THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 270, No. 22, Issue of June 2, pp. 13503–13511, 1995
Printed in U.S.A.
© 1995 by The American Society for Biochemistry and Molecular Biology, Inc.

EEA1, an Early Endosome-Associated Protein

EEA1 IS A CONSERVED α-HELICAL PERIPHERAL MEMBRANE PROTEIN FLANKED BY CYSTEINE "FINGERS" AND CONTAINS A CALMODULIN-BINDING IQ MOTIF*

(Received for publication, January 18, 1995, and in revised form, March 6, 1995)

Fi-Tjen Mu‡, Judy M. Callaghan, Olivia Steele-Mortimer§, Harald Stenmark§, Robert G. Parton§, Paul L. Campbell¶‖, James McCluskey¶, Jing-Ping Yeo, Edward P.C. Tock‡, and Ban-Hock Toh**

From the ‡Department of Pathology, National University of Singapore, Singapore 0511, Singapore, the §European Molecular Biology Laboratory, Heidelberg 69021, Federal Republic of Germany, the ¶Centre for Transfusion Medicine and Immunology, Flinders Medical Centre, Adelaide 5042, South Australia, and the Department of Pathology and Immunology, Monash Medical School, Melbourne 3181, Australia

Early endosomes are cellular compartments receiving endocytosed material and sorting them for vesicular transport to late endosomes and lysosomes or for recycling to the plasma membrane. We have cloned a human cDNA encoding an evolutionarily conserved 180-kDa protein on early endosomes named EEA1 (Early Endosome Antigen1). EEA1 is associated with early endosomes since it co-localizes by immunofluorescence with the transferrin receptor and with Rab5 but not with Rab7. Immunoelectron microscopy shows that it is associated with tubulovesicular early endosomes containing internalized bovine serum albumin-gold. EEA1 is a hydrophilic peripheral membrane protein present in cytosol and membrane fractions. It partitions in the aqueous phase after Triton X-114 solubilization and is extracted from membranes by 0.3 M NaCl. It is a predominantly α-helical protein sharing 17–20% sequence identity with the myosins and contains a calmodulin-binding IQ motif. It is flanked by metal-binding, cysteine "finger" motifs. The COOH-terminal fingers, Cys-X_2-Cys-X_{12}-Cys-X_2-Cys and Cys-X_2-Cys-X_{16}-Cys-X_2-Cys, are present within a region that is strikingly homologous with *Saccharomyces cerevisiae* FAB1 protein required for endocytosis and with *Caenorhabditis elegans* ZK632. These fingers also show limited conservation with *S. cerevisiae* VAC1, Vps11, and Vps18p proteins implicated in vacuolar transport. We propose that EEA1 is required for vesicular transport of proteins through early endosomes and that its finger motifs are required for this activity.

Endocytosis is a process whereby cells internalize extracellular molecules into cytoplasmic vesicles (Courtoy, 1991). In receptor-mediated endocytosis, ligands bind receptors on clathrin-coated pits, which invaginate forming coated vesicles; the vesicles uncoat and transport receptor-ligand complexes to early endosomes (Trowbridge *et al.*, 1993). Early endosomes are a major sorting compartment from which ligands may be released and transported to lysosomes and receptors recycled to the plasma membrane, or from which ligand-receptor complexes are transported to lysosomes, to the opposite side of polarized cells, or recycled to the plasma membrane (Shepherd, 1989). Ligands, dissociated by low endosomal pH, may be confined to the main body of endosomes and follow a nonselective default pathway to late endosomes and lysosomes (Mellman, 1993). Recycling receptors may escape the degradative pathway via tubular extensions and be sorted by specific interaction of their cytoplasmic tails with peripheral membrane proteins for targeting to distinct locations. Two models have been proposed for transport from early to late endosomes. The "vesicle shuttle model" proposes that early and late endosomes are preexisting compartments communicating via microtubule-dependent vesicular traffic (Griffiths and Gruenberg, 1991). The "maturation model" suggests that early endosomes gradually "mature" into late endosomes and lysosomes (Murphy, 1991).

Very little is known of proteins associated with early endosomes. The best characterized compartment-specific proteins are the small GTP-binding rab proteins with Rab5 localizing to early endosomes and Rab7 to late endosomes (Chavrier *et al.*, 1990; Gould, 1992). Here we report the characterization and cloning of human EEA1, a hydrophilic 180-kDa peripheral membrane protein on early endosomes. EEA1 shares 17–20% homology with the myosins and contains a calmodulin-binding IQ motif associated with these proteins. EEA1 is a predominantly α-helical protein flanked by NH_2-terminal and COOH-terminal cysteine-rich, metal-binding "finger" motifs. The COOH-terminal finger motifs are highly conserved with a peptide encoded by the ZK632 cosmid from *Caenorhabditis elegans* and by the FAB1 gene of *Saccharomyces cerevisiae*. The latter has been described as a gene required for the endocytic-vacuolar pathway and nuclear migration. Similar finger motifs, Cys-X_2-Cys-X_{12}-Cys-X_4-Cys and Cys-X_2-Cys-X_{13}-Cys-X_2-Cys-X_4-Cys-X_{38}-Cys-X_2-Cys, are present, respectively, in *S. cerevisiae* VAC1 (Weisman and Wickner, 1992) and Vps18p (Robinson *et al.*, 1991) proteins, both of which are required for vacuolar sorting. We propose that EEA1 has a role in early endosomal transport and that its conserved cysteine-rich, metal-binding fingers have a role in this process.

MATERIALS AND METHODS

Antibodies—The autoimmune serum was from a patient with subacute cutaneous systemic lupus erythematosus identified in the Monash Clinical Immunology Laboratory. Rabbit antibody to EEA1 was raised against a bacterial fusion protein incorporating the carboxyl terminus of the protein (see below). Mouse anti-Rab5 monoclonal antibodies (Chavrier *et al.*, 1990) and affinity-purified rabbit anti-Rab7 antibody were from Marino Zerial (EMBL).

* This study was supported in part by grants from the National University of Singapore, the Mary Whight Lupus Fellowship and the National Health and Medical Research Council of Australia. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) L40157.

‖ Supported by a National Health and Medical Research Council Australian Postdoctoral Fellowship.

** To whom correspondence should be addressed: Dept. of Pathology and Immunology, Monash Medical School, Commercial Rd., Prahran, Victoria 3181, Australia. Tel.: 613-276-2713; Fax: 613-529-6484.

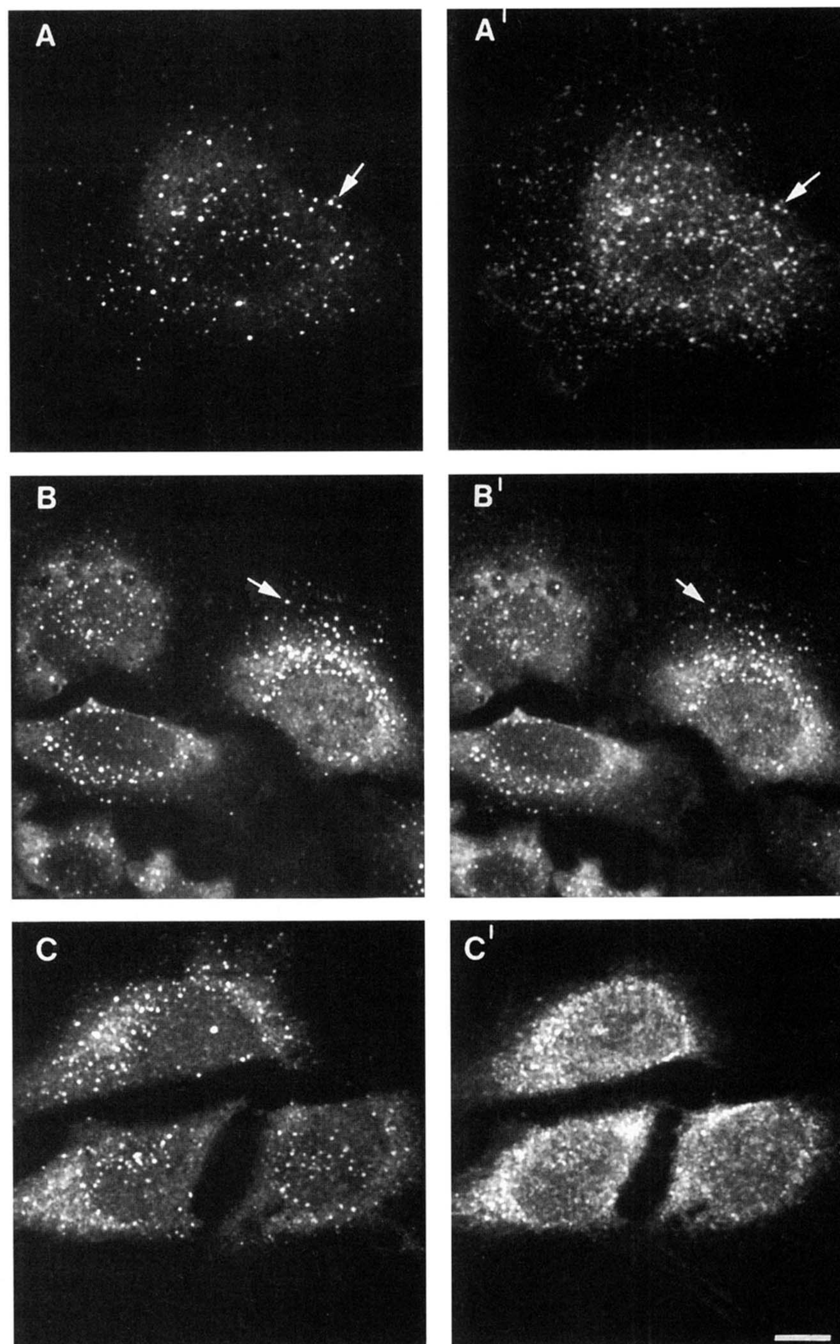

FIG. 1. **Co-localization of EEA1 with transferrin receptor and Rab5 (*arrowheads*) but not with Rab7 using confocal dual immunofluorescence microscopy.** HeLa cells (*panels A* and *A'*) or Rab5-overexpressing HeLa cell clone 2–1-6–1 (*panels B, B', C,* and *C'*) were fixed and permeabilized (see "Materials and Methods"). Cells were reacted with affinity-purified human anti-EEA1 antibody (*A–C*) and either with mouse monoclonal antibody to transferrin receptor (*A'*) or Rab5 (*B'*), or with affinity-purified rabbit anti-Rab7 antibody (*C'*). Bound antibody was traced with fluorescein isothiocyanate-labeled anti-human Ig (*A–C*) or with rhodamine-labeled anti-mouse (*A'* and *B'*) or anti-rabbit (*C'*) Ig.

Immunofluorescence—HeLa and 3T3 cells were fixed in 2–2.5% paraformaldehyde/PBS[1] for 10–20 min and permeabilized with 0.5% saponin (Sigma)/PBS for 10 min or 1% Nonidet P-40/PBS for 5 min. Fixed cells or Hep-2 cells (Kallestad), were reacted with the autoimmune serum diluted 1:200 in PBS, 0.5% saponin or with affinity-purified human or rabbit antibodies to EEA1. Cells were washed 3 times with PBS, 0.5% saponin and incubated with fluorescein isothiocyanate-conjugated sheep anti-human Ig (Wellcome).

Dual Immunofluorescence Confocal Microscopy—HeLa cells were fixed with 3% paraformaldehyde/PBS for 15 min; free aldehyde groups were quenched with 50 mM NH_4Cl/PBS; and cells were permeabilized with 0.1% Triton X-100/PBS for 4 min. The cells were reacted with anti-transferrin receptor antibody (mouse monoclonal B3–25, Boehringer Mannheim, Germany) at 1:100 dilution and the affinity-purified human autoimmune serum. After three washes in PBS, anti-transferrin receptor was visualized with rhodamine-labeled donkey anti-mouse antibody (DiaNova, Germany) at a 1:100 dilution, and EEA1 was visualized with fluorescein isothiocyanate-labeled goat anti-human Ig. Dual immunofluorescence confocal microscopy was also carried out with a Rab5-overexpressing HeLa clone, 2–1-6–1 (Stenmark *et al.*, 1994a, 1994b). Cells were reacted with mouse anti-Rab5 monoclonal antibody (1:20 dilution in 5% bovine serum), affinity-purified rabbit anti-Rab7 antibody (1:8 dilution), or the affinity-purified autoimmune serum. The cells were incubated with fluorescein isothiocyanate-labeled goat anti-human Ig, rhodamine-labeled donkey anti-mouse Ig (DiaNova), or anti-rabbit Ig (DiaNova).

Immunoelectron Microscopy—Endocytic compartments of HeLa cells were labeled with endocytic tracers (5 nm of BSA-gold for the early endosomes and 16 nm of BSA-gold for late endosomes and lysosomes) as described previously (Chavrier *et al.*, 1990) except that 5 nm of BSA-gold (A_{520} = 30) was internalized for 10 min at 37 °C. The cells were then fixed with 8% paraformaldehyde in 250 mM HEPES (pH 7.35), processed for cryosectioning, and sectioned, and the thawed sections were labeled with affinity-purified anti-EEA1 antibody followed by

[1] The abbreviations used are: PBS, phosphate-buffered saline; BSA, bovine serum albumin; PAGE, polyacrylamide gel electrophoresis; bp, base pair(s).

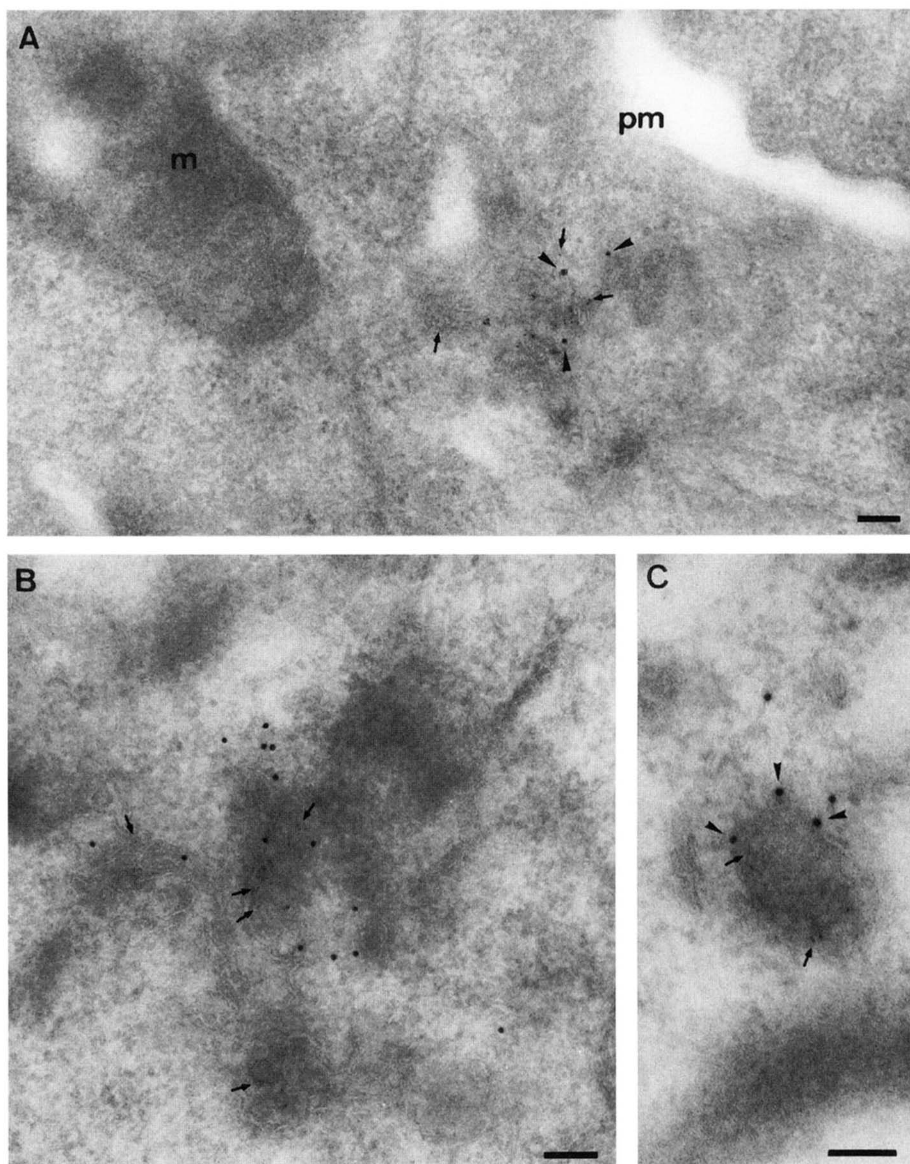

FIG. 2. **Immunoelectron microscopic localization of EEA1 in HeLa cells.** HeLa cells were incubated with 5 nm of BSA-gold to label early endosomes and 16 nm of BSA-gold to label late endosomes and lysosomes (see "Materials and Methods"). After fixation, thawed cryosections were labeled with affinity-purified antibody to EEA1 followed by 10 nm of protein A-gold. *Panel A* shows a representative image of the cell periphery. Labeling for EEA1 (10 nm of gold, *arrowheads*) co-localizes with the internalized 5 nm BSA-gold (*small arrows*). Note that there is negligible labeling for EEA1 in the cytoplasm and on mitochondria (*m*) and generally low labeling associated with the plasma membrane (*pm*). *Panels B* and *C* show higher magnification views of early endosomes. The labeling for EEA1 (10 nm of gold) is associated with the tubulovesicular early endosomes (labeled with 5 nm of BSA-gold). In *panel C*, three of the gold particles (*arrowheads*) are clearly associated with the cytoplasmic surface of the early endosome. EEA1 labeling was rarely observed on late endosomes, labeled with 16 nm of gold (results not shown). *Bars*, 100 nm.

protein A-gold (Chavrier *et al.*, 1990; Griffiths *et al.*, 1984). After embedding in methylcellulose/uranyl acetate, the sections were viewed at an accelerating voltage of 80 kV in a Zeiss transmission electron microscope.

Subcellular Fractionation—HeLa or 3T3 cells were washed 3 times with PBS, detached with a rubber policeman, and resuspended in hypotonic buffer (0.01 M NaCl, 0.003 M $MgCl_2$, 0.01 M Tris (pH 8.5) containing phenylmethylsulfonyl fluoride (1 mM), leupeptin (0.5 μg/ml), pepstatin (1.0 μg/ml), and aprotinin (60 μg/ml)) on ice for 10 min. Cells were lysed with a 21-gauge needle and sequentially centrifuged at 600 $\times g$ to pellet nuclei, 10,000 $\times g$ to pellet mitochondria, and 100,000 $\times g$ to pellet membranes. The supernatant after the last centrifugation was kept as the cytosolic fraction. The pellets, resuspended in 50 mM Hepes buffer (pH 7.6) containing 1 mM EDTA and the cytosolic fraction were stored at −70 °C. Protein content of fractions was determined using the BCA protein assay (Pierce).

Immunoblotting—Subcellular fractions of HeLa or 3T3 cells, aqueous and detergent phases of cells partitioned with Triton X-114 (Kooy *et al.*, 1992), membrane fractions extracted with 0.3 M NaCl for 30 min on ice, and affinity-purified bacterial fusion protein were separated by 7.5 or 10% SDS-PAGE under reducing conditions. Proteins were transferred to nitrocellulose and immunoblotted with affinity-purified human and rabbit antibodies to EEA1 as described previously (Yeo *et al.*, 1994). Immunoblots were also carried out with total HeLa cells and chicken fibroblasts under reducing and nonreducing conditions. Total HeLa cell lysates (100 μg) were processed for two-dimensional nonequilibrium pH gel electrophoresis/SDS-PAGE and immunoblotting as described (Mu *et al.*, 1988).

Molecular Cloning—5 $\times$ 10^5 plaque-forming units of λgt11 Hela cell cDNA expression library (Clontech) were screened with the autoimmune serum, diluted 1:1000 in PBS, 3% casein as described previously (Yeo *et al.*, 1994). Positive plaques, identified with ^{125}I-labeled protein A (2 $\times$ 10^5 cpm/ml), were purified by 3 cycles of immunoscreening. cDNA inserts were subcloned into pBluescribe (Stratagene), for double-stranded DNA sequencing by the dideoxy chain termination method. To obtain the full-length clone, a human hepatoma cDNA library (Stratagene) was screened with ^{32}P-labeled DNA probes. Positive plaques were purified and plasmid rescued according to the manufacturer's instructions. Analyses were carried out using the MacVector program (International Biotechnologies, Inc.) or using software provided by the Australian National Genome Information Service. The cDNA sequence was compared with GenBank and EMBL data bases using the FASTA program (Lipman and Pearson, 1985). The translated amino acid sequence was compared with the translated GenBank data base as well as with the Swiss-Prot protein data base. The Swiss-Prot data base was also searched using the pattern library search program of Smith and Smith (1990).

Nucleotide Blot Analysis—For DNA blots, human leucocyte genomic DNA (12 μg) was digested with restriction enzymes, subjected to electrophoresis in 0.6% agarose gel, and transferred to nitrocellulose. Filters were prehybridized for 4 h with 5 $\times$ saline/sodium phosphate/EDTA, 0.5% SDS, 0.5% skim milk, 1% SDS, 10% dextran sulfate, and 0.5% mg/ml herring sperm DNA; hybridized with ^{32}P-labeled probe (specific activity >5 $\times$ 10^8 cpm/μg DNA) overnight at 65 °C; washed with 1 $\times$ SSC, 0.5% SDS at 65 °C for 30–45 min; and autoradiographed. For RNA blots, total RNA isolated from HeLa cells was enriched for poly(A)$^+$ RNA by oligo(dT)-cellulose chromatography. Poly(A)$^+$ RNA

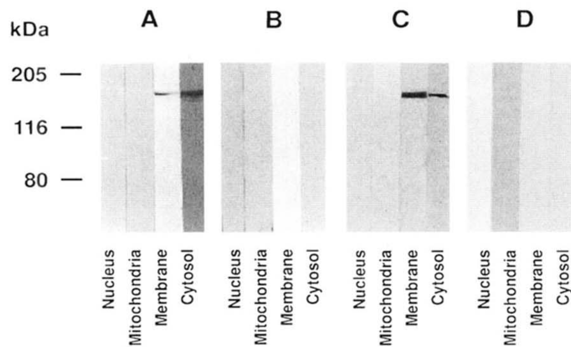

FIG. 3. **Immunoblotting of subcellular fractions of 3T3 cells with affinity-purified human or rabbit antibody to EEA1.** Nuclear, mitochondria, and cytosol fractions (10 μg/lane), or membranes (50 μg/lane) were subjected to 7.5% SDS-PAGE and transferred to nitrocellulose. Membranes were incubated with affinity-purified human (*A*) or rabbit (*C*) anti-EEA1 antibody or with normal human (*B*) or rabbit (*D*) serum. Bound antibody was detected with anti-human or anti-rabbit Ig and enhanced chemiluminescence for nuclear mitochondria, and membrane fractions or 4-choloro-1-napthol/H_2O_2 for cytosol fraction.

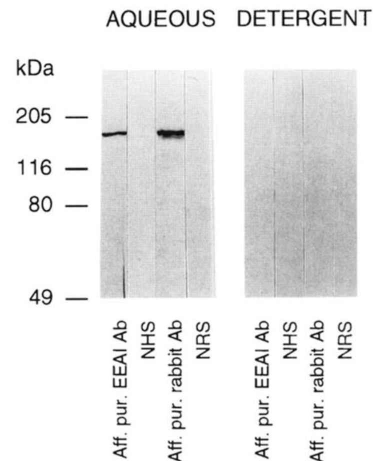

FIG. 4. **Immunoblot of Triton X-114 fractions of 3T3 cells with affinity-purified anti-EEA1 antibodies.** 3T3 cells were extracted with 0.5% Triton X-114. Aqueous and detergent fractions were subjected to 7.5% SDS-PAGE and transferred to nitrocellulose. Membranes were incubated with affinity-purified human (*Aff. pur. EEA1 Ab*) or rabbit (*Aff. pur. rabbit Ab*) anti-EEA1 antibody or with normal human (*NHS*) or rabbit (*NRS*) serum. Bound antibody was detected as described in the legend to Fig. 3.

was fractionated on 1% agarose formaldehyde-denaturing gel, transferred to nitrocellulose, and probed with ^{32}P-labeled DNA as described for DNA blots. Multiple tissue RNA blots (Clontech) were also probed.

Production of Antibody to Recombinant Fusion Protein—pFM1, containing 900 bp of coding sequence encoding the carboxyl terminus of EEA1, was subcloned into pGEX 2T (Pharmacia Biotech Inc.) vector. *Escherichia coli* were transformed with pGEX 2T or pGEX 2T containing pFM1 insert. The bacteria were induced to produce glutathione *S*-transferase or bacterial fusion protein with 1 mM isopropyl-1-thio-β-D-galactopyranoside (Kooy *et al.*, 1992). Bacteria were suspended in PBS, 1% Triton X-100, sonicated, and centrifuged at 10,000 × *g*. Soluble glutathione *S*-transferase or fusion protein in the supernatant was purified by glutathione-agarose chromatography. Rabbits were immunized with 50 μg of purified fusion protein in Freund's complete adjuvent, boosted after 4 weeks with 50 μg of protein in incomplete adjuvent and bled after 2 weeks. Rabbit sera were tested by immunofluorescence and immunoblotting of 3T3 cells.

Affinity Purification of Human and Rabbit Antibodies—1 mg of purified glutathione *S*-transferase or bacterial fusion protein were bound to 1 ml of swollen CNBr-activated Sepharose 4B beads (Pharmacia). Human autoimmune and rabbit immune sera were diluted 1:4 in PBS and recycled 3 times through the fusion protein column. In addition, the rabbit antibody was also recycled 3 times through a glutathione *S*-transferase column. The columns were washed extensively with PBS. The run-through from both columns was collected. Bound antibody, eluted with 3 M KSCN, was dialyzed against PBS and concentrated with a microconcentrator (Amicon).

RESULTS

EEA1 Co-localizes with Transferrin Receptor and with Rab5 but Not with Rab7—The autoimmune serum reacted by immunofluorescence with multiple small vesicles distributed throughout the cytoplasm of human Hep-2 and HeLa cells, mouse 3T3 cells, and chicken fibroblasts (data not shown). These results and the immunoblotting data (see below) indicate that EEA1 is evolutionarily conserved. Affinity-purified human autoantibody to EEA1 and affinity-purified rabbit antibody to a bacterial fusion protein incorporating the carboxyl terminus of EEA1 gave the same immunofluorescence reactions (data not shown). Since the cytoplasmic vesicles reactive with the affinity-purified anti-EEA1 antibodies were reminiscent of endocytic structures, we carried out experiments to identify them using dual immunofluorescence confocal microscopy and endocytic markers. In HeLa cells, EEA1 co-localized extensively with the transferrin receptor (Fig. 1, *A* and *A*′), an established marker of early endosomes (Hopkins, 1983; Schmid *et al.*, 1988). Since the image shown is a 0.5-μm optical section, plasma membrane-associated transferrin receptors were not readily detected. Localization of EEA1 to early endosomes was supported by co-localization with Rab5, a GTPase specifically localized to early endosomes (Chavrier *et al.*, 1990), in HeLa cells moderately overexpressing Rab5 (Fig. 1, *B* and *B*′), and in Madin-Darby canine kidney cells (not shown). In contrast, localization of EEA1 showed little overlap with that of Rab7 (Fig. 1, *C* and *C*′), a GTPase found on late endosomes (Chavrier *et al.*, 1990), indicating that EEA1 is not associated with the whole endocytic pathway.

EEA1 Is Associated with Tubulovesicular Early Endosomes Containing Internalized BSA-Gold—The intracellular distribution of EEA1 was examined at the ultrastructural level by immunolabeling ultrathin sections of HeLa cells, which had internalized endocytic markers. Labeling for EEA1 was predominantly associated with tubulovesicular early endosomes containing BSA-gold internalized as a fluid phase marker for 10 min at 37 °C (Fig. 2). Labeling was rarely associated with late endosomes, Golgi cisternae (results not shown), and the plasma membrane (for example, see Fig. 2*A*). In some cases, significant labeling was observed in close proximity to early endosomes but some distance away from the endosomal membrane (for example, see Fig. 2*C*). Whether this simply reflects labeling of the cytoplasmic surface of peripheral endosomal elements, which are not clearly visible in the plane of the section or of other endosome-associated material, is presently unclear.

EEA1 Is a 180-kDa Peripheral Membrane Protein—The autoimmune serum and affinity-purified antibody to EEA1 immunoblotted a 180-kDa protein in HeLa cells, 3T3 cells, and chicken fibroblasts under both reducing and nonreducing conditions (data not shown). The 180-kDa protein was detected in membrane and cytosolic subcellular fractions of HeLa or 3T3 cells (Fig. 3), exclusively in the aqueous phase of cell extracts after Triton X-114 solubilization (Fig. 4) and was extracted from membranes by 0.3 M NaCl (data not shown). These observations suggest that EEA1 is a peripheral membrane protein. The results are consistent with the prediction that EEA1 is a hydrophilic protein lacking hydrophobic transmembrane domains (see below). Immunoblots with membrane fractions were developed by enhanced chemiluminescence (Fig. 3) as the signal developed by 4-choloro-1-napthol was very weak. These results suggest that either there is a larger pool of EEA1 in the cytosol than those associated with membranes or that EEA1 is easily dissociated from membranes during fractionation. Two-

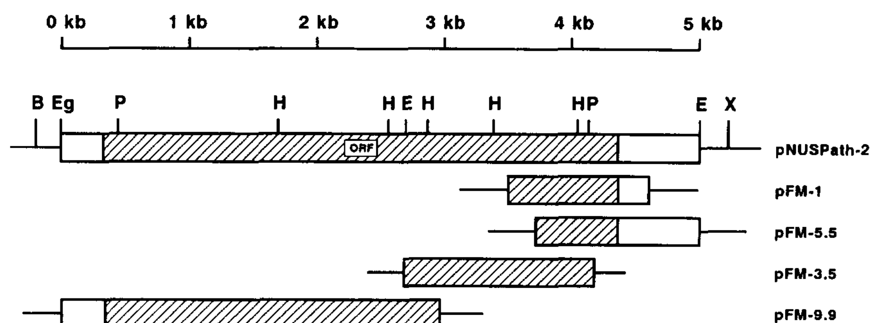

FIG. 5. **Restriction map of full-length and partial length EEA1 cDNA clones.** *B, BamHI; Eg, EagI; P, Pst; H, HindIII; E, ecoR1; X, Xho.*

dimensional immunoblots showed that EEA1 has an acidic pI (data not shown), consistent with the predicted pI of the protein (see below).

Cloning of cDNA Encoding EEA1—A single positive plaque (pFM1) containing a 1.2-kilobase insert was isolated after screening 5×10^5 plaque-forming units of a HeLa cell expression library with the autoimmune serum (Fig. 5). Two positive clones (pFM5.5 and pFM3.5) were obtained after screening a human hepatoma λ zap cDNA library with ^{32}P-labeled pFM1. A further clone (pFM9.9) was obtained from the same library after probing the library with a 500-bp *Hin*dIII fragment of pFM3.5. All four clones are probably derived from the same transcript since all overlapping regions had identical sequences. Four individual human genomic DNAs, digested with *Eco*RI, *Bam*HI, and *Hin*dIII and probed with ^{32}P-labeled pFM1 gave a single fragment after *Eco*RI and *Bam*HI digestion and two fragments after *Hin*dIII digestion (data not shown), consistent with an internal *Hin*dIII site in the cDNA. ^{32}P-labeled pFM1 hybridizes with a 9-kilobase mRNA in a poly(A)$^+$ RNA preparation from HeLa cells (data not shown). A similar sized message was found in a multiple tissue blot of skeletal muscle, heart, brain, lung, liver, and pancreas (data not shown). To exclude concatamer formation during library construction, pFM9.9, the cDNA clone containing the internal *Eco*RI site, was digested with *Eco*RI, and the fragment 5′ of this restriction enzyme site was hybridized to HeLa poly(A)$^+$ RNA or to human multiple tissue blot. The same 9-kilobase message was obtained (data not shown), indicating that the 5′ fragment is a related fragment.

Sequence Analysis of EEA1—The nucleotide sequence of pNUSPath-2 derived from the four overlapping cDNA clones comprises 4962 bp with a coding region of 4233 nucleotides (Fig. 6). The DNA sequence of EEA1 cDNA is virtually identical with that of locus HSP162 (*Homo sapiens* p162), deposited in the GenBank data base on April 26, 1994 by H. P. Seelig (accession X78998). The coding region encodes a protein of 1411 amino acids with a predicted molecular mass of 162.46 kDa and pI of 5.38. Plots of hydrophilicity (Kyte and Doolittle, 1982), surface probability (Janin *et al.*, 1978; Emini *et al.*, 1985), flexibility (Karplus and Schulz, 1985), antigenic index (Jameson and Wolf, 1988), and amphiphilicity (Eisenberg *et al*, 1984a, 1984b; von Heijne, 1986) predict a largely hydrophilic structure with short amphiphilic regions in the 15 amino-terminal amino acids and in segments centered around amino acids 515 and 645. Secondary structure predictions (Chou and Fasman, 1974a, 1974b, 1978a, 1978b; Robson and Suzuki, 1976; Garnier *et al.*, 1978) suggest a predominantly α-helical structure throughout the full length of the protein with the exception of the extreme NH_2- and COOH-terminal segments, which contain a number of proline residues and are predicted to have a high content of turns and β-sheet. It is notable that these regions contain potential metal-binding motifs (see below).

The DNA sequence of the EEA1 cDNA from positions 505 to 4095 shares 52.4% homology with a region of the *Entamoeba histolytica* myosin heavy chain gene (ENHMHCAX locus) extending from position 2921 to 6512. Optimal alignments of translated amino acid sequence corresponding to this region (EEA1 amino acids 154-1322) (Needleman and Wunch, 1970; Smith and Waterman, 1981; Rechid *et al.*, 1989) using a gap weight of 3.0 and a length weight of 0.1 gives an amino acid identity of 24.5% with 46.7% similarity when conservative substitution are accounted for. Indeed, the translated amino acid sequence of EEA1 has a low but statistically significant global homology with a large number of vertebrate and nonvertebrate myosin heavy chains from both muscle and non-muscle, typically showing amino acid identity of 17–20% across the whole protein with an additional large number of conservative substitutions. However, there is no evidence for a globular myosin "head" region in EEA1. Significant overall amino acid identity of 17.9% was also found with the human kinetochore protein CENP-E (Yen *et al.*, 1992). Homology to these proteins is scattered throughout the EEA1 polypeptide, and no large region shows greater identity overall. Statistical evaluation revealed that the homology is unlikely to be due to chance alone, since random shuffling of the myosin or CENP-E sequences generated much lesser degrees of similarity. EEA1 also contains close to its COOH terminus a motif related to the IQ motif and found in the neck regions of all myosins and in the calmodulin binding domains of non-myosin proteins. Alignment of these IQ motifs in EEA1, neuromodulin and in the myosins is shown in Fig. 7. These motifs may be involved in binding of the calmodulin/EF-hand superfamily of proteins, which includes the essential and regulatory light chains of conventional myosins (Cheney and Mooseker, 1992).

A short region of 57 amino acids at the extreme COOH terminus showed 49.1% identity with a segment of the CEZK632 locus in the GenBank data base. The sequence of this locus originated in cosmid ZK632 containing *C. elegans* sequence, which was sequenced as part of the *C. elegans* sequencing project (Sulstan *et al.*, 1992) and has not been attributed with any function. However, the implied product of the FAB1 gene, which is required for endocytic-vacuolar pathway and nuclear migration in *S. cerevisiae* (locus U01017, submitted to GenBank by A. Yamamoto and D. Koshland, 1993) also shows a striking homology with this region of EEA1, despite little homology being present elsewhere in the protein. Alignment of the corresponding regions from EEA1, CEZK632, and FAB1 is shown in Fig. 8. This region spans a potential metal binding finger present in EEA1 as Cys-X_2-Cys-X_{12}-Cys-X_2-Cys and Cys-X_2-Cys-X_{16}-Cys-X_2-Cys. It is notable that the cysteines comprising this element are absolutely conserved across the three implied proteins. EEA1 also contains a canonical zinc-finger motif of the C_2H_2 type (Klug and Rhodes, 1987; Evans and Hollenberg, 1988; Payre and Vincent, 1988; Miller *et al.*, 1985; Berg, 1988) close to the NH_2 terminus (Figs. 6 and 8).

The EEA1 protein also contains multiple consensus sequences for *N*-glycosylation, casein kinase II phosphorylation,

```
GCC GCT CGG CCC GAC CTC CTT CCC GGT CCC CTC CCG GCC CAG TCC CTC CCG CCC GCC CAC CCC GAG TAG TGA GTG GCC CCG CGC AGG GTC 90
TGG AGA GTC ACC GCG GCG CCG GGT GGT GGT TAA ACC ATG TTA AGG AGG ATT TTA CAG AGG ACT CCT GGG AGA GTT GGC TCT CAA GGT TCT 180
                                                met leu arg arg ile leu gln arg thr pro gly arg val gly ser gln gly ser 18
GAT TTA GAT TCA TCA GCA ACT CCT ATA AAC ACA GTG GAC GTC AAT AAT GAA AGC TCT TCA GAA GGT TTC ATA TGT CCC CAG TGT ATG AAA 270
asp leu asp ser ser ala thr pro ile asn thr val asp val asn asn glu ser ser ser glu gly phe ile cys pro gln cys met lys 48
GGA TCT CTT TCT GCT GAT GAA CTT TTC AAA CAT TAT GAA GCT GTT CAT GAT GCT GGT AAT GAC TCA GGT CAT GGA GGA GAG TCT AAT CTT 360
ser leu gly ser ala asp glu leu phe lys his tyr glu ala val his asp ala gly asn asp ser gly his gly gly glu ser asn leu 78
GCT TTG AAG CGA GAT GAT GTA ACA CTG CTC AGA CAA GAG GTC CAA GAC CTA CAG GCT TCA CTT AAG GAA GAA AAA TGG TAC TCG GAA GAA 450
ala leu lys arg asp asp val thr leu leu arg gln glu val gln asp leu gln ala ser leu lys glu glu lys trp tyr ser glu glu 108
TTA AAG AAG GAA TTA GAA AAA TAT CAA GGG CTG CAG CAG CAA GAG GCC AAA CCT GAT GGG TTG GTG ACT GAT TCA TCA GCA GAA CTA CAG 540
leu lys lys glu leu glu lys tyr gln gly leu gln gln gln glu ala lys pro asp gly leu val thr asp ser ser ala glu leu gln 138
TCT TTG GAA CAG CAA TTA GAA GAA GCC CAA ACA GAA AAT TTT AAT ATT AAG CAA ATG AAA GAC TTA TTT GAA CAG AAA GCA GCC CAA CTT 630
ser leu glu gln gln leu glu glu ala gln thr glu asn phe asn ile lys gln met lys asp leu phe glu gln lys ala ala gln leu 168
GCT ACT GAA ATT GCA GAT ATA AAG TCA AAG TAT GAT GAA GAA AGG AGT CTT CGA GAA GCT GCT GAA CAA AAA GTG ACA CGT CTG ACA GAA 720
ala thr glu ile ala asp ile lys ser lys tyr asp glu glu arg ser leu arg glu ala ala glu gln lys val thr arg leu thr glu 198
GAA TTA AAC AAA GAG GCA ACT GTA ATT CAA GAT CTG AAG ACG GAA CTG CTT CAG AGA CCT GGT ATA GAA GAT GTT GCC GTG CTA AAG AAA 810
glu leu asn lys glu ala thr val ile gln asp leu lys thr glu leu leu gln arg pro gly ile glu asp val ala val leu lys lys 228
GAA CTG GTC CAA GTT CAA ACA CTA ATG GAT AAC ATG ACC TTG GAA CGT GAG CGA GAA TCT GAA AAA CTC AAA GAT GAA TGC AAA AAA TTG 900
glu leu val gln val gln thr leu met asp asn met thr leu glu arg glu arg glu ser glu lys leu lys asp glu cys lys lys leu 258
CAG TCA CAA TAT GCT AGC TCA GAG GCC ACA ATA AGC CAG CTA AGG AGT GAA CTG TCC AAA GGC CCC CAG GAA GTT GCT GTA TAT GTA CAG 990
gln ser gln tyr ala ser ser glu ala thr ile ser gln leu arg ser glu leu ser lys gly pro gln glu val ala val tyr val gln 288
GAA CTA CAA AAA CTG AAA AGT TCA GTT AAT GAA TTA ACA CAA AAA AAT CAG ACC TTG ACA GAA AAC TTG CTG AAA AAA GAA CAA GAC TAT 1080
glu leu gln lys leu lys ser ser val asn glu leu thr gln lys asn gln thr leu thr glu asn leu leu lys lys glu gln asp tyr 318
ACT AAG TTA GAG GAG AAA CAT AAT GAA GAA TCT GTG AGT AAA AAG AAT ATT CAG GCA ACC CTT CAT CAA AAA GAC CTA GAT TGT CAA CAG 1170
thr lys leu glu glu lys his asn glu glu ser val ser lys lys asn ile gln ala thr leu his gln lys asp leu asp cys gln gln 348
CTT CAG TCA AGA TTG TCT GCA TCT GAA ACC TCA CTG CAT AGA ATA CAT GTA GAA CTA AGT GAA AAA GGA GAA GCT ACT CAA AAG CTC AAA 1260
leu gln ser arg leu ser ala ser glu thr ser leu his arg ile his val glu leu ser glu lys gly glu ala thr gln lys leu lys 378
GAA GAA TTA TCT GAG GTA GAG ACC AAG TAC CAG CAT CTA AAG GCG GAG TTT AAG CAG CTA CAA CAA CAG AGA GAA GAA AAG GAG CAG CAT 1350
glu glu leu ser glu val glu thr lys tyr gln his leu lys ala glu phe lys gln leu gln gln gln arg glu glu lys glu gln his 408
GGG TTA CAA CTC CAA AGT GAA ATT AAT CAA TTA CAT AGC AAA CTT CTG GAG ACA GAG CGC CAA CTA GGG GAA GCT CAT GGT AGG CTG AAG 1440
gly leu gln leu gln ser glu ile asn gln leu his ser lys leu leu glu thr glu arg gln leu gly glu ala his gly arg leu lys 438
GAA CAG AGA CAG CTT TCA AGT GAA AAG TTG ATG GAT AAA GAA CAA CAA GTG GCT GAT TTA CAA CTC AAA CTT TCT CGG TTA GAA GAG CAG 1530
glu gln arg gln leu ser ser glu lys leu met asp lys glu gln gln val ala asp leu gln leu lys leu ser arg leu glu glu gln 468
TTG AAG GAA AAA GTT ACA AAT TCT ACA GAA TTG CAG CAT CAA TTA GAT AAA ACA AAG CAA CAG CAT CAA GAA CAA CAG GCT CTT CAG CAA 1620
leu lys glu lys val thr asn ser thr glu leu gln his gln leu asp lys thr lys gln gln his gln glu gln gln ala leu gln gln 498
AGC ACC ACG GCA AAA CTT CGA GAA GCT CAG AAT GAT TTG GAA CAA GTT CTA CGT CAA ATT GGC GAT AAG GAC CAA AAG ATC CAG AAC CTT 1710
ser thr thr ala lys leu arg glu ala gln asn asp leu glu gln val leu arg gln ile gly asp lys asp gln lys ile gln asn leu 528
GAA GCT TTA TTA CAG AAG AGT AAA GAA AAT ATT TCA TTA CTA GAA AAA GAA AGA GAA GAT CTT TAT GCA AAA ATT CAG GCT GGT GAA GGA 1800
glu ala leu leu gln lys ser lys glu asn ile ser leu leu glu lys glu arg glu asp leu tyr ala lys ile gln ala gly glu gly 558
GAG ACT GCT GTT CTT AAC CAG TTA CAA GAA AAA AAC CAT ACA CTA CAG GAG CAA GTA ACT CAA CTA ACA GAG AAG CTG AAG AAT CAG TCA 1890
glu thr ala val leu asn gln leu gln glu lys asn his thr leu gln glu gln val thr gln leu thr glu lys leu lys asn gln ser 588
GAA AGT CAT AAA CAA GCC CAG GAG AAT TTG CAT GAC CAG GTA CAA GAG CAG AAG GCA CAT CTT AGA GCT GCA CAA GAC CGT GTC CTT TCC 1980
glu ser his lys gln ala gln glu asn leu his asp gln val gln glu gln lys ala his leu arg ala ala gln asp arg val leu ser 618
CTA GAA ACT AGT GTC AAT GAA TTA AAT AGT CAA TTA AAT GAA AGC AAG GAG AAG GTC TCC CAG CTT GAC ATA CAG ATT AAA GCC AAA ACC 2070
leu glu thr ser val asn glu leu asn ser gln leu asn glu ser lys glu lys val ser gln leu asp ile gln ile lys ala lys thr 648
GAA CTA TTA CTA TCA GCA GAA GCA GCA AAA ACT GCT CAA AGA GCT GAT CTT CAG AAT CAT TTG GAC ACA GCT CAA AAT GCA TTA CAA GAT 2160
glu leu leu leu ser ala glu ala ala lys thr ala gln arg ala asp leu gln asn his leu asp thr ala gln asn ala leu gln asp 678
AAA CAG CAG GAG TTA AAT AAG ATT ACT ACT CAG TTG GAT CAG GTC ACT GCA AAG TTA CAA GAC AAG CAA GAA CAT TGC AGT CAG CTG GAA 2250
lys gln gln glu leu asn lys ile thr thr gln leu asp gln val thr ala lys leu gln asp lys gln glu his cys ser gln leu glu 708
AGT CAT CTT AAA GAA TAT AAA GAG AAA TAC CTC TCT TTA GAA CAG AAA ACC GAA GAG CTA GAA GGT CAA ATT AAG AAA CTA GAA GCT GAT 2340
ser his leu lys glu tyr lys glu lys tyr leu ser leu glu gln lys thr glu glu leu glu gly gln ile lys lys leu glu ala asp 738
AGT CTT GAA GTT AAA GCA AGC AAG GAG CAG GCT TTG CAA GAT CTA CAA CAG CAA AGA CAG CTG AAC ACA GAT TTA GAG CTC AGA GCC ACA 2430
ser leu glu val lys ala ser lys glu gln ala leu gln asp leu gln gln gln arg gln leu asn thr asp leu glu leu arg ala thr 768
GAA TTG AGT AAA CAA CTT GAA ATG GAG AAG GAA ATA GTA TCC AGT ACA AGA TTG GAT CTA CAG AAA AAA TCT GAA GCC CTT GAA AGT ATC 2520
glu leu ser lys gln leu glu met glu lys glu ile val ser ser thr arg leu asp leu gln lys lys ser glu ala leu glu ser ile 798
AAG CAA AAG CTT ACC AAG CAA GAG GAA GAA AAA CAA ATC CTG AAA CAA GAT TTT GAA ACT TTA AGT CAA GAA ACA AAG ATT CAG CAT GAG 2610
lys gln lys leu thr lys gln glu glu glu lys gln ile leu lys gln asp phe glu thr leu ser gln glu thr lys ile gln his glu 828
GAA TTG AAT AAC AGA ATT CAA ACA ACA GTA ACA GAA CTA CAA AAA GTG AAA ATG GAG AAA GAA GCT TTA ATG ACA GAG CTT TCT ACA GTA 2700
glu leu asn asn arg ile gln thr thr val thr glu leu gln lys val lys met glu lys glu ala leu met thr glu leu ser thr val 858
AAG GAC AAA CTA TCA AAA GTT TCT GAT TCT TTG AAA AAC TCT AAA AGT GAA TTT GAA AAG GAG AAT CAG AAA GGA AAA GCC GCT ATA TTA 2790
lys asp lys leu ser lys val ser asp ser leu lys asn ser lys ser glu phe glu lys glu asn gln lys gly lys ala ala ile leu 888
GAC TTG GAA AAA ACT TGC AAA GAA TTA AAG CAT CAA CTT CAA GTG CAG ATG GAA AAC ACA CTT AAG GAA CAG AAG GAA CTG AAA AAG TCA 2880
asp leu glu lys thr cys lys glu leu lys his gln leu gln val gln met glu asn thr leu lys glu gln lys glu leu lys lys ser 918
CTT GAA AAA GAG AAG GAG GCT TCT CAT CAG TTG AAA TTG GAA CTC AAT TCA ATG CAG GAA CAA CTT ATA CAG GCC CAG AAT ACT TTA AAA 2970
leu glu lys glu lys glu ala ser his gln leu lys leu glu leu asn ser met gln glu gln leu ile gln ala gln asn thr leu lys 948
CAA AAT GAA AAA GAA GAG CAA CAA CTT CAG GGG AAC ATA AAT GAG CTA AAG CAA TCA AGT GAA CAG AAG AAA AAA CAA ATT GAA GCA CTC 3060
gln asn glu lys glu glu gln gln leu gln gly asn ile asn glu leu lys gln ser ser glu gln lys lys lys gln ile glu ala leu 978
CAA GGA GAG CTT AAA ATT GCT GTT TTA CAG AAG ACA GAG CTT GAG AAT AAA CTA CAG CAG CAG TTA ACA CAG GCA GCC CAG GAA CTT GCA 3150
gln gly glu leu lys ile ala val leu gln lys thr glu leu glu asn lys leu gln gln gln leu thr gln ala ala gln glu leu ala 1008
GCA GAG AAA GAG AAA ATA TCA GTA TTA CAA AAC AAC TAT GAA AAA AGT CAG GAA ACT TTC AAA CAG CTT CAA TCT GAT TTC TAT GGG AGG 3240
ala glu lys glu lys ile ser val leu gln asn asn tyr glu lys ser gln glu thr phe lys gln leu gln ser asp phe tyr gly arg 1038
GAA TCT GAA CTT CTA GCC ACC AGG CAA GAT CTT AAG TCT GTA GAA GAG AAG CTT TCT CTA GCA CAG GAG GAC TTG ATT TCA AAC AGA AAT 3330
glu ser glu leu leu ala thr arg gln asp leu lys ser val glu glu lys leu ser leu ala gln glu asp leu ile ser asn arg asn 1068
CAA ATT GGA AAT CAA AAT AAA TTG ATT CAA GAA CTG AAG ACT GCC AAG GCT ACA TTG GAG CAG GAT TCA GCA AAG AAA GAA CAG CAA TTG 3420
gln ile gly asn gln asn lys leu ile gln glu leu lys thr ala lys ala thr leu glu gln asp ser ala lys lys glu gln gln leu 1098
CAG GAG CGA TGT AAA GCA CTA CAA GAC ATT CAG AAA GAA AAG TCA CTG AAA GAA AAA GAA CTG GTA AAT GAG AAG TCT AAA TTG GCA GAG 3510
gln glu arg cys lys ala leu gln asp ile gln lys glu lys ser leu lys glu lys glu leu val asn glu lys ser lys leu ala glu 1128
ATA GAA GAA ATT AAA TGT AGA CAA GAA AAA GAA ATC ACT AAA CTA AAC GAA GAA CTC AAG TCC CAC AAA CTA GAA AGC ATA AAG GAG ATA 3600
ile glu glu ile lys cys arg gln glu lys glu ile thr lys leu asn glu glu leu lys ser his lys leu glu ser ile lys glu ile 1158
ACA AAT CTT AAA GAT GCT AAA CAG CTT CTA ATT CAG CAG AAA TTA GAA CTT CAA GGA AAA GCG GAC TCC CTG AAG GCA GCT GTT GAA CAG 3690
thr asn leu lys asp ala lys gln leu leu ile gln gln lys leu glu leu gln gly lys ala asp ser leu lys ala ala val glu gln 1188
GAG AAG AGA AAT CAG CAG ATA CTA AAA GAC CAG GTG AAA AAG GAA GAA GAG GAG CTG AAG AAA GAA TTT ATT GAG AAA GAA GCT AAG TTG 3780
glu lys arg asn gln gln ile leu lys asp gln val lys lys glu glu glu glu leu lys lys glu phe ile glu lys glu ala lys leu 1218
CAT TCC GAA ATA AAA GAA AAG GAA GTA GGA ATG AAG AAG CAT GAA GAA AAT GAG GCT AAA CTT ACC ATG CAG ATT ACA GCA TTA AAT GAA 3870
his ser glu ile lys glu lys glu val gly met lys lys his glu glu asn glu ala lys leu thr met gln ile thr ala leu asn glu 1248
AAC TTA GGC ACT GTG AAG AAG GAG TGG CAA TCT AGT CAA CGG AGA GTT AGT GAG CTT GAG AAA CAA ACG GAT GAC TTA CGG GGT GAA ATT 3960
asn leu gly thr val lys lys glu trp gln ser ser gln arg arg val ser glu leu glu lys gln thr asp asp leu arg gly glu ile 1278
GCA GTA TTA GAA GCA ACG GTT CAG AAT AAT CAA GAT GAA AGG AGA GCA CTA CTG GAA AGA TGT CTT AAA GGA GAA GGT GAA ATA GAA AAG 4050
ala val leu glu ala thr val gln asn asn gln asp glu arg arg ala leu leu glu arg cys leu lys gly glu gly glu ile glu lys 1308
CTT CAA ACC AAA GTA TTA GAA TTG CAA AGA AAG CTG GAT AAT ACA ACT GCA GCA GTG CAG GAG CTG GGC AGA GAA AAC CAA TCA CTT CAG 4140
leu gln thr lys val leu glu leu gln arg lys leu asp asn thr thr ala ala val gln glu leu gly arg glu asn gln ser leu gln 1338
ATC AAA CAT ACA CAA GCG TTG AAT AGA AAG TGG GCC GAA GAC AAT GAA GTA CAA AAC TGT ATG GCC TGT GGG AAA GGC TTT TCA GTA ACA 4230
ile lys his thr gln ala leu asn arg lys trp ala glu asp asn glu val gln asn cys met ala cys gly lys gly phe ser val thr 1368
GTG AGA CGG CAT CAC TGC CGA CAG TGT GGA AAT ATC TTC TGT GCT GAA TGT TCA GCC AAA AAT GCC TTA ACT CCT TCC TCC AAG AAG CCT 4320
val arg arg his his cys arg gln cys gly asn ile phe cys ala glu cys ser ala lys asn ala leu thr pro ser ser lys lys pro 1398
GTT CGT GTC TGT GAT GCA TGT TTC AAT GAC TTG CAA GGA TAA TGG GTT ATC ACA ACT TCA GAG TAA TAT TAC ACT AAC ATT AGA TTT TTA 4410
val arg val cys asp ala cys phe asn asp leu gln gly stop                                                                 1411
ATA AAT GTA CTT AAT AGA GGT CTT GGA CTA CTA TTT GGT TTG GAC ACT GGT TGC AAT ACT AGA CCA AAT AGG AAT TAG TAT ATT TTG GAA 4500
TGT TAT GGA TAT CAA GTA AAA CTC TTC TAT TTT TGT GAG ACT TGG GCT TAT CAT CTT TCA TTA CTT TTT TCC CAT TTA GCC CCT GGA ATA 4590
GCT TTC ATG CCA AGT TTA GAA TTA GCC AAT GAA ATG TAA ATA AAC TTT GGA CAG AAA ATA GCA ATG TAT TTT TTT AAT ATA ATT TCA TTA 4680
TTC ACA AAT GAA GAA TGC AAT TCC ATA GCT TAC TTC TTT CTC CTA AAT TTA ATG TAC AAA AAT GTA CAT GGA TGA TAT TAT CAC TTG TTG 4770
CTG TTT TTG CGA CAA TAG CAC ATT AAT TAT ATT AAA TAT TCT CTT TGT GAA GTT ATG CAC AGA ACT GTA ACA CTT TCC TCT GCC TTT TTC 4860
TCT TAT ATG TGG GTT CAT GGT TCA GTT CGA TTG TTA TGT TAT GAA TTT ACT ACT TGG GTG TTC TAA ATA CAG TTT ATC TTA AAA TAT GTA 4950
GTA ATA TGA ATT                                                                                                           4962
```

FIG. 6

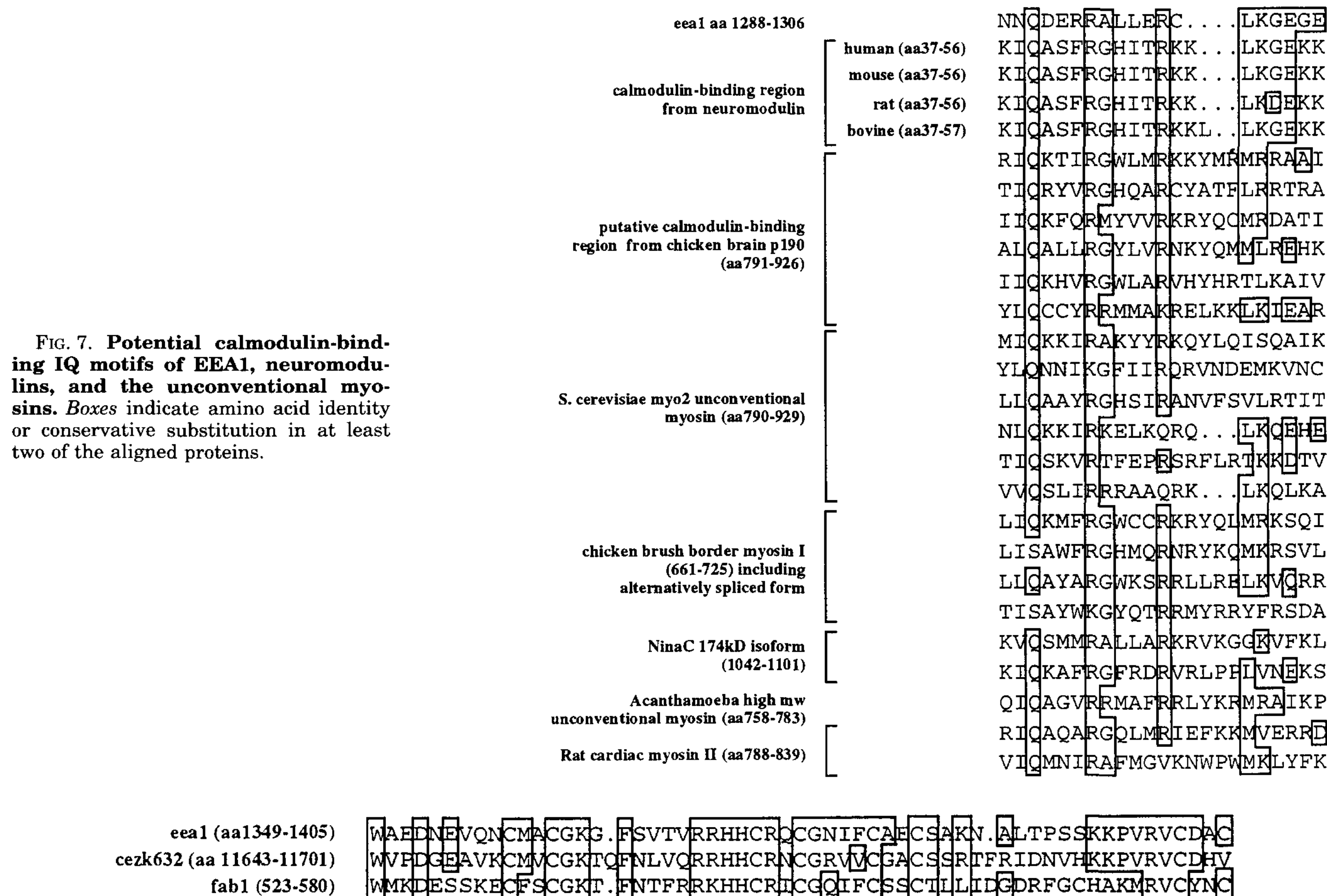

FIG. 7. **Potential calmodulin-binding IQ motifs of EEA1, neuromodulins, and the unconventional myosins.** *Boxes* indicate amino acid identity or conservative substitution in at least two of the aligned proteins.

FIG. 8. **Homology of carboxyl-terminal cysteine finger motifs of EEA1 with those of *C. elegans* ZK632 and *S. cerevisiae* FAB 1 proteins.** *Boxes* indicate amino acid identities or conservative changes between EEA1 and at least one of the other two proteins.

N-myristoylation, protein kinase C phosphorylation, tyrosine kinase phosphorylation, and for the leucine zipper structure in addition to single consensus motifs for glycosaminoglycan attachment and cAMP- and cGMP-dependent protein kinase phosphorylation. Due to a lack of specificity in these consensus motifs, the significance of this analysis is not clear, although post-translational modification of EEA1 may contribute to the difference between its observed and predicted molecular weight.

Rabbit Antiserum to Recombinant Protein Reacts with Cytoplasmic Vesicles and with a 180-kDa Molecule—The 1.2-kilobase *Eco*RI cDNA insert (pFM1) containing a 900-bp coding region was subcloned into pGEX, a vector that expresses a bacterial fusion protein at the carboxyl terminus of glutathione *S*-transferase. The size of the 60-kDa fusion protein (Fig. 9) obtained by induction with isopropyl-1-thio-β-D-galactopyranoside was in good agreement with the predicted combined size of glutathione *S*-transferase (27 kDa) and that of the partial polypeptide encoded by the cDNA clone (33 kDa). A rabbit antiserum against the fusion protein, and affinity-purified rabbit anti-fusion protein antibody, reacted with cytoplasmic vesicles and immunoblotted the fusion protein (Fig. 9*B*) and a 180-kDa 3T3 cell protein (Figs. 3 and 4). The affinity-purified rabbit antibody reacted with glutathione *S*-transferase but after purification on a glutathione *S*-transferase affinity column all reactivity was abolished (Fig. 9*C*). The affinity-purified autoantibody also reacted with the fusion protein (Fig. 9*B*), but it did not react with glutathione *S*-transferase (Fig. 9*C*). The human autoimmune serum was recycled through the glutathione *S*-transferase affinity column using the same conditions as those used with the fusion protein affinity column. Antibody titer remained unchanged after recycling through this column.

DISCUSSION

We have cloned the cDNA encoding a 180-kDa protein, named EEA1. EEA1 is evolutionarily conserved as it is present in human, mouse, and chicken cells. It associates with early endosomes since it co-localized with the transferrin receptor and Rab5 and not with Rab7. Localization to early endosomes was confirmed by ultrastructural studies showing co-localization of EEA1 with tubulovesicular early endosomes containing internalized BSA-gold. EEA1 is a hydrophilic peripheral membrane protein lacking a hydrophobic transmembrane domain. It is present in cytosolic and in membrane fractions, it fractionates exclusively in the aqueous phase after Triton X-114 solubilization, and it is extracted from membranes by 0.3 M NaCl. These properties are different from those of a 195-kDa integral membrane protein associated with early endosomes (Pitt and Schwartz, 1991). CLIP-170 is an early endosomal protein that links early endocytic vesicles to cytoplasmic microtubules by a novel motif present in a tandem repeat in the amino-terminal domain (Pierre *et al.*, 1992). Like EEA1, it has a cysteine finger

FIG. 6. **cDNA and predicted amino acid sequence of EEA1.** NH_2-terminal C_2H_2 zinc finger motif and COOH-terminal cysteine-rich metal-binding fingers are *boxed*. IQ motif is *underlined*. Segment sharing sequence similarity with *S. cerevisiae* FAB1 and *C. elegans* CEZK632 proteins is in *boldface*.

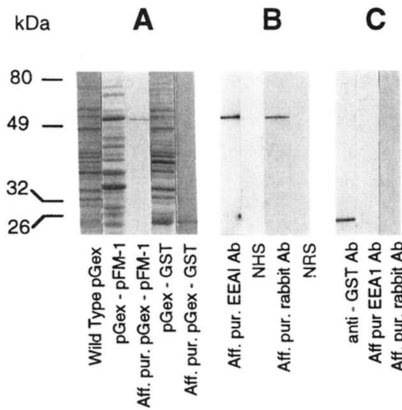

FIG. 9. **Immunoblotting of bacterial fusion protein with affinity-purified human or rabbit antibody to EEA1.** pFM1 containing 900 bp of coding sequence was subcloned into pGex 2T expression vector. Fusion protein induced by isopropyl-1-thio-β-D-galactopyranoside was purified on glutathione affinity column. *Panel A*, total proteins from noninduced wild-type pGex 2T, pGex 2T subcloned with pFM1, affinity purified pGex-pFM1, pGEX-GST, and affinity purified pGex-GST were electrophoresed on 10% SDS-PAGE and stained with Coomassie Blue. *Panel B*, immunoblots of fusion protein with affinity-purified human (*Aff. pur. EEA1 Ab*) or rabbit (*Aff. pur. rabbit Ab*) antibody to EEA1, or with normal human (*NHS*) or rabbit (*NRS*) serum. *Panel C*, immunoblots of glutathione *S*-transferase protein with rabbit anti-glutathione *S*-transferase antibody, or with affinity-purified human or rabbit antibody to EEA1. Bound antibody was detected as described in legend to Fig. 3.

motif at the carboxyl terminus. However, the finger motifs and the amino acid sequence of EEA1 are different from those of CLIP-170, and EEA1 lacks the microtubule binding motif of CLIP-170.

The sequence of EEA1 gene is virtually identical to that of HSP 162, a cDNA derived from HeLa cells (GenBank accession number X78998).[2] Indeed the sequence differences between EEA1 and HSP 162 are limited to a dinucleotide substitution giving rise to the replacement of Ala by Ser at amino-acid 277 in EEA1, as well as a silent single base substitution at nucleotide 245 in EEA1. In addition, a three-nucleotide deletion is present in the 5′ noncoding region of EEA1 as well as a two-nucleotide deletion followed by a dinucleotide substitution in the 3′ noncoding region.

The metal-binding finger motifs of EEA1, Cys-X_2-Cys-X_{12}-Cys-X_2-Cys-X_4-Cys-X_2-Cys-X_{16}-Cys-X_2-Cys, are strikingly conserved within a homologous segment of a protein encoded by the ZK632 cosmid of *C. elegans* and the *S. cerevisiae* FAB1 protein. The latter has been implicated in the endocytic-vacuolar pathway and in nuclear migration. The presence of this motif in EEA1 and FAB1, which although globally dissimilar in sequence and yet are both implicated in endocytic trafficking, suggests a common function for this region of both proteins. These observations also suggest a possible related role for the homologous protein encoded by the *C. elegans* ZK632 cosmid. Similar finger motifs are present as Cys-X_2-Cys-X_{12}-Cys-X_4-Cys in the *S. cerevisiae* VAC1 gene, which is required for vacuole inheritance and vacuole protein sorting (Weisman and Wickner, 1992) and as Cys-X_2-Cys-X_{12}-Cys-X_2-Cys in the 190-kDa bovine microtubule-associated protein-U (Aizawa *et al.*, 1990). A cysteine-rich finger motif, Cys-X_2-Cys-X_{13}-Cys-X_2-Cys-X_4-Cys-X_{38}-Cys-X_2-Cys, at the carboxyl terminus of *S. cerevisiae* Vps18p protein has been localized to the cytosolic face of vacuoles (Robinson *et al.*, 1991). A mutant in which the first cysteine of this motif was changed to serine resulted in a temperature-conditional sorting defect. Another yeast protein, Vps11, which also has a cysteine-rich motif near the carboxyl terminus, similar to Vps18p, co-localizes with Vps18p (Preston *et al.*, 1991). These observations suggest that the finger motifs are necessary for vacuolar protein sorting in yeast. On the basis of these observations, we propose that the two metal-binding cysteine finger motifs of EEA1 are required for early endosomal sorting.

The presence of a canonical C_2H_2 type zinc finger at the amino terminus of EEA1 is of unknown significance since this motif has usually been found in nucleic acid binding proteins. However, there are some 40 other proteins in the SWISS-PROT data base that are identified by this motif, and a metal binding function cannot be excluded in these cases or in the case of EEA1.

EEA1 shares 17–20% sequence identity with both conventional and unconventional myosins and contains a potential calmodulin binding IQ motif typically found in the neck region of unconventional myosins (Cheney and Mooseker, 1992). Unconventional myosins have been implicated in vesicular transport (Bretscher, 1993; Titus, 1993). However, EEA1 lacks the globular ATP and actin binding head domain to qualify it as an "unconventional" myosin. Nonetheless, it seems likely that EEA1 shares an evolutionary relationship with the unconventional myosins given its homology with these proteins and the presence of the calmodulin-binding IQ motif.

The IQ motif in EEA1 suggests that calmodulin may be implicated in the function of EEA1. However, EEA1 lacks the sequence ASF, which follows the IQ motif of neuromodulin, a 24-kDa calmodulin-binding non-myosin protein (Chapman *et al.*, 1991). The phenylalanine associates with calmodulin via a hydrophobic interaction, which is disrupted following protein kinase C-dependent phosphorylation of the adjacent serine, which introduces a strong negative charge. Thus calmodulin binds to neuromodulin in the absence of calcium and is released following calcium-dependent protein kinase C phosphorylation. Similar hydrophobic residues are present in many (although not all) of the unconventional myosins.

The function of EEA1 remains to be determined. Based on its localization to early endosomes and the similarity of its carboxyl-terminal zinc fingers to proteins regulating membrane traffic, we propose as a working hypothesis that EEA1 has a role in vesicular transport of proteins through early endosomes. The conserved metal-binding cysteine finger motifs may be crucial in this process.

Acknowledgments—We thank Marino Zerial for anti-Rab5 and anti-Rab7 antibodies and Jean Gruenberg for helpful comments and suggestions.

REFERENCES

Aizawa, H., Emori, Y., Murofushi, H., Kawasaki, H., Sakai, H., and Suzuki, K. (1990) *J. Biol. Chem.* **265,** 13849–13855
Berg, J. M. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 99–102
Bretscher, A. (1993) *Curr. Opin. Cell Biol.* **5,** 653–660
Chapman, E. R., Au, D., Alexander, K. A., Nicolson, T. A., and Strom, D. R. (1991) *J. Biol. Chem.* **266,** 207–213
Chavrier, P., Parton, R. G., Hauri, H. P., Simons, K., and Zerial, M. (1990) *Cell* **62,** 317–329
Cheney, R. E., and Mooseker, M. S. (1992) *Curr. Opin. Cell Biol.* **4,** 27–35
Chou, P. Y., and Fasman, G. D. (1974a) *Biochemistry* **13,** 211–222
Chou, P. Y., and Fasman, G. D. (1974b) *Biochemistry* **13,** 222–245
Chou, P. Y., and Fasman, G. D. (1978a) *Annu. Rev. Biochem.* **47,** 251–276
Chou, P. Y., and Fasman, G. D. (1978b) *Adv. Enymol. Relat. Areas Mol. Biol.* **47,** 45–148
Courtoy, P. J. (1991) in *Intracellular Trafficking of Proteins* (Steer, C. J., and Hanover, J. A., eds) pp. 102–156, Cambridge University Press, Cambridge, United Kingdom
Eisenberg, D., Weiss, R. M., and Terwilliger, T. C. (1984a) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 140–144
Eisenberg, D. Schwarz, E., Komaromy, M., and Wall, R. (1984b) *J. Mol. Biol.* **179,** 125–142
Emini, E., Hughes, J. V., Perlow, D. S., and Boger, J. (1985) *J. Virol.* **55,** 836–839
Evans R. M., and Hollenberg, S. M. (1988) *Cell* **52,** 1–3
Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120
Gould, B. (1992) *Semin. Cell Biol.* **3,** 301–307

[2] H. P. Seelig, unpublished results.

Griffiths, G., and Gruenberg, J. (1991) *Trends Cell Biol.* **1,** 5–9
Griffiths, G., McDowall, A., Back, R., and Dubochet, J. (1984) *J. Ultrastruct. Res.* **89,** 65–84
Hopkins, C. R. (1983) *Cell* **35,** 321–330
Jameson, B. A., and Wolf, H. (1988) *Comput. Appl. Biosci.* **4,** 181–186
Janin, J., Wodak, S., Levitt, M., and Maigret, B. (1978) *J. Mol. Biol.* **125,** 357–386
Karplus, P. A., and Schulz, G. E. (1985) *Naturwissenschaften* **72,** 212–213
Klug, A., and Rhodes, D.(1987) *Trends Biochem. Sci.* **12,** 464–469
Kooy, J., Toh, B. H., Pettitt, J. M., Erlich, R., and Gleeson, P. A. (1992) *J. Biol. Chem.* **267,** 20255–20263
Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132
Lipman, D. J., and Pearson, W. R. (1985) *Science* **227,** 1435–1441
Mellman, I. (1993) *J. Exp. Biol.* **172,** 39–45
Miller J., Mc Lachlan, A. D., and Klug, A. (1985) *EMBO J.* **4,** 1609–1614
Mu, F-T, Alderuccio, F., Toh, B. H., Gershwin, M. E., Coppel, R., and Mackay, I. R. (1988) *Clin. Exp. Immunol.* **71,** 100–106
Murphy, R. F. (1991) *Trends Cell Biol.* **1,** 77–82
Needleman, S. B., and Wunch, C. D. (1970) *J. Mol. Biol.* **48,** 443–453
Payre, F., and Vincent, A. (1988) *FEBS Lett.* **234,** 245–250
Pierre, P., Scheel, J., Rickard, J. E., and Kreis, T. E. (1992) *Cell* **70,** 887–900
Pitt, A., and Schwartz, A. L. (1991) *Exp. Cell Res.* **194,** 128–134
Preston, R. A., Manolson, M. F., Becherer, K., Weidenhammer, D., Kirkpatrick, D., Wright, R., and Jones, E. W. (1991) *Mol. Cell. Biol.* **11,** 5801–5812
Rechid, R., Vingron, M., and Argos, P. (1989) *Comput. Appl. Biosci.* **5,** 107–113
Robinson, J. S., Graham, T. R., and Emr, S. D. (1991) *Mol. Cell. Biol.* **12,** 5813–5824
Robson, B., and Suzuki, E. (1976) *J. Mol. Biol.* **107,** 327–356
Schmid, S. L., Fuchs, R., Male, P., and Mellman, I. (1988) *Cell* **52,** 73–83
Shepherd, V. L. (1989) *Trends Pharmacol. Sci.* **10,** 458–462
Smith, R. F., and Smith, T. F. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 118–122
Smith, T. F., and Waterman, M. S. (1981) *Adv. Appl. Math.* **2,** 482–489
Stenmark, H., Valencia, A., Martinez, O., Ullrich, O., Goud, B., and Zerial, M. (1994a) *EMBO J.* **13,** 575–583
Stenmark, H., Parton, R. G., Steele-Mortimer, O., Lutcke, A., Gruenberg, J., and Zerial, M. (1994b) *EMBO J.* **13,** 1287–1296
Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L., Dear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R., and Waterston, R. (1992) *Nature* **356,** 37–41
Titus, M. (1993) *Curr. Opin. Cell Biol.* **5,** 77–81
Trowbridge, I. S., Collawn, J. F., and Hopkins, C. R. (1993) *Annu. Rev. Cell Biol.* **9,** 129–161
von Heijne, G. (1986) *EMBO J.* **5,** 1335–1342
Weisman, L. S., and Wickner, W.(1992) *J. Biol. Chem.* **267,** 618–623
Yen, T. J., Li, G., Schaar, B. T., Szilak, I., and Cleveland, D. W. (1992) *Nature* **359,** 536–539
Yeo, J. P., Alderuccio, F., and Toh, B. H. (1994) *Nature* **386,** 488–491