Endolyn is a mucin-like type I membrane protein targeted to lysosomes by its cytoplasmic tail

Gudrun IHRKE,[1], Sally R. GRAY and J. Paul LUZIO

Department of Clinical Biochemistry and Wellcome Trust Centre for the Study of Molecular Mechanisms in Disease, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Hills Road, Cambridge CB2 2XY, U.K.

Endolyn (endolyn-78) is a membrane protein found in lysosomal and endosomal compartments of mammalian cells. Unlike 'classical' lysosomal membrane proteins, such as lysosome-associated membrane protein (lamp)-1, it is also present in a subapical compartment in polarized WIF-B hepatocytes. The structural features that determine sorting of endolyn are unknown. We have identified a rat endolyn cDNA by expression screening. The cDNA encodes a ubiquitously expressed type I membrane protein with a short cytoplasmic tail of 13 amino acids and many putative sites for N- and O-linked glycosylation in the predicted luminal domain. Endolyn is closely related to two human mucin-like proteins, multi-glycosylated core protein (MGC)-24 and CD164 (MGC-24v), expressed in gastric carcinoma cells and bone marrow stromal and haematopoietic precursor cells respectively. The predicted transmembrane and cytoplasmic tail domains of endolyn, as well as parts of its luminal domain, also show some similarities with lamp-1 and lamp-2. Like these and other known lysosomal membrane proteins, endolyn contains a YXXØ motif at the C-terminus of its cytoplasmic tail (where Ø is a bulky hydrophobic amino acid), but with no preceding glycine. Nonetheless, the last ten amino acids of this tail, when transplanted on to human CD8, caused efficient targeting of the chimaeric protein to endosomes and lysosomes in transfected normal rat kidney cells.

Key words: CD164, endosomes, lamp, sorting, targeting motif.

INTRODUCTION

Sorting of lysosomal membrane proteins occurs at the level of the *trans*-Golgi network (TGN) and in endosomes (reviewed in [1,2]). Newly synthesized lysosomal membrane proteins may be delivered from the TGN to endosomes and lysosomes via direct delivery pathways, which bypass the plasma membrane, or via an indirect pathway in which molecules are first transported to the cell surface and then reinternalized. Many details of these pathways, and the extent to which they are used by different lysosomal membrane proteins, are still poorly understood. At steady-state, the major lysosome-associated membrane proteins (lamps), such as lamp-1/lysosomal glycoprotein (lgp)120 and lamp-2/lgp110, are barely detectable in non-lysosomal compartments by immunocytochemistry. However, a less well characterized endosomal/lysosomal protein, endolyn-78 (endolyn), is somewhat more abundant at the plasma membrane and in the earlier parts of the endocytic pathway [3]. Recently, it was found that endolyn follows an unexpected trafficking route to lysosomes in a polarized hepatocyte-like cell line, WIF-B [4]. A significant fraction of endolyn, moving through the basolateral plasma membrane, is first transported to a subapical endosomal compartment (SAC) before being delivered to lysosomes. However, the SAC does not appear to be a major sorting station for other lysosomal membrane proteins (e.g. lamp-1) [4]. The hepatic SAC is an obligatory station through which transcytotic membrane proteins move on their way from basolateral to apical before reaching the apical cell surface [4,5]. A population of endolyn may be sorted together with transcytotic proteins into carriers that shuttle between basolateral endosomes and SAC. The pathway by which endolyn is retrieved from the SAC and transported to lysosomes has not been defined.

With the sequence of endolyn still unknown, it is unclear what structural features may determine its intracellular itinerary. All known lysosomal membrane proteins are integral membrane proteins that span the lipid bilayer one or more times (for a review see [1,2,6]). The cytoplasmic tails of these proteins are usually small, 10–20 amino acids, but contain signal motifs that interact with clathrin adaptor or adaptor-like protein complexes (AP1 and AP2 or AP3 respectively), which mediate the sorting into vesicular carriers. The core of these signals consists of either a di-leucine motif (LimpII/lgp85) or a tyrosine-based motif, YXXØ, where X is any amino acid and Ø is a bulky hydrophobic amino acid. The YXXØ motif is often preceded by a glycine (e.g. lamp-1/lgp120, lamp-2/lgp110 and limp I/CD63), which enhances the efficiency of lysosomal targeting from the TGN [7,8]. Since the trafficking of endolyn differs in some aspects from that of other lysosomal membrane proteins, we decided to clone and sequence this protein in order to determine putative sorting signals that could provide novel insights into targeting to lysosomes and, in hepatocytes, passage through the SAC.

EXPERIMENTAL

Antibodies

Rabbit polyclonal antiserum 580, prepared by immunization with a Triton X-114 extract of purified rat liver lysosomes [9], a polyclonal antibody (pAb) to the cation-independent mannose 6-phosphate receptor [10], and monoclonal antibodies (mAbs) 501 and 502 to rat endolyn [4] and 2F7.1 to TGN38 [9] have been described previously. The rat mAb to the luminal domain of the α-chain of human CD8 (Campath 8c; [11]) was a gift from Dr G.

Abbreviations used: AP, adaptor protein complex; EST, expressed sequence tag; lamp, lysosome-associated membrane protein; lgp, lysosomal glycoprotein; mAb, monoclonal antibody; NRK, normal rat kidney; ORF, open reading frame; pAb, polyclonal antibody; SAC, subapical compartment; TGN, trans-Golgi network; TM, transmembrane; MGC, multi-glycosylated core protein.

[1] To whom correspondence should be addressed (e-mail gi200@cus.cam.ac.uk).

Hale (Oxford University, Oxford, U.K.). The antibodies to rat lamp-1/lgp120 were rabbit polyclonal anti-LGP107 serum [12], generously provided by Professor K. Kato, Dr Y. Tanaka, and Professor M. Himeno (all from Kyushu University, Fukuoka, Japan), and mouse mAb GM10 [13] from Professor K. Siddle (University of Cambridge, Cambridge, U.K.) and Dr J. Hutton (University of Colorado, Denver, CO, U.S.A.). Horseradish-peroxidase-conjugated secondary antibodies were obtained from Amersham. FITC-, Texas Red- or Cy3-conjugated antibodies [$(F(ab)_2$-fragments] to mouse or rabbit IgG were from Jackson (Westgrove, PA, U.S.A.) and the Texas Red-conjugated antibody to rat IgG was from Molecular Probes. All secondary antibodies used for immunofluorescence showed minimal cross-reaction with other species, including mouse in the case of the anti-rat IgG.

cDNA library screening and DNA sequencing

All recombinant DNA procedures were carried out according to Sambrook et al. [14] unless stated otherwise. An expressed rat liver cDNA library in pUEX1 was screened with pAb 580 as described previously [15]. After colony purification, a subset of clones was screened with mAbs 501 and 502. The cDNA insert from one positive clone (Ly9) was ^{32}P-labelled by random priming (Rediprime DNA labelling system; Amersham) and used to screen a rat liver 5′-STRETCH Plus λgt11 cDNA library (RL5001b; Clontech) in order to obtain full length cDNA sequence. Of approx. 1×10^6 screened plaques, 73 gave a positive signal; of these, 13 clones were re-screened twice and their inserts further analysed by gel electrophoresis and partial sequencing. Inserts of selected clones were cloned into pBluescript II SK$^-$ (Stratagene) and both strands were fully sequenced using the Big Dye Terminator Cycle Sequencing kit (ABI–Perkin Elmer).

Northern blotting

A rat multiple tissue Northern blot (Clontech) with approx. 2 μg poly(A)$^+$ RNA per lane was probed with a ^{32}P-labelled (random priming, see above) fragment derived from endolyn clone 2 (bp 1–602, see Figure 3). The blot was exposed for 3–8 h to X-ray film (Biomax MR; Kodak), which was then scanned and processed using Adobe Photoshop (Adobe Systems, San Jose, CA, U.S.A.), or to a phosphorimaging screen for quantitative analysis on a STORM860 PhosphorImager (Molecular Dynamics).

Constructs

Pfu DNA polymerase (Stratagene) was used in all PCR reactions, and sequences of all final constructs were verified by DNA sequencing (as above). Clone 2 in pBluescript II SK$^-$ was used to amplify rat endolyn wild-type cDNA and to introduce a *Nhe*I site 27 bp 5′ of the predicted start codon and a *Not*I site 7 bp 3′ of the predicted stop codon by PCR. The *Nhe*I/*Not*I-digested PCR product was ligated into the corresponding sites in ΔpMEP4 [16]. Human CD8 α-chain cDNA in the vector S85 was kindly provided by Dr S. Munro (MRC Laboratory of Molecular Biology, Cambridge, U.K.). The CD8 cDNA sequence (used as 'wild type' in this study) contained an *Afl*II site at the end of the transmembrane (TM) region, changing Asn-His-Arg-Asn of the 'native' sequence into Lys-Arg-Leu-Lys ([17,18] respectively). The CD8 cDNA was digested with *Hin*dIII/*Xba*I and cloned into pBluescript II SK$^-$, then recloned into ΔpMEP4 using the *Hin*dIII and *Not*I sites. To construct a chimaera containing the luminal and TM domains of CD8 and the cytoplasmic tail of endolyn (CD8-En), a corresponding piece of endolyn cDNA was amplified by PCR introducing an *Afl*II site at the TM/cytoplasmic domain boundary. For this, a primer with the sequence CTATAAATTCCTTAAGTCTAAAGAACG was paired with a pBluescript II SK$^-$-specific primer binding approx. 120 bp downstream from the predicted stop codon. The resulting expressed amino acid sequence, after joining of the *Afl*II-digested CD8 and endolyn PCR product, was …TLYCKRLKSK ERNYHTL, where the underlined 10 residues were from endolyn and all other residues from CD8. The chimaeric cDNA was subcloned into ΔpMEP4 using the *Hin*dIII/*Not*I restriction sites present in both vectors.

Cell culture and transfection

HeLa and NRK (normal rat kidney) cells were grown in tissue-culture flasks or on glass coverslips as described previously [19]. Constructs in ΔpMEP4 were transfected into HeLa or NRK cells using the Fugene transfection reagent (Boehringer). Stable NRK cell lines were selected by the addition of 200 μg/ml Hygromycin B to the culture medium. Expression of exogenous protein in transiently or stably transfected cells was induced by addition of 2 μM (HeLa) or 2.5–5 μM (NRK) $CdCl_2$. In HeLa cells, $CdCl_2$ was removed after 8 h and cells were cultured in normal medium for a further 16 h. NRK cells were usually induced over a period of 16 h and then incubated without metal ions for 8 h; leupeptin (21 μM) was included during the last 6 h to reduce proteolysis of the luminal CD8 portion.

Indirect immunofluorescence microscopy

Cells grown on coverslips were fixed and permeabilized for 10 min in methanol at −20 °C, rehydrated with PBS and then labelled as described previously [4]. Confocal images from approx. 1 μm optical sections were collected on a BioRad MRC 1000 confocal microscope equipped with a Nikon Optophot-2 microscope and a 60× objective.

Purification and partial sequencing of endolyn

Highly purified rat liver lysosomal membranes were prepared according to Ohsumi et al. [20], except that the final high-salt wash was omitted. Membranes were solubilized for 30 min in 250 mM sucrose, 300 mM NaCl, 0.5% (v/v) Triton X-100, 24 mM octylglucoside, 1 mM EDTA, 25 mM sodium phosphate, pH 7.4, containing protease inhibitors (1 mM PMSF, 100 kallikrein inhibitor units/ml aprotinin, 0.4 mM benzamidine, 4 μM leupeptin) at 4 °C. Solubilized membrane proteins were used as a source of endolyn, which was purified by immunoaffinity chromatography with a mixture of mAb 501 and mAb 502 coupled to Sepharose beads (Pharmacia). Bound protein was eluted with 100 mM NaCl, 100 mM glycine, 46 mM octylglucoside, pH 2.5, and neutralized by addition of Tris base. Protein derived from approx. 2.5 livers was precipitated with trichloroacetic acid and separated by SDS/PAGE under reducing conditions. The gel was lightly stained with Coomassie Blue and the main protein band was excised. Preparation and sequencing of tryptic peptides were carried out at the Harvard Microsequencing Facility (Cambridge, MA, U.S.A.).

Sequence analysis

Hydrophobic segments in the protein sequences were identified by Kyte and Doolittle hydropathy plots [21]. Protein sequences were aligned with the Clustal W program [22] using the Blosum matrix and the following gap penalties: to open 10, to end 10, to extend 0.05, for gap separation 0.05. Cytoplasmic tail sequences were realigned manually (Gly-Gly in lamp-1). The average overall

alignment score of endolyn and lamp-1 and lamp-2 homologues was 11 % and 14 % respectively, and the average score for lamps-1 aligned with lamps-2 was 30 %. The amino acid sequences can be accessed through the Swiss Protein database under the following accession numbers: rat lamp-1, Swiss-Prot #P14562; mouse lamp-1, Swiss-Prot #P11438; human lamp-1, Swiss-Prot #P11279; rat lamp-2, Swiss-Prot #P17046; mouse lamp-2, Swiss-Prot #P17047; human lamp-2, Swiss-Prot #P13473; human CD164 [multi-glycosylated core protein (MGC)-24v], Swiss-Prot #Q04900 and human MGC-24, Swiss-Prot #O04900. The nucleotide sequence and translated amino acid sequence of mouse MGC-24v can be found in the GenBank under the GenBank accession number AB014464. The partial cDNA sequence of the putative *Drosophila* homologue of endolyn (cDNA clone LP01451) is available in the *Drosophila* database (FlyBase@fly.ebi.ac.uk; http://fly.ebi.ac.uk:7081/); it corresponds to one of several similar published expressed sequence tags (ESTs) grouped into clot 3920. Other programs and parameters used are stated in the Figure legends.

Miscellaneous

Crude lysosomal membranes used for immunoblotting were prepared as described previously [23]. Proteins were separated by SDS/PAGE and detected on immunoblots using the ECL system (Amersham).

RESULTS

Endolyn cDNA clones identified by one pAb and two mAbs in a rat liver expression library

A polyclonal antiserum (pAb 580) to lysosomal membrane proteins, previously raised in this laboratory, recognizes predominantly the lysosomal glycoprotein lamp-2 (lgp110) on immunoblots of reduced samples of rat liver lysosomes [9]. However, some additional minor bands were detected when non-reduced lysosomal membrane fractions from rat liver were separated by SDS/PAGE and immunoblotted with this antibody (Figure 1). Comparison of these minor bands with the pattern obtained with mAb 501 and mAb 502 to endolyn showed that they were virtually indistinguishable (Figure 1). Non-reduced rat liver endolyn was seen as a characteristic cluster of two to four bands of 60 to 80 kDa and a putative dimer at approx. 140 kDa.

We used pAb 580 to screen a pUEX1 rat cDNA expression library to obtain candidate clones encoding endolyn. Sequence analysis revealed that three of 18 recognized clones (Ly9, Ly11 and Ly15) showed high similarity to a human protein, MGC-24 [24]. This protein had some of the characteristics expected for endolyn, namely a peptide backbone of approx. 22 kDa and multiple putative N- and O-glycosylation sites [3], although it was described neither as an integral membrane protein nor to be present in lysosomes. Further evidence that the three clones contained endolyn cDNA was obtained by immunoblotting expressed fusion proteins with the mAbs to endolyn (Figure 2). Both mAbs specifically recognized the fusion proteins encoded by Ly9, Ly11 and Ly15. The full length cDNA of the MGC-24-like protein was obtained by screening a rat liver cDNA library

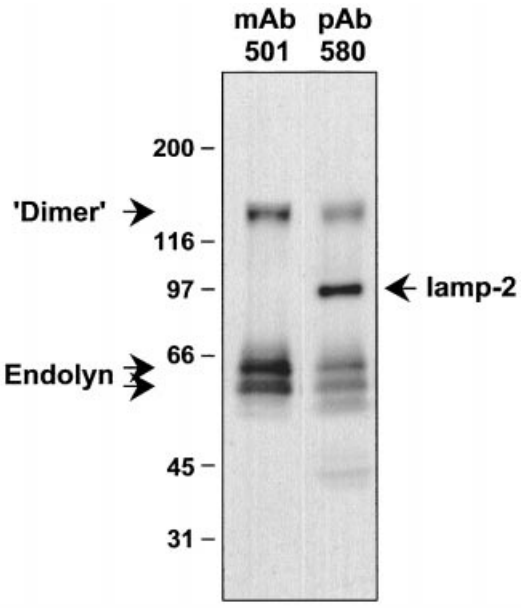

Figure 1 A polyspecific antiserum to lysosomal membrane proteins (pAb 580) recognizes lamp-2 and endolyn

Rat liver lysosomal membrane proteins were separated by SDS/PAGE under non-reducing conditions and immunoblotted with mAb 501 to endolyn (left lane) or pAb 580 (right lane). Molecular-mass markers and the positions of endolyn and the 'dimer' are shown on the left.

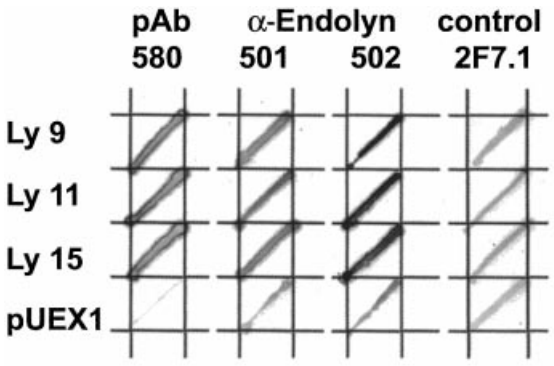

Figure 2 pAb 580 and two mAbs to endolyn react with bacterially-expressed fusion proteins

Three pUEX1 rat liver cDNA clones (Ly9, Ly11 and Ly15), identified with pAb 580, were streaked on to plates [bacteria containing empty vector served as a control (pUEX1)] and cultured under conditions that allowed expression of the fusion protein. The colonies were then transferred to nitrocellulose, lysed and immunoblotted with two different mAbs to endolyn (501 or 502) or a control antibody to an unrelated protein (mAb 2F7.1 to TGN38).

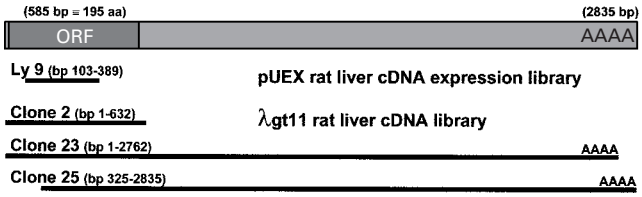

Figure 3 Identification of the full-length cDNA of rat endolyn

pUEX1 clone Ly9 was used to screen a λgt11 rat liver cDNA library. Recognized λgt11 clones 2 and 23 contained an ORF of 585 bp (195 amino acids). Clone 25 exceeded clone 23 by 73 bp, which included a standard polyadenylation site preceding a poly(A) tail. The full-length cDNA, shown at the top of the diagram, is a composite of the 5′-region contained in clones 2 and 23, and the 3′-untranslated region of clone 25. Numbers given after the clone names indicate their ends compared with the complete cDNA.

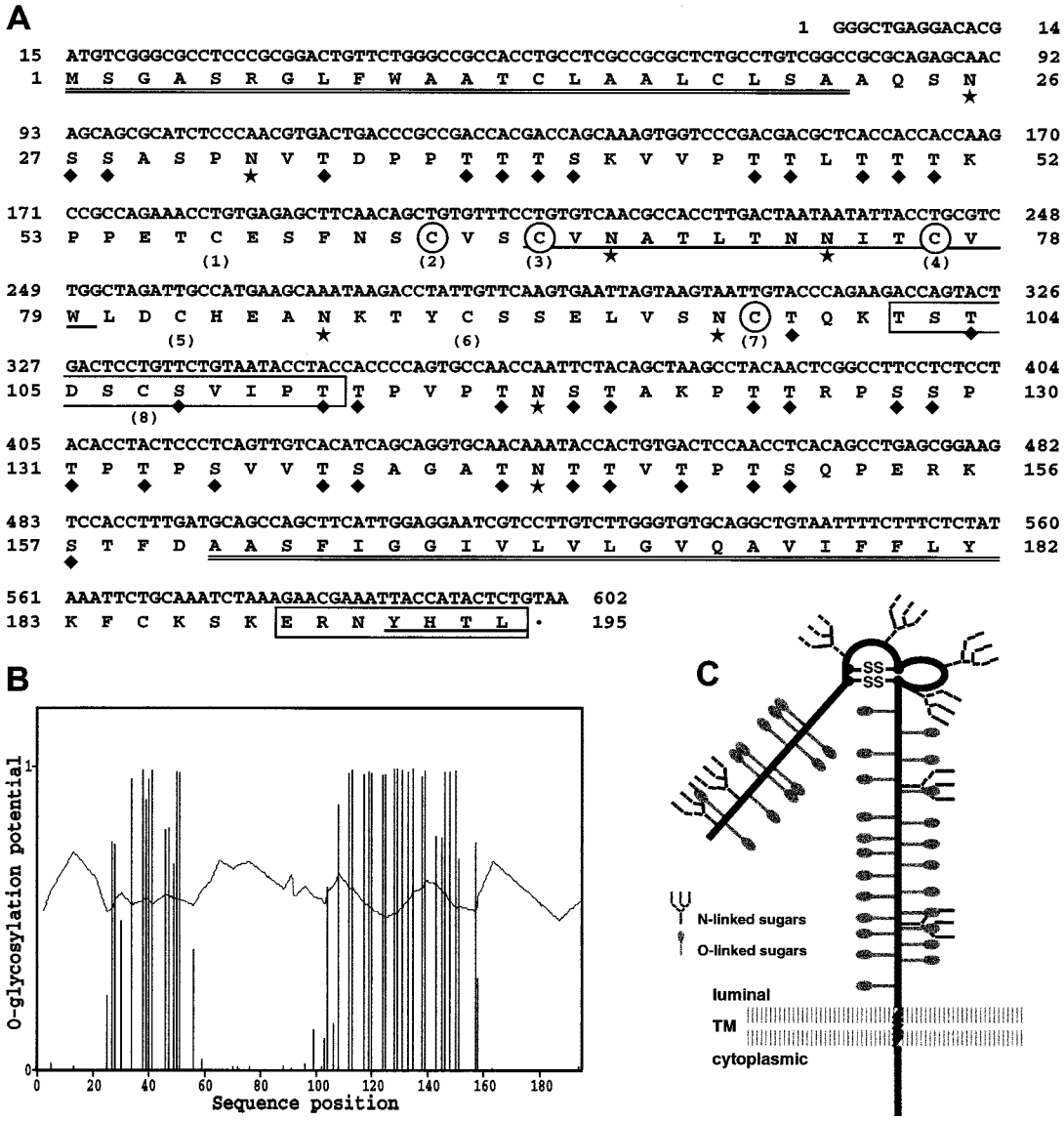

Figure 4 cDNA sequence, derived amino acid sequence and structural predictions of rat endolyn

(**A**) cDNA corresponding to the ORF and 5′-untranslated region of endolyn. Stretches of hydrophobic amino acids corresponding to the predicted signal sequence (amino acids 1–22) and the putative TM domain of the mature protein (amino acids 161–182) are doubly underlined. Boxes highlight sequences that were confirmed by tryptic peptides derived from purified rat liver endolyn. Potential N-glycosylation sites and likely O-glycosylation sites [see (**B**)] are indicated by asterisks and diamonds respectively. Cysteine residues of the predicted mature protein are numbered 1–8 in parentheses, those that may form disulphide bonds [pairing the second (Cys^{63}) and seventh (Cys^{98}) and the third (Cys^{66}) and fourth (Cys^{77}) cysteine respectively], by analogy to other proteins, are encircled. Underlined amino acids 66–79 correspond to the cytokine receptors signature 1 consensus pattern, and the underlined C-terminal four amino acids, YHTL, conform to a targeting motif of the type YXXØ (where Ø is a bulky hydrophobic amino acid). The complete nucleotide sequence of rat endolyn cDNA has been deposited in the EMBL Nucleotide Sequence Database under the accession number AJ238574. (**B**) Assignment of O-glycosylation sites by the NetOglyc 2.0 program [28]. The O-glycosylation potentials of serine and threonine residues are indicated with vertical lines, and threshold values by the horizontal line. (**C**) A diagram showing the possible topology and structure of endolyn based on the predictions stated above and in the text. The position of the N-terminal O-glycosylated domain relative to the membrane cannot be predicted from the sequence data; it may be the most distal part of the molecule protruding into the lumen.

in λgt11 using the ^{32}P-labelled Ly9 insert cDNA as probe (Figure 3). The inserts of 13 analysed clones were found to be different fragments of the same cDNA. Three clones were fully sequenced (2, 23 and 25; see Figure 3). Clone 2 contained a predicted open reading frame (ORF) of 195 amino acids and short stretches of the 5′- and 3′-untranslated regions. Clone 23 contained an identical ORF, but extended over a long 3′-untranslated region of 2160 bp followed by a poly(A) tail. Clone 25 lacked part of the ORF, but had the longest 3′-untranslated region (2218 bp) and was the only clone containing a standard polyadenylation signal (AATAAA) 25 bp upstream of the poly(A) tail. A rare polyadenylation signal (ATTAAA) was present 13 bp upstream of the poly(A) tail of clone 23. Clones 2 and 23 contained a 14 bp-long stretch of 5′-untranslated region preceding a translation initiator codon that conformed to the Kozak consensus sequence (see Figure 4A).

To verify the identity of the MGC-24-like rat membrane protein with endolyn, we compared peptide sequences obtained by microsequencing of trypsin fragments of purified rat liver endolyn (see the Experimental section) with the deduced amino

Table 1 Percentage similarity (identity) of endolyn to its homologues

Percentage similarity (identity) of the protein sequences was calculated using the GAP program for pairwise comparison of sequences (GCG package; Genetics Computer Group, Madison, WI, U.S.A.; blosum62 matrix, gap weight 8, length weight 2). mMGC, mouse MGC; hMGC, human MGC.

	mMGC-24v	CD164	hMGC-24	*Drosophila* EST
Overall homology	83 (82)	66 (64)	57 (55)	41 (40)
Luminal domain	78 (77)	57 (55)	61 (59)	34 (34)
TM domain + tail	100 (100)	100 (100)	40 (37)	69 (63)

acid sequence of the cloned protein. Two peptide sequences were obtained which completely matched the translated cDNA sequence confirming that the isolated cDNA encoded rat endolyn. The first peptide (TSTDSCSVIPT) corresponded to amino acids 102–112 of the deduced protein sequence, whereas the second peptide (ERNYHTL) was identical to its C-terminus (see Figure 4A).

Endolyn cDNA encodes a type I membrane protein containing a YXXØ motif in its cytoplasmic tail

The predicted amino acid sequence of the endolyn cDNA has the characteristics of a type I integral membrane protein. The N-terminal 22 amino acids constitute a typical signal sequence with a hydrophobic core and a putative signal sequence cleavage site between two alanine residues [25]. A second 22-amino-acid long hydrophobic stretch near the C-terminus is the predicted TM domain, which is followed by a short cytoplasmic tail of 13 amino acids (Figure 4A). The last four C-terminal amino acids, YHTL, constitute a YXXØ motif similar to those found in many

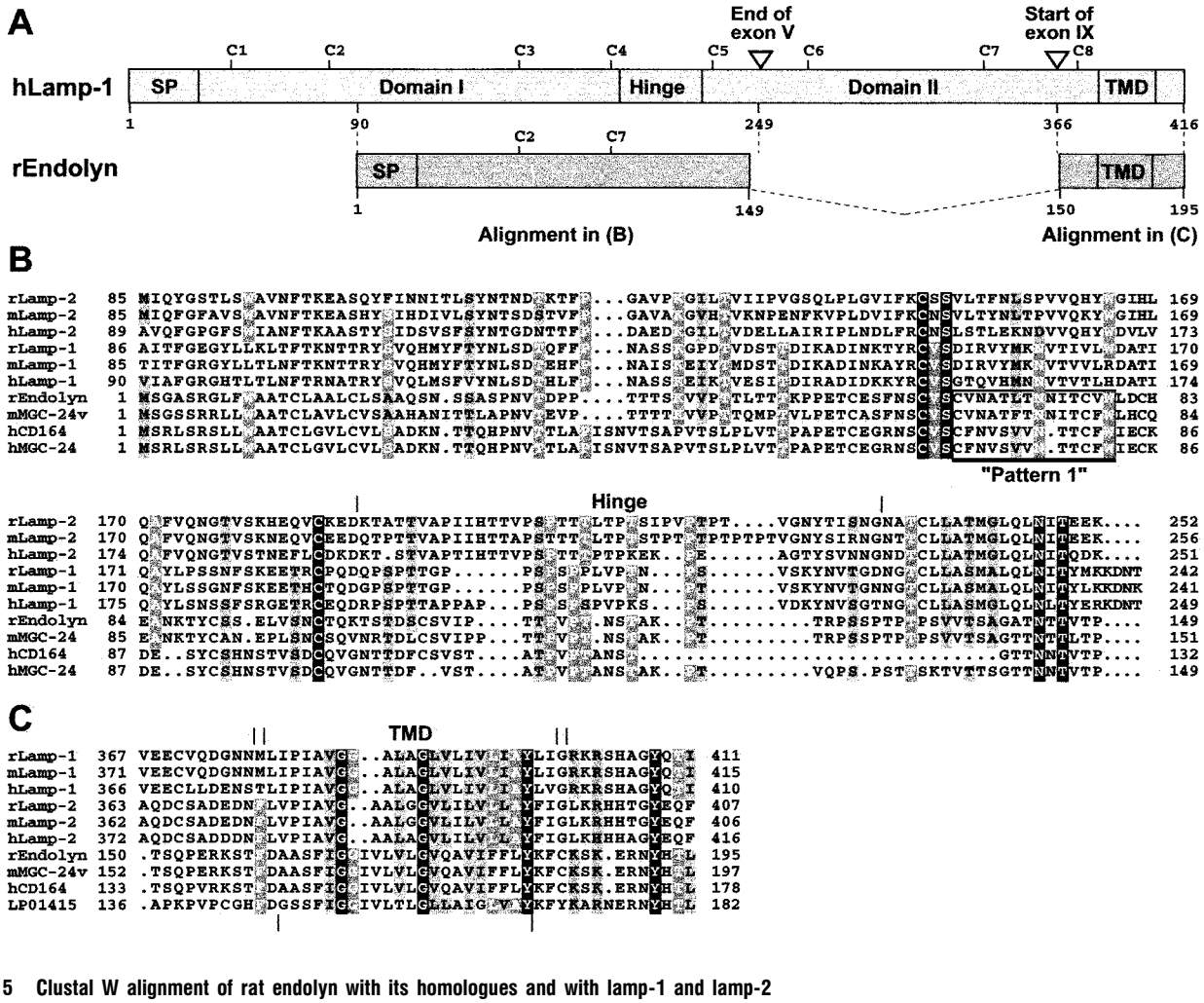

Figure 5 Clustal W alignment of rat endolyn with its homologues and with lamp-1 and lamp-2

(**A**) Schematic representation of the alignment blocks in (**B**) and (**C**) using human lamp-1 and rat endolyn as examples. The domain structure of lamp-1 corresponds to published data extracted from the Swiss-Prot database; the exon limits are taken from Fukuda [6]; SP, signal peptide; TMD, TM domain, C, cysteine; r, rat; m, mouse; h, human; vertical lines above and below the sequences in (**B**) and (**C**) demarcate domain boundaries. (**B**) The complete protein sequences of rat, mouse and human lamp-1 (rLamp-1, mLamp-1 and hLamp-1) and lamp-2(rLamp-2, mLamp-2 and hLamp-2), rat endolyn (rEndolyn), mouse MGC-24v (mMGC-24v), human CD164 (hCD164; hMGC-24v) and human MGC-24 (hMGC-24) were aligned using the Clustal W program; only the main alignment block, ending at the translated 3′ end of exon V in lamp 1 (human) and lamp-2 (human and mouse), is shown. White letters on a black background indicate identical amino acids in all sequences, white letters on a grey background or black letters on a grey background are used when at least 7/10 sequences have identical or similar amino acids respectively. (**C**) To show the similarity between the TM and tail domains of the same proteins, the last 46 amino acids of the lamps, corresponding to exon IX, were aligned with the C-terminal 46 amino acids of endolyn and its homologues containing a predicted TM region. Also included is the C-terminus of a translated EST found in the *Drosophila* database (LP01451). Note that both lamp-1 and endolyn-like sequences contain an unusual Gly–Gly motif in the TM domain. Shading as described for (**B**).

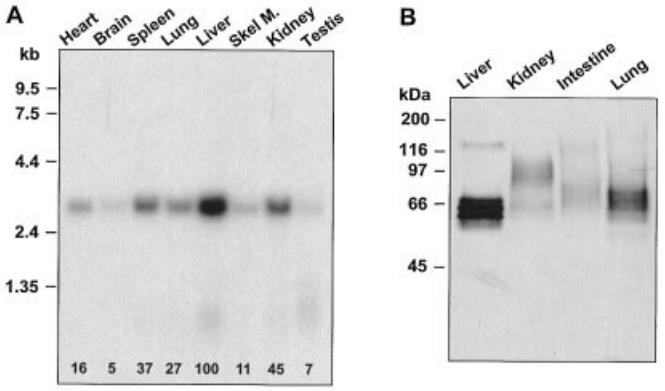

Figure 6 Distribution of endolyn in rat tissues

(**A**) Northern blot with mRNAs from eight different rat tissues (as indicated) was hybridized with the ^{32}P-labelled insert of endolyn clone 2 and exposed to X-ray film or analysed by phosphorimaging to quantify the relative intensity of radiolabelled bands. An autoradiograph from a short exposure to demonstrate the higher abundance of endolyn message in liver relative to all other tissues is shown. Endolyn mRNA was detectable in all tissues examined, including brain and testis; the numbers given at the bottom of the lanes indicate the abundance relative to the signal from liver mRNA set at 100. (**B**) Western blot of total protein fractions (non-reduced) from different tissues containing polarized cells was probed with a mixture of mAb 501 and mAb 502. Molecular-mass markers (kDa) are shown on the left.

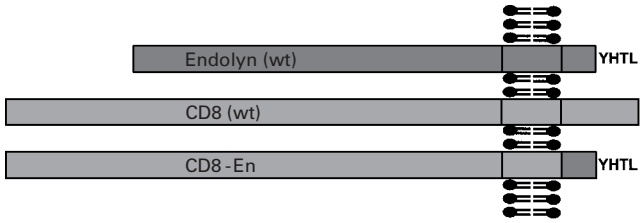

Figure 7 Diagram of the different constructs used for transfection of HeLa or NRK cells

The cDNAs of rat endolyn, human CD8 (α-chain) or a chimaera of the two (CD8–En), were cloned into the inducible mammalian expression vector ΔpMEP4. The CD8–En construct contains the luminal and TM domains of CD8 and the cytoplasmic tail of endolyn with the putative YHTL-targeting motif at its C-terminus. wt, wild type.

other membrane proteins interacting with clathrin adaptors and adaptor-like complexes [26].

Further database searches for sequences related to endolyn revealed the existence of a highly conserved mouse cDNA (mMGC-24v; Table 1 and Figure 5). Besides the human MGC-24, which does not appear to be an integral membrane protein, a second human splice variant (CD164/MGC-24v) containing the same predicted TM and tail domains as endolyn and mMGC-24v was identified recently [27] (Table 1 and Figure 5). However, this protein lacks a 20-amino-acid-long section that is found in the luminal domains of the rat and mouse sequences close to the TM domain (Figure 5B). A third splice variant of the human protein similar to endolyn and mMGC-24v over its entire length does exist, since we found an EST (AL040719 from testis) that contained an almost complete ORF including the insertion, with all other parts being practically identical to CD164 at the cDNA level. We also found a sequence closely related to endolyn in the *Drosophila* database (cDNA clone LP01451) which appears to contain a complete ORF encoding a type I transmembrane protein of approximately similar size to endolyn (Table 1 and Figure 5C).

Endolyn is a mucin-like protein with a conserved disulphide-containing motif

Closer inspection of the endolyn cDNA revealed a number of characteristic features (highlighted in Figure 4A). Most striking was the high abundance of possible O-glycosylation sites. Of the serines and threonines, 35 of 53 residues in the assumed luminal domain were found to have a high potential to be glycosylated according to the NetOglyc prediction program (Figure 4B) [28]. In addition, eight potential N-glycosylation sites were counted, four of which were positioned in the central part of the sequence containing no predicted O-glycosylation sites (between the first and the seventh cysteine residue; see Figure 4A). The serine- and threonine-rich regions found on either side of these cysteines also contain many proline residues, similar to the 'PTS regions' found in mucins. The mucins Muc5C and Muc2 are among the proteins that share a significant degree of similarity with endolyn, e.g. 52% similarity (33% identity) to Muc2 over 128 amino acids. Both human MGC-24v/CD164 and mouse MGC-24v have a similar overall organization to rat endolyn (Figure 5B), while the luminal domain of the putative *Drosophila* protein is mucin-like over its entire length (results not shown).

A search for conserved motifs shared with known proteins in the Prosite database revealed the existence of the consensus pattern C-[LVFYR]-X(7,8)-[STIVDN]-C-X-W, previously found only in proteins of the family of growth factor and cytokine receptors (PS00241, cytokine receptor signature 1). Cysteines in this motif are linked by disulphide bonds in the receptors [29]. The corresponding cysteines in the endolyn sequence [the third (Cys^{66}) and fourth (Cys^{77})] lie within the putatively N-glycosylated central part between the mucin-like regions (Figures 4A and 5B).

Endolyn is related to other lysosomal membrane proteins

The general features of endolyn, i.e. as a type I membrane protein with a heavily glycosylated luminal domain and a short cytoplasmic tail ending with a YXXØ motif, suggested that it might be related to other type I lysosomal membrane glycoproteins. Thus a sequence alignment of rat endolyn, mouse MGC-24v, human CD164 and human MGC-24 with rat, mouse and human lamp-1 and lamp-2 was carried out (Figure 5). This analysis indicated that the central part of endolyn and its homologues, the sequence surrounding and including the pattern 1 motif, may be related to domain I and the hinge region of lamp-1 and lamp-2 (Figures 5A and 5B; and see the Experimental section). Although the lamps do not contain the pattern 1 motif itself, cysteines on either side of this pattern in endolyn may correspond to the conserved Cys-3 and Cys-4 of lamp-1 and lamp-2. This assumption is also supported by the presence of two tryptophan residues at similar distances relative to these cysteines in both the endolyn homologues and lamp-2. There is an obvious sequence similarity among the TM domains and cytoplasmic tails of the lamps and the endolyn-like membrane proteins, including the putative *Drosophila* protein (Figure 5C). The last 36–38 amino acids show a similarity of 32% to 38%, which is consistent with the notion that endolyn and the 'classical' type I lysosomal membrane proteins may be evolutionarily related. All of these proteins have a C-terminal YXXØ motif but, unlike

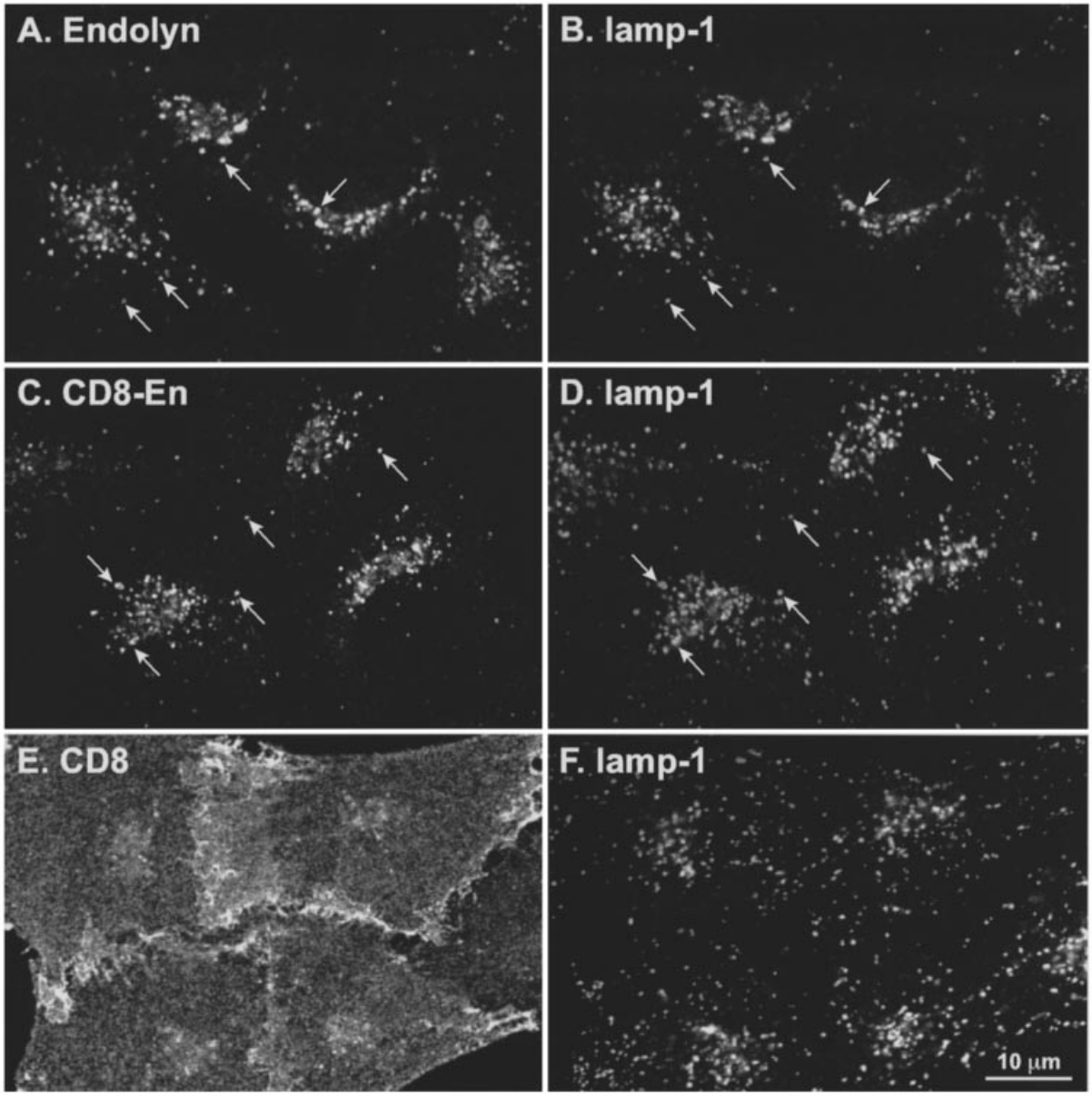

Figure 8 The cytoplasmic tail of endolyn targets CD8 to endosomes and lysosomes

The steady-state distributions of endogenous endolyn and lamp-1 in non-transfected NRK cells are shown in (**A**) and (**B**) respectively. (**C** and **D**) Stably transfected NRK cells expressing the CD8–En chimaera were induced overnight and cultured for a further 8 h in the absence of metal ions and in the presence of leupeptin for the last 6 h before fixation and permeabilization. (**E** and **F**) Stably transfected NRK cells expressing wild-type CD8 constitutively at moderate levels were not induced, but otherwise treated similarly to cells in (**C** and **D**); the addition of metal ions enhanced expression, but did not cause transport of CD8 to lysosomes (results not shown). Confocal images show the localization of CD8–En (**C**) and CD8 (**E**) visualized by a mAb to CD8. Images in the right panel show lamp-1 detected by a pAb (**B**) or a mAb (**D** and **F**). Arrows in (**A**) and (**B**) or (**C**) and (**D**) indicate some of the many structures that were positive for both endolyn and lamp-1 or CD8–En and lamp-1 respectively. (Note that the concentrations of endolyn and CD8–En at the plasma membrane relative to the concentrations in intracellular compartments were too low to be visualized by indirect immunofluorescence using permeabilized cells.)

lamp-1 and lamp-2, endolyn does not contain a glycine residue preceding the YXXØ motif.

Endolyn is ubiquitously expressed, but most highly in liver and kidney

Most known lysosomal membrane proteins are ubiquitously expressed in all mammalian tissues. Northern-blot analysis, using the clone 2 insert as a probe, showed that an endolyn transcript of 2.8–2.9 kb was present in all tissues examined (Figure 6A). The message was greatest in liver, followed by kidney, spleen and lung, all of which are tissues containing polarized epithelial cells. These data correlated well with Western blots of total protein fractions derived from various rat tissues probed with mAbs to endolyn. Endolyn protein was most highly expressed in liver, with high levels of expression found also in kidney, lung and intestine (Figure 6B). Although there was only one major mRNA species of endolyn detected by Northern blotting, the Western

blots showed the presence of a more or less well separated group of bands between 60 and 100 kDa, probably largely due to variations in the glycosylation pattern. The bands corresponding to kidney endolyn were seen at approx. 80–100 kDa, which was in agreement with the reported molecular mass of endolyn-78 in NRK cells [3]. Endolyn isoforms in all other tissues examined had a somewhat lower apparent molecular mass ($\sim$ 60–80 kDa or $\sim$ 70-90 kDa).

The cytoplasmic tail of endolyn targets a CD8 chimaera to endosomes and lysosomes

Our prime interest in studying endolyn was to understand the signals contained within its sequence that govern its trafficking throughout the cell. To confirm that wild-type endolyn, exogenously expressed in cultured cells, displayed the expected endosomal/lysosomal distribution we cloned the full-length cDNA into the inducible expression vector ΔpMEP4 and transiently transfected human HeLa cells with the construct. Endogenous human endolyn is not recognized by mAb 501 or mAb 502. Examination of the distribution of expressed rat endolyn by indirect immunofluorescence microscopy showed that the protein partially co-localized with the lysosomal enzyme cathepsin D and the cation-independent mannose 6-phosphate receptor (results not shown). This was consistent with the expected targeting of the construct to lysosomes and endosomes, similar to endogenous endolyn in rat cells.

The first question we wished to answer was whether the cytoplasmic tail containing the C-terminal YXXØ motif was sufficient to confer lysosomal sorting to a reporter molecule, despite the absence of a preceding glycine residue. We chose the human T-cell surface marker CD8 (α-chain of human CD8; [17]), also a type I membrane protein, as reporter protein, since its usefulness as a 'neutral' reporter is well established (e.g. [18,30]). We constructed a chimaera containing the luminal and TM domains of CD8 and the cytoplasmic tail of endolyn (Figure 7). This construct and, as a control, wild-type CD8 cDNA were cloned into ΔpMEP4 and the distributions of expressed proteins in stably transfected NRK cells were evaluated by double immunofluorescence confocal microscopy. The localization of the CD8–endolyn chimaera (CD8-En) was essentially indistinguishable from that of endogenous endolyn (compare Figure 8C with 8A). In both cases, there was a high degree of co-localization with lamp1 (Figures 8A and 8B, 8C and 8D). In contrast, wild-type CD8 was almost exclusively found at the cell surface and co-localization of any intracellular CD8 with lamp-1 was negligible (Figures 8E and 8F). These results show that targeting information contained within the last ten amino acids of the cytoplasmic tail of endolyn, including the C-terminal NYHTL motif, is sufficient to direct a reporter construct to endosomes and lysosomes.

DISCUSSION

Endolyn is a membrane-bound mucin related to type I lysosomal membrane proteins

The cDNA sequence of endolyn has been elusive since the biochemical and histochemical description of this highly glycosylated protein a decade ago [3]. The cDNA of rat liver endolyn, identified in this study, encodes a protein sequence with expected characteristics, namely a membrane protein with a high number of glycosylation sites and containing a lysosomal targeting motif. The derived amino acid sequence predicts a type I membrane protein which, after cleavage of the signal peptide, consists of 173 amino acids with a molecular mass of 18296 Da. This is far lower than the molecular mass of 60 to $>$ 80 kDa for mature endolyn observed by Western blotting, but in close agreement with the reported value of 22 kDa for the unglycosylated precursor [3]. The variation in molecular mass of endolyn among different tissues found by Western blotting, presumably due to differences in glycosylation, prompted us to name the protein endolyn rather than endolyn-78 [3].

Comparison with sequences available in cDNA databases show that homologues of endolyn exist in other species; however, little is known about the biochemical characteristics and the subcellular distribution of the corresponding proteins. So far, only rat endolyn has been recognized as an endosomal/lysosomal membrane protein [3,4]. There are at least three splice variants of human MGC-24: the originally described non-transmembrane variant expressed at the surface of human gastric carcinoma cells [24], CD164 (human MGC-24v), which appears to be differentially expressed at the cell surface of various lineages of haematopoietic cells [27,31] and a third protein entirely homologous with rat endolyn predicted by the presence of an EST from human testis and, presumably, more ubiquitously expressed.

Endolyn and its homologues can be defined as mucins under the following criteria [32]: first, they contain high amounts of proline, serine and threonine residues, which constitute together $>$ 40 % (20–55 %) of all amino acids and are concentrated in certain regions of the polypeptide; secondly, many serine and threonine residues within the 'PTS regions' are glycosylated, with the O-linked oligosaccharides contributing $\geqslant$ 40 % of the total mass of the protein. In recent years several other membrane-bound mucins (sialomucins) have been described; these are implicated in cell–cell interactions (for a review see [32]). A similar role, mediating the adhesion of haematopoietic progenitor cells to bone marrow stromal cells, has been suggested for CD164 [27] but, at present, there is no indication that endolyn has such a function in non-haematopoietic cells.

A surprising finding was that endolyn also shared a highly conserved motif with the growth factor and cytokine receptor superfamily (signature 1), which has, so far, been described only for members of this protein family. In these, it is one of two conserved motifs, each containing two cysteine residues implicated in the formation of disulphide bonds [33]. Since endolyn contains only one of several conserved features implicated in the formation of the cytokine binding pocket, it seems unlikely that it is able to bind any cytokine ligands via this motif. However, analogous to the receptors, the cysteines within the pattern 1 in endolyn may form a disulphide bridge and thus cause a bend in the molecule (see Figure 4C). Curiously, the pattern 1 motif lies within the suggested conserved cysteines that endolyn shares with the lamp family (see below). If both cysteine pairs in endolyn form similar disulphide bridges, this would create a stabilized loop structure, which would have a strong influence on the overall tertiary structure of the luminal domain. The assumed loop region, which possesses four of the eight potential N-glycosylation sites, may be exposed towards the cell lumen; the adjacent mucin-like domains are more likely to be rod-like [34] (see Figure 4C).

Sequence alignment with other type I lysosomal membrane proteins, in particular lamp-1 and lamp-2, suggests that endolyn is distantly related to this protein family. The luminal part of the lamps consists of two domains which are internally homologous to each other and are connected by a proline-rich, O-glycosylated hinge region. Both domains contain four conserved cysteine residues, which form two disulphide-bonded loops (for a review see [6]). Two of the cysteine residues in endolyn align with the

conserved Cys-3 and Cys-4 in domain I of lamp-1 and lamp-2. The TM and cytoplasmic domains of endolyn and these lamps (22–25 and 11–13 amino acids long respectively) share an overall sequence similarity of approx. 35%. Two other mucin-like lysosomal type I membrane proteins, related to the lamps, have been described recently, CD68 [35] and TSC403/DC-lamp [36,37]. Both proteins have a similar tail and TM region, a membrane-proximal luminal domain, related to domain II of the lamps, and an extended proline-rich and O-glycosylated membrane-distal domain. However, there is no close sequence similarity between these two proteins and endolyn.

Targeting information contained in the cytoplasmic tail directs endolyn to endosomes and lysosomes

The short cytoplasmic tail of endolyn ends in the sequence YHTL. This is a classical clathrin-coated-pit internalization signal of the type YXXØ, which would be expected to interact with the μ-subunits of clathrin-adaptor proteins (AP1 and AP2) and adaptor-like complexes (AP3 and AP4) [26,38–41]. We have shown that a short cytoplasmic tail containing the last ten amino acids of endolyn is sufficient to target a reporter plasma-membrane protein (CD8) to lysosomes. This is consistent with data for other lysosomal glycoproteins [8,42,43]. Various studies show that the tyrosine residue in the YXXØ motif is of central importance, since mutagenesis of this tyrosine to any other residue leads to accumulation of lysosomal membrane proteins at the cell surface [7,8,42,43]. More recent studies *in vitro* have shown that only the tyrosine residue and a bulky, hydrophobic amino acid in position Tyr+3 are absolutely required for the interaction of YXXØ motifs with the μ-subunits of adaptor complexes [40,44,45]. A leucine residue as the bulky hydrophobic amino acid in the Tyr+3-position, as in endolyn, enhances the affinity of the motif to all μ-subunits [46]. No other amino acid in the ERNYHTL sequence found in endolyn has been described as either especially favoured or disfavoured by different μ-subunits. Thus this motif may be able to interact, to some degree, with all known adaptor complexes *in vivo*. Recognition of targeting motifs by the lysosomal sorting machinery also depends on the correct distance (approx. 8 amino acids) from the TM domain [43,47] and is significantly enhanced when located at the very C-terminus [45,48]. Meeting both these criteria, the YHTL sequence is likely to mediate the efficient sorting of endolyn to lysosomes. The relatively greater appearance of endolyn at the plasma membrane and endocytic compartments, when compared with other lysosomal membrane proteins (e.g. lamp-1/lgp120) [3,4], suggests that a larger fraction of endolyn molecules travels via the plasma membrane before being targeted to lysosomes. Thus the absence of a glycine residue in the Tyr-1 position in endolyn, which is present in most other lysosomal membrane proteins with a YXXØ motif, may correlate with a lower sorting efficiency at the TGN [7,8]. It is presently unclear whether the YXXØ motif or other signals in the cytosolic, luminal or TM domain mediate the trafficking of endolyn to non-lysosomal compartments, such as the SAC, in polarized cells. With the cDNA of endolyn, we now have the necessary tool in hand to dissect further the structural determinants that govern the intracellular itinerary of this protein and distinguish it from that of other lysosomal membrane proteins.

This work was supported by the Wellcome Trust and Medical Research Council. We thank Dr Sean Munro for the gift of a plasmid construct encoding CD8 and Dr Geoff Hale, Professor Masaru Himeno, Professor John Hutton, Professor Keitaro Kato, Professor Ken Siddle and Dr Yoshitaka Tanaka for generously providing us with antibodies. We also thank Dr Rainer Duden and Dr Margaret S. Robinson for a critical reading of the manuscript and many helpful comments.

REFERENCES

1 Peters, C. and von Figura, K. (1994) FEBS Lett. **346**, 108–114
2 Hunziker, W. and Geuze, H. J. (1996) Bioessays **18**, 379–389
3 Croze, E., Ivanov, I. E., Kreibich, G., Adesnik, M., Sabatini, D. D. and Rosenfeld, M. G. (1989) J. Cell Biol. **108**, 1597–1613
4 Ihrke, G., Martin, G. V., Shanks, M. R., Schrader, M., Schroer, T. A. and Hubbard, A. L. (1998) J. Cell Biol. **141**, 115–133
5 Barr, V. A., Scott, L. J. and Hubbard, A. L. (1995) J. Biol. Chem. **270**, 27834–27844
6 Fukuda, M. (1994) Subcell. Biochem. **22**, 199–230
7 Harter, C. and Mellman, I. (1992) J. Cell Biol. **117**, 311–325
8 Höning, S. and Hunziker, W. (1995) J. Cell Biol. **128**, 321–332
9 Reaves, B. J., Bright, N. A., Mullock, B. M. and Luzio, J. P. (1996) J. Cell Sci. **109**, 749–762
10 Horn, M. and Banting, G. (1994) Biochem. J. **301**, 69–73
11 Bindon, C. I., Hale, G. and Waldmann, H. (1989) in Leucocyte Typing IV. White Cell Differentiation Antigens (Knapp, W., Dorken, B., Gilks, W. R., Rieber, E. P., Schmidt, R. E., Stein, H. and von dem Borne, A. E. G. K., eds.), pp. 349–350, Oxford University Press, Oxford
12 Furuno, K., Ishikawa, T., Akasaki, K., Yano, S., Tanaka, Y., Yamaguchi, Y., Tsuji, H., Himeno, M. and Kato, K. (1989) J. Biochem. (Tokyo) **106**, 708–716
13 Grimaldi, K. A., Hutton, J. C. and Siddle, K. (1987) Biochem. J. **245**, 557–566
14 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
15 Luzio, J. P., Brake, B., Banting, G., Howell, K., Bressan, G., Braghetta, P. and Stanley, K. K. (1990) Biochem. Soc. Trans. **18**, 148–149
16 Girotti, M. and Banting, G. (1996) J. Cell Sci. **109**, 2915–2926
17 Littman, D. R., Thomas, Y., Maddon, P. J., Chess, L. and Axel, R. (1985) Cell **40**, 237–246
18 Munro, S. (1995) EMBO J. **14**, 4695–4704
19 Reaves, B. J., Banting, G. and Luzio, J. P. (1998) Mol. Biol. Cell **9**, 1107–1122
20 Ohsumi, Y., Ishikawa, T. and Kato, K. (1983) J. Biochem. (Tokyo) **93**, 547–556
21 Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. **157**, 105–132
22 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) Nucleic Acids Res. **22**, 4673–4680
23 Mullock, B. M., Bright, N. A., Fearon, C. W., Gray, S. R. and Luzio, J. P. (1998) J. Cell Biol. **140**, 591–601
24 Masuzawa, Y., Miyauchi, T., Hamanoue, M., Ando, S., Yoshida, J., Takao, S., Shimazu, H., Adachi, M. and Muramatsu, T. (1992) J. Biochem. (Tokyo) **112**, 609–615
25 von Heijne, G. (1984) J. Mol. Biol. **173**, 243–251
26 Hirst, J. and Robinson, M. S. (1998) Biochim. Biophys. Acta **1404**, 173–193
27 Zannettino, A. C., Buhring, H. J., Niutta, S., Watt, S. M., Benton, M. A. and Simmons, P. J. (1998) Blood **92**, 2613–2628
28 Hansen, J. E., Lund, O., Engelbrecht, J., Bohr, H. and Nielsen, J. O. (1995) Biochem. J. **308**, 801–813
29 Fuh, G., Mulkerrin, M. G., Bass, S., McFarland, N., Brochier, M., Bourell, J. H., Light, D. R. and Wells, J. A. (1990) J. Biol. Chem. **265**, 3111–3115
30 Ponnambalam, S., Rabouille, C., Luzio, J. P., Nilsson, T. and Warren, G. (1994) J. Cell Biol. **125**, 253–268
31 Watt, S. M., Buhring, H. J., Rappold, I., Chan, J. Y., Lee-Prudhoe, J., Jones, T., Zannettino, A. C., Simmons, P. J., Doyonnas, R., Sheer, D. and Butler, L. H. (1998) Blood **92**, 849–866
32 van Klinken, B. J., Dekker, J., Buller, H. A. and Einerhand, A. W. (1995) Am. J. Physiol. **269**, G613–G627
33 Bazan, J. F. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 6934–6938
34 Carraway, K. L. and Hull, S. R. (1991) Glycobiology **1**, 131–138
35 Holness, C. L. and Simmons, D. L. (1993) Blood **81**, 1607–1613
36 Ozaki, K., Nagata, M., Suzuki, M., Fujiwara, T., Ueda, K., Miyoshi, Y., Takahashi, E. and Nakamura, Y. (1998) Cancer Res. **58**, 3499–3503
37 de Saint-Vis, B., Vincent, J., Vandenabeele, S., Vanbervliet, B., Pin, J. J., Ait-Yahia, S., Patel, S., Mattei, M. G., Banchereau, J., Zurawski, S., Davoust, J., Caux, C. and Lebecque, S. (1998) Immunity **9**, 325–336
38 Kirchhausen, T., Bonifacino, J. S. and Riezman, H. (1997) Curr. Opin. Cell Biol. **9**, 488–495
39 Marks, M. S., Ohno, H., Kirchhausen, T. and Bonifacino, J. S. (1997) Trends Cell Biol. **7**, 124–128
40 Owen, D. J. and Evans, P. R. (1998) Science **282**, 1327–1332

41 Hirst, J., Bright, N. A., Rous, B. and Robinson, M. S. (1999) Mol. Biol. Cell **10**, 2787–2802
42 Williams, M. A. and Fukuda, M. (1990) J. Cell Biol. **111**, 955–966
43 Guarnieri, F. G., Arterburn, L. M., Penno, M. B., Cha, Y. and August, J. T. (1993) J. Biol. Chem. **268**, 1941–1946
44 Boll, W., Ohno, H., Songyang, Z., Rapoport, I., Cantley, L. C., Bonifacino, J. S. and Kirchhausen, T. (1996) EMBO J. **15**, 5789–5795
45 Ohno, H., Fournier, M. C., Poy, G. and Bonifacino, J. S. (1996) J. Biol. Chem. **271**, 29009–29015
46 Ohno, H., Aguilar, R. C., Yeh, D., Taura, D., Saito, T. and Bonifacino, J. S. (1998) J. Biol. Chem. **273**, 25915–25921
47 Rohrer, J., Schweizer, A., Russell, D. and Kornfeld, S. (1996) J. Cell Biol. **132**, 565–576
48 Stephens, D. J. and Banting, G. (1998) Biochem. J. **335**, 567–572

Received 26 August 1999/30 September 1999; accepted 1 November 1999