Role of N-Linked Oligosaccharide Flexibility in Mannose Phosphorylation of Lysosomal Enzyme Cathepsin L*

Received for publication, April 1, 2002, and in revised form, August 26, 2002
Published, JBC Papers in Press, August 28, 2002, DOI 10.1074/jbc.M203097200

Jason B. Warner, Craig Thalhauser, Kai Tao, and G. Gary Sahagian‡

From the Department of Physiology, Tufts University School of Medicine, Boston, Massachusetts 02111

Mannose phosphorylation of N-linked oligosaccharides by UDP-GlcNAc:lysosomal enzyme N-acetylgalactosamine-1-phosphotransferase is a key step in the targeting of lysosomal enzymes in mammalian cells and tissues. The selectivity of this process is determined by lysine-based phosphorylation signals shared by lysosomal enzymes of diverse structure and function. By introducing new glycosylation sites at several locations on the surface of mouse procathepsin L and modeling oligosaccharide conformations for sites that are phosphorylated, it was shown that the inherent flexibility of N-linked oligosaccharides can account for the specificity of the transferase for oligosaccharides at different locations on the protein. By using this approach, the physical relationship between the lysine-based signal and the site of phosphorylation of mannose residues was determined. The analysis also revealed the existence of additional independent lysine-based phosphorylation signals on procathepsin L, which account for the low level of phosphorylation observed when the primary Lys-54/Lys-99 signal is ablated. Mutagenesis of residues that surround Lys-54 and Lys-99 and demonstration of mannose phosphorylation of a glycosylated derivative of green fluorescent protein provide strong evidence that the cathepsin L phosphorylation signal is a simple structure composed of as few as two well placed lysine residues.

The mammalian lysosomal protein targeting system has the capability of recognizing and modifying lysosomal hydrolases and growth factors from a wide range of protein families with high specificity. The molecular basis for this selectivity is due to the activity of the UDP-GlcNAc:lysosomal enzyme N-acetylgalactosamine-1-phosphotransferase (GlcNAc-1-phosphotransferase), which phosphorylates N-linked oligosaccharides of these proteins by the addition of GlcNAc-1-P. This modification begins after lysosomal proteins are separated by a similar distance (8). With this information, it will be possible to identify substrates for the transferase from information in structural data bases and to engineer phosphorylation signals on macromolecules for therapeutic purposes. In our previous study (17), we compared lysine-based signals that were identified for cathepsin L and cathepsin D, and we found that critical lysine residues were separated by a similar distance (~34 Å) in the two proteins. In this study, we further define the cathepsin L signal by determining its relationship to the site of oligosaccharide phosphorylation. We also examine the involvement of residues in the vicinity of critical lysine residues and provide additional evidence for the simplicity and generality of the signal.
EXPERIMENTAL PROCEDURES

Enzymes, Antibodies, cDNAs, and Other Reagents—Antibodies to mammalian L and green fluorescent protein (GFP) were purchased from rabbits as described previously (18). LipofectAMINE PLUS was purchased from Invitrogen. EcoScint H was acquired from National Diagnostics. The QuikChange Mutagenesis Kit and Pfu Turbo were obtained from Stratagene. DpnI and N-glycanase (peptide N-glycosidase F) were purchased from New England Biolabs. Other restriction enzymes were acquired from Invitrogen. Chemical reagents were purchased from Sigma.

Cells and Growth Conditions—COS-1 cells were purchased from the American Type Culture Collection and were cultured in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% fetal calf serum, 100 units/ml penicillin, and 100 μg/ml streptomycin and maintained at 37 °C in a humidified atmosphere of 5% CO2. Mouse epithelial cells (A549) suitable for transient or stable expression of mouse cathepsin L in mammalian cells were a gift from Dr. D. Underwood (University of California at Los Angeles). A549 cells (106) were transiently transfected with pED4neoSL-CL using the LipofectAMINE PLUS reagent according to the manufacturer’s instructions. Cultures were harvested in early exponential growth by trypsinization and centrifugation (1000 × g for 5 min).

Subcloning of Mouse Cathepsin L—A fragment containing the full coding sequence of mouse cathepsin L (19) was subcloned into the pED4neo vector (20) for expression in COS-1 cells. The pED4neo vector was modified by the insertion of a synthetic linker made from the complementary oligonucleotides 5′-ATT TTC TCG AGA CGG TGG CCG CGG CGA ATT CGT CGA CTC TAG AA-3′ and 5′-AAT TTT CTA GAG GAC TCG AAT CCG CCG CAC CAG TCT CGA GA-3′ into the EcoRI cloning site of the vector. The linker contains XhoI, AgeI, NotI, EcoRI, Sall, and XbaI sites, in 5′–3′ order, and eliminates the EcoRI site into which it was cloned. The modified vector will hereby be referred to as pED4neoSL. Mouse cathepsin L cDNA was ligated into the XhoI and XbaI cloning sites to give an expression vector, pED4neoSL-CL, suitable for transient or stable expression of mouse cathepsin L in eukaryotic cells.

Mutagenesis of Cathepsin L cDNA—All mutagenesis steps were carried out using a modification of the QuikChange Site-directed Mutagenesis kit (Stratagene). This method was chosen because of its high mutation efficiency and because it requires no specialized vectors. All oligonucleotide primers were designed using MutantMaker, a Visual Application developed by one of the authors (J. B. Warner), which expedites the task of primer selection. The program reads in protein text files, generates mutagenic primers based upon the user’s parameters, and prints an oligonucleotide order form, thereby minimizing the potential energy of the resulting structures that possess an acceptable total potential energy value. A maximum potential energy cut-off of 1000 kcal/mol was used that was found to be characteristic of Man,GlcNAc2 conformations lacking clashes and close contacts. Fifty cycles of conjugate gradient minimization were used, which brought energy values of acceptable conformations close to convergence in most cases. Conformations were generated in the context of the protein and torsion angles for rotatable bonds in the oligosaccharide and in the asparagine side chain were randomly set. Generally, less than 1% of the conformations that were generated met the energy criteria. At least 2000 acceptable conformations were obtained for each oligosaccharide. When the acceptable oligosaccharide conformations were overlaid, they formed a spheroid-like structure, such overlaps of oligosaccharide conformations are referred to as oligosaccharide clouds.

In order to assess the overlap between adjacent conformational clouds, a specialized C++ application was developed. This application determines where selected residues in separate conformational clouds can overlap in three-dimensional space. For each Man,GlcNAc2 oligosaccharide there are five mannose residues with a C-6-OH moiety, all of which are susceptible to phosphorylation in vivo (22). Minima (Xmin, Ymin, and Zmin) and maxima (Xmax, Ymax, and Zmax) are determined for the coordinates of the O-6 atom of each phosphorylatable mannose in each conformational cloud. The cube generated from these coordinates gives a simple but effective definition of the boundaries of each conformational cloud. Boundaries are generated for all clouds from phosphorylated mannosides. The program then calculates the largest minimum and smallest maximum for each coordinate to determine the overlapping region. If the free O-6 atom of a mannose residue is contained in this overlapping region, its coordinates are retained.

Generation of a Secreted, Glycosylated Form of GFP—The cDNA of proliferin-1, a secreted mannose 6-phosphate containing glycoprotein, was purchased from American Type Culture Collection and subcloned into the pED4neo expression vector. The signal sequence from proliferin-1 was amplified from pED4neo-proliferin using the following primers: 5′-GTG AAG TTC GAG AAC GGC ACC CTG GTG AAC CGC-3′ (forward primer) and 5′-GCC CGG TGC ACA GAC GGG TTA AGG-3′ (reverse primer). The PCR product was purified, cut with XhoI and Sall, and ligated into the EcoRI and Sall cloning sites of the pEGFP-N1 vector (Clontech). The modified GFP sequence was then subcloned into the pED4neoSL vector using the XhoI and NotI cloning sites. A glycosylation site was added to residues 156–158 on the secretary GFP by site-directed mutagenesis, using the following primers: 5′-GGT AAG TTC GAG AAC GGC ACC CTG GTG AAC CCG-3′ (forward primer) and 5′-GCC GTT CAC GAG GGT GCC CTC GAA CTT CAC-3′ (reverse primer). The location of the glycosylation site positions the oligosaccharide on the rim of the β-barrel structure.

RESULTS

Topography of Mannose Phosphorylation of Cathepsin L—To examine the requirement for oligosaccharide placement, glycosylation sites were placed at selected locations on the surface of cathepsin L by site-directed mutagenesis, and glycosylation and mannose phosphorylation at each site were determined after transient expression in COS-1 cells. Thirty-three glycosylation site mutations spread throughout the protein were created and tested. To maximize the number of properly folded and glycosylated proteins, surface residues in regions lacking α-helical structure were chosen for placement of the sites.

Wild-type cathepsin L contains two potential glycosylation sites, only one (Asn-221) of which is utilized in the properly folded protein (19). Because the cryptic site at Asn-288 can be
glycosylated in some modified cathepsin L proteins (19, 23), both the utilized site (Asn-221) and the non-utilized site (Asn-268) were mutated to glutamine, creating a construct encoding a nonglycosylated cathepsin L protein. When expressed, the protein was secreted poorly as expected (Fig. 2) since glycosylation is needed for proper folding of cathepsin L in the COS-1 expression system (21). This construct was used for construction of the altered glycosylation site proteins listed in Table I.

Results on the synthesis, secretion, and glycosylation of selected altered glycosylation site proteins are shown in Fig. 2, and data for all of the constructs are summarized in Table I. Synthesis and secretion of the proteins, as determined by biosynthetic labeling in the presence of NH4Cl, was used as a means of assessing whether or not altered cathepsin proteins are folded properly. NH4Cl inhibits mannose 6-phosphate receptor function and causes the quantitative secretion of newly synthesized mannose-phosphorylated proteins. We have examined previously (13, 14, 17, 19, 21) more than 100 altered cathepsin L and cathepsin D proteins using the COS cell expression system. In all cases examined, normally folded proteins, as determined by level of catalytic activity (13), susceptibility to proteolytic digestion (14, 17, 21), or susceptibility to heat denaturation (13), are secreted under these conditions, whereas abnormally folded proteins are retained within the cells and/or degraded. Of the 33 altered glycosylation site proteins created for this study, 19 were glycosylated and efficiently secreted (Fig. 2). These were retained for further analysis. Endoglycosidase H (endo H) treatment was used to examine the state of glycosylation of the proteins (Table I). Oligosaccharides on proteins that undergo mannose phosphorylation would be expected to remain sensitive to the treatment, whereas those that do not would be expected to be further processed to forms that are resistant to the treatment. Of the 19 retained proteins, 5 were sensitive to endo H treatment, 7 were partially sensitive, and 7 were resistant to the treatment.

Biosynthetic labeling with [32P]phosphate and [3H]leucine was used to determine which of the altered glycosylation site proteins were susceptible to mannose phosphorylation (Fig. 3). Five of the constructs (Asn-105, Asn-158, Asn-171, Asn-217, and Asn-229) displayed a high level of phosphorylation approaching the phosphorylation level of wild-type procathepsin L. All other constructs showed minimal (<25%) phosphorylation compared with the wild-type protein. Phosphorylation of the 5 highly phosphorylated constructs was inhibited by mutation of Lys-54 and Lys-99 to alanine (Fig. 4) indicating that all 5 constructs utilize the previously identified mannose phosphorylation signal (13, 14).

The topographical locations of engineered glycosylation sites on the surface of the protein are shown in Fig. 5. The locations of the glycosylation sites of the highly phosphorylated constructs were found to be clustered within the vicinity of Asn-221, the wild-type glycosylation site. Most of the engineered sites in the vicinity of Asn-221 were highly phosphorylated. Some sites in this region were phosphorylated weakly (Asn-104, Asn-199, and Asn-208) or not at all (Asn-156 and Asn-108). The low level of phosphorylation at these sites is attributed to site-dependent effects that reduce accessibility of the oligosaccharides to the transferase. Such effects would include protein-oligosaccharide interactions that limit oligosaccharide flexibility and site-dependent differences in oligosaccharide processing that inhibit or prevent phosphorylation.

Flexibility of N-Linked Oligosaccharide Accounts for Topography of Mannose Phosphorylation of Cathepsin L—Previous studies (22, 24) have shown that mannose-phosphorylated proteins display a surprising heterogeneity with regard to sites of phosphorylation on the oligosaccharide and the number and location of the phosphorylated oligosaccharides on the protein. To account for this apparent lack of specificity, we have proposed that selectivity of the reaction for specific mannose 6-phosphate residues on protein-linked oligosaccharides is dic-
tated in large part by the ability of those residues to migrate in three-dimensional space to the catalytic site of GlcNAc-1-phosphotransferase when the protein is bound to the transferase through its lysine-based phosphorylation signal (17). This hypothesis is supported by NMR studies, which have indicated that N-linked oligosaccharides are flexible in solution and that, although preferred conformations may exist, such conformations are short lived (25–28).

**Table I**

Synthesis, secretion, and glycosylation of altered glycosylation site constructs

| Position | Mutation | Expression† | Glycosylation‡ | Secretion§ | Endo H sensitivity¶ |
|----------|----------|-------------|----------------|-------------|---------------------|
| CL-wt    | None     | High        | Full           | High        | Sensitive           |
| CL-ng†   | NDT → QDT, NCS → QCS | Moderate | None | Low | NA |
| 22       | DQT → NQT | Moderate | Full | Moderate | Resistant |
| 24       | TFS → NPS | Moderate | None | Low | NA |
| 37       | RRL → NRS | Moderate | Partial | UD§ | NA |
| 43       | NEE → NET | Low | Full | UD | NA |
| 63       | NGE → NGT | Moderate | Partial | UD | NA |
| 68       | NGQ → NGS | Moderate | Full | High | Resistant |
| 84       | TNE → NNS | Moderate | Partial | UD | NA |
| 96       | RHQ → NRS | Moderate | Full | High | Partial |
| 98       | QKH → NKS | Moderate | Full | High | Partial |
| 104      | RLF → NLS | High | Full | Partial | NA |
| 105      | LFQ → NFS | High | Full | Partial | Sensitive |
| 106      | FQF → NQS | Moderate | None | UD | NA |
| 108      | EPL → NPS | Moderate | Partial | Low | Resistant |
| 115      | PLS → NKS | Moderate | Full | Moderate | Partial |
| 131      | NGQ → NGS | Moderate | Full | UD | NA |
| 156      | GKL → NKS | Moderate | Partial | Low | Resistant |
| 158      | LIS → NIS | High | Full | High | Partial |
| 170      | HAQ → NAS | High | Full | High | Partial |
| 179      | NGS → NGS | Moderate | Full | UD | NA |
| 199      | NES → NES | High | Full | High | Partial |
| 208      | DGS → NGS | High | Full | High | Partial |
| 217      | FAV → NAS | High | Full | High | Sensitive |
| 229      | PQF → NQS | High | Full | Moderate | Sensitive |
| 271      | SGN → NKS | High | Partial | Moderate | Sensitive |
| 273      | NLQ → NLS | Moderate | Partial | Low | Resistant |
| 286      | EGT → NCT | Moderate | Full | Low | Resistant |
| 288      | TDS → NDS | Moderate | None | Low | NA |
| 289      | DSN → NSS | Moderate | Partial | None | NA |
| 291      | NKN → NKT | Moderate | Partial | Low | Resistant |
| 293      | NYS → NKS | Moderate | Partial | UD | NA |
| 307      | GME → NMS | Low | Full | UD | NA |
| 317      | DRD → NRS | Moderate | None | UD | NA |
| 319      | DNH → NNS | Low | None | UD | NA |

† Based on the amount of cellular procathepsin L protein observed on SDS-PAGE.
‡ Based on the amount of cellular procathepsin L protein that had decreased motility on SDS-PAGE.
§ Based on the amount of labeled, secreted procathepsin L protein compared with the amount of protein labeled during the 1-h chase.
¶ Based on the amount of secreted procathepsin L protein that displayed increased motility on SDS-PAGE after treatment with endo H.
† Non-glycosylated cathepsin L construct.
§ NA, not applicable.
¶ UD, undetectable.

**Fig. 2.** Expression of cathepsin L glycosylation mutants. The original glycosylation sites were removed from wild-type procathepsin L, and new glycosylation sites were added by site-directed mutagenesis. The indicated cathepsin L constructs were expressed in COS-1 cells and split equally into two groups. One group was labeled with [35S]methionine for 1 h. The other group was similarly labeled for 1 h and chased for 6 h in media containing unlabeled methionine. Cathepsin L was immunoprecipitated from cell extracts of pulse-labeled cells and from media of chased cells. Half of the immunoprecipitated proteins were treated with endo H, and treated and untreated immunoprecipitates were subjected to SDS-PAGE and fluorography. A, fluorogram of untreated immunoprecipitates; B, fluorogram of endo H-treated immunoprecipitates. Glycosylated (G) and non-glycosylated (NG) cathepsin L proteins are marked with arrows. CL-ctrl represents the control procathepsin L, and CL-ng represents the non-glycosylated procathepsin L mutant.
To test this hypothesis, three-dimensional oligosaccharide clouds composed of compilations of randomly generated, energetically acceptable Man₈GlcNAc₂ oligosaccharide conformations were modeled at phosphorylated glycosylation sites of procathepsin L as described under “Experimental Procedures.” The graph displays mannose phosphorylation as a percentage of the wild-type cathepsin L (CL), and represents a typical experiment with three replicates.

**FIG. 3.** Mannose phosphorylation of procathepsin L glycosylation mutants. The indicated cathepsin L constructs were expressed in COS-1 cells and double-labeled with [³²P]phosphate and [³H]leucine for 6 h in the presence of 10 mM NH₄Cl. Procathepsin L protein was immunoprecipitated from media and subjected to SDS-PAGE. The protein bands were excised and quantitated in a scintillation counter. The graph displays mannose phosphorylation as a percentage of the wild-type cathepsin L (CL), and represents a typical experiment with three replicates.

**FIG. 4.** Contribution of Lys-54/Lys-99 signal to mannose phosphorylation of highly phosphorylated cathepsin L constructs. Alanine mutations were introduced at Lys-54 and Lys-99 in each highly phosphorylated cathepsin L construct. Constructs with and without these lysine mutations were expressed in COS-1 cells and double-labeled with [³²P]phosphate and [³H]leucine for 6 h in the presence of 10 mM NH₄Cl. Cathepsin L protein was immunoprecipitated from the media and subjected to SDS-PAGE. The labeled protein bands were excised and quantitated in a scintillation counter. The graph displays mannose phosphorylation as a percentage of the wild-type cathepsin L and represents the mean ± S.D. of two experiments with a total of six replicates.

**FIG. 5.** Location of the new glycosylation sites on cathepsin L. The wild-type glycosylation site (Asn-221) is shown in blue. The glycosylation sites that were added but were not significantly phosphorylated are shown in red. Glycosylation sites that are highly phosphorylated are shown in green. The α-carbons of the critical lysines, Lys-54 and Lys-99, are shown in yellow.

To test this hypothesis, three-dimensional oligosaccharide clouds composed of compilations of randomly generated, energetically acceptable Man₈GlcNAc₂ oligosaccharide conformations were modeled at phosphorylated glycosylation sites of procathepsin L as described under “Experimental Procedures.” The oligosaccharide conformations that compose each cloud represent sterically unhindered oligosaccharide conformations that were chosen based on an energy value that would exclude clashes and close contact of atoms within the oligosaccharide and between the oligosaccharide and the protein. Thus, the volume enclosed by each cloud represents the space available to one or more phosphorylatable mannose residues of the oligosaccharide attached at that location on the protein. Clouds for each glycosylation site are shown in Fig. 6. In these representations, only the positions of O-6 atoms of phosphorylatable mannose residues are shown (see Fig. 1). An overlay of all 6
including Lys-157, Lys-233, and Lys-237, contribute to this residual phosphorylation. Phosphorylation of a construct containing alanine mutations at Lys-54, Lys-99, Lys-116, Lys-157, Lys-233 and Lys-237 was >90% inhibited, indicating that these lysine residues can account for virtually all of the mannose phosphorylation of procathepsin L. The residual phosphorylation appears to result from weak lysine-based phosphorylation signals that act independently of the primary Lys-54/Lys-99 signal. The level of residual phosphorylation differs among the engineered glycosylation sites providing additional evidence that this component of the phosphorylation results from independent phosphorylation signals.

Role of Lysine Microenvironment in Mannose Phosphorylation of Cathepsin L—Surface residues within 8 Å of Lys-54 and Lys-99 on wild-type cathepsin L were mutated to alanine in order to determine the importance of the environment surrounding these lysine residues (Table II). No significant change in mannose phosphorylation was observed in single alanine mutations; however, the possibility remained that these residues had a weak interaction with the transferase and worked in concert to provide a stronger contact. In order to address this question, multiple mutations were created in single constructs so that three residues surrounding Lys-54 (construct M56A/R57A/M58A) and five residues surrounding Lys-99 (construct H97A/Q98A/H100A/K101A/K102A) were mutated. In all cases there was no significant change in mannose phosphorylation of the cathepsin L protein indicating that these residues play little if any role in the mannose phosphorylation of cathepsin L.

Glycosylation and Mannose Phosphorylation of GFP—To convert GFP into a suitable substrate for the transferase, an endoplasmic reticulum signal sequence and a glycosylation site expressing the construct were labeled with [32P]phosphate, phosphorylation for this system (20 in 239 amino acids) provides several pairs of suitably positioned lysine residues that could serve as GlcNAc-1-phosphotransferase recognition sites. The phosphorylation of a synthetic glycoprotein that could not have evolved a complex phosphorylation signal provides evidence that the signal is a relatively simple structure that may exist on a wide range of proteins regardless of their need to be targeted to lysosomes.  

ROLE OF LYSINE RESIDUES IN RESIDUAL PHOSPHORYLATION OF CATHEPSIN L—The highly phosphorylated glycosylation sites including the wild-type site (Fig. 4), as well as some of the weakly phosphorylated sites (data not shown), displayed significant levels of phosphorylation when lysine residues that compose the previously described phosphorylation signal (Lys-54 and Lys-99) were mutated to alanine. Whereas this phosphorylation represents a minor component of the phosphorylation of the wild-type protein, it does represent a substantial portion of the phosphorylation observed for two of the engineered sites (Asn-105 and Asn-229). To determine whether or not lysine residues are responsible for the residual phosphorylation of the wild-type protein, site-directed mutagenesis was carried out on a procathepsin L construct containing the wild-type glycosylation site and alanine mutations at Lys-54 and Lys-99. The results shown in Fig. 7 indicate that several lysine residues, including Lys-157, Lys-233, and Lys-237, contribute to this residual phosphorylation. Phosphorylation of a construct containing alanine mutations at Lys-54, Lys-99, Lys-116, Lys-157, Lys-233 and Lys-237 was >90% inhibited, indicating that these lysine residues can account for virtually all of the mannose phosphorylation of procathepsin L. The residual phosphorylation appears to result from weak lysine-based phosphorylation signals that act independently of the primary Lys-54/Lys-99 signal. The level of residual phosphorylation differs among the engineered glycosylation sites providing additional evidence that this component of the phosphorylation results from independent phosphorylation signals.  

DISCUSSION

Finding a common structural motif for recognition of protein substrates by GlcNAc-1-phosphotransferase has been a long-standing goal in the field of lysosomal protein trafficking. Because the motif is three-dimensional in nature and is found on
proteins of diverse origin and structure, it has been difficult to identify. Several studies have been carried out to determine which amino acid residues are necessary for mannose phosphorylation of individual proteins. Here, we provide evidence that the inherent flexibility of N-linked oligosaccharides enables these structures to be placed in a large although limited area.

FIG. 8. Secretion and mannose phosphorylation of glyco-GFP. Glyco-GFP and GFP were expressed in COS-1 cells and labeled with [35S]methionine for 6 h in the presence of 10 mM NH₄Cl. The glyco-GFP and GFP proteins were immunoprecipitated from cell extracts and media and subjected to SDS-PAGE. The protein bands were excised and quantitated in a scintillation counter. The graph displays mannose phosphorylation as a percentage of the wild-type cathepsin L phosphorylation. The control represents the K54A/K99A construct. The graph represents the mean ± S.D. of three experiments with a total of nine replicates.

TABLE II
Effect of microenvironment surrounding critical lysines 54 and 99
Surface amino acids within 8 Å of the critical lysine residues K54 and K99 were mutated to alanine. The procathepsin L constructs were expressed in COS-1 cells and double-labeled with [32P]phosphate and [3H]leucine for 6 h in the presence of 10 mM NH₄Cl. Procathepsin L protein was immunoprecipitated from the media and subjected to SDS-PAGE. The protein bands were excised, eluted in 1M NaOH, and quantitated in a scintillation counter. The table displays mannose phosphorylation as % wild-type cathepsin L control ± S.D., and represents the average of 3 experiments with a total of 9 replicates. wt is wild type.

| CL construct | Mannose phosphorylation % |
|--------------|---------------------------|
| wt           | 100.0 ± 3.4               |
| K54/K99A     | 33.4 ± 4.6                |
| Mutations surrounding Lys-54 |                   |
| R48A         | 95.6 ± 5.0                |
| M56A         | 104.2 ± 2.1               |
| R57A         | 96.5 ± 2.9                |
| M56A/R57A/M58A | 105.5 ± 6.4          |
| Mutations surrounding Lys-99 |              |
| H97A/Q98AH100A/K101A/K102A | 99.0 ± 12.0          |
| K101A        | 94.2 ± 4.0                |
| K102A        | 98.0 ± 3.1                |
| D184A        | 99.1 ± 1.5                |
| D227A        | 107.0 ± 22.0              |

FIG. 7. Alanine scanning mutagenesis of CL-K54A/K99A construct. Lysine residues in the K54A/K99A construct were mutated using site-directed mutagenesis. The cathepsin L constructs were expressed in COS-1 cells and labeled with [32P]phosphate and [3H]leucine for 6 h in the presence of 10 mM NH₄Cl. Cathepsin L proteins were immunoprecipitated from cell extracts and subjected to SDS-PAGE. The protein bands were excised and quantitated in a scintillation counter. The graph displays mannose phosphorylation as a percentage of the wild-type cathepsin L phosphorylation. The control represents the K54A/K99A construct. The graph represents the mean ± S.D. of three experiments with a total of nine replicates.
on a lysosomal protein and still have access to the catalytic domain of the transferase. By placing oligosaccharides throughout the surface of cathepsin L, it was possible to identify a region in three-dimensional space where mannose residues are likely to be phosphorylated and the spatial relationship of this site to lysine residues Lys-54 and Lys-99 of the phosphorylation signal in the protein. Knowing the placement of the oligosaccharide in reference to the signal is necessary for identifying or constructing a signal that not only allows interaction of the transferase with its substrate protein but also provides access to its N-linked oligosaccharides and phosphorylation of one or more mannose residues.

Findings of this study concerning the relationship of critical components of the cathepsin L phosphorylation signal to the N-linked oligosaccharide of the protein are summarized in Fig. 9. The distance of ~34 Å between α-carbon atoms of critical lysine residues appears to be a general feature of the phosphorylation signal. This distance was demonstrated in our earlier study (17) for cathepsin D as well as cathepsin L (14) and is consistent with results obtained for two other proteins, DNase I (15, 29) and aspartylglucosaminidase (16). In the case of aspartylglucosaminidase, mutation of two lysine residues, Lys-183 in the α subunit and Lys-214 in the β subunit, was found to inhibit mannose phosphorylation by 96%. Because aspartylglucosaminidase is expressed as a α₂β₂ heterotetramer with a phosphorylated oligosaccharide in each subunit, these residues could be used in one or more of a variety of configurations for phosphorylation. Given this caveat, Lys-183 and Lys-214 are separated by 32.71 Å, consistent with distances of 33.75 and 33.63 Å for cathepsin L and cathepsin D, respectively (14, 17).

Bovine DNase I has two oligosaccharides and requires four lysine residues, Lys-27, Lys-50, Lys-74, and Lys-124 for efficient mannose phosphorylation. Lys-27 and Lys-74 have substantial effects on phosphorylation but do not appear to belong to the same phosphorylation signal since they affect different oligosaccharides. The two other residues, Lys-50 and Lys-124, are located 34.31 Å apart, again consistent with the 34 Å inter-lysine distance. The site of mannose phosphorylation for procathepsin L, which was identified by determining the overlap of oligosaccharide clouds of highly phosphorylated glycosylation sites, is also shown in Fig. 9. The center of this region was calculated to be 26.83 Å from the closest critical lysine, Lys-99. The relationship of the site of mannose phosphorylation to the phosphorylation signal represents an intrinsic property of GlcNAc-1-phosphotransferase and should apply to other transferase substrates.

The results of this study suggest that the cathepsin L phosphorylation signal is polarized with one lysine residue (Lys-99) proximal to the site of mannose phosphorylation and the other lysine residue (Lys-54) distal. If the binding properties of these residues were equivalent, two regions of mannose phosphorylation corresponding to two transferase-binding orientations would be expected. Mutation of residues in the vicinity of critical lysine residues did not reveal involvement of other residues that would distinguish interaction of the two lysine residues with the transferase. However, it is possible that residues other than those tested could serve such a role. It is also possible that, because of the overall shape of procathepsin L, binding of the transferase to the cathepsin is sterically restricted to a single orientation or that binding in one orientation is nonproductive.

The existence of additional minor or cryptic lysine-based phosphorylation signals on cathepsin L is consistent with data gathered on other mannose-phosphorylated proteins. The chimeric studies of cathepsin D, as the authors note, can be interpreted by having two independent phosphorylation signals, instead of having a single extended signal (9). The existence of two or more signals on cathepsin D explains why mutation of individual lysine residues affects phosphorylation of the two cathepsin D oligosaccharides differently and why localized phosphorylation of engineered glycosylation sites was not observed for this protein (12, 15).

A previously unresolved issue concerning the nature of the phosphorylation signal is its level of complexity. Is the signal a complex, highly evolved structure that interacts with the transferase over an extended surface (29) or is it a relatively simple structure composed of a few well placed residues (17)? Two findings presented in this study address this issue. First, mutation of residues surrounding critical lysine residues to alanine had little if any effect on phosphorylation of cathepsin L. This supports our previous finding that lysine residues account for most if not all of the energy of interaction between the transferase and the cathepsin L phosphorylation signal (13, 14). These findings suggest that phosphorylation signals may actually be quite simple requiring as few as two well placed lysine residues as a minimal structure (14, 17). This does not mean that additional protein factors are not important in selected cases. Arginine has been shown to partially substitute...
for lysine in some contexts (15–17). Other residues in the vicinity of critical lysine residues may affect the accessibility or properties of the lysine residues. Effects such as this may explain the apparent involvement of tyrosine residues in phosphorylation of aspartylglucosaminidase and DNase I (16, 29). In theory, any residue in contact with the transferase when it is bound to the protein could affect the rate and efficiency of phosphorylation. The issue of complexity is also addressed by the finding that glyco-GFP is susceptible to mannose phosphorylation. The issue of complexity is also addressed by the finding that glyco-GFP is susceptible to mannose phosphorylation. Given the origin, subcellular location, and structure of GFP, there is no logical explanation for how a complex mannose phosphorylation signal could have evolved on this protein.

Finally, a simple phosphorylation signal is compatible with what is known about the evolution of the mannose 6-phosphate recognition system. Utilization of the mannose 6-phosphate recognition system for lysosomal targeting was a relatively late event in evolution probably occurring sometime during early vertebrate evolution. Evolution of the system would have required a genetic mechanism for generating mannose phosphorylation signals on a structurally diverse set of lysosomal hydrolases. Generation of a complex structure on a diverse set of proteins would have required extensive remodeling of the proteins. If such alterations had occurred, one would expect to observe manifestations of these changes when comparing hydrolase sequences from species that utilize the mannose 6-phosphate recognition system and those that do not. Such manifestations have not been observed. A simple phosphorylation signal, such as the one described, would allow generation of phosphorylation signals on proteins of diverse structure through one or two point mutations and would be fully compatible with evolution of the system.

Acknowledgments—We thank Laura Liscum and Michael Forgac for critical review of the manuscript.