

Lysine-based Structure Responsible for Selective Mannose Phosphorylation of Cathepsin D and Cathepsin L Defines a Common Structural Motif for Lysosomal Enzyme Targeting*

(Received for publication, February 23, 1998, and in revised form, June 4, 1998)

John W. Cuozzo, Kai Tao, Mirek Cygler‡, John S. Mort§, and G. Gary Sahagian¶

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

‡Biotechnology Research Institute, National Research Council of Canada, Montreal, Quebec H4P 2R2, Canada, and the

§Joint Diseases Laboratory, Shriners Hospital for Children, Montreal, Quebec H3G 1A6, Canada

Previous studies have shown that lysine residues on the surface of cathepsins and other lysosomal proteins are a shared component of the recognition structure involved in mannose phosphorylation. In this study, the involvement of specific lysine residues in mannose phosphorylation of cathepsin D was explored by site-directed mutagenesis. Mutation of two lysine residues in the mature portion of the protein, Lys-203 and Lys-293, cooperated to inhibit mannose phosphorylation by 70%. Other positively charged residues could not substitute for lysine at these positions, and comparison of thermal denaturation curves for the wild type and mutant proteins indicated that the inhibition could not be explained by alterations in protein folding. Structural comparisons of the two lysine residues with those required for phosphorylation of cathepsin L, using models generated from recently acquired crystal structures, revealed several relevant similarities. On both molecules, the lysine residues were positioned approximately 34 Å apart (34.06 Å for cathepsin D and 33.80 Å for cathepsin L). When the lysine pairs were superimposed, *N*-linked glycosylation sites on the two proteins were found to be oriented so that oligosaccharides extending out from the sites could share a common region of space. Further similarities in the local environments of the critical lysines were also observed. These results provide details for a common lysosomal targeting structure based on a specific arrangement of lysine residues with respect to each other and to glycosylation sites on the surface of lysosomal proteins.

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 UDP-GlcNAc, lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase, which phosphorylates *N*-linked oligosaccharides of lysosomal proteins by the addition of phospho-GlcNAc (1–5). This modification begins after lysosomal proteins are exported from the endoplasmic reticulum and is followed by removal of terminal GlcNAc moieties and binding to mannose 6-phosphate receptors as the proteins traverse the Golgi apparatus. The receptors mediate delivery of the phos-

phorylated proteins to the endosomal compartment, and from there the proteins are transported to lysosomes (for a review, see Ref. 1).

Studies focusing on the molecular basis for mannose phosphorylation have shown that the phosphotransferase recognizes a protein-based structure or “signal” on lysosomal proteins and that this recognition is required for efficient phosphorylation *in vivo* and *in vitro* (5). Expression of this phosphorylation signal requires that the protein be in its native conformation (4–6). Analysis of the cathepsin D signal, using chimeras of cathepsin D and the nonphosphorylated homologous aspartic protease pepsinogen, defined a “minimal” structure for phosphorylation involving Lys-203 and a segment of 27 amino acids in the C-terminal region of the protein (7). Both portions are contained within a 1630-Å² surface patch on cathepsin D, but it was unclear whether the entire surface was involved in recognition or just certain residues (8). Further experiments indicated that Lys-203 and other residues that were critical in the minimal structure were not as important in the context of the native protein and that additional elements in the N-terminal region of the protein also appeared to be involved (9, 10). Thus, the specific amino acids involved directly in phosphotransferase recognition of this protein remained undefined.

Using a chemical modification approach, we have demonstrated that cathepsin L, cathepsin D, and preparations containing several other mannose-phosphorylated proteins require lysine residues for efficient phosphorylation (6, 11). It was also demonstrated that the propeptide of the cathepsin L precursor (procathepsin L), the form that interacts with phosphotransferase *in vivo*, contains critical elements of the recognition structure (11). Alanine scanning mutagenesis of all cathepsin L lysine residues indicated that two lysine residues, Lys-54 and Lys-99, both of which are contained in the propeptide, play the major role in phosphorylation. Mutation of either residue effectively inhibits phosphorylation without causing noticeable changes in the overall structure of the protein (11).

The chemical modification study also indicated that lysine residues are required for efficient phosphorylation of cathepsin D (11). This result has been explored further in this study by performing alanine-scanning mutagenesis of cathepsin D lysine residues as was done previously for cathepsin L. The results indicate that two lysine residues are also the primary determinants for phosphorylation of cathepsin D. Comparison of these residues with those of cathepsin L, using recently acquired crystal structures and molecular models for the proteins, has revealed several features of these lysine residues that are common to both proteins. These features provide the basis for a general model of the phosphotransferase recognition signal for further investigation.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: Dept. of Physiology, Tufts University School of Medicine, 136 Harrison Ave., Boston, MA 02111. Tel.: 617-636-6748; Fax: 617-636-0445; E-mail: gsahagia@opal.tufts.edu.

EXPERIMENTAL PROCEDURES

Enzymes, Antibodies, cDNAs, and Other Reagents—Endoglycosidase H and thermolysin (protease X) were purchased from Genzyme and Sigma, respectively. Cathepsin D was purified from human liver by the procedure of Gulnik *et al.* (12). Two bands, corresponding to the heavy and light chains of cathepsin D, with no other apparent impurities, were observed by SDS-PAGE¹ of the purified protein followed by Coomassie Blue staining. Antiserum to human cathepsin D was raised in rabbits by Cocalico Biologicals Inc. (Reamstown, PA) and was shown to specifically immunoprecipitate cathepsin D from biosynthetically labeled mouse 66cl4 breast tumor cells and COS cells transfected with human cathepsin D cDNA. The cathepsin D inhibitor pepstatin A and pepstatin A-agarose were purchased from Sigma. Human cathepsin D cDNA was provided by Dr. Andrej Hasilik (13), and the cDNA for glycosylated human pepsinogen (mPep1,2) was provided by Dr. Stuart Kornfeld (7).

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 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.

Subcloning of Human Cathepsin D and Glycosylated Human Pepsinogen cDNAs—Both the human cathepsin D cDNA and the glycosylated human pepsinogen cDNA were subcloned into the pED4 vector (14) for expression in COS-1 cells. The pED4 vector was modified by the insertion of a synthetic linker made from the complementary oligonucleotides GTCTAGAGAATTCGTCGACT and AATTAGTCGACGAAT-TCTCTAGACTGCA between the *Pst*I and *Eco*RI cloning sites of pED4. The linker contains *Xba*I, *Eco*RI, and *Sal*I sites, in 5'-3' order, and eliminates the 3' *Eco*RI site into which it was cloned. The modified vector will be referred to as pED4L. Human cathepsin D cDNA (13) in the pHH vector was excised with *Eco*RI and ligated into the *Eco*RI site in pED4L. The orientation of the subcloned cDNA was checked by restriction enzyme digestion. The *Xba*I and *Sal*I sites that flank the *Eco*RI site in the linker were used to transfer the cathepsin D cDNA to the pAlter vector for mutagenesis (described below) and to transfer the mutated cDNAs back into the pED4L vector for expression.

Glycosylated human pepsinogen cDNA (mPep1,2) in the sp64 vector (Promega) (7) was excised and transferred to pED4L as follows. mPep1,2 was cut out using *Eco*RI, and a partial *Pst*I digest was performed to obtain a 1450-base pair segment that included the entire coding sequence. The cDNA was then inserted into the *Pst*I and *Eco*RI sites in pED4L for expression in COS cells.

Mutation of Cathepsin D cDNA—Primers were designed to substitute alanine for lysine residues in the protein. Many of the primers also included "silent" mutations that did not alter the amino acid sequence but created new restriction sites in the cDNA so that the mutant clones could be isolated rapidly. The AAA codons of Lys-P31, Lys-47, Lys-109, and Lys-281 were changed to GCA, and AAG codons of Lys-P10, Lys-P36, Lys-8, Lys-58, Lys-63, Lys-69, Lys-120, Lys-130, Lys-158, Lys-189, Lys-192, Lys-203, Lys-223, Lys-249, Lys-267, Lys-277, Lys-284, Lys-293, Lys-299 were changed to GCG. Substitution mutations of Lys-203 to arginine and Lys-293 to arginine were made by changing the AAG codons to AGG. Lys-203 to glutamate was made by changing AAG to GAG, Lys-293 to glutamine by changing AAG to CAG, and Lys-203 to histidine by changing AAG to CAT. Mutagenesis was performed by the dual primer method, simultaneously introducing a site-directed mutation and antibiotic resistance to increase the yield of colonies containing mutant cDNAs, using the commercially available *in vitro* mutagenesis kit, Altered Sites II (Promega). The cathepsin D cDNA was excised from the pED4L vector using *Xba*I and *Sal*I and inserted into the *Xba*I and *Sal*I sites in the pAlter-1 vector (Promega) for mutagenesis. Double mutants were made by using two mutagenic primers simultaneously. Triple mutants were made by using a single mutagenic primer on a template that already contained the other two mutations of interest. Mutated cathepsin D cDNAs were excised with *Xba*I and *Sal*I and transferred to the pED4L vector for expression in COS cells. Mutagenized DNAs were checked by restriction digestion and sequencing to ensure that only the expected mutations were introduced.

Protein Expression and Quantitation of Phosphorylation—COS cells were transiently transfected with pED4-cathepsin D or pED4-pepsinogen recombinants using the DEAE-dextran method as described previously (15). The transfected cells were cultured for 60 h prior to labeling and were labeled with 0.1 mCi/ml of [³⁵S]methionine in low methionine

medium (15). Unless otherwise indicated, labeling was for 6 h. Immunoprecipitation and 12.5% SDS-PAGE of procathepsin D and procathepsin D mutants was carried out using previously published methods described for cathepsin L (16).

For quantitation of phosphorylation, transfected cells were double-labeled for 6 h with both 0.1 mCi/ml [³H]leucine and 0.2 mCi/ml [³²P]phosphate, and mutant or wild type cathepsin D proteins were immunoprecipitated. Ovalbumin (10 µg) was added to each immunoprecipitate, and the samples were subjected to SDS-PAGE as above. The gels were then stained with Coomassie Blue. Ovalbumin (45,000 Da) migrated 0.5 cm below the human cathepsin D precursor in the gel system used and was used as a marker for the position of the cathepsin D bands. Cathepsin D-containing bands were excised from the wet gel, and the labeled cathepsin D proteins were extracted from the gel slices by overnight incubation on a rocking platform in 1 ml of 1 N NaOH at room temperature. After neutralization with HCl, the amount of ³²P and ³H in each band was determined by dual label scintillation counting in Ecosint A (National Diagnostics).

Pepstatin A-Agarose Affinity Chromatography—Pepstatin A-agarose affinity chromatography was performed as described by Conner (17). Aliquots (100 µl) of conditioned media from COS cells transfected with cathepsin D or cathepsin D mutant plasmids and labeled with [³⁵S]methionine were diluted to 1.0 ml with binding buffer composed of 0.2 M sodium formate, pH 3.5, 0.4 M NaCl, 0.1% Triton X-100, with or without 1 µM pepstatin A. This solution was applied to a 2-cm column of pepstatin A-agarose in a glass Pasteur pipette that had been pre-equilibrated in binding buffer containing 1 mg/ml bovine serum albumin. The column was washed with 10 column volumes of binding buffer and then eluted with the same volume of elution buffer composed of 20 mM Tris-HCl, pH8.3, 0.4 M NaCl, 0.02% Triton X-100. Both the column washes and eluates were then immunoprecipitated using cathepsin D antisera. The immunoprecipitates were subjected to SDS-PAGE and visualized by PhosphorImager analysis (Molecular Dynamics, Inc.).

Conformational Analysis of Procathepsin D Mutants—Conformations of cathepsin D and cathepsin D mutants were probed by proteolysis with thermolysin (18, 19). Wild type and mutant cathepsin D proteins were first expressed in COS cells and labeled with [³⁵S]methionine for 6 h in serum-free media using the method described above. For thermolysin treatment, 10 µl of medium containing the protein to be analyzed was incubated in a buffer containing 20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 2 mM CaCl₂, 0.1 mM EDTA, 5% glycerol, 1.4 mM β-mercaptoethanol, and 0.1 µg of thermolysin at various temperatures for 30 min, in a total volume of 50 µl. The reactions were stopped by the addition of EDTA to a final concentration of 25 mM, and the amount of remaining labeled protein was quantitated after SDS-PAGE followed by PhosphorImager analysis. Proteolysis using chymotrypsin gave results similar to those for thermolysin digestion.

Modeling of Procathepsin D and L—Human procathepsin D and mouse procathepsin L were modeled from known crystal structures using the program MODELLER (20) as part of QUANTA, a molecular modeling software package marketed by Molecular Simulations. MODELLER uses the satisfaction of spatial restraints derived from known template structures to model the three-dimensional structure of proteins with high sequence homology to the template. MODELLER is generally capable of predicting protein structure to within a root mean square (r.m.s.) deviation of approximately 1 Å when the sequences of the target and template structure display greater than 40% identity (21). Such structures are comparable with medium resolution structures obtained by x-ray crystallography.

In order to test MODELLER's ability to predict protein structure for the two cathepsins, we used it to model two known cathepsin structures: (i) the mature portion of human cathepsin D from the 1.8-Å resolution structure of porcine pepsin (22), whose sequence is 51% identical to human cathepsin D, and (ii) the mature portion of human procathepsin L from papain (1.65-Å resolution) and actinidin (1.7-Å resolution) that are 36.9 and 39.9% identical to human cathepsin L, respectively.

A total of 10 models, each beginning with different initial conformations and fitted to the calculated structural constraints by a variable target function, were generated and evaluated by MODELLER by the calculation of an objective function for each model. The objective function is inversely related to the probability that a certain model is valid based on comparison with a number of features of the template including bond lengths, bond angles, main chain dihedral angles, van der Waals contacts, C^α-C^α distances, main chain N-O distances, side chain dihedral angles, and features of disulfide bridge bonds (20).

Comparison of the best modeled structure (*i.e.* the one with the lowest objective function) with the known crystal structure for cathep-

¹ The abbreviations used are: PAGE, polyacrylamide gel electrophoresis; r.m.s., root mean square.

sin D (23) and cathepsin L (33) gave r.m.s. values of 0.845 and 1.163 Å, respectively. These values are similar to r.m.s. values obtained by Sali and Blundell (20) for models of trypsin using known structures for elastase and tonin. From this investigation, we concluded that the MODELLER program was capable of generating reasonably accurate models for the mature portions cathepsins D and L and would probably produce similarly accurate models for the proenzyme forms of these proteins.

For construction of the procathepsin D model, the 1.8-Å resolution crystal structure of porcine pepsinogen (24, 25) was used as a template. The pepsinogen sequence is 48% identical to the procathepsin D sequence and is 64% similar, based on alignment using the method of Needleman and Wunsch (26). The final model for procathepsin D based on pepsinogen was selected on the basis of having the least structural violations with respect to the original template as determined by MODELLER, using the objective function as described above for the mature cathepsin D structure. Comparison of the model to matching residues on the crystal structure of mature cathepsin D gave a r.m.s. value of 1.11 Å, similar to the 1.38-Å value reported for a previously published model of procathepsin D (28). Initially, several structures including pepsin, pepsinogen, progastricin, and mature cathepsin D were evaluated alone and in combination as templates for the procathepsin D model. Of these, pepsinogen alone was found to give the lowest r.m.s. value. Similarly, the mouse procathepsin L model was constructed from the 2.2-Å resolution crystal structure of human procathepsin L (33). The sequence of the mouse protein is 77.1% identical to the human protein. Comparison of the model with the matching residues on the crystal structure of human procathepsin L gave a r.m.s. deviation of 0.765 Å.

For examining potential hydrogen bonds in the proteins, polar hydrogens were added with QUANTA and minimized while the rest of the structure was fixed.

RESULTS

Expression and Mutagenesis of Cathepsin D—COS-1 cells were transfected with human cathepsin D cDNA, and after biosynthetic labeling with [³⁵S]methionine for 6 h, the expressed protein was immunoprecipitated from cell lysates and media using antiserum directed against human liver cathepsin D (Fig. 1A). A large portion of the labeled newly synthesized cathepsin D precursor was secreted, presumably due to saturation of intracellular mannose 6-phosphate receptors, a condition that often occurs when lysosomal proteins are overexpressed in these cells (36). Most of the expressed protein is secreted, since labeling in the presence of 10 mM ammonium chloride, which diverts newly synthesized lysosomal proteins to the constitutive secretory pathway, did not cause a further increase in secretion (data not shown). Treatment of ³²P-labeled cathepsin D with endoglycosidase H indicated that phosphorylation of the expressed protein was restricted to its oligosaccharides (Fig. 1B). Although not shown, the expressed human protein was phosphorylated as well as the low levels of endogenous cathepsin D detected in media, indicating that the COS cell phosphotransferase was capable of efficiently phosphorylating the overexpressed protein. Glycopepsinogen, a nonlysosomal aspartic protease engineered to contain an N-linked oligosaccharide (7), was inefficiently phosphorylated (<2%), indicating that nonspecific phosphorylation of aspartic proteases does not occur to any appreciable extent in this system (Fig. 1B).

Previous protein modification experiments indicated that lysine residues are required for phosphorylation of cathepsin D (11). To determine which of the 23 lysine residues were involved, each lysine residue was converted individually to alanine by site-directed mutagenesis. The resulting mutant cDNAs were transfected into COS cells and transiently expressed. Direct quantitation of the level of phosphorylation of each mutant was performed by double labeling the cells with [³H]leucine and [³²P]PO₄ (Fig. 2).

Two lysine to alanine mutations significantly decreased phosphorylation. Mutation of Lys-203 and Lys-293 reduced

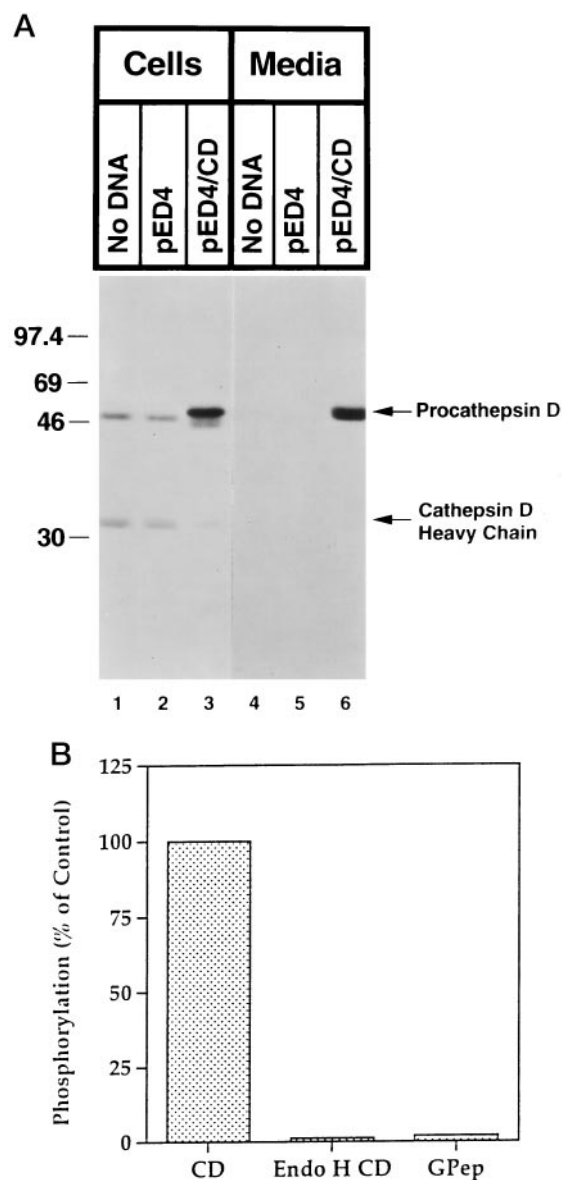


FIG. 1. Expression, secretion, and oligosaccharide phosphorylation of human cathepsin D in COS-1 cells. A, COS-1 cells were treated as follows: mock transfection with no DNA (lanes 1 and 4), transfection with mammalian expression vector, pED4 (lanes 2 and 5), or transfection with pED4 containing the human cathepsin D cDNA insert (lanes 3 and 6). Cells were labeled for 6 h with [³⁵S]methionine as described under "Experimental Procedures," and both cell lysates (lanes 1–3) and media (lanes 4–6) were subjected to immunoprecipitation using human cathepsin D antiserum. The antibody appeared to bind preferentially to the expressed human cathepsin D as seen in lane 3 by the decrease in the amount of endogenous heavy chain in the human cathepsin D expressing cell lysate immunoprecipitate. B, COS-1 cells were transfected with human cathepsin D cDNA (CD) or glycosylated human pepsinogen (GPep) and were double labeled for 6 h with [³H]leucine and [³²P]PO₄. Cathepsin D was immunoprecipitated from the cell media, incubated with or without endoglycosidase H for 17 h at 37 °C, and then subjected to SDS-PAGE. Phosphorylation was quantitated by PhosphorImager analysis. Protein levels of the same samples were assayed by SDS-PAGE followed by fluorography. For pepsinogen, the medium containing the secreted protein was subjected to SDS-PAGE, and the amount of phosphorylation relative to the amount of protein was determined by cutting the pepsinogen band from the gel and quantitating the ³²P and ³H levels by scintillation counting. The ³²P/³H ratio for pepsinogen (expressed as a percentage relative to wild type procathepsin D) is shown.

phosphorylation by 42 and 13%, respectively. As observed for the wild-type protein both mutants were efficiently expressed and secreted by the cells (data not shown).

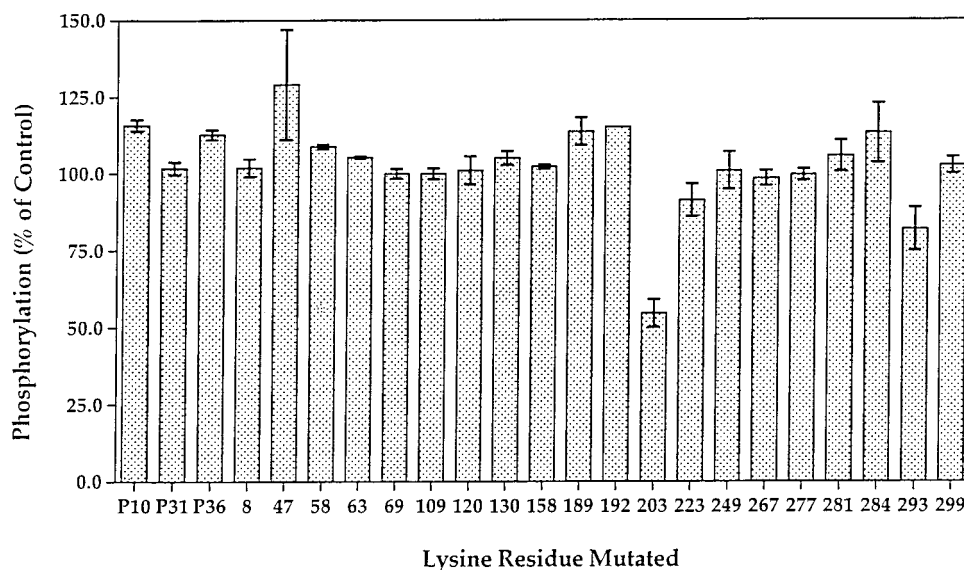


FIG. 2. **Phosphorylation of cathepsin D lysine to alanine mutants.** COS-1 cells were transfected with mutant cathepsin D cDNAs. The transfected cells were double labeled for 6 h with [^3H]leucine and [^{32}P]PO $_4$, and procathepsin D proteins secreted into the culture media were immunoprecipitated and subjected to SDS-PAGE. Bands corresponding to procathepsin D were cut from the gel, and ^{32}P and ^3H were quantitated by scintillation counting. The $^{32}\text{P}/^3\text{H}$ ratio for each mutant, expressed as a percentage relative to the wild type ratio for procathepsin D, is shown. The mutants are graphed according to the numbering of the lysine residues in the sequence of the protein, beginning with the first residue of the mature portion. Propeptide lysine residues are preceded by *P* and are numbered from the first propeptide residue. The means shown represent two separate experiments, except for the values for Lys-203 and Lys-293, which are the means of four and three experiments, respectively.

Since alanine mutations at Lys-203 and Lys-293 individually decreased phosphorylation, a Lys-203/Lys-293 double mutant (DA mutant) was constructed to test the effect of both mutations together. When the DA mutant was expressed, phosphorylation was reduced by 69% compared with wild-type cathepsin D (Fig. 3). Replacement of the lysine residues with arginine (DR mutant) did not substitute for lysines, and replacement with the homologous pepsinogen residues (Lys-203 to glutamate and Lys-293 to glutamine; DP mutant) decreased phosphorylation by 77%. The double mutant result highlights the importance of the two lysine residues by showing that they can account for a major portion of the interaction between cathepsin D and the phosphotransferase.

Various substitutions were made at position 203 to test the nature of the involvement of lysine at this position. The cationic character of lysine appears to be important for recognition, since substitution with negatively charged glutamate at position 203 causes a 15% decrease in phosphorylation, while positively charged arginine results in a 19% increase relative to the alanine mutation. Histidine had a slight inhibitory effect (Fig. 4). It is interesting to note that a similar arginine substitution at position 293 did not increase phosphorylation relative to the Lys-293 to alanine mutation. While other explanations are possible, this observation is consistent with the notion that the two lysines interact with similar but different binding sites on the phosphotransferase. Overall, while the positive charge appears to be important, there is clearly a preference for lysine at both positions.

The double lysine to alanine mutation was phosphorylated at a 31% level relative to the wild type protein. This residual phosphorylation cannot be readily explained as a nonspecific effect, since phosphorylation of a glycosylated derivative of pepsinogen, a related aspartyl protease, was barely detectable (<2% of wild type cathepsin D). The best explanation for the residual phosphorylation is that additional phosphorylation signals exist on cathepsin D. A similar conclusion was reached previously by Baranski *et al.* on the basis of their data (9). Because chemical modification of lysine residues completely eliminates phosphorylation of cathepsin D (6), we considered

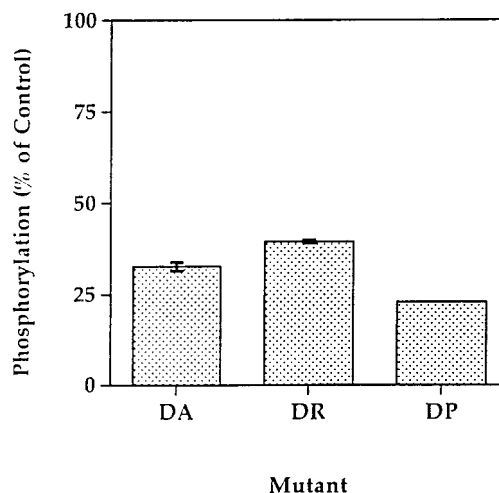


FIG. 3. **Phosphorylation of cathepsin D Lys-203 and Lys-293 double mutants.** COS-1 cells were transfected with cDNAs encoding the double mutants Lys-203 to alanine/Lys-293 to alanine (DA), Lys-203 to arginine/Lys-293 to arginine (DR), and Lys-203 to glutamate/Lys-293 to glutamine (DP). The cells were double labeled for 6 h with [^3H]leucine and [^{32}P]PO $_4$, and the level of phosphorylation of each mutant procathepsin D protein was quantitated as described in the legend to Fig. 2. The means shown represent four separate experiments for DA, two for DR, and one for DP.

the possibility that some lysine residues may function cryptically and become more prominent when the primary signal was ablated. To examine this possibility, triple lysine to alanine mutants were constructed from the double lysine mutant (DA) and examined for phosphorylation. Further decreases of 10–12% were observed for residues Lys-120, Lys-130, and Lys-223 relative to the double mutant (Fig. 5).

Because mannose phosphorylation is sensitive to the three-dimensional structure of the protein, it is important to evaluate the effects of mutations on protein folding (6, 11). Consistent with normal folding, all mutants that showed decreased phosphorylation in this study were efficiently secreted from COS

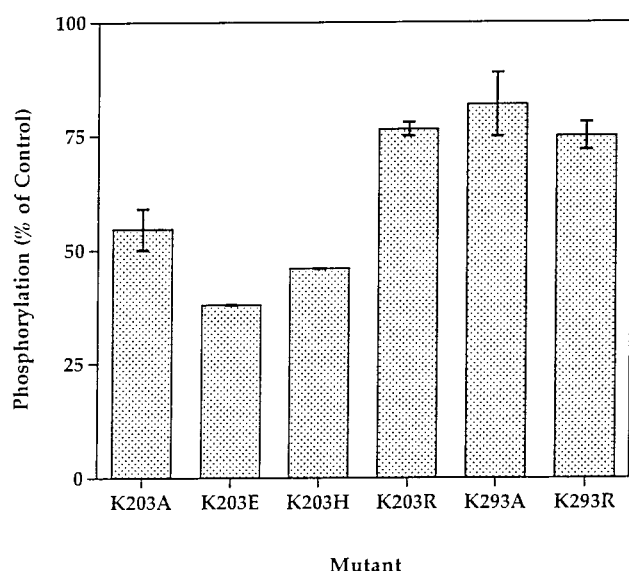


FIG. 4. **Phosphorylation of substitution mutants at positions 203 and 293.** COS-1 cells were transfected with cDNAs encoding the following substitution mutants, Lys-203 to alanine, Lys-203 to glutamate, Lys-203 to histidine, Lys-203 to arginine, Lys-293 to alanine, and Lys-293 to arginine. The cells were double labeled for 6 h with [^3H]leucine and [^{32}P]PO $_4$, and the level of phosphorylation of each mutant procathepsin D protein was quantitated as described in the legend to Fig. 2. The means shown represent two separate experiments, except for Lys-203 to alanine, which is the mean of four.

cells. All double mutants that displayed reduced phosphorylation (DA, DR, and DP) were capable of binding to a pepstatin A-agarose affinity matrix in a pepstatin and pH-dependent manner similar to wild-type cathepsin D (data not shown). These results indicate that the mutations did not cause gross alterations in protein conformation or folding.

Thermal denaturation curves for the double mutants were examined to determine whether or not the mutations had significantly affected the conformation of the protein locally. Mutational analyses of protein folding and stability have demonstrated that conformation-altering mutations produce changes in the energy required to unfold a protein and that these changes can be determined by analysis of the thermal denaturation curve of the protein (30–32). Significant destabilizing mutations will almost invariably cause shifts of several degrees that can be readily detected by a number of methods including the thermolysin digestion method used here (18, 19). As seen in Fig. 6, thermal denaturation curves for lysine double mutants with reduced phosphorylation are very similar to the curve for wild-type cathepsin D, suggesting that the mutations have had little if any effect on the conformation of the protein.

A lack of significant effect of the Lys-203 and Lys-293 mutations on folding of cathepsin D is supported by the crystallographic data (23), which indicate that both residues have high crystallographic thermal factors (B values) corresponding to relatively high mobility. The value for Lys-203 is 31.31 Å 2 , and the value for Lys-293 is 45.16 Å 2 . Exhaustive studies of stability of lysozyme mutants at nearly every amino acid (31) have shown that only mutations in residues with low mobility, which have B values well below the average for the protein (26.5 Å 2), have significant effects on the structure of the protein.

Examination of Critical Lysine Residues on Cathepsin D and Cathepsin L—An analysis of the three-dimensional structures of cathepsin D and cathepsin L was performed to determine if critical lysine residues in the two proteins share features that could account for their involvement in the phosphorylation process. Recently, crystal structures of the mature form of human cathepsin D at 2.5-Å resolution (23) and human pro-

thepsin L at 2.2-Å resolution (33) have been solved. Cathepsin D is a member of the aspartyl protease family of enzymes (23), whereas cathepsin L belongs to the cysteine protease family (33). The two families are structurally distinct, and no known structural homology exists between them. Because the phosphorylation data that had been obtained were for the proenzyme forms of human cathepsin D and mouse cathepsin L, it was necessary to construct models for these two proteins from available crystal structures. The model for human procathepsin D was constructed by homology modeling from the structure of porcine pepsinogen, and the model for mouse procathepsin L was constructed from the structure of the human protein. Construction of the models is described under "Experimental Procedures."

Comparison of the modeled and crystal structures for cathepsin L and cathepsin D (Figs. 7 and 8) showed that both are globular proteins of similar size. Human cathepsin D possesses two phosphorylated N -linked oligosaccharides (34) at Asn-70 and Asn-199 (using human cathepsin D numbering as in Ref. 35), while cathepsin L has only one at Asn-221 (36). The propeptides of both proteins maintain the proteases in inactive states by blocking access to their active sites.

Lysine residues Lys-203 and Lys-293 in the procathepsin D structure are both exposed on the surface of the proenzyme, as they are in the mature form of the protein. No evidence of hydrogen bonding between the side chains of these residues and neighboring residues were observed. As mentioned above, the side chains of both residues are fairly mobile as seen from the high B values. In procathepsin L, Lys-54 is fully exposed to solvent, while Lys-99 is partially exposed. The B values for the side chains of these residues are also relatively high, with a value for Lys-54 of 25.42 and for Lys-99 of 37.71. Although Lys-99 is only partially exposed to solvent, its mobility appears to be at least as high as that of Lys-54. With the exception of a possible hydrogen bond between the ϵ -amino group of Lys-99 and the backbone oxygen of Met-253, no direct interactions between these residues and neighboring residues were observed.

The most striking observation made was that distances between critical lysine residues in the two structures are nearly identical. In the procathepsin D model, a C $^\alpha$ –C $^\alpha$ distance of 34.06 Å between Lys-203 and Lys-293 was found. This is closely matched by the distance of 34.47 Å in the mature cathepsin D crystal structure. For procathepsin L, the C $^\alpha$ –C $^\alpha$ distance between Lys-54 and Lys-99 is 33.80 Å for both the mouse procathepsin L model and the human procathepsin L crystal structure. The difference of approximately 0.5 Å between the C $^\alpha$ –C $^\alpha$ distances for the two proteins is within the error ± 0.8 Å expected for a distance measurement between two atoms in crystal structures with resolutions on the order of 2.2–2.5 Å (37). Because of the relatively high mobility of the ϵ -amino groups compared with the main chain α -carbons, as seen by the B values, it is difficult to determine an accurate distance between the amino groups of the lysine side chains.

Similarities were also noted in the position of N -linked glycosylation sites relative to critical lysine residues. A general observation for both molecules is that one of the critical lysines is much closer to the glycosylated asparagines than the other, by nearly half the distance. In procathepsin D, the distance from the C $^\alpha$ of Lys-203 to the C $^\alpha$ of glycosylated Asn-70 is 28.97 Å, and the distance from the C $^\alpha$ of Lys-293 to the C $^\alpha$ of Asn-70 is 45.81 Å. In procathepsin L, the corresponding distances between the C $^\alpha$ of Lys-54 and the C $^\alpha$ of Lys-99 to the glycosylated Asn-221 are very similar, 25.08 and 49.48 Å, respectively. The other glycosylated asparagine residue on cathepsin D, Asn-199, is somewhat different, with a distance of 9.71 Å to the

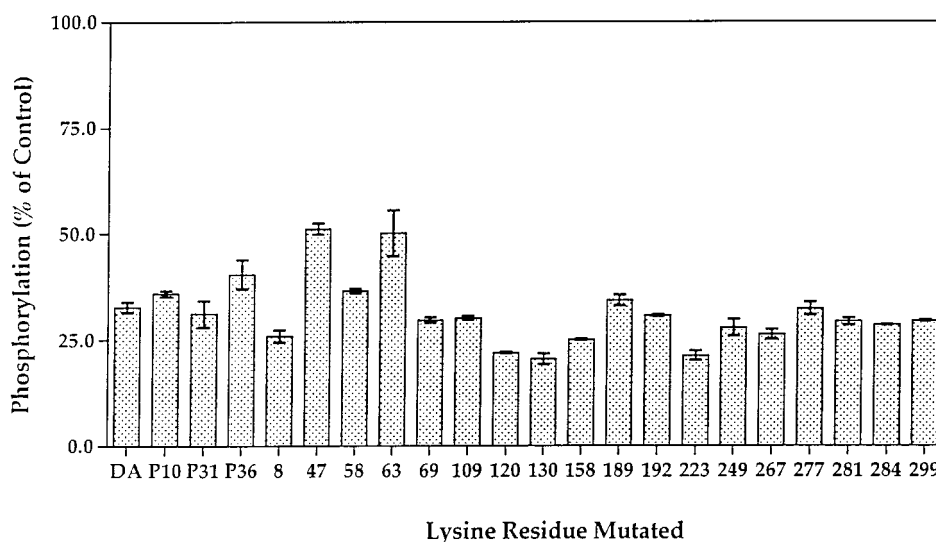


FIG. 5. **Phosphorylation of cathepsin D triple mutants.** COS-1 cells were transfected with cDNAs encoding triple mutants using the Lys-203 to alanine/Lys-293 to alanine double mutant as a template. For each mutant, a third lysine residue was mutated to alanine. The cells were double labeled for 6 h with [^3H]leucine and [^{32}P]PO $_4$, and the level of phosphorylation of each mutant procathepsin D protein was quantitated as described in Fig. 2. The mutants are graphed as in Fig. 2, with the numbering based on the position of the third mutation. The means shown represent two separate experiments, except for DA, which is the mean of four.

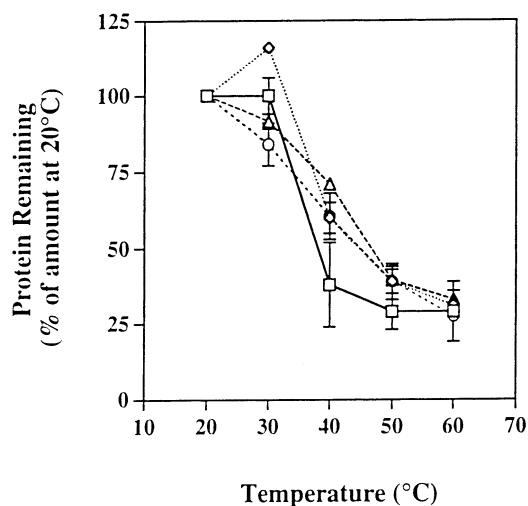


FIG. 6. **Thermolysin proteolysis of cathepsin D mutants.** Medium from cells transfected with wild-type cathepsin D (CD, □) and the mutants DA (◇), DR (○), and DP (△) was incubated for 30 min at various temperatures in the presence of thermolysin as described under "Experimental Procedures." The amount of remaining protein was quantitated by SDS-PAGE followed by PhosphorImager analysis and is normalized to the amount remaining after digestion at 20 °C for each mutant. The points and ranges indicated represent the results of two experiments.

C^α of Lys-203 and 33.38 Å to the C^α of Lys-293.

Each of the two critical lysine residues on each protein appears to reside within a particular electrostatic microenvironment on the surface of the protein. Both Lys-203 of procathepsin D and Lys-99 of procathepsin L are positioned within 5 Å of two arginine residues on the surface of the protein, Arg-141 and Arg-202 of procathepsin D and Arg-98 and Arg-101 of procathepsin L. There are no nearby acidic residues, so the overall charge in this area of the molecule should be highly positive at neutral pH. The other critical lysine residue on procathepsin D and procathepsin L, Lys-293 and Lys-54, respectively, are positioned in more acidic environments. In procathepsin D, two acidic residues, Asp-172 and Glu-288, are both within 5–6 Å of Lys-293, and in procathepsin L, one acidic residue, Glu-53, is positioned next to Lys-54. Electrostatic en-

vironments around critical lysine residues of cathepsins D and L are shown schematically in Fig. 8.

Conservation of Critical Lysine Residues in Cysteine and Aspartyl Proteases—A summary of information concerning the conservation of critical lysine residues and potential *N*-linked glycosylation sites in cysteine and aspartyl proteases is given in Table I. With the exception of bovine cathepsin D, critical lysine residues and glycosylation sites are conserved in all mammalian cathepsin D and cathepsin L proteins that have been sequenced. The lysine residues are not conserved in cathepsin proteins of more distant species that do not utilize the mannose 6-phosphate recognition system for lysosomal targeting, nor are they conserved in closely related proteins that have evolved glycosylation sites at alternate locations on the proteins. The lack of conservation in more distant aspartyl and cysteine proteases suggests that lysine residues at these positions are not required for general functioning of these enzymes but are instead related to some feature or function shared by the mammalian protein. The pattern of occurrence of critical lysines in the two cathepsins is therefore consistent with the proposed role of these residues in lysosomal targeting.

The pattern of occurrence of critical lysine residues also indicates that these residues arose relatively late in the evolution of cysteine and aspartyl proteases, around the time when mammals and other vertebrates diverged. The appearance of a phosphorylation signal at that time would be consistent with the time of emergence of the mannose 6-phosphate recognition system, which on the basis of the species distribution of the cation-independent mannose 6-phosphate receptor, would have occurred during early vertebrate evolution.

Bovine cathepsin D is an interesting exception in that of the nine mammalian cathepsin D and L proteins that have been sequenced, it is the only one for which critical lysine residues have not been conserved. This protein also possesses an additional *N*-linked glycosylation site not found in other cathepsin D proteins. An intriguing possibility is that the phosphorylation signal on this protein may have moved or shifted to accommodate the additional oligosaccharide.

DISCUSSION

Any description of the phosphotransferase recognition structure must explain both the specificity and inclusiveness of the

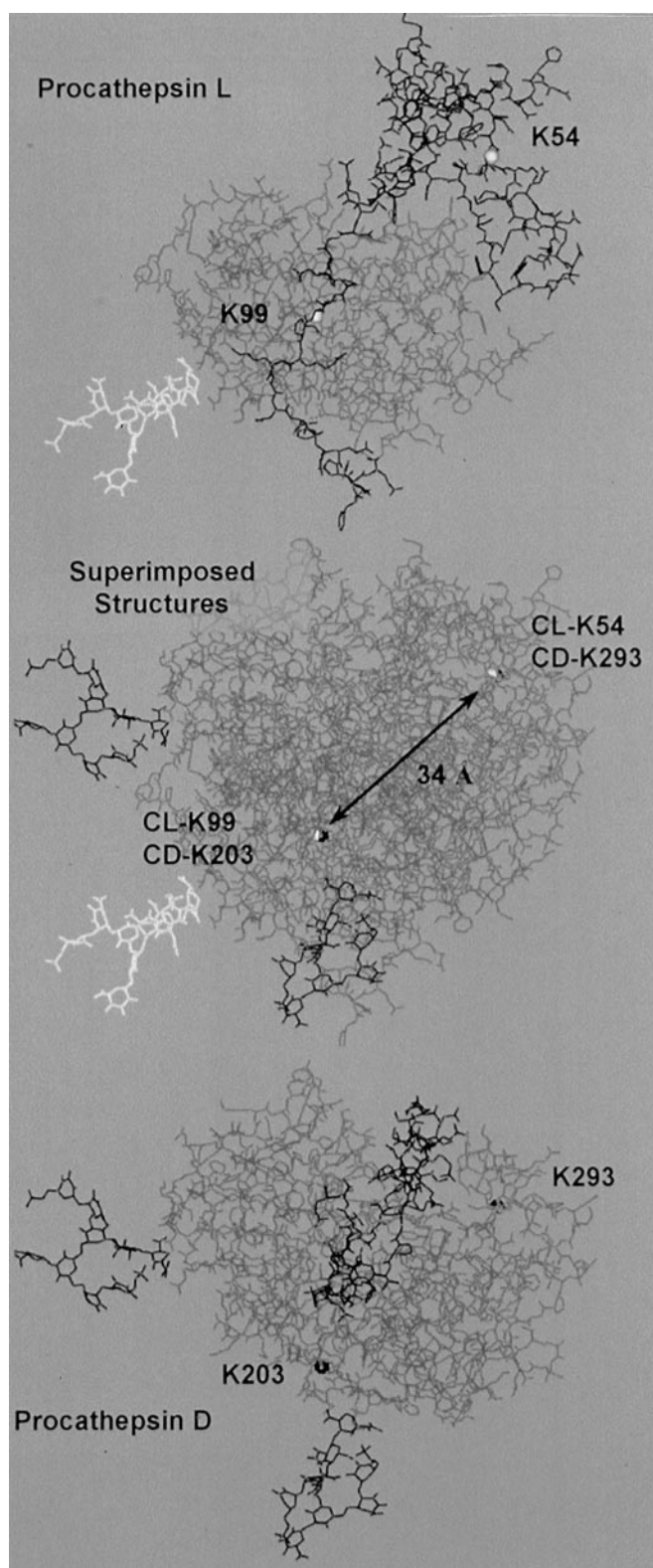


FIG. 7. Comparison of model procathepsin L and procathepsin D structures. The modeled three-dimensional structures of mouse procathepsin L (*top*) and human procathepsin D (*bottom*) are oriented to show the relative positions of the propeptides (*black*), lysine residues identified as critical for efficient phosphorylation (*white* or *black sphere*, for cathepsin L or cathepsin D, respectively), and N-linked oligosaccharides (*white* for cathepsin L, *black* for cathepsin D). The oligosaccharides shown are modeled diphosphorylated $\text{Man}_7\text{GlcNAc}_2$ structures. The *middle drawing* is a superimposition of the two cathepsin structures with the α -carbon atoms of the critical lysine residues aligned. Note that all three oligosaccharides are located toward one end of the aligned residues.

phosphorylation process. The phosphotransferase efficiently phosphorylates mannose residues only on lysosomal proteins, yet it is able to recognize a large number of proteins from a diverse range of enzyme families. A simple model for a recognition signal that could fulfill both requirements is one that consists of a few residues arranged in a particular configuration with respect to each other and to the oligosaccharide to be modified. In this model, the phosphotransferase would need only to recognize a few key residues, thus explaining its inclusiveness, and its specificity would be maintained by the unique positioning of those residues. Such a model was proposed on the basis of previous results with cathepsin L that indicated that two lysine residues in its propeptide could account fully for its interaction with the phosphotransferase (11). The results of this study support that model by demonstrating that two cathepsin D lysine residues with characteristics similar to those of cathepsin L can also account for a major portion of the phosphorylation. Comparison of the structures of the two proteins, which was made possible by the recent solution of a crystal structure for human procathepsin L (33), has allowed identification of common features of the phosphorylation signal for further analysis.

The results presented in this study demonstrate that two lysine residues in the mature portion of cathepsin D, Lys-203 and Lys-293, are primary determinants of mannose phosphorylation. These lysine residues appear to be directly involved in phosphotransferase recognition for the following reasons. (i) Lysine residues are involved in phosphorylation of all mannose-phosphorylated proteins that have been examined to date. This includes cathepsin D, cathepsin L, and phosphotransferase substrates contained in two crude lysosomal protein preparations (6). (ii) Examination of the three-dimensional structure of procathepsin D indicates that the side chains of Lys-203 and Lys-293 are accessible on the surface of the protein, are mobile, and do not appear to be involved in higher order structures. Such residues would be available for direct interaction with the phosphotransferase. (iii) The requirement for lysines at both locations is specific and cannot be effectively met by other positively charged residues. (iv) The inhibitory effect of mutations at Lys-203 and Lys-293 cannot be explained by indirect effects on protein folding. Together with similar results obtained for cathepsin L (6, 11), these findings strongly suggest that critical lysine residues of the two cathepsins interact directly with the phosphotransferase.

Are other lysine residues on cathepsin D involved in phosphotransferase recognition? Biochemical modification of lysine residues abolishes phosphorylation of cathepsin D *in vitro* (11), suggesting that lysine residues in addition to Lys-203 and Lys-293 may be involved. The results from the triple mutants assayed in this study show that when both of the critical lysine residues are mutated to alanine, mutations of other lysine residues display weak but detectable inhibitory effects. These residues might act in conjunction with Lys-203 or Lys-293, or they could form weak independent phosphorylation signals.

Substitution of Lys-203 and Lys-293 with other positively charged amino acid residues indicated that arginine can partially substitute for lysine at position 203 but not at position 293. While these results suggest that arginine may be capable of serving as a phosphorylation determinant in some contexts, previous results utilizing protein modification to assess the involvement of lysine residues have indicated that most mannose-phosphorylated proteins use lysine residues predominantly, and if arginine is used, its role is likely to be a minor one (6). It is also worth noting that placement of a negatively charged glutamate residue at position 203 of cathepsin D rendered the protein less susceptible to phosphorylation than

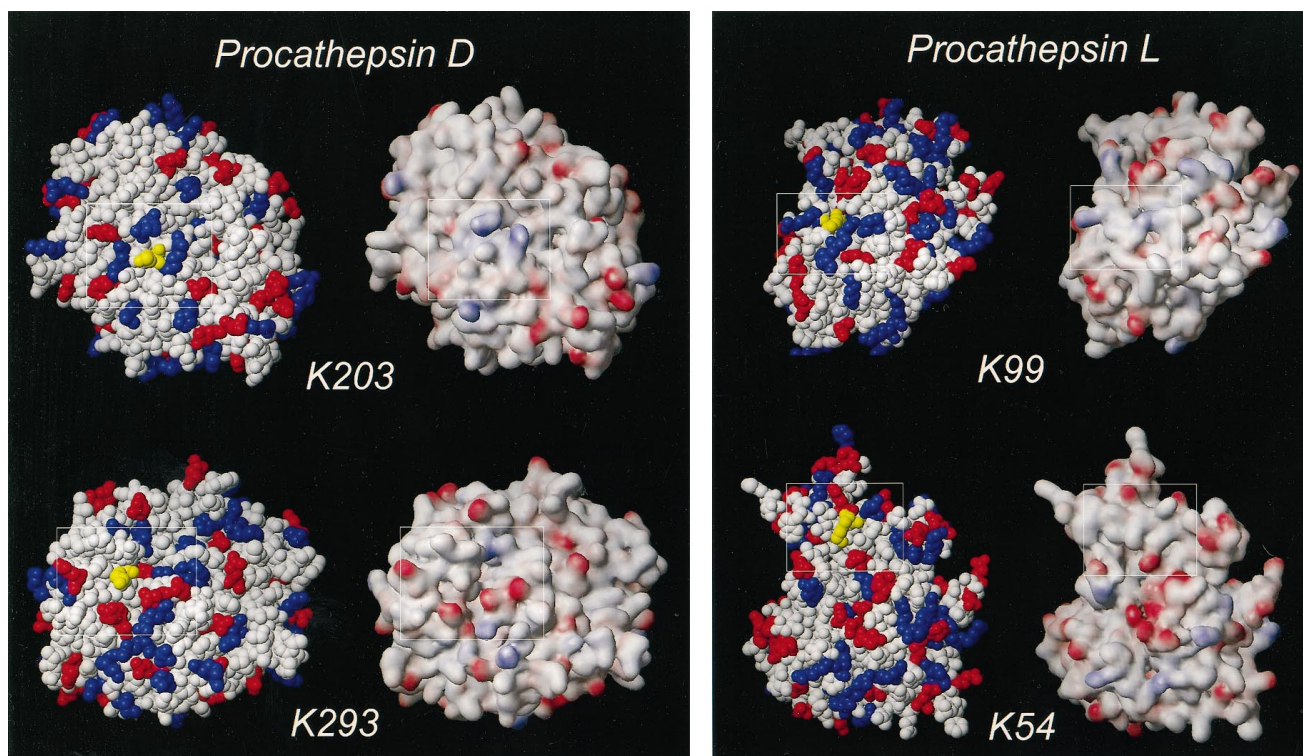


FIG. 8. **Electrostatic microenvironments around critical lysine residues of cathepsins D and L.** Shown here are space-filling representations and electrostatic potential surfaces for modeled procathepsin D (*left*) and procathepsin L (*right*). Critical lysine residues are shown in yellow, basic residues (Lys, Arg, and His) in blue, and acidic residues (Asp and Glu) in red. Corresponding microenvironments in space filling representations and potential surfaces are indicated by a box for each critical lysine residue. WebLab software (Molecular Simulations, Inc.) was used to construct the electrostatic potential surfaces. The software calculates Gasteiger charges for surface atoms and maps the electrostatic potential representing those charges to the surface.

TABLE I
Conservation of critical lysine residues

Cysteine protease	Lys-54/Lys-99 conserved?	Glycosylation sites
Cathepsin L		
Mouse	Yes	Asn-221
Rat	Yes	Asn-221
Cow	Yes	Asn-221
Human	Yes	Asn-221
Pig	Yes	Asn-221
Zebrafish	No	Asn-99
Shrimp	No	Asn-302
Silkworm	No	Asn-95
Fruit fly	No	Asn-95
Cathepsin S (human)	No	Asn-103
Cathepsin K (mouse)	No	Asn-104, Asn-213
Aspartyl protease	Lys-203/Lys-293 conserved?	Glycosylation sites
Cathepsin D		
Mouse	Yes	Asn-70, Asn-199
Rat	Yes	Asn-70, Asn-199
Human	Yes	Asn-70, Asn-199
Pig	Yes	Asn-70, Asn-199
Cow	No	Asn-70, Asn-97, Asn-199
Chicken	Yes	Asn-70, Asn-199, Asn-337
Rainbow trout	No	Asn-70
Mosquito	No	Asn-70
Cathepsin E (human)	No	Asn-27
Renin (human)	No	Asn-70, Asn-264

when alanine was placed at this location, presumably because the acidic residue interacted negatively with complementary binding site residues on the phosphotransferase.

Previous studies concerning the recognition signal on cathep-

sin D by Baranski *et al.* (7) had identified Lys-203 as a critical residue by substituting various parts of cathepsin D into the nonphosphorylated secretory protein, pepsinogen (7). Subsequent work with these chimeric proteins showed only a slight decrease in phosphorylation when Lys-293 was mutated to alanine and a more significant decrease when another lysine residue, Lys-267, was mutated to alanine (8). When the Lys-203 to alanine mutation was introduced into the complete cathepsin D protein, the protein was phosphorylated nearly as well as wild type (9) and did not show the 40–50% decrease observed in this study. Several possible explanations exist for these differences. First, the specificity of phosphorylation for the *Xenopus* oocyte expression system used in those experiments may differ significantly from the COS-1 cell system used in this study. Work with the *Xenopus* system itself has indicated species-specific differences, since it has been demonstrated that renin, which is not phosphorylated well in mammalian cells, is well phosphorylated in *Xenopus* oocytes (39). Another difficulty with the *Xenopus* oocyte system, which applies mainly to the results of the chimera studies, is that the oocytes are maintained at a lower temperature than mammalian cells and as a result may allow proteins with local structural abnormalities to traverse the secretory pathway. Such proteins may be folded well enough to be secreted but not for efficient phosphorylation. When some of the cathepsin D-pepsinogen chimeras used in the phosphorylation studies were expressed in human lymphoblasts, they exhibited temperature-sensitive phenotypes, indicating that they may contain local structural abnormalities (40).

Finally, most previous studies on cathepsin D have utilized binding to cation-independent mannose 6-phosphate receptor as a means of quantitating phosphorylation. While receptor binding provides a sensitive and relatively simple method for

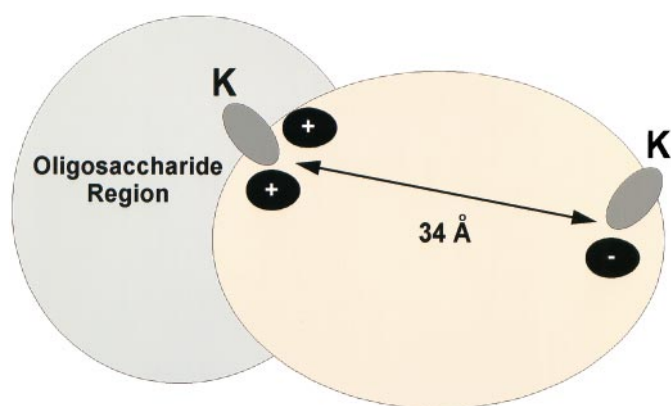


FIG. 9. **Model for the phosphotransferase recognition signal on lysosomal enzymes.** A general model is shown for the phosphotransferase recognition signal on lysosomal enzymes based on the structural data acquired for procathepsin D and procathepsin L. Two lysine residues, indicated in *blue*, are spaced 34 Å apart and are exposed on the surface of the protein. The oligosaccharides are oriented at a *right angle* to the *line* drawn between the critical lysines. The basic or acidic environment around each lysine residue is indicated by the *circled positive and negative symbols*.

assessing phosphorylation, the results are not directly related to phosphorylation. For example, at high levels of phosphorylation, the two cathepsin D oligosaccharides could maximally contain four phosphomonoester moieties. However, maximal binding to the receptor may only require two (41, 42). At lower levels of phosphorylation, a receptor binding assay would not pick up molecules with fewer than two phosphomonoesters, since two phosphomonoesters are required for efficient binding to the receptor. As a result of these effects, an assay based on receptor binding may underestimate phosphorylation at low phosphorylation levels and overestimate phosphorylation at high phosphorylation levels. Such assays are further hampered by their inability to detect phosphodiester moieties, the initial reaction product of the phosphotransferase. These problems would also apply to methods utilizing lysosomal delivery or cellular retention to estimate phosphorylation levels. The use of these methods may, in particular, explain the apparent lack of involvement of Lys-203 and Lys-293 observed in previous studies (8, 9, 50).

Examination of known and modeled cathepsin structures revealed several similarities for further testing. The most striking result was that distances between C $^{\alpha}$ atoms of critical lysines on both molecules were found to be almost exactly the same, ~ 34 Å. This is a fairly large distance relative to the total size of procathepsin D and L, which is in the range of ~ 60 – 65 Å in length along the longest axis. If one assumes that the phosphotransferase interacts primarily with the ϵ -amino groups of critical lysine residues and that lysine side chains are fully flexible, then a functional signal could be produced by lysine residues with a C $^{\alpha}$ –C $^{\alpha}$ distance that differs by as much as 3–4 Å/residue. Examination of the critical lysine residues recently reported for bovine DNase I indicated a C $^{\alpha}$ –C $^{\alpha}$ distance of 34.31 Å, very close to the distances observed here for cathepsins D and L.²

Another aspect of critical lysine residues suggested by examination of the cathepsin structures is that individual lysines

may be situated in specific microenvironments. As discussed in the previous section, both Lys-203 on cathepsin D and Lys-99 on procathepsin L are located in highly basic microenvironments on the surface of their respective proteins. Both of these residues are positioned closer to the glycosylated asparagines than the other critical lysines. The other lysine residues, Lys-293 on cathepsin D and Lys-54 on cathepsin L, appear to be situated in a more acidic environment. The finding that arginine residues have different effects when substituted for the two lysine residues on cathepsin D also suggested that there are some differences between the two lysine recognition sites. These results are consistent with a model in which the phosphotransferase binds the two critical lysine residues using different binding sites for the two interactions. If this is true, the interaction between the phosphotransferase and a phosphorylation signal on a protein would fix the position of the active site of the enzyme so that mannose residues would have to “find” the active site to be modified.

While lysine residues provide most of the energy for interaction of the phosphotransferase with its substrates, it is likely that minor interactions between the surface of the transferase and its substrates also play a role. Weak interactions are likely to occur wherever the two proteins come in contact, and they could act in a positive or negative manner to alter the overall interaction between the phosphotransferase and its substrate. In contrast to the critical lysine interactions, such minor interactions, if they exist, would be expected to differ from one protein to the next.

A unique aspect of the phosphorylation process is the attenuated selectivity with which mannose residues are modified once a protein has been targeted for phosphorylation. Thus, proteins that possess a phosphorylation signal are able to undergo phosphorylation at several different positions on one or more oligosaccharides (51, 52). While it has been suggested that the phosphotransferase may be mobile during the reaction process, this effect is most likely related to the flexibility of *N*-linked oligosaccharides.

A number of studies have focused on the three-dimensional structure of *N*-linked oligosaccharides (for a review, see Ref. 43). Free and protein-bound oligosaccharides display considerable flexibility as a result of rotation about glycosidic bonds and other bonds involved in linkages between residues. Although hydrogen bonds and steric interactions between residues can reduce chain mobility and produce preferred conformations, most oligosaccharides remain flexible (44, 45). Protein-bound oligosaccharides display reduced mobility at the site of attachment to the protein, and flexibility can be reduced further by steric and/or chemical interactions between the oligosaccharide and the protein (46, 47). However, in most cases, decreased mobility is restricted to core residues in the base of the oligosaccharide, and branch residues remain flexible. Preliminary reconstruction of *N*-linked oligosaccharides on modeled cathepsin D and cathepsin L structures has indicated that oligosaccharide flexibility may account completely for the pattern of phosphorylation observed for these proteins.³ Further work is needed to verify this hypothesis.

The model for the phosphorylation signal based on the results of this study consists of two lysine residues, exposed on the surface of the protein, that are spaced apart by 34 Å and positioned in a specific orientation relative to the target oligosaccharide (Fig. 9). The importance of lysine microenvironments for expression of critical lysine function has not yet been determined. However, as discussed above, the structures suggest that these environments may be a factor and that one

² After completion of this study, Nishikawa *et al.* (53) provided evidence that mannose phosphorylation of DNase I is mediated by a simple lysine-based structure similar to that described previously for cathepsin L (6, 11) and here for cathepsin D. Examination of the DNase I crystal structure (54) reveals an inter-lysine distance of 34.3 Å, a distance very similar to inter-lysine distances described in this study for cathepsin D and L.

³ J. W. Cuozzo and G. Gary Sahagian, unpublished observation.

environment may need to be more basic than the other.

One of the main questions concerning lysine residues as primary binding sites for the phosphotransferase is how the enzyme makes contact with both lysine residues, at 34 Å apart, and the oligosaccharides. For both cathepsins, there is a substantial amount of surface area between the contact points. In cathepsin D, the surface between the critical lysines is convex, while the surface between the critical lysines on cathepsin L is relatively flat. Further work must be done to address this question, but some initial direction can be found in a recent study reporting the purification and subunit analysis of the bovine phosphotransferase (48). The bovine phosphotransferase is a 540-kDa complex consisting of six subunits, two disulfide-linked homodimers of 166- and 51-kDa subunits, and two noncovalently attached 56-kDa subunits. The phosphotransferase, at over 10 times the molecular mass of the individual cathepsins, appears to be easily capable of covering the area necessary for the recognition and mannose phosphorylation of these molecules. Another study has demonstrated that the 166-kDa subunit contains the active site (49). It will be of interest to determine which subunit or subunits are involved in the recognition of lysosomal enzymes.

Another question that remains is why mutation of the critical lysines on procathepsin L each reduces phosphorylation by 80% or more (11), while the procathepsin D lysines have smaller effects that are additive. We suspect that this is a manifestation of the existence of additional phosphorylation signals in the cathepsin D protein. For example, one could envision weak signals being produced by the presence of lysine residues or even arginine residues appropriately positioned elsewhere in the protein. Additional signals such as these could serve to improve phosphorylation of suboptimally placed oligosaccharides by providing alternate modes for phosphotransferase binding.

Further work is needed to explore and test features of the recognition structure identified in this study. It will be necessary to determine if other residues are involved in recognition and, if so, what relationships exist between these residues, the oligosaccharides, and the lysine residues already identified. The final test of these features will be to reconstruct the signal *de novo* on a nonphosphorylated protein. By this approach, the contribution of the structural elements defined in this study to the phosphorylation signal, including the specific distance of approximately 34 Å between two lysine residues, the orientation of the lysine residues with respect to the oligosaccharides, and the specific microenvironments of the individual lysines, can be individually evaluated.

Acknowledgments—We thank Dr. Andrej Hasilik for providing the cDNA for human cathepsin D, Dr. Stuart Kornfeld for providing the cDNA for glycosylated human pepsinogen, and Drs. John Mort and Mirek Cygler for providing coordinates for the crystal structure of human procathepsin L prior to public release. We also thank Dr. David Ricklin (Department of Pathology, New England Medical Center) for providing human liver for these studies.

REFERENCES

- Kornfeld, S., and Mellman, I. (1989) *Annu. Rev. Cell Biol.* **5**, 483–525
- Hasilik, A., Waheed, A., and von Figura, K. (1981) *Biochem. Biophys. Res. Commun.* **98**, 761–767
- Varki, A., and Kornfeld, S. (1981) *J. Biol. Chem.* **255**, 8398–8401
- Reitman, M. L., and Kornfeld, S. (1981) *J. Biol. Chem.* **256**, 11977–11980
- Lang, L., Reitman, M., Tang, J., Roberts, R. M., and Kornfeld, S. (1984) *J. Biol. Chem.* **259**, 14663–14671
- Cuozzo, J. W., and Sahagian, G. G. (1994) *J. Biol. Chem.* **269**, 14490–14496
- Baranski, T. J., Faust, P. L., and Kornfeld, S. (1990) *Cell* **63**, 281–291
- Baranski, T. J., Koelsch, G., Hartsuck, J. A., and Kornfeld, S. (1991) *J. Biol. Chem.* **266**, 23365–23372
- Baranski, T. J., Cantor, A. B., and Kornfeld, S. (1992) *J. Biol. Chem.* **267**, 23342–23348
- Dustin, M. L., Baranski, T. J., Sampath, D., and Kornfeld, S. (1995) *J. Biol. Chem.* **270**, 170–179
- Tao, K., Stearns, N. A., Wu, Q.-L., Young, W., and Sahagian, G. G. (1995) *J. Biol. Chem.* **270**, 15611–15619
- Gulnik, S., Baldwin, E. T., Tarasova, N., and Erickson, J. (1992) *J. Mol. Biol.* **227**, 265–270
- Horst, M., and Hasilik, A. (1991) *Biochem. J.* **273**, 355–361
- Kaufman, R. J., Davies, M. V., Wasley, L. C., and Michnick, D. (1991) *Nucleic Acids Res.* **19**, 4485–4490
- Kaufman, R. (1990) *Methods Enzymol.* **185**, 487–511
- Tao, K., Stearns, N. A., Dong, J., Wu, Q.-L., and Sahagian, G. G. (1994) *Arch. Biochem. Biophys.* **311**, 19–27
- Conner, G. E. (1989) *Biochem. J.* **263**, 601–604
- Hecht, M. H., Nelson, H. C. M., and Sauer, R. T. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 2676–2680
- Hecht, M. H., Sturtevant, J. M., and Sauer, R. T. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81**, 5685–5689
- Sali, A., and Blundell, T. L. (1993) *J. Mol. Biol.* **234**, 779–815
- Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995) *Proteins* **23**, 318–326
- Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S., and James, M. N. (1990) *J. Mol. Biol.* **214**, 143–170
- Baldwin, E. T., Bhat, T. N., Gulnik, S., Hosur, M. V., Sowder, R. C., Cachau, R. E., Collins, J., Silva, A. M., and Erickson, J. W. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6796–6800
- James, M. N. G., and Sielecki, A. R. (1986) *Nature* **319**, 33–38
- Sielecki, A. R., Fujinaga, M., Read, R. J., and James, M. N. (1991) *J. Mol. Biol.* **219**, 671–692
- Needleman, S. B., and Wunsch, C. D. (1970) *J. Mol. Biol.* **48**, 443–453
- Cooper, J. B., Khan, G., Taylor, G., Tickle, I. J., and Blundell, T. L. (1990) *J. Mol. Biol.* **214**, 199–222
- Koelsch, G., Metcalf, P., Vetricka, V., and Fusek, M. (1995) in *Aspartic Proteinases: Structure, Function, Biology and Biomedical Implications* (Takahashi, K., ed) pp. 273–278, Plenum Press, New York
- Deleted in proof
- Matthews, B. W. (1996) *FASEB J.* **10**, 35–41
- Matthews, B. W. (1993) *Annu. Rev. Biochem.* **62**, 139–160
- Alber, T., Dao-pin, S., Wilson, K., Wozniak, J. A., Cook, S. P., and Matthews, B. W. (1987) *Nature* **330**, 41–46
- Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J. S., and Cygler, M. (1996) *EMBO J.* **15**, 5492–5503
- Hasilik, A., and Neufeld, E. F. (1980) *J. Biol. Chem.* **255**, 4946–4950
- Faust, P. L., Kornfeld, S., and Chirgwin, J. M. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 4910–4914
- Stearns, N. A., Dong, J. M., Pan, J.-X., Brenner, D. A., and Sahagian, G. G. (1990) *Arch. Biochem. Biophys.* **283**, 447–457
- Fersht, A. (1977) *Enzyme Structure and Mechanism*, pp. 1–46, 2nd Ed., W. H. Freeman and Co., New York
- Deleted in proof
- Faust, P. L., Chirgwin, J. M., and Kornfeld, S. (1987) *J. Cell Biol.* **105**, 1947–1955
- Glickman, J. N., and Kornfeld, S. (1993) *J. Cell Biol.* **123**, 99–108
- Tong, P. Y., Gregory, W., and Kornfeld, S. (1989) *J. Biol. Chem.* **264**, 7962–7969
- Dong, J., and Sahagian, G. G. (1990) *J. Biol. Chem.* **265**, 4210–4217
- Woods, R. J. (1995) *Curr. Opin. Struct. Biol.* **5**, 591–598
- Cumming, D. A., Shah, R. N., Krepinsky, J. J., Grey, A. A., and Carver, J. P. (1987) *Biochemistry* **26**, 6655–6663
- Cumming, D. A., and Carver, J. P. (1987) *Biochemistry* **26**, 6664–6676
- Wyss, D. F., Choi, J. S., Li, J., Knoppers, M. H., Willis, K. J., Arulanandam, A. R., Smolyar, A., Reinherz, E. L., and Wagner, G. (1995) *Science* **269**, 1273–1278
- Wyss, D. F., Choi, J. S., and Wagner, G. (1995) *Biochemistry* **34**, 1622–1634
- Bao, M., Booth, J. L., Elmendorf, B. J., and Canfield, W. M. (1996) *J. Biol. Chem.* **271**, 31437–31445
- Bao, M., Elmendorf, B. J., Booth, J. L., Drake, R. R., and Canfield, W. M. (1996) *J. Biol. Chem.* **271**, 31446–31451
- Schorey, J. S., Fortenberry, S. C., Chirgwin, J. M. (1995) *J. Cell Sci.* **108**, 2007–2015
- Varki, A., and Kornfeld, S. (1983) *J. Biol. Chem.* **258**, 2808–2818
- Dittmer, F., Pohlmann, R., and von Figura, K. (1997) *J. Biol. Chem.* **272**, 852–858
- Nishikawa, A., Gregory, W., Frenz, J., Cacia, J., and Kornfeld, S. (1997) *J. Biol. Chem.* **272**, 19408–19412
- Lahm, A., and Suck, D. (1991) *J. Mol. Biol.* **222**, 645–667

**Lysine-based Structure Responsible for Selective Mannose Phosphorylation of
Cathepsin D and Cathepsin L Defines a Common Structural Motif for Lysosomal
Enzyme Targeting**

John W. Cuozzo, Kai Tao, Mirek Cygler, John S. Mort and G. Gary Sahagian

J. Biol. Chem. 1998, 273:21067-21076.
doi: 10.1074/jbc.273.33.21067

Access the most updated version of this article at <http://www.jbc.org/content/273/33/21067>

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 50 references, 26 of which can be accessed free at
<http://www.jbc.org/content/273/33/21067.full.html#ref-list-1>