The Basolateral Targeting Signal in the Cytoplasmic Domain of Glycoprotein G from Vesicular Stomatitis Virus Resembles a Variety of Intracellular Targeting Motifs Related by Primary Sequence but Having Diverse Targeting Activities*

(Received for publication, January 25, 1994, and in revised form, March 23, 1994)

D'Nette C. Thomas and Michael G. Roth†
From the Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75235-9038

Using systematic site-directed mutagenesis, the basolateral targeting signal in the cytoplasmic domain of glycoprotein G from vesicular stomatitis virus (VSV G) has been localized to an 11-amino acid sequence, which contains two essential residues and a third that makes a minor contribution. A tyrosine at position 19 of the 29-residue carboxyl-terminal cytoplasmic tail is the most important residue and cannot be replaced by other aromatic amino acids, while an isoleucine at position 22, 3 residues carboxyl-terminal to this tyrosine, is also critical but can be replaced by other aliphatic residues. Additionally, an arginine at position 16 makes a minor contribution. Therefore the crucial elements of this targeting signal can be represented by the sequence Y-X-X-aliphatic. While earlier investigation has suggested similarity between basolateral targeting and internalization signals, alignment of this sequence with other cytoplasmic targeting signals suggests the existence of a broad class of homologous targeting motifs that direct protein delivery to a variety of cellular locations. This in turn suggests the existence of a family of homologous receptors, distributed throughout the cell, which differ in their affinity for subsets of these targeting sequences.

The plasma membrane of a polarized epithelial cell is typically divided into two domains with distinct compositions (1). Understanding the establishment of the non-uniform distribution of proteins in epithelial cell membranes is an active area of research (1–3). In the Madin-Darby canine kidney (MDCK) cell line, apical and basolateral proteins apparently co-migrate through the exocytic pathway until they reach the TGN, where they are "sorted" into separate exocytic vesicles destined for either the apical or the basolateral domain of the cell surface (4, 5). Recent experiments have suggested that short amino acid sequences in the cytoplasmic domains of transmembrane proteins function as basolateral targeting signals (6–12). Most of these signals resemble cytoplasmic internalization signals in that they contain important aromatic residues. Occasionally basolateral targeting signals are found in the same region of the primary sequence as internalization signals; however, in each case where signals have been more carefully investigated, basolateral targeting and internalization signals have proven to be distinct (8, 9, 13–15). A number of internalization signals have been analyzed in detail, leading to several proposals for consensus internalization motifs (16–19). However, in the case of basolateral targeting, only two such signals, one in the polymeric immunoglobulin receptor (pIg receptor) and one in lysosomal acid phosphatase, have been extensively analyzed and they appear to share no sequence similarity (9, 20). Before any consensus can be proposed for a basolateral targeting motif, a larger number of these signals need to be precisely defined.

Several earlier experiments suggested that the glycoprotein G from vesicular stomatitis virus (VSV G) might contain cytoplasmic basolateral targeting information (21, 22). The existence of a dominant basolateral targeting signal in the cytoplasmic domain of this protein was conclusively established when the tail sequence was shown to be sufficient to confer basolateral targeting upon a normally apical protein (10). The targeting information was shown to be completely dependent upon a unique tyrosine at position 19 of the 29-residue carboxyl-terminal tail (10). Additionally, high internalization rates measured for VSV G in other experiments were interpreted to mean that this cytoplasmic sequence also contained a coated pit localization signal (23, 24), which was postulated to involve the sequence YTDI (19). However, a complication of these experiments was that the internalization rate of G was measured after fusion of virus with the cell membrane, perhaps allowing aggregated patches of viral G protein to yield an artificially high internalization rate. Subsequent experiments, finding a much lower internalization rate for VSV G (10, 25), as well as a negligible effect of mutation of the tyrosine on the internalization rate (10), suggested that this protein did not contain a coated pit localization signal at all. Therefore, the identification of basal targeting information in the cytoplasmic tail of VSV G provided an opportunity for a detailed analysis of a basal targeting sequence, as well as the potential to identify characteristics that distinguish these signals from apparently similar internalization signals.

Mutational analysis performed in a chimera containing the cytoplasmic domain of VSV G and the luminal and transmembrane domains of the normally apical influenza hemagglutinin (HA) has allowed identification and characterization of a sequence which directs basal targeting of this chimera. The signal identified shows homology with many cytoplasmic targeting signals that are able to direct targeting to a variety of cellular locations in addition to the basal membrane and coated pits.

**MATERIALS AND METHODS**

**Construction of Mutants**—The gene for the G protein investigated in this work was originally derived from the San Juan strain of the Indi-
ana serotype of VSV (26). Construction of cDNAs encoding the HHG chimera and the HA<sup>ww</sup> construct has been previously reported (10, 27). The entire coding sequence of these genes was subcloned into the expression vector pCB6 (7) under the control of the cytomegalovirus immediate-early promoter. The HHG Y19S mutant was constructed using the M13-based mutagenesis protocol of Kunkel (28) modified slightly to use the plasmid pUC19. Site-directed mutations were introduced using the QuikChange Kit (Stratagene) according to the manufacturer's instructions. Mutants were constructed by one of several PCR-based mutagenesis techniques (30–32) using either an HHG or HHG R26t gene (where t = stop) as a PCR template and the Vent<sup>™</sup> polymerase (New England Biolabs) as the PCR enzyme. For the truncated constructs, codons following the last amino acid of each new construct were mutated to encode stop codons. Constructs containing internal deletions were made using mutagenic deoxyoligonucleotides with homology to the template on either side of the sequence to be deleted and lacking the sequence of the intended deletion. Point mutants were made with deoxyoligonucleotides designed to change the coding sequence of each codon individually, and in the case of the degenerate mutagenesis at positions Tyr-19 and Ile-22, a mixture of oligonucleotides was used to produce a variety of amino acids at each position. For each of these mutants, a 500-base pair restriction fragment generated from digestion of the mutagenesis product was subcloned into the expression vector pCB6 containing the remnant wild-type (wt) mutations as a cloning vector. These fragments were inserted into a plasmid carrying the entire HHG cytoplasmic domain of the region of the genes encoding the entire HHG cytoplasmic domain. Additionally, for seven of the mutants generated by PCR techniques, the complete 500-base pair subcloned fragment was sequenced to check for second site mutations; however, as no second site mutations were found, this was not considered to be a problem with the techniques used. All sequencing was performed according to the protocol supplied with the Sequenase kit version 2.0 (U. S. Biochemical Corp.). During the time course of targetting experiments, a similar proportion of each mutant (and wild-type) protein was processed to its fully glycosylated form and expression level was observed. Controls showed that this treatment did not affect the polarity of the wild-type HHG construct in these experiments, the uninfected monolayers of cells were washed several times with phosphate-buffered saline (138 mM NaCl, 2.7 mM KCl, 8.1 mM Na<sub>HPO</sub<sub>4</sub>, 1.2 mM KH<sub>2</sub>P<sub>O</sub><sub>4</sub>, pH 7.4) containing 1 mM Mg<sup>2+</sup> and 0.1 mM Ca<sup>2+</sup>, incubated for 30–45 min in Dulbecco's modified Eagle's medium (DMEM, Life Technologies, Inc.) containing 20 μg/ml soybean trypsin inhibitor (Sigma) to inhibit trypsin on both cell surfaces and any in endosomes. This medium was removed and replaced with ice-cold DMEM also containing 100 μg/ml soybean trypsin inhibitor, and the cells were rocked gently on ice for 30–45 min. The cells were then transferred to another set of tubes containing an equal volume of ice-cold lysis buffer (50 mM Tris-HCl, pH 8.0, 1% Nonidet P-40, 0.1% SDS, 5 μg/ml aprotinin; Sigma) containing 100 μg/ml soybean trypsin inhibitor. Cells were scraped into the lysis, which was centrifuged for 30 min at 12,000 × g. One-half to two-thirds of the supernatant was transferred to another set of tubes containing an equal volume of NETgel (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 150 mM NaCl, 0.25% gelatin, 0.05% Nonidet P-40, 0.01% NaN<sub>3</sub> + 3% bovine serum albumin, 100 μl of a 10% solution of protein A-Sepharose (Pharmacia Biotech), and 0.5 μl of anti-HA rabbit serum, which will bind to the external domain of HA present in all constructs. After allowing antibody and protein A binding to occur during gentle rocking either overnight at 4 °C or for 4 h at room temperature, the Sepharose was washed and proteins were eluted and analyzed by SDS-polyacrylamide gel electrophoresis as previously described (10). Gels were exposed to phosphor image screens to quantify the bands of the uncleaved proteins and their two cleavage products, HA1 and HA2 (model 400E Phosphorimager, Molecular Dynamics, Sunnyvale, CA). An autoradiograph of a sample gel can be found in Ref. 10.

The relative amount of each chimera delivered to the basal as opposed to the apical surface during the 1-h incubation was calculated as follows. The percentage of cleaved protein recovered as HA1 and HA2 when trypsin was present in either the apical or the basal medium of the two separate monolayers was first corrected for the level of endogenous cleavage measured in the third sample incubated without trypsin. This correction assumed that endogenous cleavage did not occur preferentially at either the apical or basolateral surface. The amount of trypsin-specific cleavage at the basolateral surface was then expressed as a percentage of total (apical + basolateral) protein. The numbers reported represent the percent of newly exocytosed protein that was initially delivered to the basolateral domain of the cell surface. Unless otherwise specified, reagents were purchased from Sigma.

RESULTS

The Basolateral Targeting Information in the Cytoplasmic Domain of VSV G Is Completely Contained within an 11-Amino Acid Sequence Spanning Tail Residues Lys-14 through Met-24—Mutational analysis of wild-type VSV G to define the basolateral targeting signal in the cytoplasmic domain of this protein was complicated by the possibility, supported by earlier experiments, that the luminal domain of this protein contained redundant basolateral targeting information (10, 33, 34). Therefore, mutagenesis was performed on a chimera, HHG, containing the cytoplasmic domain of VSV G and the luminal and transmembrane domains of the normally apical influenza HA. An added advantage of using this chimera was the availability of an assay that uses trypsin cleavage to monitor surface arrival of the protein. This assay, used in the past to monitor the surface arrival of a variety of HA constructs (7, 10), has proven to be more sensitive and more reproducible than techniques that require saturable, tight binding of a reagent such as an antibody or biotin to monitor protein distribution (35). To define the region of the G tail containing the basolateral targeting information, the HHG gene was mutated to encode a series of truncations and internal deletions in the G cytoplasmic domain and the mutated genes were subcloned into the expression vector pCB6 under the control of the cytomegalovirus early promoter. This DNA was transfected into MDCK cells and neomycin-resistant transfected cells were selected using G418. To monitor the exocytic targeting of the resulting proteins in the transfected MDCK cells, the cells were cultured on filters where they form polarized monolayers in which the apical and basolateral surface domains are separated by tight junctions.
The exocytic proteins were metabolically labeled with \(^{35}S\) by a brief incubation of the cells in \(^{35}S\)-labeled media. The cells were then incubated at 37 °C with trypsin in the media to allow the labeled protein to proceed through the exocytic pathway until it reached the cell surface and encountered the protease. In the presence of trypsin, HA\(^{1-13}\) and chimeras of HA containing foreign cytoplasmic domains were cleaved into two tryptic fragments that remained membrane-bound. After this incubation, the trypsin was inhibited, the cells were lysed, and the HHG construct present in the lysates was immunoprecipitated using antibodies to the external domain of HA. The precipitated proteins were viewed by SDS-polyacrylamide gel electrophoresis and autoradiography, and the polarity of surface delivery of each protein was determined by comparing the amount of tryptic fragments produced when trypsin was present in the basal medium of one monolayer to the amount produced when trypsin was present in the apical medium of a separate monolayer. Standard deviations are given when \(n\) (number of experiments) is greater than or equal to 3. The last column contains either \(n\) or actual experimental values when \(n = 2\). The sequence located beneath the series of constructs illustrates the region determined by this analysis to encode basolateral targeting information.

An Alanine Scan of Residues 14-24 Indicates That Tyr-19 and Ile-22 Are the Most Critical Residues in the Basal Targeting Signal, with Arg-16 Playing a Minor Role—After identification of the region of the tail containing basal targeting information, a set of point mutants was constructed to determine which amino acids within this 11-residue sequence contributed to the targeting signal. This series of mutations, as well as those to be described subsequently, were made in the background of R26t. To determine which residues played a positive role in targeting, each of the 11 residues between Lys-14 and Met-24 was individually changed to alanine and the polarity of surface arrival of each mutant protein was measured as described above. As illustrated in Fig. 2A, only 3 of the residues in this sequence seem to have a significant effect on the basal targeting activity, as measured by the “alanine scan.” Tyr-19 and Ile-22 both appear to be essential residues in this signal, as mutation of either one alone reduced the basal targeting of the protein to the background level of HA\(^{1-13}\). Arg-16 also makes a small but significant contribution to the basal targeting activity. Additionally, 3 of the residues neighboring the critical tyrosine and isoleucine, Thr-20, Asp-21, and Glu-23, were all changed to alanine in the same protein to determine what contribution the

### Table: Sequence and Polarity of Surface Delivery of HHG Proteins Bearing Deletions That Localize the Basal Targeting Information

| Construct | Percent Basal | Percent Apical |
|-----------|---------------|---------------|
| HHG       | 98 ±1        | 0 ±1          |
| HA\(^{1-13}\) | 26 ±3      | 74 ±3         |
| R29t      | 92 ±4        | 8 ±4          |
| MEA1      | 93            | 91 ±4         |
| I22t      | 83 ±3        | 17 ±3         |
| R16t      | 45            | 55 ±4         |
| Δ-1-13    | 90 ±4        | 10 ±4         |
| Δ-6-13(26t) | 96 ±1     | 4 ±1          |
| Δ-1-13    | 58            | 42 ±2         |

**Fig. 1. Sequence and Polarity of Surface Delivery of HHG Proteins Bearing Deletions That Localize the Basal Targeting Information.** The cytoplasmic domain and last 3 residues of the transmembrane domain of HHG are illustrated at the top of the figure. The bold black residues and block bars represent sequence originally derived from VSV G, while the outlined residues and the gray bars represent sequence derived from the transmembrane domain of HA. The carboxyl-terminal cytoplasmic tail of G has been numbered from 1 to 29 starting with the first amino acid of the cytoplasmic domain. For each construct, bars represent sequences retained while lines represent sequences deleted. The number reported as “percent basal” is the fraction of newly exocytosed protein delivered directly to the basal surface during a 1 h incubation, calculated as described under “Materials and Methods.” Standard deviations are given when \(n\) (number of experiments) is greater than or equal to 3. The last column contains either \(n\) or actual experimental values when \(n = 2\). The sequence located beneath the series of constructs illustrates the region determined by this analysis to encode basolateral targeting information.
**Aliphatic Residues**—To mutant illustrated across the horizontal delivered directly to the basal surface during a 1-h incubation. 

**Targeting Activity of the Tyrosine-Isoleucine Pair.** These effects sequence is listed along the Number of trials for each mutant are as follows: K14A, the phenylalanine 19 mutant indicates that the tyrosine in this was 86 that could substitute for tyrosine at position 19. The low basal targeting activity of the serine 19 and valine 19 mutants rules that these mutations at Tyr-19 was also tested in the full-length HHG and R26t, the polarity of the triple mutant was largely due to the loss of Arg-16 and not to diminished positive charge in this region. This KKR triple mutant was also constructed in the full-length HHG protein and, as reported for mutations made at the position of tyrosine 19, had essentially the same effect in the R26t construct and the full-length HHG. 

One additional mutant in the background of R26t, R16L, gave an interesting result. The substitution of a leucine for the arginine creates a protein that is significantly more basal (86 ± 2% basal, data not shown) than one containing an alanine at this position (73% basal, Fig. 2A), and in fact the leucine seems to be almost as functional as the wild-type arginine (93% basal, Fig. 2).

**DISCUSSION**

**The Basolateral Targeting Signal in the VSV G Cytoplasmic Tail Fits the General Motif of Y-X-Alphabetic**—Using deletion analysis, the basal targeting information encoded in the cytoplasmic domain of VSV G has now been shown to span residues Lys-14-Met-24 provided that a certain minimal separation between the beginning of the signal and the transmembrane domain is maintained. The last two residues of this region, 23 and 24, can be changed to alanine without affecting basal targeting. Deletion of these two residues however, by truncation down to overall physical properties of this region might make to the targeting activity of the tyrosine-isoleucine pair. These effects might not be detected by the mutation of individual residues to alanine. However, as illustrated in Fig. 2B, mutation of all 3 residues at once had no effect on the targeting activity, suggesting that these residues do not make any positive contribution to basal targeting.

Finally, the Y19A and I22A mutations were also made in the full-length HHG protein, which is 3 residues longer than the R26t construct, to compare the effect of these mutations in the two different backgrounds. While the Y19A mutation had the same effect on basal targeting activity in HHG and R26t, the I22A mutation had a less severe effect in the full-length protein; R26t I22A was only 29% basal (Fig. 2B), while HHG I22A was 86 ± 1% basal (data not shown).

**There Is a Strict Requirement for Tyrosine at Position 19, while the Isoleucine at Position 22 Can Be Replaced by Other Aliphatic Residues**—To investigate the requirements of the signal for specific amino acid side chains at the two critical positions identified by the alanine scan, a variety of amino acids were tested for their ability to function at the position of isoleucine 22. However, as illustrated in Fig. 3B, in the basal targeting signal of VSV G there is a strong preference for an extended aliphatic side chain at this position, as only valine and leucine retain wild-type basal targeting activity. Phenylalanine, tyrosine, and cysteine function at a suboptimal level, while small, polar, or charged residues function poorly or not at all.

**The Positively Charged Character of the Sequence Lys-14-Arg-16 Does Not Make a Significant Contribution to the Basal Targeting Activity Encoded in the G Tail**—The presence of 3 contiguous positively charged residues, Lys-14, Lys-15, and Arg-16, located within the 11-residue sequence containing targeting information suggested the possibility that the highly charged character of this region might contribute to the basal targeting signal. Therefore all three of these residues were changed to alanine at once in the R26t protein. However, as illustrated in Fig. 2B, the effect of this triple mutation on basal targeting was very similar to the effect of mutating Arg-16 alone (see Fig. 2A), suggesting that the polarity of the triple mutant was largely due to the loss of Arg-16 and not to diminished positive charge in this region. This KKR triple mutant was also constructed in the full-length HHG protein and, as reported for mutations made at the position of tyrosine 19, had essentially the same effect in the R26t construct and the full-length HHG.
residue 22, does have a minor effect on targeting, suggesting that a certain minimal sequence length is required for optimal signal function. Point mutants throughout the entire Lys-14 - Met-24 sequence in the R26t protein suggest that critical residues of the signal include Tyr-19 and Ile-22 with Arg-16 making a minor, but significant contribution. Additionally, the tyrosine is completely essential and cannot be replaced, while the isoleucine can be replaced equally well by either valine or leucine. Therefore, the critical part of this signal seems to be represented by the sequence Y-X-X-aliphatic. The involvement of the tyrosine is fully consistent with earlier data, which indicated that this tyrosine is absolutely required for the tail of VSV G to retain its basal targeting activity in the wild-type VSV G protein (10).

Although the tyrosine and the isoleucine appear to be the most critical elements of the signal, Arg-16 does make a definite contribution as judged by the slight decrease in basal polarity of the R16A mutant. Interestingly, while the presence of an alanine significantly impairs basal targeting, a leucine is almost as functional as the wild-type arginine, suggesting that the function of the arginine is not solely to provide a positive charge, but that the size, shape, or hydrophobicity of the side chain is important. The possibility that the contribution of the positive charge is minor is supported by the observation that removal of all three positive charges upstream of the tyrosine in the KKR mutant has a relatively small effect on basal targeting. The critical elements of the basal targeting signal in the cytoplasmic domain of G defined by this work are summarized in Fig. 4.

Although most of this mutational analysis was done in the HHG R28t construct, a selected set of point mutations, including KKR, I22A, and each Tyr-19 mutant, were also tested in the full-length HHG protein, which is 3 residues longer. Only the I22A mutant gave a significantly different result in the two backgrounds. It is possible that the longer tail is structurally more stable and allows the hydrophobic character of alanine to substitute for isoleucine. This substitution may also be tolerated in the full-length protein due to the presence of additional minor interactions between the three extra residues and the signal "receptor."

An Alignment of This Targeting Signal with Other Cytoplasmic Targeting Signals Suggests the Existence of a Broad Class of Targeting Motifs—As illustrated in Table I, the basolateral targeting signal in the cytoplasmic tail of VSV G has similarity to a large number of cytoplasmic targeting determinants with the more general motif of aromatic-X-X-large hydrophobic. In addition to basal targeting and coated pit localization, some of these sequences have been shown to be necessary for targeting to the lysosome, the endoplasmic reticulum, and the TGN. However, in at least four cases where sequences encode more than one type of targeting information, the overlapping signals can actually be distinguished by mutagenesis (see Table II). For example, mutation of certain residues in the aligned region of lysosomal acid phosphatase caused a differential effect on basal targeting and internalization indicating that there are two different signals in this sequence (9). The basal targeting and internalization signals in this case can be distinguished by a lysosomal targeting signal by mutation of the neighboring glycine (6), and the internalization signal in TGN 38 can be distinguished from the TGN targeting information by mutation of an arginine (36). We have now made a similar observation for the overlapping basolateral targeting and internalization signals in the cytoplasmic domain of the influenza hemagglutinin mutant HA Y543,F546.5

One explanation for overlapping, but distinct, targeting motifs is that different signals share a common structure and rely on individual amino acid side chains to encode specificity. However, results for VSV G indicated that this is not necessarily the case for all basal targeting and internalization motifs. Mutation of Ile-22 to phenylalanine creates the sequence of an optimal internalization signal in the G tail based on the alignment in Table I, as well as on previous proposals for internalization signal consensus sequences (17–19, 37). However neither this nor any of nine other point mutations tested caused more than a 2-fold increase in the normally low internalization rate of VSV G (data not shown). This suggests that the structure of the G tail required for basal targeting is distinct from structures that mediate internalization.

There are several basal targeting sequences and many internalization signals that do not display the motif shown in Table I. Two basal targeting signals in the cytoplasmic domain of the low density lipoprotein receptor (13) and one in the polyimmunoglobulin receptor (20) each contain an important aromatic amino acid but cannot be aligned with the sequences illustrated. Additionally, there are a number of internalization signals that have an essential aromatic residue but have a polar or charged residue in place of the second large hydrophobic amino acid, as well as a subset of internalization signals that have similarity to the N-P-X-Y internalization signal of the low density lipoprotein receptor (38, 39). However, in vitro competition experiments have suggested that, in the case of the internalization signals, this last class of signal may compete with those presented in Table I for binding to the same cytoplasmic receptor molecules (40–42). Additionally, recent analysis of the internalization signal in the influenza hemagglutinin mutant HA Y549 (18) suggests that degenerate internalization signals can actually function quite well. Finally, there is evidence that multiple internalization signals in the same protein function at different efficiencies and act cooperatively (14, 38, 39). Therefore it seems likely that the receptors for cytoplasmic internalization signals are actually capable of binding a much more degenerate set of sequences than Table I would indicate and that other targeting receptors may share this ability to recognize a diverse set of primary sequences.

There are two distinct, although non-exclusive, mechanisms by which degenerate signals might be recognized and then bound by cytoplasmic receptors. In one mechanism, a signal would be recognized by its inherent structure and, in another, some feature of a flexible sequence of amino acids would initiate binding by a receptor and be productively bound only if it could acquire a particular “induced” structure upon binding. In the "inherent structure" mechanism, signal degeneracy would result from different primary sequences which form similar secondary structures, while in the “induced structure” mechanism, signal degeneracy would result from the tolerance of a variety of non-interfering amino acid side chains in the signal binding domain of a receptor molecule. Current data are insufficient to distinguish between these two mechanisms.

5 H. Y. Naim, D. C. Thomas, and M. G. Roth, manuscript in preparation.
**TABLE I**

| Protein                  | Cytoplasmic targeting sequence | Sorting activity present: | Endocytosis activity present: | Other targeting activity or localization | ref. |
|--------------------------|-------------------------------|---------------------------|--------------------------------|------------------------------------------|------|
| **I. Basolateral targeting signals** |                               |                           |                                |                                          |      |
| Gwt                      | KKRQIYTDEIM                    | yes                       | no                             |                                          | 10, this work |
| FcR                      | EAENTITYSSLKH                  | yes                       | yes                            |                                          | 6    |
| hNGFR                    | KRNWSLYSSLPYAK                 | yes                       | probably                      |                                          | 49   |
| ASGPR H1                 | EYQDLQML                      | yes                       | yes                            |                                          | 11,50 |
| HA Y543,F546             | NGSLQYRIFI                    | yes                       | yes                            |                                          | 7,18, M. Roth, in prep. |
| **II. Lysosomal targeting signals** |                               |                           |                                |                                          |      |
| LAP                      | PPGVRHVAD                      | yes                       | yes                            | lys                                      | 9,51 |
| lgp120 (LAMP 1)          | RKRSAGEYQTI                    | yes                       | yes                            | lys                                      | 6,52,53,54 |
| LAMP 2                   | LKHHAGYEQF                     | probably                  | probably                      | lys                                      | 55,56 |
| CD3 gamma                | YQPLK                         | ?                         | yes                            | lys                                      | 57   |
| CD3 delta                | YQPLR                         | ?                         | yes                            | lys                                      | 57   |
| **III. Assorted targeting signals** |                               |                           |                                |                                          |      |
| CD3 epsilon              | KGQRDLYSGL                     | ?                         | ?                              | ER                                       | 58   |
| GLUT 4                   | MPSGFGQAI                      | yes                       | yes                            | stor. ves.                               | 59   |
| TGN 38/41                | SDYQFL                        | yes                       | yes                            | TGN                                      | 36,60,61, pers. comm. ** |
| **IV. Other endocytic signals** |                               |                           |                                |                                          |      |
| FcRIII gamma             | YETL                           | ?                         | yes                            |                                          | 62   |
| FcRIII gamma             | YTGL                           | ?                         | yes                            |                                          | 62   |
| CI MPR                   | KYSKV                          | ?                         | yes                            |                                          | 17   |
| CD MPR                   | YRGV                           | ?                         | yes                            |                                          | 63   |
| CD MPR                   | FPHLAF                         | ?                         | yes                            |                                          | 63   |
| RM                       | YENF                           | ?                         | yes                            |                                          | 64   |
| plgR                     | LAYSAF                         | no                        | yes                            |                                          | 14,65 |
| HA+8                     | YDYKSF                         | probably                  | yes                            | lys                                      | 66   |
| tFR                     | LSYTRF                         | no                        | yes                            |                                          | 37,67,68,69,70 |
| glyc. Y106               | SDYKPL                         | ?                         | yes                            |                                          | 16   |
| EGF                    | FYRAL                           | ?                         | yes                            |                                          | 39   |
Does a Class of Related Cytoplasmic Adapter Molecules Mediate a Wide Variety of Protein Trafficking Events?—After the initial identification of the first cytoplasmic internalization motifs, it was proposed that cytoplasmic proteins, later called adapters, were binding to these sequences to promote the clustering of proteins in coated pits (43–45). It now seems likely that there are a number of adapter molecules, functioning at various stages of the exocytic and endocytic pathway, that recognize a variety of targeting motifs and whose function is to sort the cargo proteins into the proper trafficking vesicles at the right time. To visualize how small sequence differences between motifs could generate specificity for different adapters, an analogy may be made to the binding of different classes of peptides by different major histocompatibility complex alleles where specificity “pockets” in the otherwise similar binding grooves discriminate between different peptides by their ability to accommodate particular amino acid side chains (46).

In the case of the internalization signals, there is substantial evidence that implicates the AP2 “adaptors” of the plasma membrane clathrin coat as playing the role of the cytoplasmic receptor/adapter (40–42, 47, 48). The possibility that AP2 is the sorting adapter at the plasma membrane responsible for recruiting proteins into coated pits raises the possibility that the related AP1 adaptors of the Golgi clathrin coat might play a role in the recognition of the basolateral targeting motifs in the TGN. However, the AP1 adaptins do not appear to be able to bind cytoplasmic sequences that encode basal targeting information (40, 47). Given the relatedness of the sequences of a number of sorting signals, there must be families of related sorting adapters that are currently unidentified. To understand the molecular basis for intracellular protein traffic, it will be necessary to identify, purify, and clone a variety of these cytoplasmic targeting molecules.

Acknowledgment—We thank Dr. Colleen B. Brewer for experimental advice and assistance in preparation of the manuscript.

REFERENCES

1. Caplan, M. & Matlin, K. S. (1986) in Functional Epithelial Cells in Culture (Matlin, K. S. & Valenzilch, J. D., eds) pp. 71–127, Alan R. Liss, Inc., New York.
2. Simons, K. & Wandinger-Ness, A. (1990) Cell 62, 207–210.
3. Mostov, K. A. (1985) in Proc. Natl. Acad. Sci. U. S. A. 82, 9885–9889.
4. Fuller, S. D., Bravo, R. & Simons, K. (1985) EMBO J. 4, 297–307.
5. Wandinger-Ness, A., Bennett, M. K., Astyan, C. & Simons, K. (1990) J. Cell Biol. 116, 577–583.
6. Hunziker, W., Harter, C., Matter, K. & Mellman, I. (1991) Cell 66, 907–920.
7. Brewer, C. B. & Roth, M. G. (1991) J. Cell Biol. 114, 413–421.
8. Casanova, J. E., Apodaca, G. & Mostov, K. E. (1991) Cell 66, 65–75.
9. Pril, V., Leimbach, L., van Figuera, R. & Peters, C. (1991) EMBO J. 12, 2181–2193.
10. Thomas, D. C., Brewer, C. B. & Roth, M. G. (1993) J. Biol. Chem. 268, 3313–3320.
11. Geffen, I., Fuhrer, C., Leitinger, B., Weiss, M., Hugell, K., Griffiths, G. & Spiess, M. (1993) J. Biol. Chem. 268, 20772–20777.
12. Naftz, S. J. & Isacke, M. C. (1993) J. Cell Biol. 121, 1299–1310.
13. Matter, K., Hunziker, W. & Mellman, I. (1992) Cell 71, 741–753.
14. Okamoto, C. T., Shia, S. P., Bird, C., Mostov, K. E. & Roth, M. G. (1992) J. Biol. Chem. 267, 9595–9602.
15. Darcangelo, C., Le Breve, A., Rothenberger, S., Iacoppeta, B. & Kuhn, L. C. (1993) EMBO J. 12, 1713–1721.
16. Kitastki, N. S., Thomas, D. & Roth, M. G. (1993) J. Cell Biol. 121, 1393–1407.
17. Canfield, W. M., Johnson, K. F., Ye, R. D., Gregory, W. & Kernofsd, S. (1991) J. Biol. Chem. 266, 5662–5668.
18. Nain, H. Y. & Roth, M. G. (1994) J. Biol. Chem. 269, 3928–3933.
19. Trowbridge, I. S., Collawn, J. F. & Hopkins, C. R. (1995) Annu. Rev. Cell Biol. 11, 129–151.
20. Aroeti, B., Kosen, P. A., Kuntz, I. D., Cohen, F. E. & Mostov, K. E. (1994) J. Cell Biol. 123, 1149–1160.
21. Paddock, L., Woodgett, J. & Rose, J. K. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 2765–2769.
22. Brown, D. A., Crise, B. & Rose, J. K. (1989) Science 245, 1496–1501.
23. Passon, C., Oriagbor, W. & Simons, K. (1984) J. Cell Biol. 96, 795–802.
24. Gruenberg, J. & Howell, K. E. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 5738–5742.
25. Gotlibich, T. A., Ivanov, I. E., Adesnik, M. & Subatini, D. D. (1993) J. Cell Biol. 120, 695–710.
26. Rose, J. K. & Gallione, C. F. (1981) J. Biol. Chem. 256, 519–528.
27. Lazarovits, J., Shia, S. F., Kitastki, N., Lee, M. S., Bird, C. & Roth, M. G. (1990) J. Biol. Chem. 265, 4760–4767.
28. Kunkel, T. A. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 488–492.
29. Zoller, M. J. & Smith, M. (1984) DNA 3, 479–489.
30. Kadowaki, H., Kadowaki, T., Wodswand, F. E. & Taylor, S. I. (1989) Gene (Amst.) 76, 161–166.
31. Sarkar, G. & Sommer, S. S. (1995) Biotechnology 8, 404–407.
32. Slapert, J., Wroehng, J. & Remelt, R. (1992) Nucleic Acids Res. 20, 624.
33. MacQueen, N., Nayak, D. P., Stephens, E. B. & Compana, R. W. (1997) J. Biol. Chem. 272, 16233–16240.
34. Compton, T., Ivanov, I. E., Gottlich, T., Rindler, M., Adesnik, M. & Subatini, D. (1993) Proc. Natl. Acad. Sci. U. S. A. 96, 4112–4116.
35. Gottardi, C. J. & Caplan, M. J. (1993) Science 260, 552–554.
36. Humphreys, J. S., Peters, P. J., Yuan, L. C. & Bonifacino, J. S. (1993) J. Cell Biol. 120, 1123–1135.
37. Collawn, J. F., Stangel, M., Kuhn, L. A., Eseewogu, V., Jing, S., Trowbridge, I. S. & Tainer, J. A. (1990) Cell 60, 1061–1072.
38. Rajagopalan, M., Neidig, J. L. & McClain, D. A. (1995) J. Biol. Chem. 269, 23066–23073.
39. Chang, C. P., Lazar, C. S., Walsh, B. J., Komuro, M., Collawn, J. F., Kuhn, L. A., Tainer, J. A., Trowbridge, I. S., Parpahur, M. G., Rosenfeld, M. G., Wiley, H. S. & Gill, G. N. (1995) J. Biol. Chem. 268, 19312–19320.
40. Pearse, B. M. (1986) EMBO J. 5, 3301–3306.
41. Glickman, J. N., Combeir, E. & Pearse, B. M. (1989) EMBO J. 8, 1041–1047.
42. Belzzer, I. P. & Spiess, M. (1991) EMBO J. 10, 3735–3742.
43. Davis, C. G., van Driel, I. R., Russell, D. W., Brown, M. S. & Goldstein, J. L. (1987) J. Biol. Chem. 262, 4075–4082.
44. Lazarovits, J. & Roth, M. (1998) Cell 93, 743–752.
VSV G Basolateral Targeting Signal

45. Chen, W. J., Goldstein, J. L. & Brown, M. S. (1990) J. Biol. Chem. 265, 3118–3123
46. Barber, L. D. & Parham, P. (1993) Annu. Rev. Cell Biol. 9, 163–206
47. Sosa, M. A., Schmidt, B., von Figura, K. & Hille-Rehfeld, A. (1995) J. Biol. Chem. 268, 12537–12543
48. Sorokin, A. & Carpenter, G. (1993) Science 261, 612–615
49. LeBivic, A., Sambuy, Y., Patzak, A., Patil, N., Chao, M. & Rodriguez-Boulan, E. (1991) J. Cell Biol. 115, 607–618
50. Fuhrer, C., Geffen, I. & Spiess, M. (1991) J. Cell Biol. 114, 423–432
51. Peters, C., Braun, M., Weber, B., Wendland, M., Schmidt, B., Pohlmann, R., Waheed, A. & von Figura, K. (1996) EMBO J. 9, 3497–506
52. Williams, M. A. & Fukuda, M. (1996) J. Cell Biol. 111, 955–966
53. Mathews, P. M., Martinie, J. B. & Fambrough, D. M. (1993) J. Cell Biol. 118, 1027–1040
54. Guernier, F. G., Arterburn, L. M., Penno, M. B., Cha, Y. & August, J. T. (1993) J. Biol. Chem. 268, 1941–1946
55. Nabi, I. R., Le Bivic, A., Pambrough, D. & Rodriguez-Boulan, E. (1991) J. Cell Biol. 115, 1373–1384
56. Nabi, I. R. & Rodriguez-Boulan, E. (1993) Mol. Biol. Cell 4, 627–635
57. Letourneur, F. & Klausner, R. (1992) Cell 69, 1143–1157
58. Mallabharrena, A., Fresno, M. & Alarcon, B. (1992) Nature 357, 593–596
59. Piper, R. C., Tai, C., Kulesza, P., Pang, S., Warnock, D., Baenziger, J., Slot, J., W. Geuze, H. J., Puri, C. & James, D. E. (1993) J. Cell Biol. 121, 1221–1232
60. Bos, K., Wraith, C. & Stanley, K. K. (1993) EMBO J. 12, 2219–2228
61. Wang, S. H. & Hong, W. (1990) J. Biol. Chem. 265, 22853–22862
62. Amigorena, S., Salamero, J., Davoust, J., Fridman, W. H. & Bonnerot, C. (1992) Nature 358, 337–341
63. Johnson, K. F., Chan, W. & Kornfeld, S. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 10010–10014
64. Ozaki, K., Itoh, N. & Kawasaki, T. (1993) J. Biochem. (Tokyo) 113, 271–276
65. Breitfeld, P. P., Casanova, J. E., McKinnon, W. C. & Mostov, K. E. (1990) J. Biol. Chem. 265, 13753–13757
66. Roth, M. G., Henis, Y. I., Brewer, C. B., Kristakos, N. T., Shia, S., Lazarovits, J., Fire, E., Thomas, D. C. & Zwart, D. E. (1993) in Cell Biology and Biotechnology (Oka, M. S. & Rupp, R. G., eds) pp. 137–156, Springer-Verlag, New York
67. McGraw, T. E. & Maxfield, F. R. (1990) Cell Regul. 1, 369–377
68. McGraw, T. E., Pytowski, B., Arzt, J. & Ferrone, C. (1991) J. Cell Biol. 112, 853–861
69. Jing, S., Spencer, T., Miller, K., Hopkins, C. & Towbridge, I. S. (1990) J. Cell Biol. 110, 283–294
70. Giriones, N., Alvarez, E., Seth, A., Lin, I. M., Latour, D. A. & Davis, R. J. (1991) J. Biol. Chem. 266, 19006–19012