Characteristics of the Tyrosine Recognition Signal for Internalization of Transmembrane Surface Glycoproteins

Nicholas T. Ktistakis, D'Nette Thomas, and Michael G. Roth

Department of Biochemistry, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas 72935-9038

Abstract. A tyrosine residue in the cytoplasmic domain of a class of cell surface receptors is necessary, but not sufficient, for internalization through coated pits. To identify the amino acid context enabling a tyrosine to serve as a signal for endocytosis, we mutated the short cytoplasmic domain of a mutant influenza virus hemagglutinin that is competent for internalization, HA-Y543, and determined the effect of each change on internalization. From these results and a comparison of sequences of other proteins recognized by coated pits, a "tyrosine internalization signal" was proposed. Site-directed mutagenesis was employed to insert complete, or incomplete "tyrosine internalization signals" into the cytoplasmic domain of a protein normally not endocytosed, human glycophorin A. Only the complete signal caused internalization of mutant glycophorins by coated pits. The signal is formed by a short amino acid sequence, with polar or basic residues preferred at certain positions on either side of the tyrosine. Amino acids, which in proteins of known structure are frequently found in turns, are clustered near the tyrosine on the side of the signal nearest the transmembrane domain.

THE plasma membranes of animal cells contain small specialized domains, the clathrin-coated pits, designed for the rapid internalization of cell surface receptors. Coated pits appear to sort transmembrane surface proteins into at least three classes: those that localize with high affinity to coated pits and are internalized rapidly (Goldstein et al., 1985), those that can enter coated pits slowly and are internalized at approximately the same rate as the bulk of membrane (Roth et al., 1986), and proteins that are excluded from coated pits (Bretscher et al., 1980; Roth et al., 1986). The information necessary to specify rapid internalization through coated pits resides in the cytoplasmic domain of at least some transmembrane proteins. Deletion of the cytoplasmic domain of all receptors studied to date results in proteins that travel to the cell surface but are internalized very slowly (Lehrman et al., 1985; Prywes et al., 1986; Mostov et al., 1986; Rothenberg et al., 1987; Miettinen et al., 1989; Lobel et al., 1989). Domain switching experiments have shown that when the cytoplasmic domains of proteins with high affinity for coated pits are transferred to a protein normally excluded from coated pits, the resulting chimeric proteins are efficiently internalized.

Although the cytoplasmic epitope recognized by coated pits has not been identified, several lines of evidence suggest that a relatively short stretch of amino acids containing a tyrosine is necessary and sufficient to specify internalization. Internalization of the receptors for LDL or for mannose-6-phosphate (M-6-P) is not affected by deletion of the distal half of the cytoplasmic domain (Davis et al., 1987; Lobel et al., 1989). Mutation of a single tyrosine within the stretch of cytoplasmic amino acids required for internalization of the receptors for LDL, M-6-P, or transferrin (Tf), results in receptors unable to be efficiently internalized (Davis et al., 1987; Lobel et al., 1989; Jing et al., 1990; McGraw and Maxfield, 1990). Introduction of a tyrosine residue at position 543 of the short cytoplasmic domain of the influenza virus hemagglutinin (HA)[1] converts HA from a protein that is ordinarily excluded from coated pits into one that enters coated pits and is rapidly internalized (Lazarovits and Roth, 1988). Significantly, introduction of tyrosines at positions 540, 545, and 546 of the HA cytoplasmic domain does not result in endocytosed proteins (Lazarovits and Roth, 1988; and this work). In addition, the LDL receptor cytoplasmic domain contains three tyrosines, but only one (at position 807) participates in signaling internalization of this protein (Davis et al., 1987). Thus, in addition to the presence of a tyrosine, other features of the amino acid sequence of the cytoplasmic domain are necessary to allow a protein to enter the endocytic pathway.

To identify the amino acid context required for a tyrosine to function as an endocytosis signal, the amino acid sequence surrounding tyrosine 543 in the HA cytoplasmic domain was systematically mutated and the resulting proteins were analyzed for their ability to be internalized. From the results of these experiments, and from comparison of amino acid sequences of the region surrounding tyrosines in the cytoplasmic domains of proteins known to be internalized through coated pits, we have identified a "tyrosine internali-

1. *Abbreviation used in this paper*: HA, hemagglutinin.

zation signal" common to many receptors. We have introduced this tyrosine signal into the cytoplasmic domain of glycophorin A, a protein that we show is normally excluded from coated pits in CV-1 cells, and have demonstrated that the mutant glycophorin A carrying this signal is efficiently internalized through clathrin-coated pits. Conversely, mutations that create tyrosines in two sequences predicted not to form "tyrosine internalization signals" in glycophorin A failed to allow internalization.

Materials and Methods

Cells, Cell Culture, and Radiolabeling of Proteins

CV-1 cells were maintained in DME supplemented with 10% Serum Plus (Hazleton Research Products, Lenexa, KS). Actively growing cells were infected in suspension with recombinant SV40 virus stocks for 30 min on ice and then plated on 6-well plates for experiments at 30 or 42 h after infection. To label newly synthesized proteins, cells were starved for 30 min with DME without methionine and cysteine and then were radiolabeled with 50 μCi of Translabel (ICN Radiochemicals, Irving, CA) per 35-mm plate in the same medium. At the end of the pulse, the radioactive medium was replaced with complete DME and the cell monolayers were incubated at 37°C for various times.

Recombinant DNA Techniques

The construction of SV40 vectors encoding the wild-type or mutant Japan HA genes has been described previously (Doyle et al., 1985; Lazarovits and Roth, 1988). The gene encoding human glycophorin A, obtained from Dr. Minoru Fukuda (Siebert and Fukuda, 1986), was excised from pBluescript with Eco RI for subcloning into mp18, and with Cla I and Bam HI for subcloning into the SV40 late-replacement vector pKSVE. For subcloning of glycophorin A after mutagenesis in mp18, we replaced the wild type Sac I and Bam HI fragment of glycophorin in pKSVE with the corresponding fragment of the mutated glycophorin from mp18. Oligonucleotide-directed mutagenesis was performed as described in the Mutagene instruction manual (Bio-Rad Laboratories, Richmond, CA) (the method of Kunkel, 1985). With the exception of the mutations at position 544 of HA, all mutations were produced in duplicate on independently isolated templates. For mutagenesis at position 544, we synthesized an oligonucleotide that contained an equimolar mixture of all four bases at positions one and two of codon 544. These mutations were produced on a single template. The sequence of all mutants was established by the method of Sanger et al. (1977) using the Sequenase sequencing kit (United States Biochemical Corp., Cleveland, OH) following the manufacturer's instructions. After subcloning and before the production of virus stocks, the identity of the mutants was reconfirmed in most cases by sequencing the double-stranded expression vector pKSVE.

Internalization Assays

The presence of mutant proteins in intracellular vesicles diagnostic of an endocytosis-competent phenotype was established by immunofluorescence as described in Lazarovits and Roth (1988). The rate of internalization of mutant HAs was measured using the antibody internalization assay described in Lazarovits and Roth (1988). To determine the cellular location of mutant HA proteins at steady state, we took advantage of the fact that in CV-1 cells, HA is produced in an uncleaved form that trypsin quantitatively cleaves into two fragments, HA1 and HA2. After a 10-min pulse with ^{35}S methionine and cysteine, the newly synthesized HA was chased to the cell surface for 2 h. To measure the amount of HA that arrived at the cell surface during the chase period, one half of the samples were treated with 5 μg/ml trypsin in the chase medium while the other half were not. To measure the amount of HA that was internal at the end of the chase, all samples were treated with 100–200 μg/ml of trypsin on ice 30 min before neutralization of trypsin with equal amount of soybean trypsin inhibitor, followed by lysis and immunoprecipitation. An internalization index was defined as the ratio of uncleaved HA to the total amount of HA (uncleaved + HA1 + HA2) for the samples that received trypsin on ice at the end of the chase. This number was corrected for HA that had not yet arrived at the cell surface by 2 h after synthesis, which is given by the ratio of uncleaved HA versus total HA for the samples that received trypsin during the chase at 37°C. For glycophorin A, which is degraded by trypsin, the assay for location at steady state was as described for HA, except that the internalization index was determined as the ratio of glycophorin A recovered after trypsin treatment to the amount of protein recovered in the absence of trypsin. As in the case of HA, a correction was made for protein that might not have arrived at the cell surface during the 2-h chase. For each of the proteins in this study, 95% reached the cell surface by 2 h after synthesis. For the purposes of comparing various mutants, in each experiment HA wt (or glycophorin wt) and HA-Y543 were included as negative and positive controls. Results of internalization assays for HA mutants were normalized to that measured for HA-Y543 in the same experiment. Values for HA-Y543 varied <10% among all experiments.

EM

CV-1 cells infected with various recombinant SV40 viruses were fixed at 40-h postinfection and were labeled first with antiglycophorin A followed by protein A conjugated to 10-nm colloidal gold (Janssen Life Science Products, Piscataway, NJ). The labeled monolayers were then processed for EM using the propylene oxide lifting method as previously described (Roth et al., 1986). Samples were photographed at 80 kV using a Jeol JEM 100-SX microscope.

Other Methods

Restriction enzyme digestions using enzymes from New England Biolabs (Beverly, MA) or Bethesda Research Laboratories (Bethesda, MD) were performed following the manufacturer's instructions. Other DNA manipulations were as described in the Molecular Cloning manual (Maniatis et al., 1982). Immunoprecipitations, analysis by SDS-PAGE (Laemmli, 1970), fluorography, and autoradiography were performed as described in Lazarovits and Roth (1988) using polyclonal anti-HA antibodies. For detection of glycophorin, antibodies against the whole molecule (commercially available from Research Plus, Inc., Bayonne, NJ) or against the cytoplasmic domain (kindly donated by Dr. M. Fukuda, La Jolla Cancer Research Foundation, La Jolla, CA) were used. Quantitation of autoradiograms was performed with a Molecular Dynamics 300A computing densitometer (Sunnyvale, CA), usually by the volume integration subroutine. Care was taken to use autoradiograms exposed within the linear response range of the X-AR film (Eastman Kodak Co., Rochester, NY).

Results

Mutagenesis of the Cytoplasmic Domain of HA-Y543

To identify neighboring amino acids that might contribute to the function of a tyrosine in an internalization signal, we compared the amino acid sequences surrounding tyrosines in the cytoplasmic domains of proteins known to be internalized through coated pits and found that at several positions there was a significant preference for two classes of amino acids (discussed below). Using this analysis as a guide, several of the amino acids surrounding tyrosine 543 were altered, with special emphasis on the polar or charged residues at positions 538, 542, and 544. The cytoplasmic domain sequences of the resulting mutant proteins are shown in Fig. 1.

Alteration of a single amino acid may have a "global" effect on the structure of a protein, complicating the determination of the cause of any change in function of the mutant protein. To avoid this problem, we introduced identical changes into both the wild-type and HA-Y543 sequences and compared structural properties, as well as the endocytosis phenotype of both sets of mutants (for a list of structural assays see Lazarovits et al., 1990). No gross structural alterations in the mutant HAs were created by amino acid changes in the cytoplasmic domain. In Fig. 1, mutations that were also made on the wild type background with a cysteine at position 543 are denoted with an asterisk. The region of HA mutated in these experiments spans amino acids 538–547. When single amino acid changes were made in the wild-type HA se-

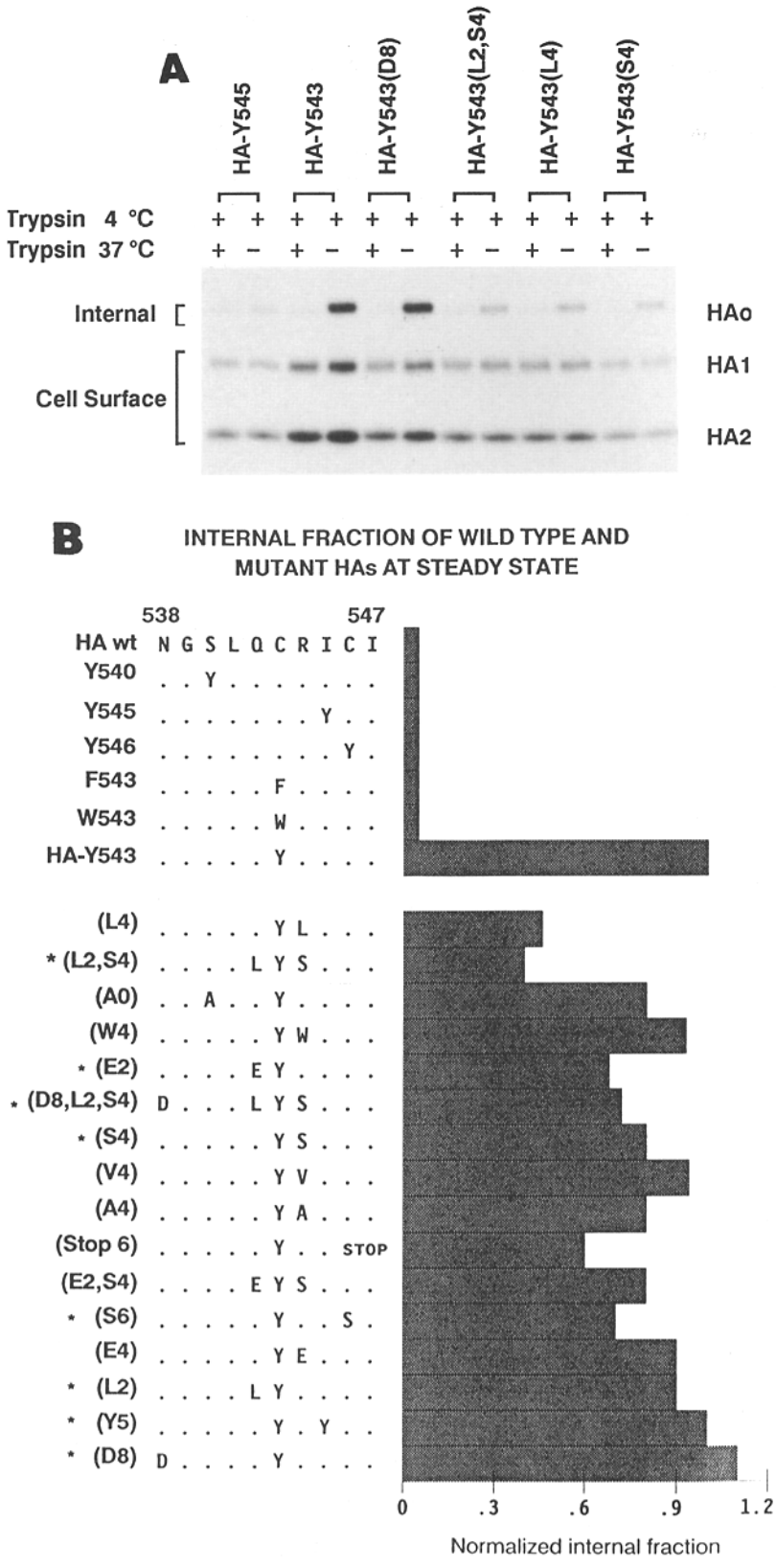


Figure 1. Distribution of wild-type and mutant HA molecules at steady state. (*A*) CV-1 cells expressing various HA mutant proteins were radiolabeled for 10 min, chased for 2 h, and subjected to digestion with trypsin during or at the end of the chase as indicated. The proteins were collected by immunoprecipitation and analyzed by PAGE and fluorography. Shown are some representative mutants ranging from very low, *HA-Y545*, to high *HA-Y543*, (*D8*), internalization. (*B*) The steady-state internalization of wild type and mutant HAs is expressed as a fraction of the value for HA-Y543 determined in the same experiment. Each value is the average of at least three experiments and the variation from one experiment to another was <10%. * indicates that the mutation was also made in the wild-type background.

quence, the resulting proteins were designated by the single letter code for the amino acid replacement, followed by the amino acid position, as in HA-Y543. When additional changes were made in the HA-Y543 sequence, this nomenclature was shortened to parentheses enclosing the single letter code for the amino acid inserted, followed by the last digit of the amino acid position. For example, (L4) designates a leucine at position 544 in the HA-Y543 background. In Fig.

1 the names of the mutants are shown to the left, followed by the sequence of the cytoplasmic region of HA, with only changes in amino acids shown for the mutant proteins.

Several Amino Acids in the HA Cytoplasmic Domain Contribute to the Ability of Tyrosine 543 to Signal Internalization

Two assays were used to evaluate the ability of HA mutants to be internalized. In the first, HAs at the cell surface were identified by their accessibility to trypsin added to the culture medium. Although in some cell types endogenous proteases cleave each polypeptide of the HA trimer at a single site, producing two disulfide-bonded subunits, HA1 and HA2, in CV-1 cells this cleavage occurs only if exogenous proteases are added. Addition of trypsin to CV-1 cells expressing properly folded HAs rapidly cleaves essentially all cell surface HA into its subunits, which can be quantitatively recovered by immunoprecipitation. To determine the proportion of various HA mutants residing in endocytic compartments at steady state, protein synthesized in the presence of ^{35}S-methionine and cysteine was chased to the cell surface and allowed to equilibrate between cell surface and endocytic compartments. 2 h postsynthesis, cell monolayers were treated with extracellular trypsin at 2°C, a temperature at which endocytosis is inhibited. The ratio of cleaved to total HA was determined by immunoprecipitation with anti-HA antibodies, PAGE, and autoradiography (Fig. 1 *A*). As a control to determine the residual proportion of HAs that had not left the exocytic compartments or were otherwise inaccessible to trypsin, for a sample of each mutant, trypsin was included in the 37°C chase medium to cleave all protein that appeared on the cell surface. The proportion of uncleaved HAs in these controls (never more than 5%) was subtracted from the proportion uncleaved by trypsin at 2°C.

An advantage of this assay is that it does not require labeling of HAs at the cell surface until the end of the assay when protein traffic has ceased. In addition, the assay is particularly good for distinguishing proteins that are completely incompetent for endocytosis from those with incomplete internalization signals that, nevertheless, slowly enter coated pits. However, the steady-state distribution of a protein depends upon rates of internalization, recycling to the surface, and delivery to degradative compartments; thus, this assay determined whether a mutant HA entered the endocytic pathway, but did not directly measure its affinity for coated pits.

In Fig. 1 *B*, the portion of wild-type and mutant HA proteins that was internal at steady state is expressed as a fraction of the value for HA-Y543 (the actual value for HA-Y543 is 40% internal). As reported previously, the only single amino acid change in the wild-type background that induced HA to be internalized was the introduction of a tyrosine at position 543. Other aromatic amino acids at that position did not substitute, nor did tyrosines at positions 540, 545, and 546 (Fig. 1 *B*, first seven mutants). In contrast, all the mutants on the HA-Y543 background (Fig. 1 *B*, second set) were internalized to various extents. The least amount of internal protein was observed when both amino acids adjacent to Y543 were changed (L2 and S4) or when a leucine was introduced at position 544 (L4). Introduction of other amino acids with large, hydrophobic side chains, (W4), (S4), (V4), and (A4), at position 544 caused a less severe reduction. Changes on both sides of the tyrosine reduced the amount of internal protein at steady state (AO), (E2), (S6), and (STOP 6). Most importantly, a change that caused a relatively severe reduction in the internal protein at steady state, (L2, S4) could be partially reversed by changes at other positions, (D8, L2, S4) and (E2, S4), that increased the hydrophilicity of the region (Fig. 1 *B*).

To evaluate the rate at which mutant HAs were internalized, protein that had been labeled with ^{35}S methionine and cysteine in vivo was allowed to reach the cell surface and then the cells were chilled on ice to stop endocytosis. HAs at the cell surface were bound by anti-HA antibodies at 0–2°C. After removal of unbound antibodies, the cell monolayers were incubated at 37°C for various intervals to allow endocytosis to resume. At the end of an interval of internalization, cells were chilled on ice to stop endocytosis and HAs at the cell surface were marked by cleavage into HA1 and HA2 by trypsin added to the ice-cold culture medium. Antibody–HA complexes were immunoprecipitated with Protein A sepharose and the fraction of uncleaved (internal) HA was determined by PAGE, fluorography, and densitometry (Fig. 2). This assay measured the rate at which each HA was internalized and reflects the relative affinities of the proteins for coated pits.

The rates of internalization of wild-type and mutant HA proteins during the first 10 min after warming to 37°C were expressed as fractions of the rate of internalization of HA-Y543 and are shown in Fig. 2 *B*. Comparison of mutations in the HA-Y543 background revealed that amino acids at positions 540, 542, and 544 were important for the ability of the tyrosine signal to function for internalization, but that the relation between changes in these amino acids was complex. Large hydrophobic amino acids were not tolerated at position 544 (mutants (L4), (W4), (V4), and (A4), Fig. 2 *B*). Substitution of alanine for serine at position 540 (A0) had an equally severe effect (Fig. 2 *B*). Reduction of the polar character of the sequence near tyrosine 543 impaired internalization. Thus, mutant (L2, S4) is internalized more slowly than the more polar mutations in mutants (E2, S4) and (D8, L2, S4) (Fig. 2 *B*). In addition, a modest but reproducible increase in internalization rate was observed for mutant (D8) compared to HA-Y543 itself. However, increasing hydrophilicity did not necessarily improve function of the signal, as can be seen by the reduction in internalization rate caused by changing glutamine 542 to glutamic acid (E2, Fig. 2 *B*), and by the fact that although mutant (E2, S4) was less charged than (E2), which has an arginine at position 544, it was internalized slightly faster. Replacing arginine 544 with an amino acid of opposite charge (E4) was surprisingly well tolerated. Taken together, these results suggested that amino acids surrounding the tyrosine were important for contributing to both a polar character and particular secondary structure, but none of the amino acids in the six positions changed were as critical as the tyrosine for signal function.

In addition, several mutations that had modest effects on the proportion of protein that was internal at steady state, nevertheless, dramatically reduced the rate of internalization (compare (W4) and (V4) in Figs. 1 *B* and 2 *B*). This observation suggested that these mutations affected a second step in the intracellular traffic of these proteins, by decreasing either the rate or the extent to which they recycled to the plasma

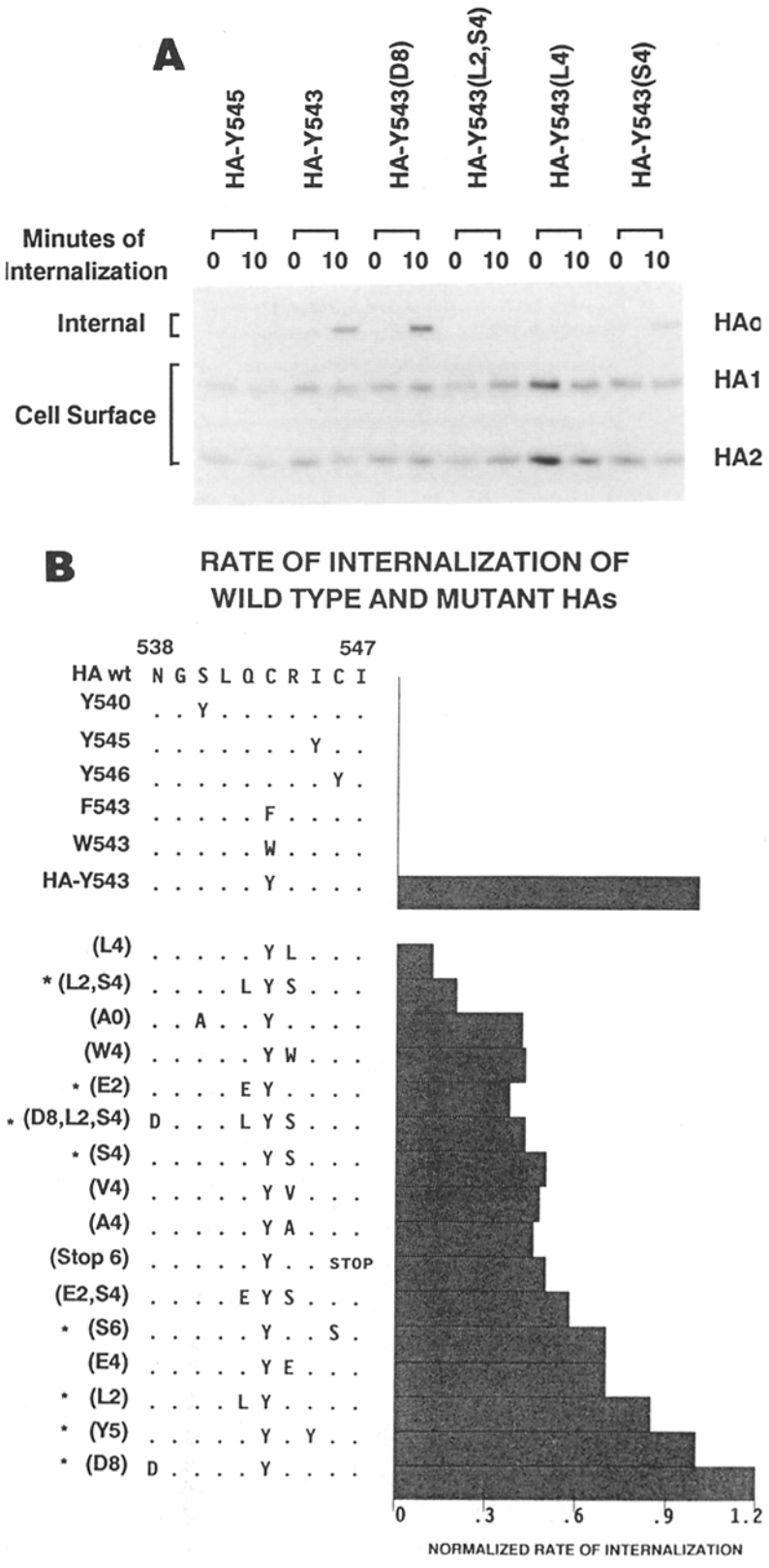


Figure 2. Internalization rate of wild-type and mutant HA molecules after binding antibody at the cell surface. (*A*) CV-1 cells expressing various mutant proteins were radiolabeled for 10 min and chased for 2 h. At the end of the chase, HAs at the cell surface were bound with anti-HA antibodies at 2°C, unbound antibodies were removed, and the HA-antibody complexes were allowed to internalize at 37°C as indicated. Following internalization, the surface HA was cleaved with trypsin and all HA-antibody complexes were captured with protein A sepharose and analyzed by PAGE and fluorography. Shown are representative mutants ranging from no, *HA-Y545*, to high, *HA-Y543 (D8)*, internalization. (*B*) The initial rate of internalization of wild-type and mutant HAs is expressed as a fraction of the value for HA-Y543. Each mutant was analyzed in at least three experiments and the internalization intervals were usually 0, 10, and 20 min. The values shown are averages for internalization during the first 10 min at 37°C and variation within these experiments was never >10%. * indicates that the mutation was also made in the wild-type background.

membrane after internalization. The basis for this second change is currently under investigation and will be the subject of a future report.

In experiments in which tyrosines have been removed from the cytoplasmic domains of receptors, or in which amino acids surrounding tyrosines have been mutated (Davis et al., 1987; Jing et al., 1990; McGraw and Maxwell, 1990), effects on internalization have ranged from less than two to

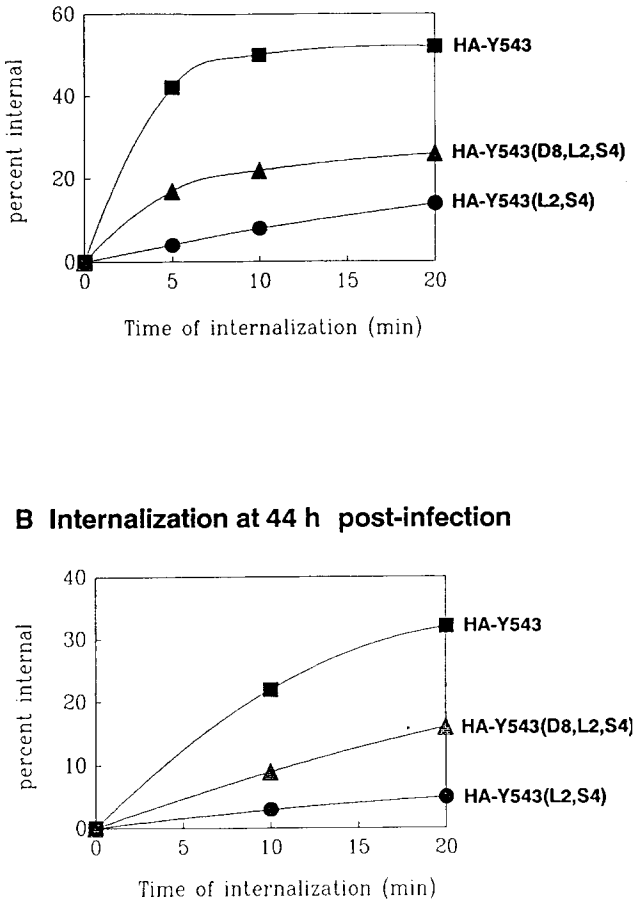


Figure 3. Internalization of three mutant HAs at two different periods postinfection with recombinant SV40 virus vectors. Identical infection and labeling conditions were used for the two experiments. The proportion of HAs internalized was determined as shown in Fig. 2 *A* and plotted as a function of chase time at 37°C.

approximately sevenfold, similar to the results reported above for mutants of HA-Y543. Since several of our mutants showed changes of internalization of twofold or less, it was important to know how variations in the amount of protein expressed for different mutants within a given experiment might effect our results. To investigate this possibility, we made use of the fact that the amount of protein expressed in CV-1 cells infected with recombinant SV40 vectors varies enormously as a function of time after infection (Gething and Sambrook, 1981). Fig. 3 shows the results of an antibody internalization assay performed with parallel samples, one at 22 and the other at 44 h after infection. At 22 h after infection expression of the protein has just become detectable. At the later period, the amount of HA present, almost all of which is at the plasma membrane, has increased eightyfold as detected by Western blot analysis (data not shown). At earlier times, both the rate and extent of internalization were greater than at later times, consistent with an interpretation that the apparent rate of endocytosis decreased as coated pits became saturated. However, the relationship between the three mutant proteins shown remained essentially the same, indicating that comparisons of the rates of internalization of various mutants were valid measures of their relative affinity for coated pits, as long as care was taken in each assay to express similar amounts of protein per cell.

Two Aspects of the Tyrosine Internalization Signal

Our mutagenesis of HA-Y543 was guided by the observation that in the region surrounding tyrosines in the cytoplasmic domains of several receptors there was a significant conservation of basic or polar residues. However, the observation that the arginine at position 544 could be replaced by a residue of opposite charge without destroying the capacity for endocytosis did not fit our original hypothesis and stimulated a second comparison of the nature of sequences surrounding tyrosines. In proteins of known amino acid sequence that enter coated pits, there is at least one tyrosine in the cytoplasmic domain, but the position of the tyrosine in the linear amino acid sequence, with respect to either the end of the hydrophobic transmembrane sequence, or the end of the polypeptide chain, is not fixed. Of these proteins, even among those with cytoplasmic domains that are relatively short and have no enzymatic function, the cytoplasmic domain varies in length from 10 to 164 amino acids. No amino acid identities are apparent when the sequences surrounding tyrosines in these short cytoplasmic domains are compared to each other (Fig. 4 *A*; also discussed by Davis et al., 1987), indicating that any common structural motif surrounding the tyrosine will be formed by allowing more than one type of amino acid at each position.

Fig. 4 *A* presents an alignment of the cytoplasmic amino acid sequences of thirteen proteins known, or in one case, ELAM-1, suspected, to be endocytosed. In the first seven sequences, the tyrosine chosen has been shown to be important for internalization. Of the other six sequences, only the FcR and pIgR contain more than one tyrosine. For the FcR, the sequence shown is contained within a region demonstrated to be important for endocytosis (Miettinen et al., 1989). For the pIgR, data from Breitfield et al. (1990) and ourselves (Shia, S. P., and M. G. Roth, unpublished observations), suggests that both cytoplasmic tyrosines contribute to the endocytosis of pIgR. The membrane proximal tyrosine of that protein is included in Fig. 4 *A*; the sequence surrounding the other tyrosine has no features in common with the others shown in the figure (see Discussion). For 12 of the 13 sequences shown, the tyrosine is predicted to be located near a local maximum for surface probability and hydrophilicity when analyzed by the program "peptide structure" (Wolf et al., 1988). The exception is the tyrosine of the FcR, which has a slightly negative surface prediction.

All sequences were oriented with the transmembrane-proximal residues to the left: for most proteins this means NH_2- to COOH-terminus, but for some (marked *) it means COOH- to NH_2-terminus (see Discussion). The number of amino acids used for the alignment was based on two types of observations. (*a*) Only three amino acids on the membrane-distal side of the tyrosine were included because the lysosomal proteins (Lgp-As and Lgp-Bs) contain only three, and the asialoglycoprotein receptors only four residues on that side. (*b*) For the transmembrane-proximal side of the tyrosine, eight amino acids before the tyrosine were included because that position marks the junction with the transmembrane domain for the lysosomal membrane proteins and prob-

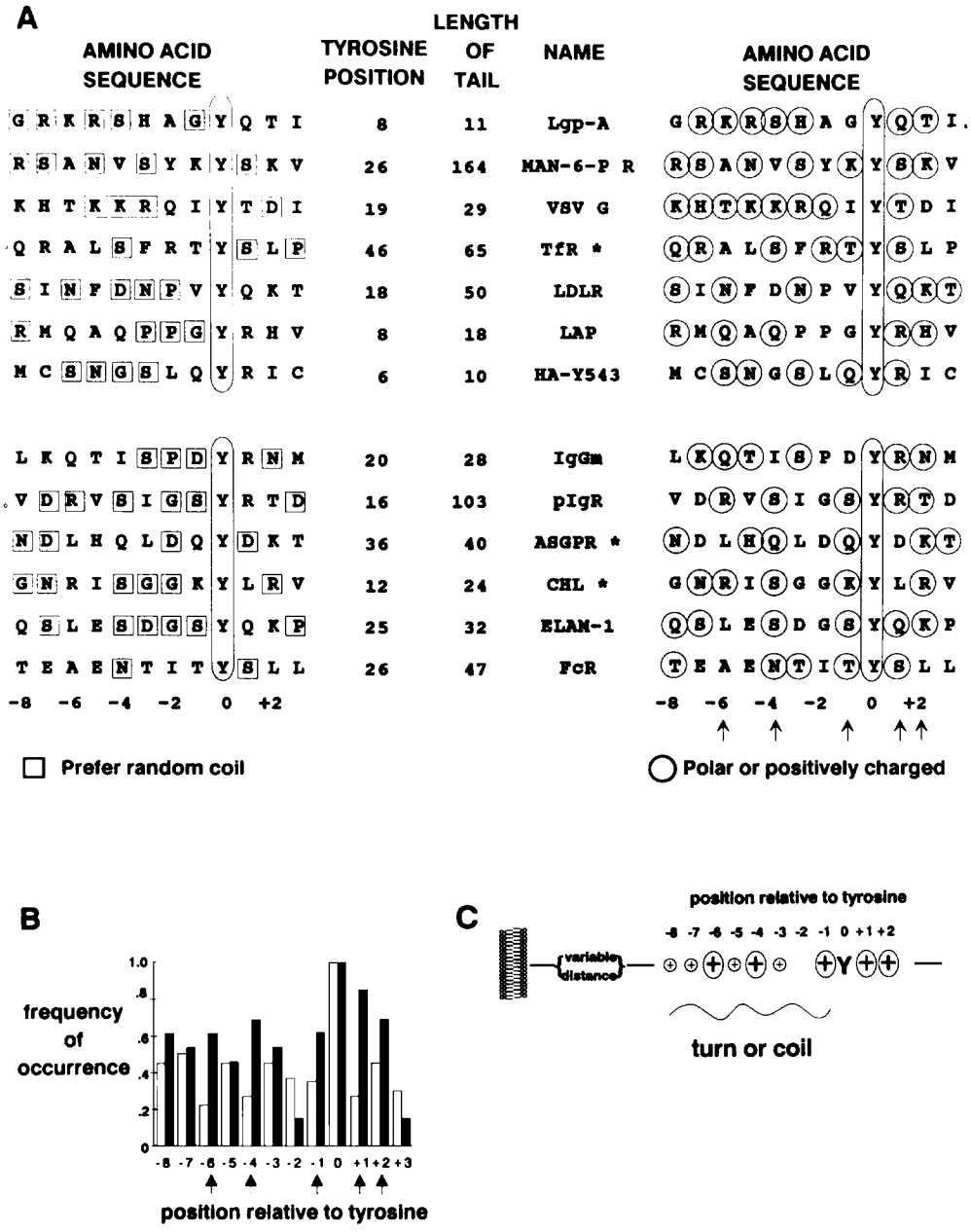


Figure 4. Common motifs in the cytoplasmic sequences of endocytosed proteins. (*A*) Alignment of the cytoplasmic sequences of endocytosed proteins. Sequences of 12 amino acids from the cytoplasmic domains of 13 proteins are shown with the membrane proximal residues to the left. The position of the tyrosine with respect to the first cytoplasmic amino acid as well as the total number of amino acids of the domain are indicated. Unless marked (*), the orientation is from NH_2- to COOH-terminus. The sequences are shown twice; once to emphasize residues that favor a random coil conformation (*left side*) and once to indicate residues that are polar or positively charged (*right side*). The arrows beneath the sequences on the right indicate positions where polar or positively charged amino acids are most abundant. The sequences shown are according to the following references: Lgp-A, Kornfeld and Mellman, 1989; Man-6-P R, Lobel et al., 1987; VSV G, Rose and Gallione, 1981; TfR, McClelland et al., 1984; LDLR., Yamamoto et al., 1984; LAP, Pohlmann et al., 1988; HA-Y543, Lazarovits and Roth, 1988; IgG_m, Kataoka-Yamawaki et al., 1982; pIgR, Mostov et al., 1984; ASGPR, Spiess et al., 1985; ELAM-1, Bevilacqua et al., 1989; CHL, Drickamer, 1981; FcR, Lewis et al., 1986. (*B*) The frequency of occurrence of polar or positively charged amino acids for the sequences shown in *A* (*black bars*) is compared to the frequency of occurrence in an alignment of 40 tyrosines from the external domains of the same thirteen proteins (*white bars*). Arrows indicate positions where the increased frequency is significant with a confidence level of >95%. (*C*) A schematic representation of the endocytosis signal. The size of the symbols for polar and basic residues indicates relative importance.

ably for HA-Y543 as well (see Lazarovits et al., 1990). To discuss relationships between these sequences, we have designated the position of the tyrosine as "0" and number positions in both directions starting from the tyrosine. Amino acids located between the tyrosine and the transmembrane domain are given negative numbers and those located on the membrane-distal side are given positive numbers.

Two patterns emerge from this alignment. The stretch of amino acids before the tyrosine is rich in residues (G, N, S, P, D, R) which are frequently found in a random coil conformation or in turns (Gibrat et al., 1987). When the sequences are considered as a whole, 46% of the amino acids in this region favor a random coil conformation (Fig. 4 *A*, *left side*). This percentage is significantly higher (with a confidence level of >99.99% using the binomial distribution) than the percentage of such residues expected for a random amino acid arrangement (27% would be expected using the amino acid frequencies in human proteins [Lathe, 1985]). In 11 of 13 sequences these amino acids occur in a group of two to four, but no spatial conservation is apparent when comparing one sequence to another. On the membrane-distal side of the tyrosine there is no significant preference for residues favorable to random coil conformation.

The region near the tyrosine, with the exception of position -2, is also rich in polar (S, T, Q, N) or positively charged (R, K, H) residues (Fig. 4 *A*, *right side*). The frequency of

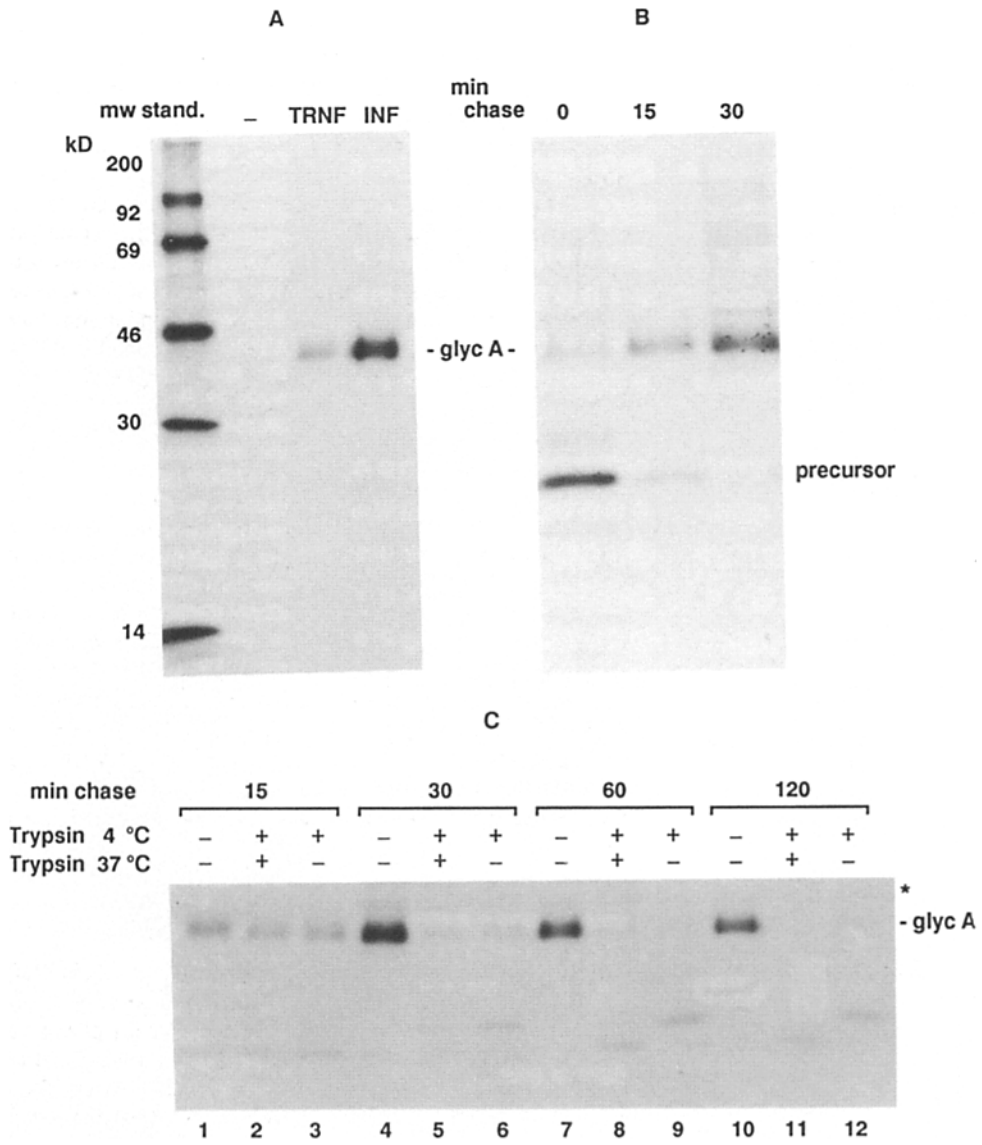


Figure 5. Expression, biosynthesis, and transport of wild-type glycophorin A in CV-1 cells after radiolabeling, immunoprecipitation, PAGE, and fluorography. (*A*) CV-1 cells were infected with a recombinant SV-40 virus expressing HA (−) or glycophorin A (*INF*), or were transfected with a glycophorin A plasmid vector (*TRNF*). 42 h after infection or transfection, cells were radiolabeled for 10 min and chased for 2 h. (*B*) CV-1 cells expressing wild-type glycophorin A were radiolabeled as described above and chased for the times shown. Under these conditions, the half-life of glycophorin A in CV-1 cells is 4 h. (*C*) Triplicate samples of CV-1 cells expressing glycophorin A were radiolabeled for 10 min and chased for the times shown. The asterisk in these and subsequent photographs indicates a 46-kD protein which appears to coprecipitate nonspecifically with glycophorin.

positively charged or polar residues for each position ranges from 46 to 85%, with the positions +2, +1, −1, −4, −6, and −8 most prominent, where the percentages are 69, 85, 62, 69, 62, and 62%, respectively (Fig. 4 *B*, *black bars*). According to the frequencies with which these residues are found in human proteins (R = 5.6%, K = 7.0%, H = 2.5%, S = 8.1%, T = 5.6%, Q = 4.5%, and N = 3.5% [Lathe, 1985]), the expected frequency for a "random" sequence is 36%. In the 13 amino acid sequences compared, 62% of the amino acids are polar or positively charged, an increase over random occurrence that is significant at a >99.99% confidence level by the binomial distribution. For individual positions, the preference for polar or positively charged amino acids is significant with a confidence level of >95% for positions −6, −4, −1, +1, and +2. There was no significant preference or exclusion of negative charges at any position. Alignment of 40 sequences surrounding tyrosines found in the external domains of the proteins listed in Fig. 4 *A* results in an overall abundance of polar or positively charged residues of 34%, and frequencies at positions −6, −4, −1, +1, and +2 of 22, 27, 35, 27, and 45%, respectively (Fig. 4 *B*, *white bars*). Thus, the feature recognized by coated pits for efficient internalization appeared to consist of a tyrosine that was (*a*) preceded by a short region in which helical or sheet structure is suppressed; and (*b*) surrounded by polar or positively charged amino acids, preferentially located at positions +2, +1, −1, −4, and −6 (Fig. 4 *C*).

The lack of conservation of primary sequence around tyrosines, and the observation that single amino acid changes at one position of the HA-Y543 sequence could be compensated by changes at a second position, indicated that to test our hypothesis, additional mutagenesis of the HA-Y543 sequence would require a large number of multiple amino acid changes. However, a more powerful test of the signal proposed in Fig. 4 *C* would be to use recombinant techniques to build an internalization sequence into the cytoplasmic domain of a protein that lacked one. In its simplest form, this would require identification of positions within the cytoplasmic amino acid sequences of proteins that are excluded from coated pits where a tyrosine could be introduced to create an internalization signal, and, conversely, where tyrosines would not have that function. As a test of our hypothesis, we

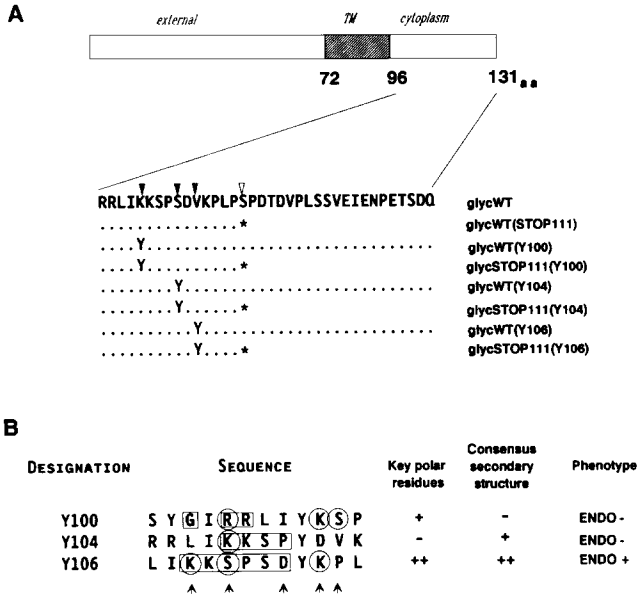


Figure 6. Mutagenesis of the cytoplasmic domain of glycophorin A. (*A*) The sequence of 36 amino acids of the cytoplasmic domain of glycophorin A is shown, as well as the sequences of all mutants created. Arrows indicate positions where tyrosines or a stop codon were introduced. The orientation of the sequence is from NH_2- to COOH-terminus. (*B*) Alignment of the sequences surrounding the three tyrosines introduced in the cytoplasmic domain of glycophorin A. Polar or basic amino acids at conserved positions are circled and those favoring random coil are boxed.

identified a second protein that was excluded from coated pits in CV-1 cells, and converted it into one that was endocytosed by introducing a tyrosine residue into a sequence of the cytoplasmic domain that had the proper context.

Human Glycophorin A Is Not Endocytosed When Expressed in CV-1 Cells

Glycophorin A, the major sialoglycoprotein of the erythrocyte cell membrane, contains no cytoplasmic tyrosines (Tomita and Marchesi, 1976). When CV-1 cells were either transfected with a plasmid vector or infected with recombinant SV40 virus carrying a cDNA for glycophorin A (Siebert and Fukuda, 1986), we were able to immunoprecipitate a protein of the apparent molecular mass of glycophorin, 43 kD, using antibodies specific for glycophorin A (Fig. 5 *A*, TRNF, INF). No protein was immunoprecipitated under the same conditions from CV-1 cells that were infected with SV40 virus encoding HA (Fig. 5 *A*, −). Glycophorin A labeled with ^{35}S methionine and cysteine for 10 min was first detected as a precursor with an apparent molecular mass of 22 kD (Fig. 5 *B*, 0-min chase) that was rapidly converted to the highly glycosylated, fully processed form of molecular mass 43 kD (Fig. 5 *B*, 15- and 30-min chase). Glycophorin A arrived at the surface of CV-1 cells within 30 min after synthesis and had a half-life on the cell surface of ~4 h. Our antiserum for glycophorin also immunoprecipitated an unrelated polypeptide that migrated as a sharp band slightly above glycophorin A during electrophoresis on polyacrylamide gels (marked with an asterisk in Fig. 5). This polypeptide of apparent molecular mass of 46 kD did not appear to associate specifically with glycophorin A, as the ratio of this band to glycophorin varied from experiment to experiment, and did not correlate with the endocytic phenotype of any of the mutant glycophorin A molecules.

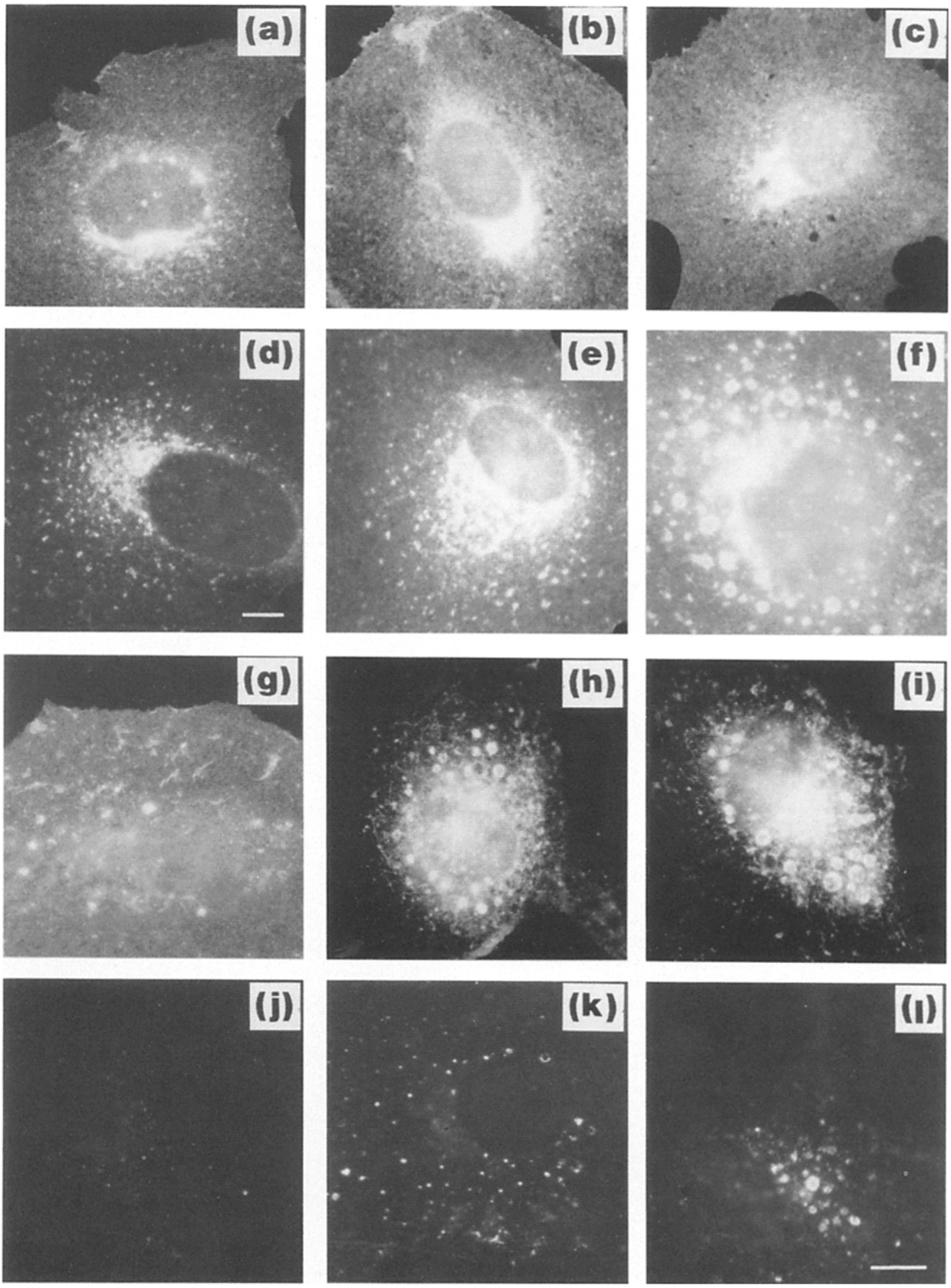
To determine whether the wild-type glycophorin was endocytosed in CV-1 cells, triplicate cell cultures were labeled with radioactive methionine and cysteine and chased for various periods (Fig. 5 *C*). For each chase period, trypsin was included in the chase medium of one sample so that any protein arriving at the cell surface would be digested. Another sample was treated with trypsin at 2°C at the end of the chase so that any proteins that were internal at that time would be protected from digestion. A third sample was not treated with trypsin. Most of glycophorin arrived at the cell surface by 30 min after synthesis, as evidenced by the loss of protein in the sample containing trypsin in the chase medium (Fig. 5 *C*, compare lanes *4* and *5*). After arrival at the cell surface, none of the protein was internalized as is indicated by the absence of bands corresponding to intact glycophorin A in lanes *9* and *12* of Fig. 5 *C*. The lower-mobility bands seen in the samples treated with trypsin (lanes *5*, *6*, *8*, *9*, *11*, *12*) correspond to tryptic fragments of glycophorin encompassing the transmembrane and cytoplasmic domains that could be immunoprecipitated with antibodies specific for the glycophorin cytoplasmic tail (data not shown). The observation that all of the newly synthesized glycophorin A arrived at the cell surface but was not internalized was reinforced by experiments employing immunofluorescence and EM (see below).

Mutagenesis of the Cytoplasmic Domain of Glycophorin A

The wild-type glycophorin A cytoplasmic domain is composed of 36 amino acids and lacks a tyrosine (Fig. 6 *A*). Amino acids at positions 100, 104, and 106 of the glycophorin cytoplasmic domain were converted to tyrosines, since these positions were located in sequences that fulfilled either one or both of our postulates concerning the tyrosine recognition signal (Fig. 6 *B*). The mutant glycWT(Y106) was expected to be internalized since (*a*) immediately upstream of the tyrosine there are four contiguous amino acids (including a proline) which favor random coil; and (*b*) in three of the five most important neighboring positions there are polar or positively charged residues (positions +1, −4, and −6). In contrast, the mutant glycWT(Y104) fulfilled the secondary structure requirement but only one of the five "critical" positions held a polar or positively charged residue (−4); whereas the glycWT(Y100) mutant contained polar or positively charged residues at three of the five critical positions (+2, +1, −4), but had only one residue favoring random coil conformation on the membrane-proximal side of the tyrosine. The identical mutations were introduced into a gene truncated at position 111 to produce glycophorins lacking the carboxy-terminal 20 amino acids, a region thought to interact with the cytoskeleton in erythrocytes (Anderson and Lovrien, 1984). The mutant proteins were expressed in CV-1 cells and their ability to be internalized was determined.

Only GlycWT(Y106) and GlycSTOP111(Y106) Are Endocytosed

When CV-1 cells expressing wild-type and mutant glycopho-


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)

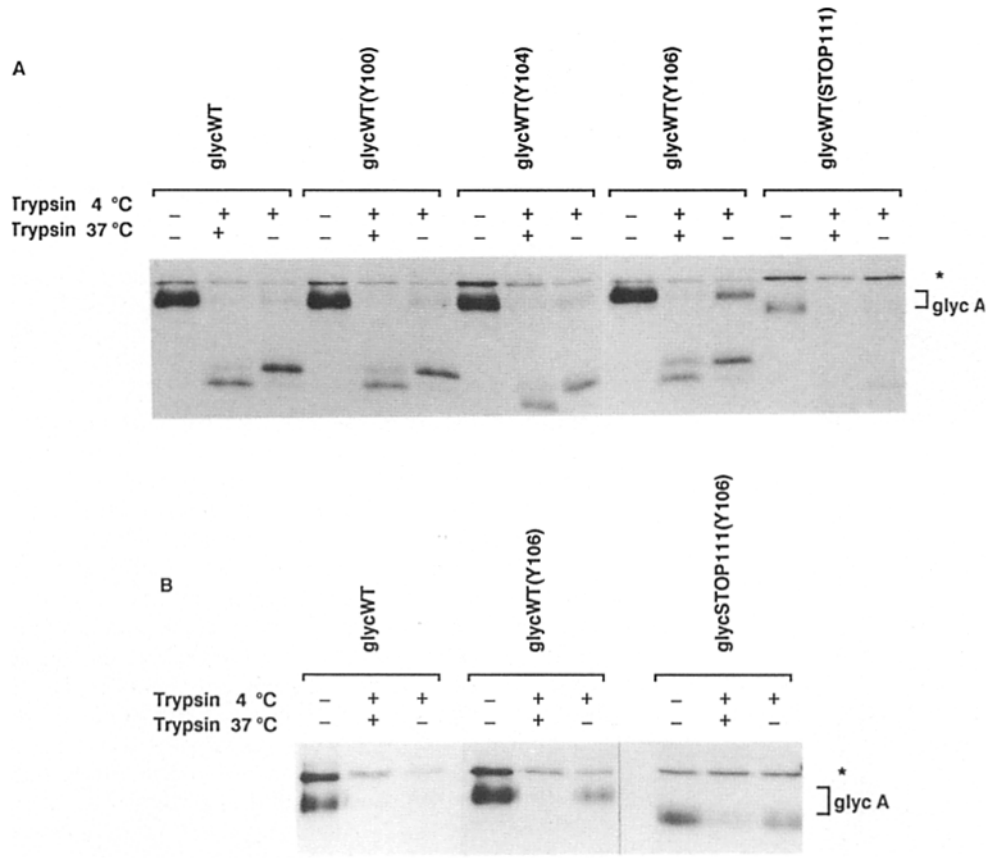


Figure 8. Distribution of wild-type and mutant glycophorin A molecules at steady state. CV-1 cells expressing the wild-type and mutant glycophorin A molecules were radiolabeled for 10 min and chased for 2 h in triplicate as indicated. The presence of an intact glycophorin A band in samples that received trypsin at 4°C indicates protein that has been internalized. (*A*) 44 h; (*B*) 40 h postinfection. * indicates a 46-kD protein unrelated to glycophorin.

rins were fixed and examined by indirect immunofluorescence, those expressing the wild-type protein showed a diffuse pattern of fluorescence superimposed upon a more concentrated perinuclear staining (Fig. 7 *a*) that was identical to the pattern obtained with cells expressing wild-type HA. Double infections with HA and glycophorin A followed by double staining with FITC and Texas red resulted in complete colocalization of the two proteins, suggesting that glycophorin A, like HA, is excluded from coated pits (data not shown). The pattern of fluorescence of the mutants GlycWT(Y100) and GlycWT(Y104) (Fig. 7, *b* and *c*, respectively) as well as of GlycWT(STOP111), GlycSTOP111(Y100), and GlycSTOP111(Y104) (data not shown) was identical to the wild-type glycophorin, suggesting that the mutants containing incomplete internalization signals failed to be endocytosed. In contrast, cells expressing GlycWT(Y106) or GlycSTOP111(Y106) contained a punctate pattern of immunofluorescence in addition to the diffuse and perinuclear staining patterns (Fig. 7, *d*, *e*, and *f*). To determine whether this pattern of immunofluorescence was caused by entry of these mutants into endocytic vesicles, cells expressing wild-type and mutant glycophorins were treated with cycloheximide to allow newly synthesized proteins to exit the exocytic pathway and were then treated with chloroquine (in the continued presence of cycloheximide) to prevent the degradation of internalized proteins. The results of such an experiment for the wild-type glycophorin and for GlycWT(Y106) and GlycSTOP111(Y106) are shown in Fig. 7 (*g–i*). In the presence of the two drugs, wild-type glycophorin A was cleared from the exocytic pathway (as evidenced by the absence of any perinuclear staining) and very few vesicles with immunoreactive material were observed (Fig. 7 *g*). Cells expressing glycophorins with tyrosines at positions 100 and 104 exhibited a pattern of immunofluorescence identical to the wild-type (not shown). In contrast, much of the immunoreactivity of cells expressing GlycWT(Y106) and GlycSTOP111(Y106) was observed in large vesicles and the diffuse cellular staining was reduced, indicating that the protein was being depleted

Figure 7. Indirect immunofluorescence of wild-type and mutant glycophorin A molecules expressed in CV-1 cells. (*a–f*) 40 h after infection, the cells were fixed, permeabilized, and treated sequentially with rabbit antiglycophorin A, biotin-conjugated goat anti-rabbit, and Texas red-conjugated avidin. (*a*) Wild type glycophorin A; (*b*) GlycWT(Y100); (*c*) GlycWT(Y104); (*d* and *e*) GlycWT(Y106); (*f*) GlycSTOP111(Y106). (*g–i*) Before fixation the cells were treated at 37°C with 100 μg/ml of cycloheximide for 2 h followed by 100 μg/ml of cycloheximide plus 50 μM chloroquine for 4 h, and labeled as above. (*g*) Wild type glycophorin A; (*h*) GlycWT(Y106); (*i*) GlycSTOP111(Y106). (*j–l*) Living cells were incubated with antiglycophorin A antibodies at 37°C for 30 min to allow internalization of antibody–glycophorin complexes. Unbound antibodies were then washed off and the cells were fixed, permeabilized, and stained with FITC-conjugated goat antirabbit. (*j*) Wild-type glycophorin A; (*k*) GlycWT(Y106); (*l*) GlycSTOP111(Y106). Bars, 5 μm.

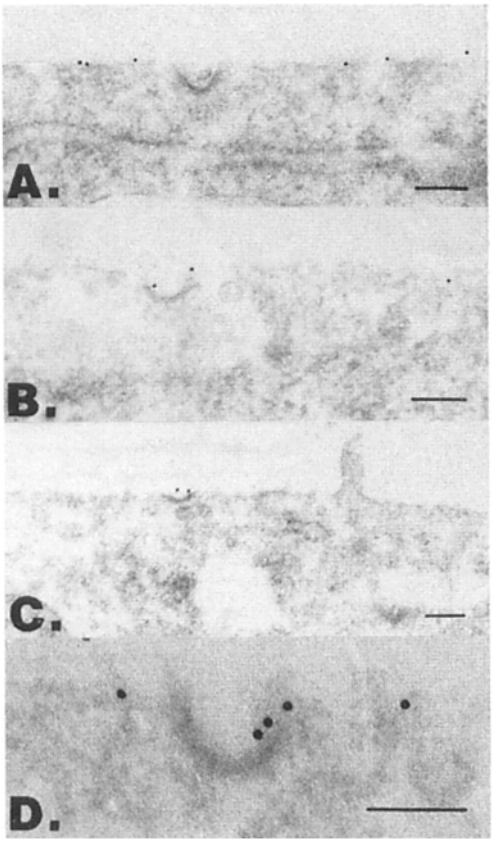


Figure 9. EM of CV-1 cells expressing wild-type and mutant glycophorins. CV-1 cells expressing (*A*) glycWT; (*B* and *D*) glycWT(Y106); and (*C*) glyWTSTOP111(Y106) were labeled on the surface with antiglycophorin antibodies followed by protein A-10-nm gold conjugates. The monolayers were embedded in plastic, sectioned, and stained with uranyl acetate before examination in the electron microscope. Bars, 100 nm.

from the plasma membrane and trapped in endocytic vesicles (Fig. 7, *h* and *i*). To demonstrate internalization directly, CV-1 cells expressing glycophorin were incubated at 37°C with antiglycophorin antibodies for 30 min. After removal of unbound antibodies, the cells were fixed, permeabilized, and stained with fluorescent second antibody (Fig. 7, *j–l*). Cells expressing GlycWT(Y106) and GlycSTOP111(Y106) accumulated immunoreactive material in intracellular vesicles (*k* and *l*), whereas cells expressing the wild-type glycophorin did not (*j*). Thus, GlycWT(Y106) and GlycSTOP-111(Y106), which contain a tyrosine within a context predicted to form an internalization signal, were internalized, whereas mutants of glycophorin containing tyrosines in unfavorable positions were not. Like HA, glycophorins lacking a tyrosine internalization signal remained at the cell surface even when bound by polyvalent antibodies.

For all three sets of cells the diffuse fluorescence at the cell surface was greatly diminished after 30 min incubation at 37°C. This was due to the elution of antibodies from unfixed glycophorin, a property of our polyclonal antibody that prevented the measurement of the rate of internalization of mutant glycophorins by the antibody internalization assay. However, the steady-state distribution of the mutant glycophorin A proteins was determined using degradation by extracellular trypsin as a means to differentiate internal and cell surface proteins. This assay is identical to that shown in Fig. 1 *A* for mutant HAs, with the exception that, since the external domain of glycophorin A is degraded by trypsin, it is necessary to compare the percentage of the protein surviving treatment with trypsin to the amount recovered in the absence of trypsin treatment (Materials and Methods). The results of some of these experiments are shown in Fig. 8. The absence of intact glycophorin A in samples that received trypsin at 37°C during the chase indicated that the wild-type and mutant proteins had arrived at the cell surface by 2 h after synthesis (Fig. 8 *A*, second lane in each set of mutants). When trypsin was added at 2°C at the end of the chase period, two classes of proteins were evident (Fig. 8 *A*, third lane in each set of mutants). Wild-type glycophorin, glycWT(Y100), glycWT(Y104), and glycWT(STOP111) were digested to >95% by extracellular trypsin; thus, an assay that detects internalization even of proteins with incomplete internalization signals (shown for HA, compare Fig. 1 *B* and Fig. 2 *B*) indicated that these proteins were not internalized. The same was true for glycSTOP111(Y100) and glycSTOP111(Y104), data not shown. In contrast, at the end of the chase a portion of glycWT(Y106) was protected from extracellular trypsin indicating that some of this mutant protein was internal (Fig. 8 *A*). At 40 h after infection, 30% of GlycWT(Y106) or GlycSTOP111(Y106) was internal at steady state as compared to <3% for the wild-type protein and the other tyrosine mutants (Fig. 8 *B*).

Table I. Immunogold Labeling of Wild-type and Mutant Glycophorins

Name	No. of cell profiles	Coated pits per profile	Gold particles per profile	Gold in coated pits	Coated pits labeled with gold
					%
GlycWT	23	1.7 (1–3)	24.7 ± 19.0	0.09 (0–1)	5
Y100	7	1.6 (1–2)	17.3 ± 4.2	0.14 (0–1)	11
Y104	16	1.4 (1–3)	17.4 ± 10.3	0.06 (0–1)	4
Y106	16	1.4 (1–3)	10.9 ± 6.2	1.0 (0–4)	71
Y106STOP111	15	1.8 (1–4)	11.7 ± 5.2	1.13 (0–4)	63

CV-1 cells expressing wild-type or mutant glycophorins were fixed and labeled with antiglycophorin antibodies and protein-A colloidal gold. For cells expressing each type of glycophorin, the percent of clathrin-coated pits that were labeled with gold was determined on cell profiles examined by EM. Numbers in parenthesis indicate ranges of values. Standard deviations are given for numbers of gold particles per cell profile. The percentage of labeled coated pits on cells expressing glycophorins bearing the Y106 mutation differs significantly from those labeled on cells expressing wild-type glycophorin (>99.9% confidence level by the two-sample *t* test). The difference in percentage of coated pits labeled on cells expressing glycY100 or glycWT is not significant ($P > 5.0$, two-sample *t* test).

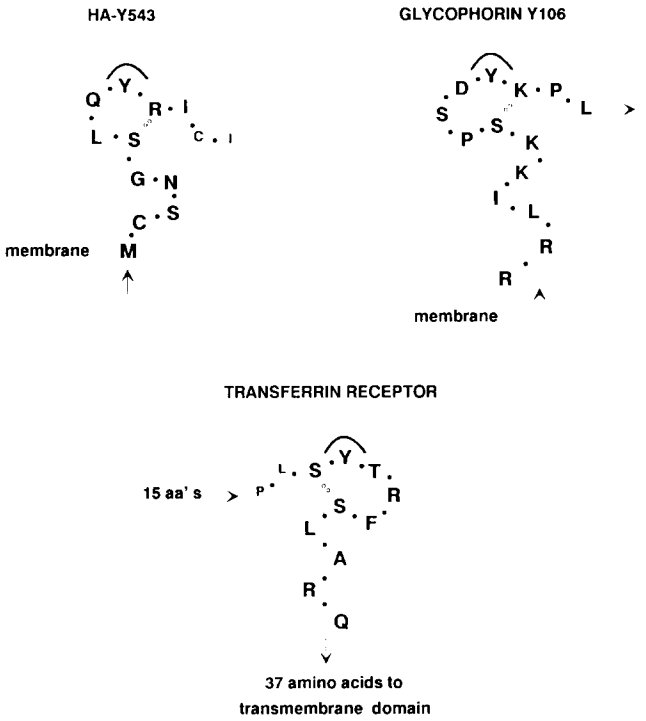


Figure 10. Schematic representation of tyrosine internalization signals. The signal is composed of a loop containing tyrosine to one side. Hydrogen bonding of residues on either side of the tyrosine may stabilize the loop. The arrowhead indicates the polarity of the polypeptide chain.

GlycWT(Y106) and GlycSTOP111(Y106) Enter Coated Pits

To establish that the mutant glycophorins with a tyrosine at position 106 were internalized through clathrin-coated pits, cells expressing various mutants or the wild-type protein were fixed, labeled first with antiglycophorin antibodies and then with protein A gold, and prepared for EM. Coated pits on cells expressing GlycWT(Y106) or GlycSTOP111(Y106) (Fig. 9, *B–D*) were labeled with the gold probe more frequently than coated pits on cells expressing wild-type glycophorin (Fig. 9 *A*). The numbers of coated pits that were labeled or unlabeled in thin sections of cells expressing each type of glycophorin were counted (Table I). Cells expressing glycWT(Y106) or glycSTOP111(Y106) had on average >60% of the coated pits decorated with gold. Thus, the mutation that resulted in an increased internal fraction of glycophorin A at steady state also allowed the protein to enter coated pits. In contrast, only 5–11% of the coated pits were labeled on profiles of cells expressing glycWT, glycY100, or glycY104, suggesting that the smaller internal fraction of these proteins at steady state was caused by their inability to efficiently localize in coated pits for internalization.

Discussion

Classes of genetically unrelated proteins appear to be identified for common intracellular sorting events in at least two different ways. Either proteins can contain a discrete surface feature that is bound by cellular proteins that specify an interaction with a common cellular mechanism, or proteins can share a common physical property that, under the proper conditions, allows them to form an aggregate, complex, or membrane domain. For receptors that enter coated pits, there is evidence supporting each of these two mechanisms (Goldstein et al., 1985; Schlesinger, 1980; Zidovetski et al., 1981). Strong evidence has been reported recently that many receptors require specific cytoplasmic amino acid sequences for localization to coated pits and internalization. In addition, Pearse has presented data that peptides that have some of these cytoplasmic sequences can bind coated pit proteins in vitro (Pearse, 1988). All of these data suggest that many cell surface proteins are sorted into coated pits through a specific binding event with a protein associated with the clathrin coat. However, no strong sequence identities or homologies other than a cytoplasmic tyrosine have been observed to suggest that all of these receptors share a common mechanism for entry into coated pits. This raises the possibility that there may be many different signals for recognition by coated pits, which in turn implies the existence of discrete and possibly noncompeting binding sites. Alternatively, the lack of amino acid identities surrounding tyrosines in the known or suspected "tyrosine internalization signals" of receptor cytoplasmic domains could be explained by the presence of a common surface feature distinctive for its size, location, secondary structure, and tyrosine, but not its primary sequence. The apparent degeneracy of primary sequence of such a signal would be analogous to that observed in the hydrophobic signal peptides for import into the ER, or the NH_2-terminal amphipathic signals for import into mitochondria.

We have proposed features common to a "tyrosine internalization signal" which appears to be sufficient to direct certain transmembrane glycoproteins into the endocytic pathway, and necessary for the internalization of a variety of cell surface proteins. The N-P-X-Y sequence shown to be important for endocytosis of the LDL receptor (Chen et al., 1990) contains several important features of this signal. A previous proposal of a consensus feature for internalization (Vega and Strominger, 1988) differs from our proposal in most important aspects. The generic internalization signal that we propose consists of no more than 8–10 cytoplasmic amino acids with a tyrosine residue present in the membrane-distal portion of the sequence. A statistically significant preference for certain classes of amino acids at specific positions surrounding the tyrosine suggests that the signal is oriented with respect to the membrane, is independent of the polarity of the polypeptide chain, and, for both type I and type II transmembrane proteins, extends from six amino acids membrane proximal to two residues membrane distal to the tyrosine. This statistical analysis is supported by recent results (Jing et al., 1990) published while this manuscript was in preparation, that show that a sequence of 10 amino acids containing a tyrosine in the penultimate position and including the region that we identify as likely to contain a "tyrosine recognition signal" for the transferrin receptor (Fig. 4 *A*) is, in fact, sufficient to specify internalization of that receptor in the absence of other cytoplasmic sequences.

Within the internalization signal, the position of the tyro-

sine is critical with respect to surrounding amino acids. In some proteins the tyrosine can be substituted by phenylalanine (Davis et al., 1987; McGraw et al., 1990), although this is not the case for HA. The results of our mutagenesis of HA, and also experiments with the LDL receptor (Davis et al., 1987; Chen et al., 1990), demonstrate that the adjacent amino acids play an important role in presenting the tyrosine. Our results indicate, however, that these residues may not be individually necessary, since for HA-Y543, many single mutations had modest effects and some mutations that reduced the rate of internalization were compensated by mutations at a second position. Two characteristics of the region surrounding the tyrosine appear to be of primary importance: (*a*) there is a preference for residues that break regular structure to be positioned membrane proximal to the tyrosine; and (*b*) there is a preference for polar or positively charged residues (negative charges are allowed but not preferred) at conserved positions surrounding the tyrosine (Fig. 4 *C*). The two amino acids that break regular structure may serve to disrupt the effect of the transmembrane sequences that probably initiate or dominate the folding of adjacent cytoplasmic sequences. This may explain the observation that the signal appears oriented with respect to the membrane rather than to the polarity of the polypeptide chain. In addition, we believe that these amino acids promote the formation of a specific structure other than alpha-helix or beta-sheet that is required for the tyrosine to be recognized by elements of the coated pits. Most of the internalization signals shown in Fig. 4 *A* (the exception is the FcR) are predicted to contain a turn in the region membrane proximal to the tyrosine when the sequences are analyzed by both the Chou-Fassman and Garnier-Robson-Osguthorpe secondary structure prediction algorithms, and we suggest that the tyrosine recognition signal is, in fact, a small surface loop. The polar and basic residues which are components of the signal could participate in its recognition either by electrostatic or hydrogen-bonding interactions. Our observation that changing the arginine at position +1 of the HA-Y543 signal to glutamic acid produced only a modest reduction in the rate of internalization of HA-Y543 suggests that hydrogen bonding rather than electrostatic interactions are important for the recognition of the signal by its receptor. This hydrogen bonding could occur either between the signal and its receptor, with signals differing in the strength of the interaction depending upon number of bonds formed, or it is possible that the loop is stabilized by hydrogen bonds between residues on either side of the tyrosine, most commonly at positions +1 and −4, as indicated in the cartoon in Fig. 10 for HA-Y543, glycWT(Y106) and the transferrin receptor. The transferrin receptor is included in this figure to emphasize the point that similar loops can be formed in chains running with opposite polarity. In this view, the critical determinants of the signal are a secondary structure that exposes the tyrosine to its receptor, with polar residues at certain positions (Fig. 4 *C*) promoting higher affinity binding. Our analysis does not exclude the possibility that there is more than one class of signal, and thus, more than one signal receptor within coated pits. Even for the sequences exhibiting the common elements that we have described, it is possible that there are related families of signals, each having a different conservation of primary sequence elements and binding to related, distinct receptors.

We have only considered proteins with short cytoplasmic domains and no enzymatic function. Another class of proteins internalized through coated pits includes growth factor receptors, such as those for EGF or insulin, with large cytoplasmic domains that have tyrosine kinase activity. For this class of proteins it has not been shown that tyrosines are involved in signaling internalization. Recently, Chen et al. (1989) described a 48 amino acid long stretch in the cytoplasmic domain of the EGF receptor that is critical for calcium signaling as well as for high- and low-affinity internalization. Although this region contains tyrosines, the context in which they occur is not similar to the one that we describe here but is characterized by a stretch of negatively charged residues. Negatively charged residues are also found preceding the tyrosine at position 734 of the poly-Ig receptor, a residue that has been shown to be important for allowing internalization of that receptor in MDCK cells (Breitfield et al., 1990). In a chimeric HA-pIgR protein, either the membrane proximal tyrosine included among the sequences in Fig. 4 *A*, or the membrane distal tyrosine at 734 were sufficient for internalization (S.-P. Shia and M. Roth, unpublished results). Thus, the available evidence indicates that there must be at least two independent determinants for efficiently directing proteins into the endocytic pathway. It is formally possible that the same protein might contain more than one type of signal and that a single signal within a protein could be cryptic or active, depending upon the degree to which it is exposed.

We thank C. Brewer for reading the manuscript and M. Fukuda for the cDNA for glycophorin-A and for antiserum specific for the cytoplasmic domain of the protein. A. McDowall and H. Arizpe provided expert help with the EM and Katrina Latham made one of the HA mutants.

This work was supported in part by grants GM37547 from the National Institutes of Health and 890757 from the American Heart Association (AHA). M. G. Roth is an Established Investigator of the AHA.

Received for publication 7 May 1990 and in revised form 26 June 1990.

References

Anderson, R. A., and R. E. Lovrien. 1984. Glycophorin is linked by band 4.1 protein to the human erythrocyte membrane skeleton. *Nature (Lond.).* 307:655–658.

Bevilacqua, M. P., S. Stengeli, M. A. Gimbrone, and B. Seed. 1989. Endothelial leukocyte adhesion molecule 1. An inducible receptor for neutrophils related to complement regulatory proteins and lectins. *Science (Wash. DC).* 243:1160–1165.

Breitfield, P. P., J. E. Casanova, W. C. McKinnon, and K. E. Mostov. 1990. Deletions in the cytoplasmic domain of the polymeric immunoglobulin receptor differentially affect endocytotic rate and postendocytotic traffic. *J. Biol. Chem.* In press.

Bretscher, M. S., J. N. Thomson, and B. M. Pearse. 1980. Coated pits act as molecular filters. *Proc. Natl. Acad. Sci. USA.* 77:46–49.

Chen, W., C. Lazar, K. Lund, J. Welsh, C.-P. Chang, G. Walton, C. Der, H. S. Wiley, G. N. Gill, and M. G. Rosenfeld. 1989. Functional independence of the epidermal growth factor receptor from a domain required for ligand-induced internalization and calcium regulation. *Cell.* 59:33–43.

Chen, W. J., J. L. Goldstein, and M. S. Brown. 1990. NPXY, a sequence often found in cytoplasmic tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor. *J. Biol. Chem.* 265:3116–3123.

Davis, C. G., I. R. van Driel, D. W. Russell, M. S. Brown, and J. L. Goldstein. 1987. The low density lipoprotein receptor: identification of amino acids in cytoplasmic domain required for rapid endocytosis. *J. Biol. Chem.* 262: 4075–4082.

Doyle, C., M. Roth, J. Sambrook, and M.-J. Gething. 1985. Mutations in the cytoplasmic domain of the influenza virus hemagglutinin affect different stages of intracellular transport. *J. Cell Biol.* 100:704–714.

Drickamer, K. 1981. Complete amino acid sequence of a membrane receptor for glycoproteins. *J. Biol. Chem.* 256:5827–5839.

Gething, M.-J., and J. Sambrook. 1981. Cell-surface expression of influenza haemagglutinin from a cloned DNA copy of the RNA gene. *Nature (Lond.).* 293:620–625.

Gibrat, J.-F., J. Garnier, and B. Robson. 1987. Further developments of pro-

tein secondary structure prediction using information theory: new parameters and consideration of residue pairs. *J. Mol. Biol.* 198:425–443.

Goldstein, J. L., M. S. Brown, R. G. W. Anderson, D. W. Russell, and W. J. Schneider. 1985. Receptor-mediated endocytosis: concepts emerging from the LDL receptor system. *Annu. Rev. Cell Biol.* 1:1–39.

Jing, S., T. Spencer, K. Miller, C. Hopkins, and I. S. Trowbridge. 1990. Role of the human transferrin receptor cytoplasmic domain in endocytosis: localization of a specific signal sequence for internalization. *J. Cell Biol.* 110:283–294.

Kataoka-Yamawaki, Y., S. Nakai, T. Miyata, and T. Honjo. 1982. Nucleotide sequences of gene segments encoding membrane domains of immunoglobulin gamma chains. *Proc. Natl. Acad. Sci. USA.* 79:2623–2627.

Kornfeld, S., and I. Mellman. 1989. The biogenesis of lysosomes. *Annu. Rev. Cell Biol.* 5:483–526.

Kunkel, T. A. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA.* 82:488–492.

Laemmli, U. K. 1970. Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature (Lond.).* 227:680–685.

Lathe, R. 1985. Synthetic oligonucleotide probes deduced from amino acid sequence data: theoretical and practical considerations. *J. Mol. Biol.* 183: 1–12.

Lazarovits, J., and M. Roth. 1988. A single amino acid change in the cytoplasmic domain allows the influenza virus hemagglutinin to be endocytosed through coated pits. *Cell.* 53:743–752.

Lazarovits, J., S.-P. Shia, N. Ktistakis, M.-S. Lee, C. Bird, and M. G. Roth. 1990. The effects of foreign transmembrane domains on the biosynthesis of the influenza virus hemagglutinin. *J. Biol. Chem.* 265:4760–4767.

Lehrman, M. A., J. L. Goldstein, M. S. Brown, D. W. Russell, and W. J. Schneider. 1985. Internalization-defective LDL receptors produced by genes with nonsense and frameshift mutations that truncate the cytoplasmic domain. *Cell.* 41:735–743.

Lewis, V. A., T. Koch, H. Plutner, and I. Mellman. 1986. A complementary DNA clone for a macrophage-lymphocyte Fc receptor. *Nature (Lond.).* 324:372–375.

Lobel, P., N. M. Dahms, J. Breitmeyer, J. M. Chirgwin, and S. Kornfeld. 1987. Cloning of the bovine 215-kDa cation-independent mannose 6-phosphate receptor. *Proc. Natl. Acad. Sci. USA.* 84:2233–2237.

Lobel, P., K. Fujimoto, D. Ye, G. Griffiths, and S. Kornfeld. 1989. Mutations in the cytoplasmic domain of the 215 kD mannose 6-phosphate receptor differentially alter lysosomal enzyme sorting and endocytosis. *Cell.* 57: 787–796.

Maniatis, T., E. Fritsch, and J. Sambrook. 1982. Molecular Cloning Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.

McClelland, A., L. C. Kuhn, and F. H. Ruddle. 1984. The human transferrin receptor gene: genomic organization, and the complete primary structure of the receptor deduced from a cDNA sequence. *Cell.* 39:267–274.

McGraw, T. E., and F. R. Maxfield. 1990. Human transferrin receptor internalization is partially dependent upon an aromatic amino acid on the cytoplasmic domain. *Cell Regulation.* 1:369–377.

Miettinen, H. M., J. K. Rose, and I. Mellman. 1989. Fc receptor isoforms exhibit distinct abilities for coated pit localization as a result of cytoplasmic domain heterogeneity. *Cell.* 58:317–327.

Mostov, K. E., M. Friedlander, and G. Blobel. 1984. The receptor for transepithelial transport of IgA and IgM contains multiple immunoglobulin-like domains. *Nature (Lond.).* 308:37–43.

Mostov, K. E., A. B. de Kops, and D. L. Deitcher. 1986. Deletion of the cytoplasmic domain of the polymeric immunoglobulin receptor prevents basolateral localization and endocytosis. *Cell.* 47:359–364.

Pearse, B. M. F. 1988. Receptors compete for adaptors found in plasma membrane coated pits. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:3331–3336.

Pohlmann, R., C. Krentler, B. Schmidt, W. Schroder, G. Lorkowski, J. Culley, G. Mersmann, C. Geier, A. Waheed, S. Gottschalk, K.-H. Grzeschik, A. Hasilik, and K. von Figura. 1988. Human lysosomal acid phosphatase: cloning, expression and chromosomal assignment. *EMBO (Eur. Mol. Biol. Organ.). J.* 7:2343–2350.

Prywes, R., E. Lioneh, A. Ullrich, and J. Schlessinger. 1986. Mutations in the cytoplasmic domain of EGF receptor affect EGF binding and receptor internalization. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:2179–2190.

Rose, J. K., and C. J. Gallione. 1981. Nucleotide sequences of the mRNAs encoding the vesicular stomatitis virus G and M proteins determined from cDNA clones containing the complete coding regions. *J. Virol.* 39:519–528.

Roth, M. G., C. Doyle, J. Sambrook, and M.-J. Gething. 1986. Heterologous transmembrane and cytoplasmic domains direct functional chimeric influenza virus hemagglutinins into the endocytic pathway. *J. Cell Biol.* 102: 1271–1283.

Rothenberg, S., B. S. Iacopetta, and L. C. Kuhn. 1987. Endocytosis of the transferrin receptor requires the cytoplasmic domain but not its phosphorylation site. *Cell.* 49:423–431.

Sanger, F., S. Nicklen, and S. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5471.

Schlesinger, J. 1980. The mechanism and role of hormone-induced clustering of membrane receptors. *Trends. Biochem. Sci.* 5:210–214.

Siebert, P. D., and M. Fukuda. 1986. Isolation and characterization of human glycophorin A cDNA clones by a synthetic oligonucleotide approach: Nucleotide sequence and mRNA structure. *Proc. Natl. Acad. Sci. USA.* 83:1665–1669.

Spiess, M., A. L. Schwartz, and H. F. Lodish. 1985. Sequence of human asialoglycoprotein receptor cDNA. *J. Biol. Chem.* 260:1979–1982.

Tomita, M., and V. T. Marchesi. 1975. Amino-acid sequence and oligosaccharide attachment sites of human erythrocyte glycophorin. *Proc. Natl. Acad. Sci. USA.* 72:2964–2968.

Vega, M. A., and J. L. Strominger. 1989. Constitutive endocytosis of HLA class I antigens requires a specific portion of the intracytoplasmic tail that shares structural features with other endocytosed molecules. *Proc. Natl. Acad. Sci. USA.* 86:2688–2692.

Wolf, H., S. Modrow, M. Motz, B. A. Jameson, G. Hermann, and B. Foertsch. 1988. An integrated family of amino acid sequence analysis programs. *Comput. Appl. Biosci.* 4:187–192.

Yamamoto, T., C. G. Davis, M. S. Brown, W. J. Schneider, M. L. Casey, J. L. Goldstein, and D. W. Russell. 1984. The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA. *Cell.* 39:27–38.

Zidovetski, R., Y. Yarden, J. Schlessinger, and T. M. Tovin. 1981. Rotational diffusion of epidermal growth factor complexes to cell surface receptors reflects rapid microaggregation and endocytosis of occupied receptors. *Proc. Natl. Acad. Sci. USA.* 78:6981–6985.