Prime region subsite specificity characterization of human cathepsin D: The dominant role of position 128

BRIAN M. BEYER¹ AND BEN M. DUNN

Department of Biochemistry and Molecular Biology, College of Medicine, University of Florida,
Gainesville, Florida 32610-0245

(RECEIVED May 29, 1997; ACCEPTED September 24, 1997)

Abstract

In order to contribute to our understanding of cathepsin D (CatD) active site specificity, two series of chromogenic octapeptides with systematic substitutions in positions P_2' and P_3' were synthesized. This panel was characterized with native human liver cathepsin D (nHuCatD) and yielded information concerning specificity trends within the S_2' and S_3' subsites. The pepstatin inhibited crystal structure of nHuCatD (Baldwin et al., 1993) was then utilized in conjunction with these subsite preference data to identify residues suspected of contributing to "prime" side subsite specificity. These residues were targeted for site-directed mutagenesis using the re-engineered recombinant model, "short" pseudocathepsin D (Beyer & Dunn, 1996). As a result of these analyses it was determined that prime region subsites do contribute to the unique specificity of human CatD. Furthermore, it was ascertained that the poly-proline loop does not have an active role in S_3' subsite specificity. Lastly, it appears that Ile128 has a dominant role on S_2' subsite specificity whereas Val130 does not.

Keywords: aspartic proteinase; cathepsin D; chromogenic substrate; prime region subsites; "short" pseudocathepsin D; substrate specificity

The abundant lysosomal endopeptidase cathepsin D (CatD), has been shown to be involved in protein turnover (Barrett, 1977) and has been postulated to play a role in prohormone and antigen processing (Pillai & Zull, 1986; Diment et al., 1988). CatD activity has been correlated with tumor invasiveness (Aaltonen et al., 1995; Rochefort, 1990) and the release of β -amyloid peptides from the amyloid precursor in Alzheimer's disease (Siman et al., 1993; Lador et al., 1994). With increasing interest in this enzyme, the work presented here focused on defining the subtle differences in ligand-enzyme interactions that exist between human CatD and other human aspartic proteinases.

Previous reports from this laboratory (Scarborough et al., 1993; Scarborough & Dunn, 1994) describing the substrate specificity of nHuCatD focused on S_2 and S_3 subsite binding requirements. It was determined from these analyses that the conditions in the S_3 subsite were favorable for hydrophobic interactions. Furthermore, in examining the S_2 subsite, the overall preferences observed were clear: a hydrophobic or Glu residue was optimal in the P_2 position, whereas a positively charged side-chain in position P_2 yielded a substrate that was strongly rejected.

Previous reports on the specificity of the prime region of CatD (van der Drift & van Noort, 1989) implicated that additional contributions to specificity may lie in the S_2' subsite and reported a strong preference for a positively charged P_2' residue in protein substrates. With this in mind, and the general lack of knowledge of prime region specificity of CatD, the approach of subsite characterization using synthetic peptides with systematic replacements was extended into the prime region subsites. These new series were based on the excellent CatD substrate Lys-Pro-Ile-Glu-Phe*Nph-Arg-Leu (* = point of cleavage, Nph = *p*-nitro-phenylalanine), with systematic replacement of the P_2' Arg or the P_3' Leu with a variety of amino acid substitutions.

The pepstatin inhibited crystal structure of nHuCatD (Baldwin et al., 1993) was then utilized in conjunction with the resulting preference data to identify residues suspected of contributing to active site specificity. The resulting residues were then targeted for site-directed mutagenesis utilizing the re-engineered recombinant model "short" pseudocathepsin D. This recombinant system was chosen because the resulting protein, "short" pseudocathepsin D, exhibits kinetic properties that are virtually identical to nHuCatD and is therefore ideal for the analysis of structure/function relationships (Beyer & Dunn, 1996). Presented here are results detailing the active role of the prime region subsites S_2' and S_3' in nHuCatD active site specificity. Furthermore, we report a detailed investigation analyzing the contributions of various prime region subsite residues to subsite specificity.

Reprint requests to: Ben M. Dunn, Department of Biochemistry and Molecular Biology, College of Medicine, University of Florida, Gainesville, Florida 32610-0245; e-mail: bdunn@college.med.ufl.edu

¹Current address: Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, New Jersey 07033-0539.

Results

P₂' substrate series

Peptide derivatives with substitutions in the P₂' position were analyzed and determined to be ≥95% pure by HPLC and MALDI-TOF. The solubility of this series varied considerably and was troublesome due to the replacement of the positively charged P₂' Arg residue. Hydrophobic residues in this position decreased solubility in distilled water by approximately ten-fold (determined by amino acid analysis). In order to analyze those less soluble substrates, derivatives with Leu, Ile, Ser, Ala or Val in the P₂' position were initially dissolved in 100 μL of 50% formic acid. Aliquots of the substrates were then taken and analyzed by MALDI-TOF to ensure that the integrity of the peptide had not been compromised. The dissolved substrates were then brought up to a final volume of 1 mL with distilled water and submitted for amino acid analysis. Following quantitation, initial rates and K_m values of several substrates that had been dissolved using both the 50% formic acid as well as 100% distilled water method were compared to ensure that the presence of formic acid did not affect the kinetic properties of nHuCatD. The resulting parameters proved identical and confirmed that the use of formic acid did not affect the results of the kinetic assay. N-terminal sequencing indicated that peptide hydrolysis occurred at the anticipated point of cleavage (Phe*Nph).

S₂' subsite preferences

In analyzing the kinetic properties of this series with nHuCatD (Table 1), it was determined that the newly designed peptides yielded a variety of excellent substrates. Upon closer analysis, a variety of trends were identified. The data suggested that the S₂' subsite has a preference for positive charged P₂' side-chains. An Arg or Lys in the P₂' position provided the best substrates within the series with k_{cat}/K_m values of 1.8 and 2.0 μM⁻¹ s⁻¹, respectively. Furthermore, this confirmed the previously observed kinetic data from van der Drift and van Noort that positively charged P₂' side-chains are preferred in substrates of CatD.

In looking at the kinetic trends of the peptides containing hydrophobic side-chains at position P₂' (Leu, Ile, and Val), an increase in the hydrophobicity of the side-chain led to an increase in the efficiency of catalysis. k_{cat}/K_m values for these substrates were respectable with values ranging from 0.7 to 1.5. The accommoda-

tion of long or bulky hydrophobic groups suggests that the S₂' subsite is of significant size. Also, smaller side-chains (Ser and Ala) within this series provided good substrates with k_{cat}/K_m of approximately 1.0. On the other hand, the presence of a Glu in position P₂' yielded a poor substrate for nHuCatD whereas an Asp in this position created a substrate that was not cleaved under standard assay conditions. The enzyme concentration was increased ten-fold in additional efforts to study the P₂' Asp peptide, in attempts to generate initial rates to permit K_m determinations. However, it was not possible to obtain reliable kinetic parameters for this substrate; this may reflect the hydrophobic nature of the S₂' subsite.

P₃' substrate series

Peptide derivatives with substitutions in the P₃' position were analyzed and determined to be ≥95% pure by HPLC and MALDI-TOF. Fortunately, this series was readily soluble in distilled water and did not present the unique problems associated with solubility found in the P₂' series.

S₃' subsite preferences

The P₃' peptide series provided several excellent substrates for nHuCatD (Table 2) with hydrolysis occurring exclusively between Phe*Nph. Differences in specificity constants (k_{cat}/K_m values) were not as notable as those observed for the P₂' series; however, some distinct trends were evident upon closer inspection. There was a preference for peptides with large hydrophobic side-chains in the P₃' position. The four best substrates in this series contained a Leu, Val, Ile and Nle with k_{cat}/K_m values ranging from 1.4 to 1.9. Following this, the aromatics exhibited k_{cat}/K_m values of approximately 1.0. It also appears that this subsite will accommodate smaller nonpolar residues in that an Ala at P₃' provided a good substrate as well. These data point to the potential hydrophobic nature of this subsite.

On the other hand, in moving away from the large hydrophobics and aromatics, charged side-chains at position P₃' provided poorer substrates with rates of catalysis that were lower than the rest of the series. An Asp or Arg at this position yielded k_{cat}/K_m values of 0.5 and 0.3, respectively. Although these values were respectable, they were four to six-fold lower than those observed for the better substrates. Interestingly, a serine substitution within this series

Table 1. Kinetic parameters of nHuCatD for substrate series with systematic substitutions in the P₂' position

P ₅	P ₄	P ₃	P ₂	P ₁	P ₁ '	P ₂ '	P ₃ '	k_{cat} (s ⁻¹)	K_m (μM)	k_{cat}/K_m (μM ⁻¹ s ⁻¹)
K	P	I	E	F	X ^a	R	L	50 ± 4	28 ± 2	1.8 ± 0.2
K	P	I	E	F	X	K	L	76 ± 8	38 ± 6	2.0 ± 0.2
K	P	I	E	F	X	L	L	18 ± 2	12 ± 2	1.5 ± 0.1
K	P	I	E	F	X	I	L	42 ± 5	35 ± 4	1.2 ± 0.1
K	P	I	E	F	X	S	L	21 ± 2	20 ± 1	1.0 ± 0.1
K	P	I	E	F	X	A	L	36 ± 4	39 ± 5	0.9 ± 0.1
K	P	I	E	F	X	V	L	34 ± 3	51 ± 9	0.7 ± 0.1
K	P	I	E	F	X	E	L	17 ± 2	73 ± 15	0.2 ± 0.1
K	P	I	E	F	X	D	L	—	—	<0.01

^aX = *p*-nitrophenylalanine.

Table 2. Kinetic parameters of nHuCatD for substrate series with systematic substitutions in the P₃' position

P ₅	P ₄	P ₃	P ₂	P ₁	P ₁ '	P ₂ '	P ₃ '	k_{cat} (s ⁻¹)	K_m (μ M)	k_{cat}/K_m (μ M ⁻¹ s ⁻¹)
K	P	I	E	F	X ^a	R	L	54 ± 4	27 ± 4	1.9 ± 0.3
K	P	I	E	F	X	R	V	58 ± 5	36 ± 3	1.6 ± 0.2
K	P	I	E	F	X	R	I	115 ± 12	75 ± 9	1.5 ± 0.3
K	P	I	E	F	X	R	Nle	130 ± 13	91 ± 9	1.4 ± 0.3
K	P	I	E	F	X	R	F	100 ± 10	103 ± 12	1.0 ± 0.2
K	P	I	E	F	X	R	Y	110 ± 11	133 ± 17	0.8 ± 0.1
K	P	I	E	F	X	R	A	21 ± 3	31 ± 4	0.7 ± 0.1
K	P	I	E	F	X	R	D	30 ± 5	59 ± 7	0.5 ± 0.1
K	P	I	E	F	X	R	R	10 ± 1	40 ± 6	0.3 ± 0.1
K	P	I	E	F	X	R	S	—	—	<0.01

^aX = *p*-nitrophenylalanine.

provided the only peptide that was not cleaved. With an efficient rate observed for an Ala in this position but not a Ser, the integrity of the peptide was rechecked by mass spectrometry, HPLC and amino acid analysis. Upon confirmation, enzyme concentrations were increased ten-fold; however, even this condition did not yield reliable kinetic data for this substrate.

Identification of potential interactions

The newly designed substrate series provided data concerning the specificity requirements of the S₂' and S₃' subsites of nHuCatD. From these data, it was clear that there are substantial effects on the efficiency of cleavage arising from interactions in the S₂' and S₃' regions of the active site. In order to interpret the subsite preferences obtained, amino acid sequences surrounding potential points of contact in the "prime" region of the active site were identified for nHuCatD and a variety of other aspartic proteinases. This information (Table 3) was utilized to identify regions of variability that may affect subsite specificity. One noticeable region contains the residues in a loop around position 128. This provides a "back wall" for the S₂' subsite and may therefore explain the binding preferences of this subsite.

As an experiment in modification of specificity and to evaluate our hypothesis of the origin of specificity, residues Ile 128 and Val 130 were targeted for site-directed mutagenesis. An Arg was substituted at each of these positions in an effort to improve the binding and cleavage of the P₂' Asp substituted peptide substrate. Residue 128 was targeted first for change, as it appeared to be the closer of the two to the P₂' side-chain and was pointed into the

aqueous environment of the cleft. This meant that the Ile128Arg mutant should have been able to refold correctly. Position 130 was also targeted for replacement with an Arg residue and was characterized along with the Ile128 mutant.

Another area of variability suspected of dictating specificity is the region from 289–300. This sequence has been designated the poly-proline loop due to the presence of a unique -Pro-Pro-Pro-sequence between positions 292 and 294. This structural component is only shared with one other aspartic proteinase, renin, and is thought to contribute to the specificity of the S₃' subsite. In the comparison shown in Table 3 it was observed that the corresponding sequence in pepsin was -Pro-Thr-Ser-. In order to investigate the influence of this variability on specificity, we introduced the corresponding pepsin sequence into "short" rHuCatD. This sequence was chosen because of the wide range of specificity exhibited by pepsin in the S₃' subsite (Dunn et al., 1995). This change also provided an opportunity to analyze the effect this region has on the refolding of rHuCatD. The importance of proline *cis/trans* isomerization is well documented (Houry et al., 1995; Jaenicke, 1995), and therefore it was thought that the succession of prolines in this region might contribute to the low recovery rate of approximately 1% in "short" rHuCatD. In comparison, the refolding rate of pepsin is significantly better, approximately 10%. Consequently, the yield of protein recovered from expression and refolding protocols was carefully monitored in addition to specificity characterization.

Processing of S₂' and S₃' subsite mutants

Following solubilization and refolding of the mutant proteins, the emergence of hydrolytic activity was monitored with the peptide Lys-Pro-Ile-Glu-Phe-Nph-Arg-Leu. This sequence was originally utilized for the introduction of the new cleavage junction of "short" rHuCatD (Beyer & Dunn, 1996) that effectively removes 12 of the otherwise eighteen retained pro-segment residues present in WT pseudocathepsin D. Therefore, mutant activity against this peptide would suggest that the cleavage junction, essential for processing to the shortened pseudoform, was also recognized as a favorable substrate. As a result of this analysis, cleavage of the peptide was observed with rates, however, that were decreased in comparison with wild-type recombinant enzyme. Activity against this sequence was encouraging, however, N-terminal sequencing was utilized to

Table 3. Amino acid sequence surrounding potential points of contact in the "prime" region of the active site

	35	73	128	188	289
Porcine pepsin	SSN	ITY	ISA	GYW	MDVPTSSG.DLWI
nHuCatD	SSN	IHY	ISV	AYW	NDIPPPSG.PLWI
HuCatE	SSN	IQY	LAV	AYW	LDIHPPAG.PLWI
Rhizopus pepsin	SSD	ISY	IITTV	GWV	NFDFA.....I

confirm processing at the engineered cleavage junction for all S_2' and S_3' mutants. Therefore, the active site mutations did not interfere with processing at the cleavage junction and yielded mutants with N-terminal sequences identical to WT "short" rHuCatD.

Activities of S_2' mutants

Prior to purification, an aliquot of the refolding solution of Ile128Arg was assayed with the P_2' Asp peptide. The amount of activity observed for this mutant is shown in Figure 1 in comparison with wild-type "short" rHuCatD at identical protein concentrations. The observed decrease in absorbance is indicative of hydrolytic activity toward this peptide. Following this observation, the Ile128Arg mutant was purified using affinity chromatography and initial rates were calculated for this peptide in order to calculate a K_m value. Because the mutation at position 128 clearly altered the specificity of this subsite, the entire panel of P_2' peptides described in Table 1 was analyzed once again. Upon analyzing Figure 2, it was evident that in mutating position 128, we created an enzyme that exhibited decreased catalytic efficiency (specificity constant, k_{cat}/K_m) for all substrates with the exception of the P_2' Asp or Glu peptides. There was a five to seven fold decrease in the specificity constant values observed for positively charged residues and k_{cat}/K_m parameters for hydrophobic side-chains were approximately one half of those exhibited by nHuCatD (Table 1). In turning to the negatively charged P_2' peptides, an increase in k_{cat}/K_m for both the P_2' Asp and P_2' Glu containing substrates was noted. Only a slight enhancement of catalytic efficiency was observed for the P_2' Glu peptide, but a dramatic increase was observed in the rate of catalysis of the P_2' Asp peptide. Whereas this substrate was not cleaved at measurable rates by nHuCatD, we were able to obtain a k_{cat}/K_m value of $0.2 \pm 0.1 \mu\text{M}^{-1} \text{s}^{-1}$ ($k_{cat} = 10 \pm 1 \text{s}^{-1}$ and $K_m = 54 \pm 5 \mu\text{M}$) for the Ile128Arg mutant.

The Val130Arg "short" rHuCatD mutant exhibited altered kinetic properties in comparison with nHuCatD. However, the dif-

Fig. 1. Comparison of activities for WT "short" pseudocathepsin D (Δ) and Ile128Arg "short" pseudocathepsin D ($\bullet$) with the P_2' Asp containing peptide Lys-Pro-Ile-Glu-Phe*Nph-Asp-Leu (* = point of cleavage, Nph = *p*-nitrophenylalanine). Activity is indicated by a decrease in absorbance and was monitored at identical protein concentrations for both recombinant enzymes.

Fig. 2. Graphical representation depicting the differences in kinetic parameters (k_{cat}/K_m) for the S_2' subsite mutants (Ile128Arg and Val130Arg) and nHuCatD with the S_2' substrate series.

ferences observed here were not as dramatic as those of the Ile128Arg mutant. The parameters of this mutant with the P_2' peptide series are also listed in Table 2. Once again, there was a significant reduction, of approximately 50%, in the cleavage rates of peptides containing positively charged side-chains and the catalytic efficiencies for peptides containing hydrophobic residues in this position were decreased (15–80%) as well. In contrast with the kinetic properties exhibited by Ile128Arg, the peptide containing an Asp at position P_2' was not cleaved and provided the most notable difference between the two mutants.

Activity and refolding of S_3' mutant

Amounts of correctly folded proteolytically active material were quantitated by affinity chromatography and protein concentration determinations. The results indicated that upon loading 17.20 mg onto the affinity column following attempted refolding, 240 μg of correctly folded proteolytically active mutant rHuCatD was recovered. This resulted in a refolding efficiency of 1.4% and was a modest improvement over the efficiency observed with WT "short" rHuCatD of 1.1%. These results indicated that mutagenesis within the poly-proline loop alone did not dramatically increase the refolding efficiency of rHuCatD. Kinetic parameters of the P_3' substrate series were determined with this mutant (Table 4) and indicated that the poly-proline loop does not play a role in the specificity of the S_3' subsite. A clear preference for hydrophobic residues was observed with the mutant parameters being slightly lower than those observed for nHuCatD. Furthermore, a clear dislike for charged residues was observed and is consistent with those values observed for native material.

Discussion

The first series of peptides synthesized to explore prime region specificity requirements of a number of aspartic proteinases contained a Lys in the P_2 position (Dunn et al., 1995). This series proved ineffective for the investigation of nHuCatD prime region specificity because of the inability of nHuCatD to accommodate a positively charged side-chain in position P_2 (Scarborough et al., 1993). Consequently, two new series were designed based on the excellent nHuCatD substrate Lys-Pro-Ile-Glu-Phe*Arg-Leu.

Fig. 3. Molecular model depicting the potential interactions between an Asp or Arg at position P_2' of a substrate and the side-chains lining the S_2' subsite. The inhibitor pepstatin is displayed in capped sticks and the flap (residues 72–81) is in purple. The S_2' subsite is lined with several amino acids of hydrophobic nature (Ile128, Val130, and Ile73). Note that the side chain of the Asp in position P_2' comes into close proximity with these residues while the Arg turns up and therefore interacts with solvent at the top of this view.

In analyzing the data from both series of peptides, we have confirmed prior observations (van der Drift & van Noort, 1989) that there are substantial effects upon the efficiency of cleavage arising from interactions in the S_2' and S_3' regions of nHuCatD. In analyzing the S_2' subsite of nHuCatD, three residues were identified (Ile128–Ser129–Val130) that provide a “back wall” for the S_2' subsite. The preference of this enzyme for hydrophobic residues in the P_2' position might be explained by the hydrophobicity of this loop. Furthermore, the reduction of at least 100-fold in catalytic efficiency for nHuCatD with a peptide containing an Asp residue in the P_2' position of a substrate might be explained by the largely hydrophobic character as well. The

negative charge likely to be present on the Asp at the pH of the assay (3.7), could potentially be buried within a pocket composed mostly of hydrophobic residues (Fig. 2), and would therefore create an unfavorable condition (Fig. 3). An Arg or Lys residue, on the other hand, is long enough to turn “up” and allow the positively charged group to reach the surface of the enzyme so as to interact with the aqueous surroundings.

In attempting to define the dominating features of the S_3' subsite, the studies of other related aspartic proteinases were taken into consideration (Rao et al., 1993). From these analyses, it has been shown that specificity pockets are also influenced by the elabora-

Table 4. Comparison of kinetic parameters for nHuCatD and “Loop” mutant for substrates with systematic substitutions in position P_3'

P_5	P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'	nHuCatD k_{cat}/K_m ($\mu M^{-1} s^{-1}$)	“Loop” mutant k_{cat}/K_m ($\mu M^{-1} s^{-1}$)
K	P	I	E	F	X ^a	R	L	1.9 ± 0.3	1.3 ± 0.2
K	P	I	E	F	X	R	V	1.6 ± 0.2	1.0 ± 0.1
K	P	I	E	F	X	R	I	1.5 ± 0.2	1.2 ± 0.1
K	P	I	E	F	X	R	Nle	1.4 ± 0.2	0.6 ± 0.1
K	P	I	E	F	X	R	F	1.0 ± 0.1	0.7 ± 0.1
K	P	I	E	F	X	R	Y	0.8 ± 0.1	0.6 ± 0.1
K	P	I	E	F	X	R	A	0.7 ± 0.1	1.0 ± 0.1
K	P	I	E	F	X	R	D	0.5 ± 0.1	0.2 ± 0.1
K	P	I	E	F	X	R	R	0.3 ± 0.1	0.3 ± 0.1
K	P	I	E	F	X	R	X	<0.01	<0.01

^aX = *p*-nitrophenylalanine.

tion of loops on the periphery of the binding cleft in renins and is most marked at S_3' and S_4' . For instance, in endotheiapepsin, which has been the subject of the greatest number of studies (Dealwis et al., 1994; Bailey et al., 1993), different conformations of the inhibitor side-chain at position P_3' are adopted and the residue at P_4' is generally disordered. In contrast, these side-chain residues are clearly defined in inhibitor complexed structures of mouse renin (Dealwis et al., 1994) and is mainly a consequence of a unique structural feature known as the poly-proline loop. This feature is shared by nHuCatD (Baldwin et al., 1993) and serves to form a secondary flap (the first being formed by residues 72–81) within the C-terminal domain. Thus, in nHuCatD, the cleft is covered by flaps from both domains rather than from the N-terminal domain alone as in other pepsin-like aspartic proteinases.

This sequence of successive proline residues was therefore thought to play a role in the observed specificity of this subsite. The non-polar nature of this region might explain why hydrophobic or aromatic side-chains were preferred in the P_3' position. Furthermore, charged residues coming into close contact with this loop may, therefore, prove unfavorable, thus contributing to the reductions seen in catalytic efficiencies of nHuCatD for an Asp or Arg in P_3' .

With the mutagenesis candidates in mind discussed above, site-directed mutagenesis using PCR was utilized using the expression vector encoding the gene for the recombinant protein "short" pseudocathepsin D. However, the danger in modifying subsite specificity is that the introduced cleavage junction might, as a consequence, not be cleaved. This would prevent conversion of the expressed proenzyme (prosegment is required for folding) into the active, self-processed species. Although specificity was successfully modified in each of the mutants (Fig. 2 and Table 4), N-terminal sequencing confirmed that processing occurred exclusively at the engineered cleavage junction of "short" pseudocathepsin D and therefore validated the use of this re-engineered recombinant system for structure/function studies. The choice of the engineered cleavage junction was based on a very good oligopeptide substrate. Thus, mutagenesis of the type reported here, while causing alterations in specificity, does not affect specificity so dramatically as to eliminate the ability of self-activation.

The mutagenesis of residue 128 from an Ile to an Arg clearly identified the dominant role this position plays in active site specificity. The introduction of a positive charge at this point modified subsite preferences in a predictable fashion. Cleavage rates of the P_2' positively charged residues decreased significantly and can be attributed to an unfavorable electrostatic interaction between the positively charged substrate side-chain and Arg128. The fact that this peptide is accommodated to some degree points to the inherent flexibility of the side chains and the size of the subsite. Secondly, the decrease in catalytic efficiency for hydrophobic residues in position P_2' can also be attributed to the presence of the positively charged Arg at position 128. The hydrophobic residue would be buried within a pocket containing a positively charged residue and would therefore be unfavorable. Lastly, the cleavage of the P_2' Asp peptide can be attributed to a favorable electrostatic interaction between the Arg at position 128 and the negatively charged side-chain of the peptide.

In turning to position 130, we observed that this residue only plays a moderate role in S_2' subsite specificity. Mutagenesis of this position decreased the catalytic efficiency of both positively charged as well as hydrophobic substrate residues in this position. However, cleavage of the P_2' Asp peptide was not observed even at

concentrations of enzyme up to 10-fold of those sufficient for cleavage of all other substrates.

Based on the kinetic evidence presented, the poly-proline loop (Baldwin et al., 1993) does not appear to play a role in the specificity of the S_3' subsite. Modifying the "short" rHuCatD sequence of -Pro-Pro-Pro- to -Pro-Thr-Ser- did not appear to have an effect, in that hydrophobic P_3' side-chains were still preferred and charged residues produced the lowest catalytic efficiency values within the series. This observation confirms speculation, based on the crystal structure (Aguilar et al., 1995), that the proline-rich loop of nHuCatD occupies a position further away from bound inhibitors than in renin and therefore does not come into contact with P_3' side-chains. The well-defined interactions of P_3' , described in the mouse renin complex (Dealwis et al., 1994), explain the significant affinity when inhibitors have phenylalanine or tyrosine at P_3' as well as the importance of a P_3' residue for catalytic cleavage of a substrate by renin. Lastly, based on refolding efficiencies, it appeared that the loop may have only a modest impact on refolding in that the mutagenesis to the pepsin sequence did increase the efficiency by only 25%. Further experimentation is essential in order to identify the rate limiting step in the refolding pathway.

Materials and methods

Enzymes and reagents

Enzymes for molecular biology were purchased from the Promega Corporation (Madison) and used according to the manufacturer's instructions. Sequencing radiochemicals were obtained from DuPont NEN (Wilmington), Sequenase kit was purchased from United States Biochemical (Cleveland), Life Technologies Inc. (Grand Island) provided bacterial growth media, ultra-pure urea was from GIBCO-Bethesda Research Laboratory, pepstatinyl agarose was from Sigma (St. Louis) and other biochemicals came from Fisher. Native human liver cathepsin D was purchased from ART Biochemicals (Athens). The synthetic oligonucleotides were synthesized by the University of Florida Interdisciplinary Center for Biotechnology Research (UF-ICBR) DNA Synthesis Core Facility using an Applied Biosystems DNA 394 synthesizer. Chromogenic substrates were synthesized by the solid phase method using an Applied Biosystems Model 430A or 432A in the UF-ICBR Protein Chemistry Core Facility. The purity of the peptides ($\geq 95\%$) was verified by HPLC and by using a Perceptive Biosystems VoyagerTM RP mass spectrometer (MALDI-TOF). N-terminal sequence analysis was also performed by UF-ICBR Protein Chemistry Core Facility with Applied Biosystems 470A and 473A Protein Sequencers. All media recipes were from Molecular Cloning: A Laboratory Manual (Maniatis et al., 1982).

Peptide substrates and native enzyme

Stock peptide solutions were made (except as noted in Results for the P_2' substitutions) in filtered distilled water and quantified by amino acid analysis. The purity of the peptides ($\geq 95\%$) was verified by HPLC and MALDI-TOF analysis. Native human liver cathepsin D was dissolved in ice cold distilled water and aliquots were stored in 10% glycerol at -70°C for future use.

Hydrolysis of substrates

The kinetic parameters k_{cat} , K_m , and k_{cat}/K_m were determined by monitoring the decrease in the average absorbance from 284 to

324 nm upon substrate hydrolysis between Phe*Nph. This assay has been extensively described (Scarborough et al., 1993; Dunn et al., 1994) and values were determined in 0.1 M sodium formate pH 3.7 at 37 °C. The amount of active enzyme was determined by competitive titration with the inhibitor (Hui et al., 1993) Ac-NA-Nva-AHPPA-Nva-NA-NH₂⁽¹⁾ (a generous gift from Kwan Hui, Lilly Research Laboratories) under assay conditions of 2% DMSO.

Modeling of nHuCatD complex

To aid in the interpretation of the kinetic data, molecular modeling techniques were utilized in order to identify potential interactions between substrate side-chains and various subsite residues. Molecular modeling was performed using the Sybyl Software package (Tripos Associates Inc, St. Louis) on a Silicon Graphics Indigo system. Models were constructed using the high resolution three-dimensional structure of human liver cathepsin D complexed with pepstatin (Baldwin et al., 1993). This structure possessed an overall free energy of 500.1 kcal/mol and was therefore optimized before initiation of the modeling process. Energy minimization techniques were utilized to obtain a structure with a global free energy of -302.3 kcal/mol. The Mutate Residue command within the BIOPOLYMER suite of SYBYL was then utilized for the substitution of an Asp or Arg at position P₂' of the complexed inhibitor. Following this, energy minimization techniques were repeated (charges and solvent were not included) in the specific locality of the P₂' side-chain to relieve steric clashes arising between nonbonded interactions. No changes were observed in the protein chain following minimization of the P₂' Asp or Arg.

Active site mutagenesis

In attempts to modify active site specificity, site-directed mutagenesis by PCR (Ho et al., 1989) using Taq Polymerase and the expression vector encoding the gene for the recombinant protein "short" pseudocathepsin D (Beyer & Dunn, 1996) were employed with slight modifications (Scarborough & Dunn, 1994). For mutagenesis, three sets of primers were utilized targeting positions Ile128, Val130, and Pro292-Pro294 for mutagenesis. The resulting PCR products and the WT "short" pseudocathepsin D gene were then restriction digested and ligated together. Lastly, the entire coding region was dideoxy-sequenced to confirm the presence of the mutations and to ensure no misincorporations from Taq polymerase.

Expression, refolding, purification, and N-terminal characterization of mutant rHuCatD

Mutant recombinant protein was overexpressed in *E. coli*, purified in the form of insoluble inclusion bodies, refolded by rapid dilution, and purified via affinity chromatography as previously described (Beyer & Dunn, 1996; Conner & Richo, 1992). Active fractions were pooled and mass spectrometry along with N-terminal sequencing were utilized to confirm processing occurred at the engineered cleavage junction of "short" pseudocathepsin D following incubation at pH 3.7. To determine if molecular masses corresponded with processing at this point, samples were first mixed 1:1 with a saturated solution of 3,5-dimethoxy-4-hydroxy cinnamic acid in 70% acetonitrile and then analyzed using a Perceptiv Biosystems VoyagerTM RP mass spectrometer (MALDI-

TOF). For N-terminal sequencing, purified recombinant enzyme was subjected to SDS-PAGE, transferred to ImmobilonTM-P and N-terminal residues were identified using an Applied Biosystems 470A protein sequencer using gas-phase chemistry by University of Florida Protein Chemistry Core.

Acknowledgments

We extend our appreciation to the University of Florida Protein Chemistry Core staff. This work was supported by NIH grant DK18865 to B.M.D.

References

- Aaltonen M, Lipponen P, Kosma M, Aaltomaa S, Syrjanen K. 1995. Prognostic value of cathepsin-D expression in female breast cancer. *Anticancer Res* 15: 1033-1038.
- Aguilar CF, Dhanaraj V, Guruprasad K, Dealwis C, Badasso M, Cooper JB, Wood SP, Blundell TL. 1995. Comparisons of the three-dimensional structures, specificities and glycosylation of renin, yeast proteinase A and cathepsin D. In: Takahashi K, ed. *Aspartic proteinases: Structure, function, biology, and biomedical implications* 362. New York: Plenum Press. pp 155-166.
- Bailey D, Cooper JB, Veerapandian B, Blundell TL, Atrash B, Jones DM, Szelke M. 1993. X-ray crystallographic studies of complexes of pepstatin A and a statine containing human renin inhibitor with endothiapepsin. *Biochemical J* 289:363-71.
- Baldwin ET, Bhat TN, Gulnik S, Hosur MV, Sowder RC, Cachau RE, Collins J, Silva AM, Erickson JW. 1993. Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design. *Proc Natl Acad Sci* 90:6769-6800.
- Barrett AJ. 1977. Cathepsin D and other carboxyl proteinases. In: *Proteinases of mammalian cells and tissues*. Amsterdam: Elsevier. pp 209-248.
- Beyer BM, Dunn BM. 1996. Self activation of recombinant human cathepsin D at a newly engineered cleavage junction: "Short" pseudocathepsin D. *J Biol Chem* 271:15590-15596.
- Conner GE, Richo G. 1992. Isolation and characterization of a stable activation intermediate of the lysosomal aspartyl protease cathepsin D. *Biochemistry* 31:1142-1147.
- Dealwis CG, Frazao C, Badasso M, Cooper JP, Tickle IJ, Driessen H, Blundell TL, Murakami K, Miyazaki H, Suciras-Diaz J. 1994. X-ray analysis at 2.0 Å resolution of mouse submaxillary renin complexed with a decapeptide inhibitor CH-66, based on the 4-16 fragment of rat angiotensinogen. *J Mol Biol* 236:342-360.
- Diment S, Leech M, Stahl P. 1988. Cathepsin D is membrane-associated in macrophage endosomes. *J Biol Chem* 263:6901-6908.
- Dunn BM, Scarborough PE, Davenport R, Swietnicki W. 1994. Analysis of proteinase specificity by studies of peptide substrates: The use of UV and fluorescence spectroscopy to quantitate rates of enzymatic cleavage. In: Pennington M, Dunn BM, eds. *Peptide analysis protocols*. Totowa, New Jersey: Humana Press Inc. pp 225-243.
- Dunn BM, Scarborough PE, Lowther T, Rao-Naik C. 1995. Comparison of the active site specificity of the aspartic proteinases based on a systematic series of peptide substrates. In: Takahashi K, ed. *Aspartic proteinases: Structure, function, biology, and biomedical implications* 362. New York: Plenum Press. pp 1-10.
- Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77:51-59.
- Houry WA, Rothwarf DM, Scheraga HA. 1995. The initial step in the conformational folding of disulfide-intact ribonuclease A. *Nat Struct Biol* 2:495-503.
- Hui KY, Hermann RB, Manetta JV, Gygi T, Angleton EL. 1993. Model peptides to study the effects of P2 and P3 substitutions in statine-containing HIV proteinase inhibitors. *FEBS Lett* 327:355-360.
- Jaenicke R. 1995. Folding and association versus misfolding and aggregation of proteins. *Protein Folding* 23:1221-1345.
- Ladror US, Snyder SW, Wang GT, Holzman TF, Krafft GA. 1994. Cleavage at the amino and carboxyl termini of Alzheimer's amyloid-beta by cathepsin D. *J Bio Chem* 269:18422-18428.
- Maniatis T, Fritsch EF, Sambrook J. 1982. *Molecular cloning. A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Lab.
- Pillai S, Zull J. 1986. Production of biologically active fragments of parathyroid hormone by isolated Kupffer cells. *J Biol Chem* 261:14919-14923.
- Rao C, Scarborough PE, Kay J, Batley B, Rapundalo S, Klutchko S, Taylor MD, Lunney EA, Humblet CC, Dunn BM. 1993. Specificity in the binding of

- inhibitors to the active site of human/primate aspartic proteinases: Analysis of P2-P1-P1'-P2' variation. *J Med Chem* 36:2614-2620.
- Rocheffort, H. 1990. Mechanism of the overexpression of the cathepsin D gene in breast cancer and consequences in the metastatic process. *Breast Can Res and Treatment* 16:1-13.
- Scarborough PE, Dunn BM. 1994. Redesign of the substrate specificity of human cathepsin D: The dominant role of position 287 in the S2 subsite. *Protein Eng* 7:495-502.
- Scarborough PE, Guruprasad K, Topham C, Richo GR, Conner GE, Blundell TL, Dunn BM. 1993. Exploration of subsite binding specificity of human cathepsin D through kinetics and rule-based molecular modeling. *Protein Sci* 2:264-276.
- Siman R, Mistretta JT, Durkin MJ, Savage LT, Trusko S, Scott RW. 1993. Processing of the beta-amyloid precursor. Multiple proteases generate and degrade potentially amyloidogenic fragments. *J Biol Chem* 268:16602-16609.
- van der Drift ACM, van Noort JM. 1989. The selectivity of cathepsin D suggests an involvement of the enzyme in the generation of T-cell epitopes. *J Biol Chem* 264:14159-14165.