other hand, helix 8 of the ENTH domains could also be displaced and adopt a conformation similar to that of helix 8 in the VHS domains. The epsin-1 ENTH domain has been shown to interact with the promyelocytic leukemia  $\text{Zn}^{2+}$  finger protein, a transcription factor (8). This interaction could occur through the solvent-exposed side of helix 8, which is lined with several charged and polar residues, or through another part of the ENTH domain surface that has no analogue in the VHS domains. However, if ENTH helix 8 is displaced, helices 2 and 4 would be exposed and could also be utilized for this protein-protein interaction.

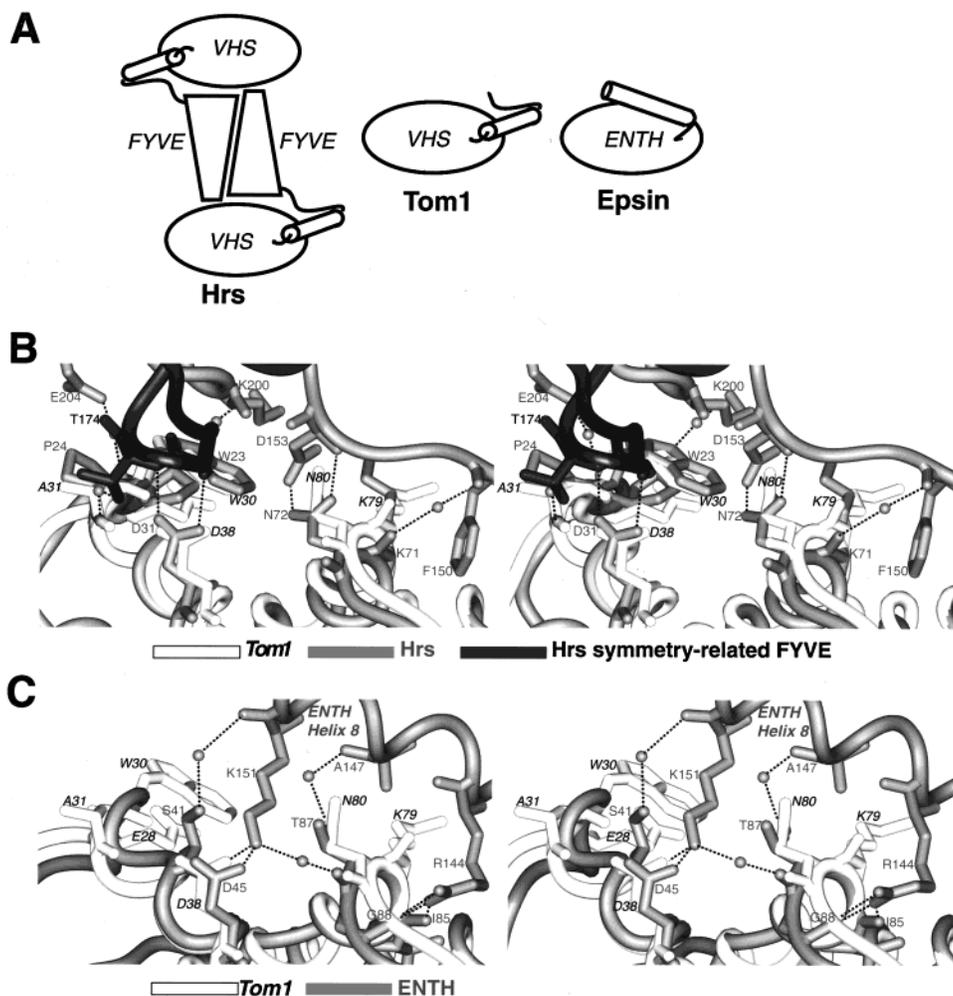


FIGURE 6: (A) Schematic of the protein interface site in the Hrs VHS–FYVE heterodimer, Tom1 VHS, and the epsin-1 ENTH domain. Each FYVE domain makes intra- and intermolecular contacts with the associated VHS domains. Helix 8 is shown explicitly. (B) Wall-eye stereoview of interactions between residues on the Hrs VHS domain and FYVE domains, and their Tom1 counterparts. Tom1 VHS residues and the backbone are shown in white. Tom1 VHS residues are labeled in black italics. The Hrs backbone and side chains are colored gray. Hrs residues are labeled with gray text. The backbone and side chains of the symmetry-related FYVE domain are colored black. Residues of the symmetry-related FYVE domain are labeled in plain black text. Hrs water molecules are shown as gray spheres. (C) Interactions between helices 2, 4, 5, and 8 of the ENTH domain, and their counterparts in Tom1. ENTH backbone and side chains are colored gray. ENTH residues are labeled with gray text. Water molecules in the ENTH structure are shown as gray spheres. The Tom1 backbone and residues are colored white, and residues are labeled with black italicized text, as in panel B. Dashed lines represent putative hydrogen bonds (3.2 Å cutoff). Note the overlap between ENTH helix 8 and the interdomain linker and the FYVE domain of Hrs. Rendered using SPOCK (59).

The interfaces observed with the protein contact site are notable for the relatively small number of direct polar interactions (Figure 6B,C). There is just one salt bridge in the ENTH domain and just one direct side chain–side chain polar hydrogen bond in the Hrs VHS–FYVE interface. All other side chain–side chain interactions are made through intervening water molecules. The remaining interactions are between side chain and backbone atoms or involve only backbone atoms. The apparent low specificity of these interactions lends support to the possibility that the interactions observed with helices 2 and 4 of the VHS and ENTH domains are breakable and could therefore be used to interact reversibly with a variety of intra- as well as interprotein binding partners.

*Extended N-Terminal Sequence and NPF Motif of Tom1 VHS.* The N-terminal loop of Tom1 VHS contains the sequence Asn-Pro-Phe (NPF). In certain components of the endocytic machinery such as intersectin (47), the epsins

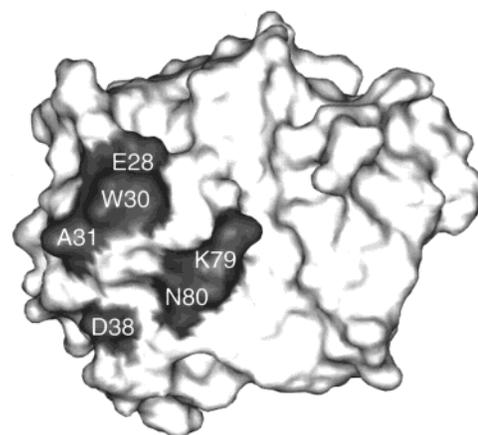


FIGURE 7: Surface of the Tom1 VHS domain, colored to emphasize potential points of contact with other domains or proteins. The molecule is in the same orientation as in Figure 2A. Rendered using SPOCK (59).

(30), and synpatojanin-1 (48), the NPF sequence acts as a binding site for Eps15 homology (EH) domains (49). The VHS domains of the Tom1, EAST, and STAM proteins also contain NPF motifs at their N-termini. We attempted to model the entire VHS domain or the N-terminal loop from Tom1 into the defined NPF binding sites of the EH1 or EH2 domains of Eps15 (50, 51). It was not possible to avoid steric clashes between the domains when docking the N-terminal loop (data not shown). In addition, we performed GST pull-down assays between the Tom1 VHS domain and GST-EH1 or GST-EH2 fusions (data not shown). The assays revealed no binding between these domains. EH domains demonstrate a preference for binding peptides that contain Ser and/or Thr residues immediately preceding the NPF motif (47, 52); these are not present in the Tom1 sequence. We conclude that the NPF motif of Tom1-VHS does not bind EH domains.

*Do VHS Domains Bind Membranes?* The conservation of a basic surface in the region surrounding helix 3 is the only striking similarity between the otherwise divergent molecular surfaces of the VHS and ENTH domains. Conserved basic surfaces are characteristic of membrane targeting domains (53). The positively charged surface can come into apposition with the negatively charged membrane surface, while a hydrophobic protrusion serves as a second anchor by penetrating into the membrane interior. In membrane binding domains such as the pleckstrin homology (PH), C2, and FYVE domains, the positively charged surface contains a stereospecific binding pocket for negatively charged membrane lipids. In the PH and FYVE domains, arginine and lysine residues extend outward from the domain and make electrostatic interactions with the negatively charged head-groups of phosphoinositide lipids (54–56). The surface of the Tom1 VHS domain (Figure 2A) is characterized by a positively charged patch. Arg52, Lys55, and Lys56 in this pocket protrude outward from the domain, as does the nearby Lys48, which is also well-conserved among VHS domains. Such a positioning of the VHS at the membrane surface leaves the negatively charged surface on the other side of the domain exposed and away from the membrane. This portion of the surface would be free to interact with other proteins or domains. Several VHS and ENTH domain-containing proteins are known to localize to intracellular vesicles or to the surfaces of organelles. Given the lack of any other known common functional feature in this class of proteins, the most parsimonious explanation for the conserved basic surface is that this region is involved in membrane binding.

In the structure of the Hrs FYVE domain (7), the relative positions of the VHS and FYVE domains would not allow the positive patch on the VHS domain and the phosphoinositide binding pocket of the FYVE domain to lie against the membrane surface simultaneously. As discussed above, however, the interactions between the FYVE and VHS domains appear to be potentially breakable. The flexibility of the interdomain linker could then allow both domains to come into proper contact with the membrane surface. This could be a way to regulate the membrane binding affinity of proteins containing tandem VHS and FYVE domains.

Intracellular localization studies suggest that VHS domain-containing proteins are typically localized to cellular membranes. These proteins contain other domains that can

function as membrane localization domains in two ways. They may bind membrane lipids directly, as FYVE domains do. Alternatively, they may bind other proteins that are themselves localized to the membrane. This appears to be the role of the ITAM domains of EAST and STAM. A construct of the EAST SH3 and ITAM domains localizes to perinuclear vesicles in the absence of the VHS domain (29). The full-length EAST protein, or the VHS domain by itself, localizes to the plasma membrane. The VHS domain of the GGAs is diffusely localized to intracellular membranes when expressed by itself (25). In contrast, specific localization of the GGAs to the trans-Golgi network is mediated by their GAT, or ARF-binding, domains (23). The VHS domain may function as a low-efficiency membrane-binding domain. As has been reported for the PH domains of dynamin (57), oligomerization of VHS domains or VHS domain-containing proteins might be necessary to drive membrane binding. A preference for binding specific intracellular membranes may arise from other domains on the same protein. Alternately, both VHS and other membrane localization domains on the same protein may have specific targets. High-affinity binding of the protein may only occur at membranes that contain all the required ligand targets, leading to high selectivity in intracellular localization.

#### ACKNOWLEDGMENT

We gratefully acknowledge the assistance of Zbigniew Dauter and K. R. Rajashankar at beamline X9B, NSLS. Thanks to Jonathan Y. S. Ho, Matthew Pearson, and Yosuke Tsujishita for help with data collection. We also thank Beverly Wendland, Scott Emr, and Juan Bonifacino for insightful discussions.

#### REFERENCES

- Owen, D. J., Vallis, Y., Noble, M. E., Hunter, J. B., Dafforn, T. R., Evans, P. R., and McMahon, H. T. (1999) *Cell* 97, 805–815.
- Marsh, M., and McMahon, H. T. (1999) *Nature* 285, 215–220.
- Sutton, R. B., Fasshauer, D., Jahn, R., and Brünger, A. T. (1998) *Nature* 395, 347–353.
- Rice, L. M., and Brünger, A. T. (1999) *Mol. Cell* 4, 85–95.
- Ostermeier, C., and Brünger, A. T. (1998) *Cell* 96, 363–374.
- Misra, S., and Hurley, J. H. (1999) *Cell* 97, 657–666.
- Mao, Y., Nickitenko, A., Duan, X., Lloyd, T. E., Wu, M. N., Bellen, H., and Quioco, F. A. (2000) *Cell* 100, 447–456.
- Hyman, J., Chen, H., Di Fiore, P. P., De Camilli, P., and Brünger, A. T. (2000) *J. Cell Biol.* 149, 537–546.
- Lohi, O., and Lehto, V.-P. (1998) *FEBS Lett.* 440, 255–257.
- Burd, C. G., and Emr, S. D. (1998) *Mol. Cell* 2, 157–162.
- Gaullier, J. M., Simonsen, A., D'Arrigo, A., Bremnes, B., Stenmark, H., and Aasland, R. (1998) *Nature* 394, 432–433.
- Patki, V., Lawe, D. C., Corvera, S., Virbasius, J. V., and Chawla, A. (1998) *Nature* 394, 433–434.
- Wurmser, A. E., Gary, J. D., and Emr, S. D. (1999) *J. Biol. Chem.* 274, 9129–9132.
- Bean, A. J., Seifert, R., Chen, Y. A., Sacks, R., and Scheller, R. H. (1997) *Nature* 385, 826–829.
- Lohi, O., and Lehto, V.-P. (1998) *FEBS Lett.* 436, 419–423.
- Benmerah, A., Lamaze, C., Begue, B., Schmid, S. L., Dautry-Varsat, A., and Cerf-Bensussan, N. (1998) *J. Cell Biol.* 140, 1055–1062.
- Benmerah, A., Bayrou, M., Cerf-Bensussan, N., and Dautry-Varsat, A. (1999) *J. Cell Sci.* 112, 1303–1311.
- Benmerah, A., Poupon, V., Cerf-Bensussan, N., and Dautry-Varsat, A. (2000) *J. Biol. Chem.* 275, 3288–3295.

19. Takeshita, T., Arita, T., Higuchi, M., Asao, H., Endo, K., Kuroda, H., Tanaka, N., Murata, K., Ishii, N., and Sugamura, K. (1997) *Immunity* 6, 449–457.
20. Asao, H., Sasaki, Y., Arita, T., Tanaka, N., Endo, K., Kasai, H., Takeshita, T., Endo, Y., Fujita, T., and Sugamura, K. (1997) *J. Biol. Chem.* 272, 32785–32791.
21. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadmodar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J. M. (2000) *Nature* 403, 623–627.
22. Boman, A. L., Zhang, C. j., Zhu, X., and Kahn, R. A. (2000) *Mol. Biol. Cell* 11, 1241–1255.
23. Dell'Angelica, E. C., Puertollano, R., Mullins, C., Aguilar, R. C., Vargas, J. D., Hartnell, L. M., and Bonifacino, J. S. (2000) *J. Cell Biol.* 149, 81–93.
24. Hirst, J., Lui, W. W. Y., Bright, N. A., Totty, N., Seaman, M. N. J., and Robinson, M. S. (2000) *J. Cell Biol.* 149, 67–79.
25. Poussu, A., Lohi, O., and Lehto, V.-P. (2000) *J. Biol. Chem.* 276, 7176–7183.
26. Page, L. J., Sowerby, P. J., Liu, W. W. Y., and Robinson, M. S. (1999) *J. Cell Biol.* 146, 993–1004.
27. Burk, O., and Klempnauer, K.-H. (1999) *Biochim. Biophys. Acta* 1446, 243–252.
28. Oh, I.-H., and Reddy, E. P. (1999) *Oncogene* 18, 3017–3033.
29. Lohi, O., Poussu, A., Meriläinen, J., Kellokumpu, S., Wasenius, V.-M., and Lehto, V.-P. (1998) *J. Biol. Chem.* 273, 21408–21415.
30. Rosenthal, J. A., Chen, H., Slepnev, V. I., Pellegrini, L., Salcini, A. E., Di Fiore, P. P., and De Camilli, P. (1999) *J. Biol. Chem.* 274, 33959–33965.
31. Wendland, B., Steece, K. E., and Emr, S. D. (1999) *EMBO J.* 18, 4383–4393.
32. Sheffield, P., Garrard, S., and Derewenda, Z. (1999) *Protein Expression Purif.* 15, 34–39.
33. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* 276, 307–326.
34. 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., Pannu, N. S., et al. (1998) *Acta Crystallogr.* D54, 905–921.
35. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr.* A47, 110–119.
36. Engh, R., and Huber, R. (1991) *Acta Crystallogr.* A47, 392–400.
37. Tronrud, D. E. (1996) *J. Appl. Crystallogr.* 29, 100–104.
38. Brünger, A. T. (1992) *Nature* 355, 472–474.
39. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* 24, 946–950.
40. Chook, Y. M., and Blobel, G. (1999) *Nature* 399, 230–237.
41. Holm, L., and Sander, C. (1998) *Nucleic Acids Res.* 26, 316–319.
42. Boriack-Sjodin, P. A., Margarit, S. M., Bar-Sagi, D., and Kuriyan, J. (1998) *Nature* 394, 337–343.
43. Huber, A. H., Nelson, W. J., and Weis, W. I. (1997) *Cell* 90, 871–882.
44. Féthiere, J., Eggimann, B., and Cygler, M. (1999) *J. Mol. Biol.* 288, 635–647.
45. Park, H.-W., Boduluri, S. R., Moomaw, J. F., Casey, P. J., and Beese, L. S. (1997) *Science* 275, 1800–1805.
46. Dalgarno, D. C., Botfield, M. C., and Rickles, R. J. (1997) *Biopolymers* 43, 383–400.
47. Yamabhai, M., Hoffman, N. G., Hardison, N. L., McPherson, P. S., Castagnoli, L., Cesareni, G., and Kay, B. K. (1998) *J. Biol. Chem.* 273, 31401–31407.
48. Haffner, C., Takei, K., Chen, H., Ringstad, N., Hudson, A., Butler, M. H., Salcini, A. E., Di Fiore, P. P., and De Camilli, P. (1997) *FEBS Lett.* 419, 175–180.
49. Salcini, A. E., Confalonieri, S., Doria, M., Santolini, E., Tassi, E., Minenkova, O., Cesareni, G., Pelicci, P. G., and Di Fiore, P. P. (1997) *Genes Dev.* 11, 2239–2249.
50. de Beer, T., Carter, R. E., Loebel-Rice, K. E., Sorkin, A., and Overduin, M. (1998) *Science* 281, 1357–1360.
51. Whitehead, B., Tessari, M., Carotenuto, A., van Bergen en Henegouwen, P. M. P., and Vuister, G. W. (1999) *Biochemistry* 38, 11271–11277.
52. Paoluzi, S., Castagnoli, L., Lauro, I., Salcini, A. E., Coda, L., Fre, S., Confalonieri, S., Pelicci, P. G., Di Fiore, P. P., and Cesareni, G. (1998) *EMBO J.* 17, 6541–6550.
53. Hurley, J. H., and Misra, S. (2000) *Annu. Rev. Biophys. Biomol. Struct.* 29, 49–79.
54. Ferguson, K. M., Lemmon, M. A., Schlessinger, J., and Sigler, P. B. (1995) *Cell* 83, 1037–1046.
55. Baraldi, E., Djinovic Carugo, K., Hyvönen, M., Lo Surdo, P., Riley, A. M., Potter, B. V. L., O'Brien, R., Ladbury, J. E., and Saraste, M. (1999) *Structure* 7, 449–460.
56. Kutateladze, T. G., Ogburn, K. G., Watson, W. T., de Beer, T., Emr, S. D., Burd, C. G., and Overduin, M. (1999) *Mol. Cell* 3, 605–611.
57. Klein, D. E., Lee, A., Frank, D. W., Marks, M. S., and Lemmon, M. A. (1998) *J. Biol. Chem.* 273, 27725–27733.
58. Kraulis, P. (1991) *J. Appl. Crystallogr.* 24, 946–950.
59. Christopher, J. A. (1997) *SPOCK*, <http://quorum.tamu.edu/jon/spock>.