Exploration of subsite binding specificity of human cathepsin D through kinetics and rule-based molecular modeling

PAULA E. SCARBOROUGH,¹ KUNCHUR GURUPRASAD,²
CHRIS TOPHAM,² GARY R. RICH,³ GREGORY E. CONNER,³
TOM L. BLUNDELL,² AND BEN M. DUNN¹

¹ Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, Florida 32610

² ICRF Unit of Structural Molecular Biology, Birkbeck College, University of London,
Malet Street, London WC1E 7HX, United Kingdom

³ Department of Anatomy and Cell Biology, University of Miami School of Medicine, Miami, Florida 33101

(RECEIVED May 19, 1992; REVISED MANUSCRIPT RECEIVED October 12, 1992)

Abstract

The family of aspartic proteinases includes several human enzymes that may play roles in both physiological and pathophysiological processes. The human lysosomal aspartic proteinase cathepsin D is thought to function in the normal degradation of intracellular and endocytosed proteins but has also emerged as a prognostic indicator of breast tumor invasiveness. Presented here are results from a continuing effort to elucidate the factors that contribute to specificity of ligand binding at individual subsites within the cathepsin D active site.

The synthetic peptide Lys-Pro-Ile-Glu-Phe*Nph-Arg-Leu has proven to be an excellent chromogenic substrate for cathepsin D yielding a value of $k_{cat}/K_m = 0.92 \times 10^{-6} \text{ s}^{-1} \text{ M}^{-1}$ for enzyme isolated from human placenta. In contrast, the peptide Lys-Pro-Ala-Lys-Phe*Nph-Arg-Leu and all derivatives with Ala-Lys in the P₃-P₂ positions are either not cleaved at all or cleaved with extremely poor efficiency. To explore the binding requirements of the S₃ and S₂ subsites of cathepsin D, a series of synthetic peptides was prepared with systematic replacements at the P₂ position fixing either Ile or Ala in P₃. Kinetic parameters were determined using both human placenta cathepsin D and recombinant human fibroblast cathepsin D expressed in *Escherichia coli*. A rule-based structural model of human cathepsin D, constructed on the basis of known three-dimensional structures of other aspartic proteinases, was utilized in an effort to rationalize the observed substrate selectivity.

Keywords: aspartic proteinase; cathepsin D; chromogenic assay; kinetics; molecular modeling; substrate specificity; synthetic substrate

Human cathepsin D, an intracellular aspartic proteinase, has been thought to function primarily in the normal degradation of cellular and phagocytosed proteins. Synthesized as a proenzyme followed by cotranslational removal of the signal peptide and glycosylation, the proenzyme is targeted to the lysosomal compartment. Glycosylated procathepsin D is further processed to the mature single-chain and finally two-chain active forms of cathepsin D that reside in lysosomes (Erickson et al., 1987; Hasilik, 1992). Characteristic of the aspartic pro-

teinases, cathepsin D exhibits an acidic pH optimum in vitro.

In addition to normal physiological functions, some pathophysiological correlations have been observed. Increased levels of procathepsin D mRNA and overexpression of procathepsin D have been correlated with the frequency of metastasis in human breast cancer tissues (Capony et al., 1989; Spyrtos et al., 1989; Thorpe et al., 1989; Tandon et al., 1990) and have also been observed in certain neurodegenerative conditions (Hallpike et al., 1970; Whitaker & Seyer, 1979; Yankner et al., 1990).

With the escalating interest in this member of the aspartic proteinase family, we report here on a continuing study to understand the specific structure-activity relationships of cathepsin D. In-depth kinetic studies have

Reprint requests to: Ben M. Dunn, Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, Florida 32610.

Abbreviations: Nph, *p*-nitrophenylalanine; rms, root mean square; SCRs, structurally conserved regions; SVRs, structurally variable regions; CAPS, 3-(cyclohexylamino)-1-propanesulfonic acid.

been used to suggest specific interactions that may be involved in the active site of the enzyme based on computer modeling of the predicted tertiary structure (Fig. 1). Previous studies have demonstrated the value of rule-based structural modeling in exploring subsite interactions within the active sites of related aspartic proteinases (Blundell et al., 1987a,b; Lunney et al., 1991). Interactions at subsites within the active site cleft have been shown to discriminately facilitate the catalysis of favorable substrates by aspartic proteinases (Dunn et al., 1986, 1987; Pohl & Dunn, 1988; Konvalinka et al., 1990; Richards et al., 1990). In order that active site requirements may be examined, human fibroblast procathepsin D has been expressed in *Escherichia coli* (Conner et al., 1989), refolded from solubilized inclusion bodies, and purified using a pepstatin affinity column (Conner & Richo, 1992). The resulting enzyme is identical to native enzyme isolated from human placenta with respect to thermal and pH stability and fluorescence spectrum (Conner & Richo, 1992), although the recombinant enzyme is neither glycosylated nor fully processed to the mature single- and two-chain forms of cathepsin D. To confirm that the resulting recombinant enzyme is an acceptable model for studying substrate specificity in vitro, kinetic properties of recombinant cathepsin D were compared to those of native placental cathepsin D. Kinetic parameters were determined for a series of synthetic chromogenic peptides derived from a parent peptide with the following sequence, which has been shown (Dunn et al., 1986) to be a superior substrate for human cathepsin D ($k_{cat}/K_m = 0.92 \times 10^{-6} \text{ s}^{-1} \text{ M}^{-1}$).

P₅ P₄ P₃ P₂ P₁ *P₁' P₂' P₃'

Lys-Pro-Ile-Glu-Phe*Nph-Arg-Leu

The peptides, designed to completely extend the length of the active site cleft, have incorporated at the P₁*P₁' scissile bond two aromatic residues, -Phe*Nph-, which anchor the peptide into the respective S₁-S₁' subsites of the enzyme. Systematic substitutions of residues are introduced in flanking positions.

In earlier studies (Scarborough et al., 1991), peptides synthesized as possible substrates for a number of aspartic proteinases gave extremely poor results for human cathepsin D (Table 1). All derivatives with -Ala-Lys- in the P₃-P₂ positions were either not cleaved at all or were cleaved with extremely poor efficiency as compared to Lys-Pro-Ile-Glu-Phe*Nph-Arg-Leu. It was hypothesized that a positively charged residue is not acceptable in the S₂ subsite and that the more bulky Ile residue in P₃ of the parent sequence may also contribute to binding. The new series of peptides described here was synthesized to further detail the requirements of the S₂ and S₃ subsites of the cathepsin D active site.

Combining substrate specificity results with structural modeling of cathepsin D, residues in the primary sequence may be selected for mutagenesis studies. The ultimate goal of this continuing work is to identify specific residues in cathepsin D, which dictate binding requirements of the active site.

Results

Kinetics

Tables 2 and 3 list kinetic parameters derived from chromogenic kinetic assays with the new series of peptides, in order of descending specificity constant (k_{cat}/K_m) for the

Fig. 1. Schematic representation showing arrangement of α -helices, β -sheets, and loop regions in the model produced using program SETOR (Evans, 1992). The catalytic aspartate side chains (red), four disulfide bridges (yellow), and the side chains of the substrate from P₃ to P₁' are shown.

Table 1. Earlier specificity results for human cathepsin D^a

Peptide sequence								Substrate specificity $\times 10^{-6}$ ($V_{\max}/K_m$) ^b		
P ₅	P ₄	P ₃	P ₂	P ₁	*P ₁ '	P ₂ '	P ₃ '	Placenta	Spleen	Recombinant
Lys-Pro-Ile-Glu-Phe*Nph-Arg-Leu								77.5 $\pm$ 2.86	53.6 $\pm$ 4.09	40.8 $\pm$ 3.77
Asp-Pro-Ala-Lys-Phe*Nph-Arg-Leu								4.63 $\pm$ 0.39	1.90 $\pm$ 0.20	3.35 $\pm$ 0.11
Leu-Pro-Ala-Lys-Phe*Nph-Arg-Leu								3.94 $\pm$ 0.25	1.51 $\pm$ 0.15	1.64 $\pm$ 0.05
Arg-Pro-Ala-Lys-Phe*Nph-Arg-Leu								0.86 $\pm$ 0.07	0.45 $\pm$ 0.04	0.51 $\pm$ 0.02
Ser-Pro-Ala-Lys-Phe*Nph-Arg-Leu								2.49 $\pm$ 0.10	0.84 $\pm$ 0.09	1.17 $\pm$ 0.03
Lys-Pro-Ala-Lys-Phe*Nph-Arg-Leu								NC ^c	NC	NC
Lys-Pro-Ala-Lys-Ala*Nph-Arg-Leu								NC	NC	NC
Lys-Pro-Ala-Lys-Leu*Nph-Arg-Leu								NC	NC	NC
Lys-Pro-Ala-Lys-Val*Nph-Arg-Leu								NC	NC	NC
Lys-Ala-Ala-Lys-Phe*Nph-Arg-Leu								NC	NC	NC
Lys-Asp-Ala-Lys-Phe*Nph-Arg-Leu								NC	NC	NC
Lys-Leu-Ala-Lys-Phe*Nph-Arg-Leu								NC	NC	NC
Lys-Arg-Ala-Lys-Phe*Nph-Arg-Leu								NC	NC	NC
Lys-Ser-Ala-Lys-Phe*Nph-Arg-Leu								NC	NC	NC

^a Scarborough et al. (1991).^b Velocities were measured in units of AU s⁻¹, then converted to M s⁻¹; substrate concentrations were 5–300 μ M; enzyme quantities required for assays were in the range of 0.015–0.33 μ g in 250 μ L total volume.^c In 250- μ L reactions containing up to 5 μ g enzyme and 400 μ M peptide and monitored for up to 30 min at 37 °C, no measurable cleavage was observed.

placental enzyme. Initial rates of cleavage at varying concentrations of peptide were determined for both the native placental and recombinant cathepsin D preparations, and values for K_m and k_{cat} were calculated. In all cases, high-performance liquid chromatography (HPLC) separation was utilized to demonstrate that cleavage occurred exclusively at the -Phe*Nph- peptide bond, in agreement with results previously reported for porcine pepsin and other enzymes (Dunn et al., 1986; Pohl & Dunn, 1988). The Michaelis constants and trends in substrate specific-

ities seen for the recombinant enzyme are comparable to those of the native preparation, with most of the variation contributed by lower k_{cat} values for the nonglycosylated, partially processed enzyme when compared to parallel values for the native cathepsin D.

Overall, the series of peptides reported here yielded several excellent substrates for cathepsin D. More importantly, some P₂ subsite preferences were elucidated. Hydrophobic substitutions in P₂ were accommodated well by the enzyme. The longer Leu (peptides 1 and 12)

Table 2. Specificity of human cathepsin D for P₂ derivatives with Ile in P₃

Peptides								Placental cathepsin D ^a			Recombinant cathepsin D ^a		
P ₅	P ₄	P ₃	P ₂	P ₁	*P ₁ '	P ₂ '	P ₃ '	K_m (μ M)	k_{cat} (s ⁻¹)	k_{cat}/K_m (s ⁻¹ μ M ⁻¹)	K_m (μ M)	k_{cat} (s ⁻¹)	k_{cat}/K_m (s ⁻¹ μ M ⁻¹)
1	Lys-Pro-Ile-Leu-Phe*Nph-Arg-Leu							28 $\pm$ 9	45 $\pm$ 5	1.61 $\pm$ 0.18	40 $\pm$ 9	19 $\pm$ 2	0.48 $\pm$ 0.05
2	Lys-Pro-Ile-Glu-Phe*Nph-Arg-Leu							50 $\pm$ 4	46 $\pm$ 2	0.92 $\pm$ 0.03	40 $\pm$ 8	18 $\pm$ 1.4	0.45 $\pm$ 0.04
3	Lys-Pro-Ile-Ile-Phe*Nph-Arg-Leu							26 $\pm$ 8	22 $\pm$ 3	0.85 $\pm$ 0.10	41 $\pm$ 9	6.7 $\pm$ 0.7	0.16 $\pm$ 0.02
4	Lys-Pro-Ile-Ala-Phe*Nph-Arg-Leu							41 $\pm$ 7	31 $\pm$ 2	0.76 $\pm$ 0.05	41 $\pm$ 5	4.5 $\pm$ 0.2	0.11 $\pm$ 0.005
5	Lys-Pro-Ile-Val-Phe*Nph-Arg-Leu							46 $\pm$ 5	25 $\pm$ 1	0.54 $\pm$ 0.02	61 $\pm$ 11	9.3 $\pm$ 0.8	0.15 $\pm$ 0.01
6	Lys-Pro-Ile-Ser-Phe*Nph-Arg-Leu							62 $\pm$ 9	24 $\pm$ 2	0.39 $\pm$ 0.03	84 $\pm$ 16	11 $\pm$ 1	0.13 $\pm$ 0.01
7	Lys-Pro-Ile-Asp-Phe*Nph-Arg-Leu							56 $\pm$ 11	15 $\pm$ 1	0.27 $\pm$ 0.02	64 $\pm$ 14	5.3 $\pm$ 0.7	0.08 $\pm$ 0.01
8	Lys-Pro-Ile-Gln-Phe*Nph-Arg-Leu							54 $\pm$ 15	11 $\pm$ 2	0.20 $\pm$ 0.04	43 $\pm$ 9	1.7 $\pm$ 0.2	0.04 $\pm$ 0.005
9	Lys-Pro-Ile-Arg-Phe*Nph-Arg-Leu							72 $\pm$ 18	6 $\pm$ 1	0.08 $\pm$ 0.01	49 $\pm$ 8	1.7 $\pm$ 0.2	0.04 $\pm$ 0.004
10	Lys-Pro-Ile-Lys-Phe*Nph-Arg-Leu								NC ^b			NC	
11	Lys-Pro-Ile-His-Phe*Nph-Arg-Leu								NC			NC	

^a Velocities were measured in units of AU s⁻¹ then converted to M s⁻¹; substrate concentrations were 5–300 μ M; enzyme concentrations required for assays were in the range of 1–6 nM.^b In 250- μ L reactions containing up to 5 μ g enzyme, 400 μ M peptide, and monitored for up to 30 min at 37 °C, no measurable cleavage was observed.

Table 3. Specificity of human cathepsin D for P_2 derivatives with Ala in P_3

Peptides								Placental cathepsin D ^a			Recombinant cathepsin D ^a		
P ₅	P ₄	P ₃	P ₂	P ₁	*P ₁ '	P ₂ '	P ₃ '	K_m (μ M)	k_{cat} (s ⁻¹)	k_{cat}/K_m (s ⁻¹ μ M ⁻¹)	K_m (μ M)	k_{cat} (s ⁻¹)	k_{cat}/K_m (s ⁻¹ μ M ⁻¹)
12	Lys-Pro-Ala-Leu-Phe	Nph-Arg-Leu						63 ± 15	63 ± 8	1.00 ± 0.13	127 ± 34	40 ± 7	0.32 ± 0.06
13	Lys-Pro-Ala-Glu-Phe	Nph-Arg-Leu						95 ± 45	35 ± 10	0.37 ± 0.11	83 ± 22	6.5 ± 2	0.08 ± 0.02
14	Lys-Pro-Ala-Asp-Phe	Nph-Arg-Leu						5 ± 3	1.5 ± 0.1	0.30 ± 0.02	28 ± 5	1.0 ± 0.1	0.03 ± 0.004
15	Lys-Pro-Ala-Ala-Phe	Nph-Arg-Leu						72 ± 13	11 ± 1	0.15 ± 0.01	116 ± 15	5.7 ± 0.5	0.05 ± 0.004
16	Lys-Pro-Ala-Ser-Phe	Nph-Arg-Leu						41 ± 12	3 ± 0.4	0.07 ± 0.01	74 ± 17	2 ± 0.3	0.03 ± 0.004
17	Lys-Pro-Ala-Arg-Phe	Nph-Arg-Leu							NC ^b			NC	
18	Lys-Pro-Ala-Lys-Phe	Nph-Arg-Leu							NC			NC	

^a Velocities were measured in units of AU s⁻¹ then converted to M s⁻¹; substrate concentrations were 5–300 μ M; enzyme concentrations required for assays were in the range of 1–6 nM.

^b In 250- μ L reactions containing up to 5 μ g enzyme, 400 μ M peptide, and monitored for up to 30 min at 37 °C, no measurable cleavage was observed.

as well as β -branched Ile and Val (peptides 3 and 5, respectively) were all acceptable. Interestingly, whereas a Glu residue in P_2 (peptides 2 and 13) always yielded an excellent substrate, substituting an Asp (peptides 7 and 14) generally was less favorable. In both cases the substitution of Ser into P_2 (peptides 6 and 16) gave a poorer substrate. The most dramatic results were those seen with peptides substituted with a positively charged residue in P_2 . Peptide 9 (P_2 Arg) in Table 2 underwent cleavage at rates barely sufficient for K_m and k_{cat} determination, and for all other peptides containing a positively charged residue in P_2 (peptides 10 and 11 in Table 2 and peptides 17 and 18 in Table 3), cleavage was abysmal. Products of overnight reactions (containing up to 5 μ g cathepsin D and 250 μ M peptide at 37 °C) were separated by HPLC and identified by amino acid analysis. Results demonstrated that the Lys-containing peptides 10 and 18 are eventually cleaved when presented to overwhelming amounts of enzyme; however, a significant amount of the His-containing peptide 11 *remained* uncleaved even under these extreme conditions.

Modeling

To aid in interpretation of the kinetic data, a model of cathepsin D has been constructed using the rule-based comparative approach encoded in the COMPOSER suite of computer programs (see coordinates on Diskette Appendix). The model was based on the three-dimensional structures of pepsin, chymosin, renin, and other homologous aspartic proteinases defined by high-resolution X-ray analysis. In order to define the residues involved in subsite specificity, the inhibitors CH66 (trimethylacetyl-His-Pro-Phe-His-Leu- ψ (CH(OH)CH₂)-Leu-Tyr-Tyr-Ser-NH₂) and H256 (Pro-Thr-Glu-Phe-(CH₂NH)-Phe-Arg-Glu) and the substrate peptide Lys-Pro-Ile-Glu-Phe*Nph-Arg-Leu have been modeled into cathepsin D

using the coordinates of inhibitor complexes of renin and other aspartic proteinases as a guide.

A schematic representation of the model of human cathepsin D showing the overall topology and the arrangement of α -helices and β -strands, along with the side chains of the substrate from P_3 to P_1' modeled into the active site is shown in color in Figure 1 (see also Kinemage 1). Asp 32 and Asp 215 (shown in red) in the model are involved in the catalytic activity. The fourth disulfide bond, unique to human cathepsin D, formed between Cys 26, which lies on a β -sheet and the other cysteine that lies in the insertion region (see SVR 7 in Table 4) is also shown. These cysteines are accessible to solvent. The model was compared with other known structures in the aspartic proteinase family using the program COMPARE (Sali et al., 1990; Zhu et al., 1992). This program equivalences similar structural features in proteins and produces an alignment of the sequences based on this comparison (see Blundell et al., 1991). The result is a schematic representation of α -helices and β -strands and the nature of the equivalent structural features such as accessibility of the side chain or type of hydrogen bond (Overington et al., 1990, 1992). In the cathepsin D model, with the exception of the residues in the active site cleft shown in Table 5, most of the conserved amino acids are inaccessible to solvent and, together with a number of buried polar groups that are hydrogen bonded, they contribute to a close-packed core. This gives confidence in the alignment used in the model and the quality of the model constructed.

A stereo view of the superposition of C α atoms using X-ray structures of porcine pepsin, bovine chymosin, mouse renin, yeast proteinase A, and the model of human cathepsin D is shown in Figure 2. The availability of several aspartic proteinase structures with ~50% sequence identity is a good basis for comparative modeling. Figure 2 highlights the structurally conserved core in the aspartic protease family. This can be further seen from

Table 4. Selection of fragments for structurally variable regions (SVRs) in the model

SVR No.	SVR ^a	Fragment in SVR	Fragment selected for SVR	PDB code of protein containing SVR ^b
1	Gly 0–Pro 1	GP	AP	1TIM
2	Asn 8–Tyr 14	NYMDAQY	NYLDTEY	5PEP
3	Thr 22–Pro 24	TPP	TPP	4CMS
4	Lys 46–Ala 49	KLLDIA	SRLYTA	MREN
5	Lys 66–Gly 68	KNG	ATS	5PEP
6	Gly 76–Leu 80	GSGSL	GTGSM	5PEP
7	Pro 91a–Val 94	PCQSASSASALGGV	PGADPPKKRKEAML	2PLV
8	Val 130–Val 133	VNNV	ASGA	5PEP
9 ^c	Ser 156–Gly 163	SRDPDAQPGG	DRDG–QES	4CMS
			PVD	1LDB
10	Tyr 174–Ser 178	YYKGS	YYTGS	5PEP
11	Thr 185	T	T	4CMS
12	Ala 200–Ser 201	AS	DG	5PEP
13	Pro 224–Glu 227	PVDE	PTSA	5PEP
14	Pro 239–Met 246	PLIQGEYM	QNQYGEFD	4CMS
15	Leu 255–Ala 257	LPA	MPT	4CMS
16	Thr 275–Leu 283	TLKVSQAGKTLCL	DFGPSTGSSSCF	4APE
17	Met 289–Trp 299	MDIPPPSGPLW	MDIPPPTGPTW	MREN
18	Arg 326a–Leu 326b	RL	RR	4FDR

^a Residues are numbered according to porcine pepsin.

^b Protein Data Bank (PDB) codes: 1TIM, triose phosphate isomerase; 5PEP, porcine pepsin; 4CMS, bovine chymosin; MREN, mouse renin; 2PLV, polio virus; 1LDB, lactate dehydrogenase; 4APE, endothiapepsin; 4FDR, dihydrofolate reductase. Coordinates for mouse renin have been deposited in the PDB (Dhanaraj et al., 1992).

^c SCRs on either side of SVR 9 were extended by selecting the first four and last three residues from the corresponding fragment of bovine chymosin (4CMS). A shorter fragment consisting of three residues was then selected from lactate dehydrogenase to connect the ends and complete the construction of SVR 9.

the pairwise comparisons of the sequence identities and the rms differences in the structural model of human cathepsin D with other aspartic proteases, calculated by least-squares superposition of cathepsin D on each of the structures of aspartic proteinases defined by X-ray anal-

ysis and shown in Table 6. The very close similarity of the model of cathepsin D with the structures of pepsin and chymosin partly reflects the fact that the SCRs of cathepsin D resemble those of pepsin and chymosin and have been modeled on the basis of the pepsin and chymosin

Fig. 2. Stereo view showing superposition of the C_α atoms of porcine pepsin (yellow), bovine chymosin (green), mouse renin (blue), yeast proteinase A (violet), and the model of human cathepsin D (red). Regions that are superposed are seen in white.

Table 5. Specificity pockets in models of cathepsin D with inhibitors in the active site^a

Pocket	CH-66 ^b model	H-256 ^c model	Substrate ^d model
S ₆	M10, A12, L276	—	—
S ₅	Nil	Nil	—
S ₄	S219 , L220	S219	S219
S ₃	Q13 , S77, T111, F117, G217 , T218 , S219	A12, Q13 , S77, G217 , T218 , S219	Q13 , S77, F117, G217 , T218 , S219
S ₂	S77, G217 , T218 , V222 , M289 , I300	S77, G217 , T218 , V222 , M287, M289	S77, G217 , T218 , V222 , M287, M289
S ₁	V30 , D32 , G34, Y75, S77, I120, D215 , G217 , T218	V30 , D32 , S77, F117, D215 , G217 , T218	V30 , D32 , S77, F117, D215 , G217 , T218
S ₁ '	G34, Y189, I213, D215 , T218 , I291 , I300	G76, D215 , T218 , M289, I291 , I300	G76, D215 , T218 , M289, I291 , I300
S ₂ '	G34 , I128 , V130 , Y189	G34 , I73, I128 , V130	G34 , I73, I128 , V130
S ₃ '	G76, Y189, I291, P292	H74, G76, M289, D290, P292	H74, D290, I291, P292
S ₄ '	P294	—	—

^a The distance cutoff was 4.0 Å; residues are numbered according to porcine pepsin (Protein Data Bank [Bernstein et al., 1977; Abola et al., 1987] code 5PEP); bold lettering indicates residues common to a given pocket in each model.

^b CH-66: trimethylacetyl-His-Pro-Phe-His-Leu-ψ(CH(OH)CH₂)-Leu-Tyr-Tyr-Ser-NH₂.

^c H-256: Pro-Thr-Glu-Phe-(CH₂NH)-Phe-Arg-Glu.

^d Substrate: Lys-Pro-Ile-Glu-Phe*Nph-Arg-Leu.

structures alone. In fact, a greater number of SCRs have been selected from pepsin, with which there is slightly higher sequence similarity.

Although the model has approximately the same per-

centage of C_α atoms topologically equivalent with pepsin (85.9%), chymosin (83.9%), and yeast proteinase A (81.3%), the rms difference with yeast proteinase A is almost double (1.0 Å) than that with either pepsin (0.52 Å)

Table 6. Pairwise percentage sequence identities and pairwise root mean square (rms) differences of structurally equivalent C_α positions, for a model of human cathepsin D (CATD) as compared with other aspartic proteinases of known three-dimensional structure, along with similar comparisons for each of the aspartic proteinases.

	Protein code								
	5PEP	4CMS	2APR	3APP	4APE	HREN ^b	MREN ^b	YPRA ^c	CATD ^d
Resolution (Å)	2.3	2.2	2.5	1.8	2.1	2.8	1.9	3.5	—
Amino acids	326	320	325	323	30	331	331	329	348
Protein code									
5PEP	100	0.84 (93.4)	1.13 (78.1)	1.11 (71.5)	0.99 (58.7)	1.01 (84.5)	1.02 (78.2)	1.00 (89.3)	0.52 (85.9)
4CMS	59.4	100	1.06 (81.2)	1.09 (74.9)	1.09 (61.8)	0.97 (80.6)	1.07 (79.1)	1.06 (86.9)	0.66 (83.9)
2APR	38.5	34.1	100	0.96 (79.8)	1.01 (74.8)	1.11 (68.8)	1.02 (62.8)	1.17 (71.4)	1.09 (74.4)
3APP	31.9	29.1	40.6	100	0.95 (79.6)	1.17 (67.6)	1.23 (65.8)	1.20 (68.6)	1.12 (68.1)
4APE	27.9	26.6	38.2	54.8	100	1.02 (57.4)	1.07 (57.4)	1.04 (58.9)	0.99 (57.1)
HREN	39.9	37.8	25.5	22.3	23.3	100	0.68 (94.5)	1.06 (85.4)	0.97 (81.0)
MREN	42.9	38.4	28.9	24.5	23.3	71.3	100	1.01 (82.9)	1.07 (81.0)
YPRA	40.2	37.5	36.6	26.0	27.7	38.0	38.9	100	1.00 (81.3)
CATD	50.6	47.8	32.9	26.0	28.5	46.2	50.2	45.6	100

^a Values in the lower triangular matrix represent the sequence identities, and values in the upper triangular matrix represent the rms differences between the two proteins being compared. Values within parentheses are the percentage of C_α equivalents. C_α atoms that lie within 3.0 Å distance of each other in the optimally superposed proteins are considered to be the structurally equivalent residues.

^b PDB codes: 5PEP, porcine pepsin; 4CMS, bovine chymosin; 2APR, rhizopuspepsin; 3APP, penicillopepsin; 4APE, endothiapepsin. Coordinates for mouse renin (MREN) and human renin (HREN) (Dhanaraj et al., 1992) have been deposited with the Brookhaven Protein Data Bank. (PDB; Bernstein et al., 1977 and Abola et al., 1987).

^c Coordinates for yeast proteinase A (YPRA) (Carlos et al., unpubl.).

^d Human cathepsin D model (CATD). Code for other proteins is the same as in the PDB.

or chymosin (0.66 Å). This is expected, not only because the yeast proteinase structure was not used in the modeling but also because the X-ray analysis is at a much lower resolution. The structures of human and mouse renins also have a high percentage of residues that are defined as topologically similar; this is expected from the high sequence percentage identity. It is interesting that the sequence similarity with renin is in the longer loops to a greater extent than for pepsin and chymosin; this emphasizes the importance of using renin structures, especially in the region of the disulfide loop, Lys 46–Ala 49, and the long loop, Met 289–Trp 299, which contains many prolines including cis-peptide conformations. Despite choosing fragments from mouse renin to construct two loops in the model, the rms difference in the values of the model with mouse and human renins is relatively large as compared with pepsin or chymosin reflecting the lower sequence identities in the SCR regions of the renins when compared to cathepsin D. The conformation of the polyproline loop is important for binding in the pockets S_1' – S_4' as shown by the structure of mouse renin complex (Dhanaraj et al., 1992), and this interaction may explain the existence of a similar structure in the renins and cathepsin D.

Most of the loops have been selected from the aspartic proteinases (see Table 4). The region containing residues Ser 156–Gly 163 has been modeled on chymosin to form a collar and then a further insertion has been modeled from lactate dehydrogenase (Table 4). This loop is far away from the active site cleft and is unlikely to affect substrate specificity.

The disulfide Cys 26–Cys 91b in human cathepsin D is not found in other aspartic proteinases of defined three-dimensional structure. It is associated with an insertion of 11 residues in the region between Val 91 and Gly 92, which is unique to human cathepsin D. The strands immediately preceding and following the large insertion are similar in cathepsin D to those of other aspartic proteinases and are conserved structurally in the enzymes of known structure. Retention of the positions of these strands in cathepsin D requires the insertion to be on the surface. Although there is some evidence for a short amphipathic helix within the loop, the large number of hydrophilic residues in the inserted region, together with the ease with which they can be removed by proteolysis, is consistent with their being exposed to solvent. We modeled this unusual region of cathepsin D using a fragment selected from polio virus (see Table 4). However, on comparison with the sequence of porcine cathepsin D (which has a serine instead of a cysteine at 91b), it has been proposed earlier that the possible sites for proteolytic cleavage of single chain cathepsin D are between Gln and Ser and between Ala and Leu in the insertion region. The confidence in modeling this region and that around Pro 159–Pro 161 is lower than elsewhere, as these fragments were selected from members outside the homologous family.

The prolines at 150 and 161 are likely to have a different conformation from other aspartic proteases. Because seven residues in the region of insertion between Val 91–Gly 92 are cleaved out in the mature enzyme and because the large number of polar residues implies a flexible structure, we have not included these residues in Figure 2. The inserted loop occupies a position close to the N-terminus. Therefore, the propeptide of the enzyme precursor may affect the proteolysis of this insertion. Both loops Val 91–Gly 92 and Pro 159–Pro 161 are distant from the active site cleft and will not affect the conclusions concerning specificity.

Discussion

The current data taken together with earlier kinetic results (Scarborough et al., 1991) demonstrate that both the native placental and nonglycosylated recombinant enzyme preparations exhibit similar trends in substrate preference with respect to P_2 and P_3 residues with perhaps two exceptions. Peptide 14 yields anomalous results that will be discussed below. The other peptide, which yields differing results for the two enzyme samples, is the Ala-containing peptide 4. However, the differences in the values of k_{cat}/K_m among peptides 3, 4, 5, and 6 for both the recombinant as well as the native placental enzymes are not large enough to show a definitive order of preference. Examining the series as a whole, the overall range of preferences is clear: The enzyme prefers a hydrophobic residue in P_2 , accepts residues capable of hydrogen bonding, and strongly rejects positively charged amino acids in the P_2 position. There is also a preference of Ile in P_3 over the smaller Ala. It may therefore be concluded that the recombinant enzyme presents an acceptable model for in vitro studies of substrate specificity.

The lower specific activity, a mean difference in k_{cat} values of approximately threefold, observed for the nonglycosylated, partially processed recombinant enzyme when compared to the native preparation from placenta may reflect the contribution of glycosylation to proteolytic activity or enzyme stability. Evidence has been reported from other studies (Aikawa et al., 1990) that glycosylation may enhance the proteolytic activity of a related aspartic proteinase, *Mucor* rennin. Alternatively, it may be suggested that the nonglycosylated form is more sensitive to autoproteolysis. A total disappearance of recombinant cathepsin D has been observed during attempts to process to fully mature enzyme. The incomplete processing of pseudocathepsin D to the mature enzyme may also contribute to the small difference in specific activity reported here.

Previous substrate studies (Table 1; Scarborough et al., 1991) examined a series of peptides that were derived from the parent peptide Lys-Pro-Ala-Lys-Phe-Nph-Arg-Leu that was shown by Pohl and Dunn (1988) to be an excellent substrate for the highly homologous porcine

pepsin. When compared to reactions with Lys-Pro-Ile-Glu-Phe*Nph-Arg-Leu, it was determined that either or both of the replacements in P_3 and P_2 were significantly affecting catalysis by cathepsin D. It appeared that human cathepsin D either would not accommodate a positively charged residue in P_2 or had a significant preference in P_3 for Ile rather than Ala. In order to resolve these inter-related questions, a new series of peptides was synthesized that systematically varied the residue in P_2 , fixing either Ile (Table 2) or Ala (Table 3) in P_3 . In contrast with the original series, the new peptides yielded many excellent substrates for human cathepsin D. Most of the variation in preference was reflected in values of k_{cat} ; however, the peptides with hydrophobic residues in P_2 generally exhibited slightly better Michaelis constants as well.

In looking at the peptide series as a whole, an increase in the hydrophobicity of the side chain at P_2 led to improved efficiency of catalysis. The accommodation of long or bulky hydrophobic groups suggests that the S_2 subsite is of a significant size. Table 5 describes S_2 as overwhelmingly hydrophobic, with the side chains of Met 289, Ile 300, Val 222, and Met 287 available to associate closely with the P_2 residue of the substrate or inhibitor. Met 287 displays a similar spatial orientation to the Glu 287 residue, which has been proposed to be involved in an electrostatic interaction in the S_2 subsite of the highly homologous enzyme porcine pepsin (Pohl &

Dunn, 1988). The methionine substitution in the human cathepsin D sequence may significantly increase the hydrophobicity of S_2 , evidenced by the values of k_{cat}/K_m determined for peptides containing hydrophobic residues Leu, Ile, Val, and Ala in P_2 (peptides 1, 3, 4, and 5 in Table 2 and peptides 12 and 15 in Table 3). Figure 3 illustrates the proposed interactions between the side chain of the P_2 residue and the two methionines in the S_2 pocket.

The absence of a negative charge at position 287 may also greatly influence the observed substrate specificity in serving to benefit the binding of the Glu-containing substrates. It was, therefore, predicted that the substitution of Asp into P_2 would yield peptides that would exhibit kinetic parameters similar to those containing Glu. The results were puzzling for peptide 7 but especially for peptide 14. Both peptides exhibited similar specificity constants; however, the binding of the peptides appeared to be very different. Peptide 14, with -Ala-Asp- in P_3 - P_2 , exhibited an apparent Michaelis constant that was anomalously low and a very slow turnover rate when compared to peptide 7, which has -Ile-Asp-. This is most likely indicative of nonproductive interactions for the -Ala-Asp-peptide.

The data observed for the peptides with charged residues in P_2 (negative or positive) may reveal more about electrostatic interactions and geometry in the S_2 subsite.

Fig. 3. A view of interactions in the S_2 subsite of the human cathepsin D/substrate model. The side chain of the P_2 residue (Glu 351) of the substrate (shown in yellow) may rest between the long hydrophobic arms of the two methionine residues (green). Met 287 is spatially equivalent to Glu 287 of porcine pepsin.

The catalytic Asp 215 is also among the enzyme residues surrounding the S₂ subsite (the distance between the β -carbon of the P₂Glu and OD2 of Asp 215 in the model is 6.2 Å). Green et al. (1990) proposed that ion pairing between the P₂ side chain of a renin substrate and the renin catalytic Asp 226 (renin numbering; corresponds to Asp 231 in cathepsin D) "could perturb the ionization of the active site aspartyls." They suggested that a positive charge near the catalytic pair of aspartic acids of renin may serve to lower the pK_a of Asp 226 thereby increasing the pK_a of the other catalytic Asp. This effect may in turn raise the optimum pH of the proteolytic reaction catalyzed by renin. A similar explanation may apply to the results seen with cathepsin D action on substrates with positively charged residues in P₂. First, considering the peptides with Arg, Lys, or His in P₂ (peptides 9, 10, 11, 17, and 18), Arg is mildly acceptable exhibiting observable cleavage (however slight) in the chromogenic assay. Lys in P₂ did not exhibit measurable cleavage in the assay, but when incubated overnight with large amounts of enzyme, the appropriate products of cleavage were detectable in HPLC peaks as revealed by amino acid analysis. Finally the His-containing peptide 11, even in overnight reactions was not significantly cleaved. These results suggest that possibly the geometry and number of rotatable bonds of the positively charged side chain may exert an effect on the catalytic efficiency of the enzyme; therefore, the fact that Arg and Lys were marginally more acceptable than His in the P₂ position may be attributed to a greater hydrophobic character, and perhaps a greater flexibility, of the longer arms of the Lys and Arg side chains and, therefore, an ability to better interact with the hydrophobic residues lining the pocket. Similarly the comparison of the Glu- and Asp-containing peptides (2, 7, 13, and 14) may suggest that the longer Glu side chain, with one more methylene group than the shorter Asp, is able to better associate with the hydrophobic groups lining the S₂ subsite thereby not disturbing the delicate interactions that drive catalysis.

The k_{cat} parameter is, in general, the contributor of most of the variation seen in Tables 2 and 3. In extensive studies of porcine pepsin, Pohl and Dunn (1988) found that in a large series of peptides, the k_{cat} parameter was generally invariant, especially when the pH dependence of the kinetic parameters was examined, while large changes were observed in the apparent K_m as well as in k_{cat}/K_m . In contrast, in results reported here for human cathepsin D, the catalytic efficiency, represented by k_{cat} , is exquisitely sensitive to substitution in position P₂. This might imply that the transition state for the catalytic reaction in the case of cathepsin D requires further positional adjustments while the pepsin reaction would require little or no such realignment.

Turning to the S₃ subsite of human cathepsin D, in comparing pairs of peptides with the same P₂ substitution and either Ile or Ala in P₃ (Tables 2 and 3, respec-

tively), there was an overall preference for Ile in P₃ as seen in the values of k_{cat}/K_m . For example, peptide 1 with -Ile-Leu- in P₃-P₂ exhibited a specificity constant ($k_{cat}/K_m = 1.61 \times 10^{-6} \text{ s}^{-1} \text{ M}^{-1}$), which was measurably greater than that of peptide 12 with -Ala-Leu- ($k_{cat}/K_m = 1.00 \times 10^{-6} \text{ s}^{-1} \text{ M}^{-1}$). A similar comparison can be made between the peptides that were substituted at P₂ with Glu, Ala, Ser, and Arg with either Ile or Ala fixed in P₃ (peptides 2, 4, 6, and 9 in Table 2 compared to 13, 15, 16, and 17 in Table 3, respectively).

Figure 4 is a view of the S₃ subsite of the modeled cathepsin D/inhibitor complex. The residues of particular interest are Ala 114 and Val 30. The corresponding porcine pepsin residues are Tyr 114 and Ile 30, respectively. Another significant difference in sequence not shown in the figure is a Thr residue that aligns with Phe 111 in porcine pepsin. The smaller amino acids present in cathepsin D should serve to increase the size of the S₃ pocket substantially. Previous studies (Pohl & Dunn, 1988) have revealed that Ala is the optimal residue in P₃ for porcine pepsin. It might then be suggested that peptides with the bulky hydrophobic Ile side chain in P₃ may anchor strongly into the cathepsin D S₃ subsite, whereas Ala-containing peptides may be more loosely bound.

Efforts are now underway to produce site-directed mutants of Met 287, Ala 114, Thr 111, and Val 30 to evaluate the hypotheses presented here.

Conclusions

The chromogenic synthetic substrates and sensitive assay utilized in these studies have made it possible to gather evidence on interactions between peptide and enzyme within the active site of human cathepsin D. Lacking an X-ray crystal structure for this particular member of the aspartic proteinase family, we have made use of a rule-based model of human cathepsin D that has proven to be critical in this continuing study. These results further support earlier studies suggesting that electrostatic interactions within the active site cleft of aspartic proteinases can be critical with respect to substrate specificity in either a favorable or an unfavorable manner. Additional studies utilizing kinetics, modeling, and mutagenesis should continue to reveal residues in the cathepsin D sequence that contribute to the specificity of this enzyme.

Materials and methods

Analyses

Human fibroblast procathepsin D was expressed in and isolated from *E. coli* cells in the form of insoluble inclusion bodies that were stored in 1-mg aliquots at -70°C . In order to recover active enzyme, inclusion bodies were solubilized and reduced at a concentration of 2 mg/mL in 50 mM CAPS, 50 mM β -mercaptoethanol, and 8 M

Fig. 4. A view of the S_3 subsite of the human cathepsin D/substrate model looking from the prime end of the substrate along the backbone toward the P_3 residue, Ile 350 (yellow). The characteristic Gln 13 (white; corresponding to Glu 13 in porcine pepsin) interacts with the Ile side chain. Val 30 and Phe 117 seem to form the boundary between S_3 and S_1 . In this view, Ala 114 is on the upper boundary of the pocket.

urea, pH 10.7, and incubated at room temperature for 30 min. After being centrifuged in a Speed Vac for 10 min, the sample was added dropwise to water to a concentration of 0.5 nmol/mL, and the pH was adjusted to 8.7. After a 2-h incubation, glutathione was added to a concentration of 11 mM (10 mM reduced, 1 mM oxidized). After 2 more hours the pH was lowered to 3.7 and after 48 h, the refolding mixture was applied onto a pepstatinyl-agarose affinity column equilibrated with 0.1 M sodium formate buffer, pH 3.5, then eluted with 0.1 M sodium phosphate buffer, pH 8.0. Column fractions containing hydrolytic activity were pooled and stored at -70°C .

For the purposes of purifying placental cathepsin D, frozen human placental tissue was thawed and homogenized in 10 mM Tris-HCl, pH 7.4, and 0.5% Brij 35 in a Waring blender. The homogenate was centrifuged at 10,000 rpm for 30 min in a Sorvall GSA rotor. The pH of the supernatant was adjusted to 3.7 with 5.7 N HCl, and sodium acetate was added to a concentration of 0.1 M. Following a 30-min incubation at 0°C , the precipitate formed was removed by a 30-min centrifugation at 10,000 rpm. The acidic supernatant was applied to a pepstatinyl-agarose column equilibrated with 0.1 M sodium acetate, pH 3.5, 0.1% Brij 35, and 1 M NaCl, and washed until the OD_{280} of the effluent returned to baseline. The col-

umn was eluted with 50 mM Tris-HCl, pH 8.6, 0.1% Brij 35, and 1 M NaCl. The eluent was dialyzed against 10 mM sodium phosphate, pH 7.0, 0.1% Brij 35, and 0.1 M NaCl, and then applied to a 1-mL DEAE-Sephacel column equilibrated with the same buffer. Cathepsin D was recovered in the breakthrough fraction.

All peptide substrates were synthesized and purified by HPLC as described previously (Konvalinka et al., 1990; Richards et al., 1990). Concentrations of substrate solutions and products of cleavage were verified by amino acid analysis.

In chromogenic activity assays the decrease in absorbance in the range of 284–324 nm was followed on a Hewlett-Packard 8452A diode array spectrophotometer equipped with a multi-cell transport and thermostatted at 37°C . The absorbance readings taken at 0.1-s intervals every 2 nm over this range were averaged. Each cuvette was sampled five times (total time, 0.5 s), and those readings were averaged to give each time point. For each cuvette, the readings were repeated with a cycle time of 17.2 s. Data were typically collected for 1,000 s. The initial velocity was calculated from the slope during the linear phase of the reaction and plotted versus substrate concentration at the start of the reaction. These data were fitted to the standard Michaelis-Menten equation by Mar-

quardt analysis to yield calculated values of $V_{\max}$ and K_m . All K_m and $V_{\max}$ values were determined in 0.1 M sodium formate at pH 3.7. Enzyme preparations were titrated with a tightly binding inhibitor.

Modeling

A rule-based approach to comparative protein modeling (Blundell et al., 1987b; Sali et al., 1990) encoded in COMPOSER was used to model human cathepsin D. This method depends on the assembly of three sets of fragments of protein structures (Sutcliffe et al., 1987a,b): (1) regions conserved in the homologous family that usually compose the secondary structural elements such as helices and strands that make up the "framework," (2) regions outside the "framework" selected from a database of loop substructures using a distance filter in a similar way to Jones and Thirup (1986), and (3) side chains selected using rules derived from the analysis of homologous structures.

Selection of homologous structures and construction of framework

The sequences of aspartic proteinases of known three-dimensional structure were aligned together with the sequence of human cathepsin D using the multiple sequence alignment program MALIGN (Johnson, unpubl.). The pairwise sequence identities, shown in Table 6, indicate that porcine pepsin (with 50.6% sequence identity with cathepsin D; Cooper et al., 1990), bovine chymosin (47.8%; Newman et al., 1991), yeast proteinase A (45.6%; Carlos et al., 1992, unpubl.), and mouse renin and human renin (50.2% and 46.2%, respectively; Dhanaraj et al., 1992) are candidates for construction of the framework. However, at the time of modeling, the renin structures were incompletely refined, and the structure of yeast proteinase A was at low resolution (~ 3.3 Å). Therefore, only the three-dimensional structures of porcine pepsin and bovine chymosin were used to model the framework of human cathepsin D. These were superposed as rigid bodies by first specifying the active site residues Asp 32–Gly 34 and Asp 215–Gly 217 as topologically equivalent. An iterative reweighted least-squares procedure was then used to define the topologically equivalent residues in the two proteins using a distance cut-off value of 3 Å. This procedure determined the SCRs or "framework," which comprised a series of strands and/or helices with occasional connecting turns that are conserved in both pepsin and chymosin.

The sequence of human cathepsin D was then aligned to a template representing the sequences of the known structures in the framework region. Fragments from pepsin and chymosin with the closest similarity to the model were least-squares fitted to the framework in order to give the model real protein geometry. The SCRs in the model and the proteins from which they were selected are shown in Table 7.

Selection of the SVRs

In order to model the SVRs, protein fragments, which are of the same length as the loop to be modeled and which are compatible with the geometry of a three-residue overlap in each adjacent SCR, were first identified by searching a local distance matrix database of all structures including members of the same family. Several types of filters were then successively applied in the selection of the most appropriate loop fragment to be annealed to the model core. A novel ring closure algorithm was used in which the main-chain torsion angles of the loop are tweaked, causing minimal disturbance to the structurally conserved core (Eisenmenger, unpubl.). The most appropriate fragment is generally the one that (1) has the smallest rms differences in the C_α positions in the loop during ring closure, (2) has maximum sequence identity to the fragment that is being modeled, (3) has smallest changes in the torsion angles that accompany the annealing of the fragment on to the core, (4) has no steric clashes with rest of the protein, and (5) satisfies constraints of disulfides where applicable. Fragments were selected from a non-homologous family only when an appropriate fragment within the homologous family was not available. The SVRs selected in the model are shown in Table 4.

The original model was constructed using fragments from models of human and mouse renins previously constructed on the basis of other aspartic proteinases (Blundell et al., 1983; Sibanda et al., 1984; Frazao & Topham, unpubl.), as the coordinates of mouse and human renins were not available at that time. However, the close sequence similarity of cathepsin D and renin in the regions

Table 7. Selection of fragments for structurally conserved regions (SCRs) in the model

SCR no.	SCR ^a	Length of SCR	Protein (PDB code) used in "framework"
1	Ile 2–Lys 7	6	Porcine pepsin (5PEP)
2	Tyr 15–Gly 21	7	Porcine pepsin (5PEP)
3	Gln 25–Cys 45	21	Porcine pepsin (5PEP)
4	Cys 50–Val 65	16	Porcine pepsin (5PEP)
5	Thr 69–Tyr 75	7	Bovine chymosin (4CMS)
6	Ser 81–Val 91	11	Porcine pepsin (5PEP)
7	Lys 95–Ser 129	35	Porcine pepsin (5PEP)
8	Leu 137–Leu 155	22	Porcine pepsin (5PEP)
9	Glu 164–Lys 173	10	Porcine pepsin (5PEP)
10	Leu 179–Val 184	6	Porcine pepsin (5PEP)
11	Arg 186–Val 199	14	Porcine pepsin (5PEP)
12	Gly 202–Gly 223	22	Porcine pepsin (5PEP)
13	Val 228–Val 238	11	Bovine chymosin (4CMS)
14	Ile 247–Thr 254	8	Bovine chymosin (4CMS)
15	Ile 258–Tyr 274	17	Porcine pepsin (5PEP)
16	Ser 284–Gly 288	5	Porcine pepsin (5PEP)
17	Ile 300–Ala 326	27	Porcine pepsin (5PEP)

^a Residues are numbered according to porcine pepsin (PDB [Bernstein et al., 1977; Abola et al., 1987] code 5PEP).

Lys 46–Ala 49 and Met 289–Trp 299 and the unusual conformation, particularly of the proline-rich region, encouraged us to remodel these on the basis of renin (Dhanaraj et al., 1992) when the coordinates were refined. The region Thr 275–Leu 283 (SVR 16 in Table 4) was modeled from endothiapepsin (Blundell et al., 1990), as the corresponding region in the renins was poorly defined and there is no loop of equivalent length in either pepsin or chymosin.

Side chain modeling

The side chains were added using a set of environment-dependent rules (Sutcliffe et al., 1987b).

Refinement of the model

The model generated was refined using (Hermans & McQueen, 1974) options in FRODO (Jones, 1978) to regularize the geometry over the entire molecule, particularly in regions where SVRs were connected to the SCRs, where longer side chains have replaced shorter side chains, and where insertions/deletions have been interactively modeled on graphics. The model was then refined using energy minimization techniques to relieve steric clashes arising between nonbonded interactions using SYBYL software (supplied by TRIPOS). Over 200 cycles of refinement were carried out. The coordinates of the model have been deposited in the Brookhaven Protein Data Bank (Bernstein et al., 1977; Abola et al., 1987).

Inhibitor/substrate modeling

In order to identify the residues that contribute to individual specificity subsites, we constructed three models of human cathepsin D complexed with ligands. The three-dimensional structures of mouse renin complexed with a decapeptide inhibitor CH-66 (Dhanaraj et al., 1992; see Table 5) and endothiapepsin complexed with the pepsin inhibitor H-256 (Cooper et al., 1987; see Table 5) were used as guides to model the respective inhibitors into the active site of human cathepsin D. CH-66 was chosen because it is the longest inhibitor defined by X-ray analysis (as a complex with mouse renin) and was valuable in defining sites S_6 to S_4' . H256 was chosen because it contains a Pro at P_4 , Glu at P_2 , Phe at P_1 and P_1' , and Arg at P_2' as in the substrate used. The inhibitor modeling was achieved by superposing the enzyme–inhibitor complexes onto the model of cathepsin D and transferring the coordinates of the inhibitor to the model of cathepsin D. Only small adjustments were required to optimize the fit, and these were achieved using FRODO and then energy minimization. These models of the inhibitor complexes of cathepsin D were then used to model the sequence of the substrate Lys-Pro-Ile-Glu-Phe*Nph-Arg-Leu into cathepsin D with a $-C(OH)_2-NH_2$ group at the position of the scissile bond to represent an intermediate state. Specificity pockets for the models complexed with CH-66, H256, and the substrate are listed in Table 5.

Computations

All computations were performed on either a VAX 11/750 or a micro VAX II. Interactive modeling was performed using the graphics package FRODO on an Evans and Sutherland PS390 system. Energy minimization was carried out on a Silicon Graphics personal IRIS workstation.

Acknowledgments

We express our gratitude to Ruth Davenport and Joseph Geber for peptide synthesis and HPLC analysis of cleavage products, and to Benne Parten, Rachelle Wright, Jennifer Pritchard, and Charles Provenzano for substrate preparation and amino acid analysis. Also we thank Dr. Mark Johnson, Dr. John Overington, Dr. Andrej Sali, Dr. Gordon Louie, and Zhanyang Zhu for their help and advice. The inhibitor used in enzyme titrations was kindly provided by Dr. Tomi K. Sawyer. This work was supported by NIH grant DK18865 to B.M.D., grants from Parke-Davis and the Imperial Cancer Research Fund to T.L.B., and NIH grant GM35812 and F91UM-1 from the American Cancer Society to G.E.C.

References

- Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F., & Weng, J. (1987). Protein Data Bank. In *Crystallographic Databases—Information Content, Software Systems, Scientific Applications*, (Allen, F.H., Bergerhoff, G., & Sievers, R., Eds.), pp. 107–132. Data Commission of the International Union of Crystallography, Bonn.
- Aikawa, J., Yamashita, T., Nishiyama, M., Horinouchi, S., & Beppu, T. (1990). Effects of glycosylation on the secretion and enzyme activity of *Mucor* rennin, and aspartic proteinase of *Mucor pusillus*, produced by recombinant yeast. *J. Biol. Chem.* 265, 13955–13959.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Jr., Rodgers, J.R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112, 535–542.
- Blundell, T.L., Cooper, J., & Foundling, S.I. (1987a). On the rational design of renin inhibitors: X-ray studies of aspartic proteinases complexed with transition-state analogues. *Biochemistry* 26, 5585–5590.
- Blundell, T.L., Cooper, J.B., Sali, A., & Zhu, Z.-Y. (1991). Comparisons of the sequences, 3-D structures and mechanisms of pepsin-like and retroviral aspartic proteinases. In *Structure and Function of the Aspartic Proteinases: Genetics, Structures and Mechanisms* (Dunn, B.M., Ed.), pp. 443–453. Plenum Press, New York.
- Blundell, T.L., Jenkins, J.A., Sewell, B.T., Pearl, L.H., Cooper, J.B., Tickle, I.J., Veerapandian, B., & Wood, S.P. (1990). X-ray analyses of aspartic proteinases. The three-dimensional structure at 2.1 Å resolution of endothiapepsin. *J. Mol. Biol.* 211, 919–941.
- Blundell, T.L., Sibanda, B.L., & Pearl, L.H. (1983). Three-dimensional structure, specificity and catalytic mechanism of renin. *Nature* 304, 273–275.
- Blundell, T.L., Sibanda, B.L., Sternberg, M.J.E., & Thornton, J.M. (1987b). Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* 326, 347–352.
- Capony, F., Rougeot, C., Montcourrier, P., Cavailles, V., Salazar, G., and Rochefort, H. (1989). Increased secretion, altered processing, and glycosylation of pro-cathepsin D in human mammary cancer cells. *Cancer Res.* 49, 3904–3909.
- Conner, G.E., & Richo, G. (1992). Isolation and characterization of a stable activation intermediate of the lysosomal aspartyl protease cathepsin D. *Biochemistry* 31, 1142–1147.
- Conner, G.E., Udey, J.A., Pinto, C., & Sola, J. (1989). Non-human cells correctly sort and process the human lysosomal enzyme cathepsin D. *Biochemistry* 28, 3530–3533.

- Cooper, J.B., Foundling, S., Hemmings, A., Blundell, T.L., Jones, D.M., Hallet, A., & Szelke, M. (1987). The structure of a synthetic pepsin inhibitor complexed with endothiapepsin. *Eur. J. Biochem.* **169**, 215–221.
- Cooper, J.B., Khan, G., Taylor, G., Tickle, I.J., & Blundell, T.L. (1990). X-ray analysis of aspartic proteinases, part II: The 3D-structure of the hexagonal crystal form of porcine pepsin at 2.3 Å resolution. *J. Mol. Biol.* **214**, 199–222.
- Dhanaraj, V., Dealwis, C.G., Frazao, C., Badasso, M., Sibanda, B.L., Tickle, I.J., Cooper, J.B., Driessen, H.P.C., Newman, M., Aguilar, C., Wood, S.P., Blundell, T.L., Hobart, P.M., Geoghegan, K.F., Ammirati, M.J., Danley, D.E., O'Connor, B.A., & Hoover, D.J. (1992). X-ray analyses of peptide-inhibitor complexes define the structural basis of specificity for human and mouse renins. *Nature* **357**, 466–472.
- Dunn, B.M., Jimenez, M., Parten, B.F., Valler, M.J., Rolph, C.E., & Kay, J.K. (1986). A systematic series of synthetic chromophoric substrates for aspartic proteinases. *Biochem. J.* **237**, 899–906.
- Dunn, B.M., Valler, M.J., Rolph, C.E., Foundling, S.I., Jimenez, M., & Kay, J. (1987). The pH dependence of the hydrolysis of chromogenic substrates of the type, Lys-Pro-Xaa-Yaa-Phe-(NO₂)Phe-Arg-Leu, by selected aspartic proteinases: Evidence for specific interactions in subsites S₃ and S₂. *Biochem. Biophys. Acta* **913**, 122–130.
- Erickson, A.H., Conner, G.E., & Blobel, G. (1987). Biosynthesis of lysosomal proteinase: Polypeptide chain. In *Lysosomes, Their Role in Protein Breakdown* (Glaumann, H. & Ballard, F.J., Eds.), pp. 115–161. Academic Press, London.
- Evans, S.V. (1992). SETOR: Hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics*, in press.
- Green, D.W., Aykent, S., Gierse, J.K., & Zupc, M.E. (1990). Substrate specificity of recombinant human renal renin: Effect of histidine in the P₂ subsite on pH dependence. *Biochemistry* **29**, 3126–3133.
- Hallpike, J.F., Adams, C.W.M., & Bayliss, O.B. (1970). Histochemistry of myelin. VIII. Proteolytic activity around multiple sclerosis plaques. *Histochem. J.* **2**, 199–208.
- Hasilik, A. (1992). The early and late processing of lysosomal enzymes: Proteolysis and compartmentation. *Experientia* **48**, 130–151.
- Hermans, J., Jr. & McQueen, J.E., Jr. (1974). Computer manipulation of (macro)molecules with the method of local change. *Acta Crystallogr.* **30**, 730–739.
- Jones, T.A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272.
- Jones, T.A. & Thirup, S. (1986). Using known substructures in protein model building and crystallography. *EMBO J.* **5**, 819–822.
- Konvalinka, J., Strop, P., Velek, J., Cerna, V., Kostka, V., Phylip, L.H., Richards, A.D., Dunn, B.M., & Kay, J. (1990). Subsite preferences of the aspartic proteinase from the human immunodeficiency virus, HIV-1. *FEBS Lett.* **268**, 35–38.
- Lunney, E.A., Humblet, C.C., Repine, J.T., Blundell, T.L., Cooper, J.B., & Sibanda, B.L. (1991). Molecular modeling of renin inhibitor P₂ substituents. In *Structure and Function of the Aspartic Proteinases: Genetics, Structures and Mechanisms* (Dunn, B.M., Ed.), pp. 391–395. Plenum Press, New York.
- Newman, M., Saftro, M., Frazao, C., Khan, G., Zdanov, A., Tickle, I.J., Blundell, T.L., & Andreeva, N. (1991). X-ray analyses of aspartic proteinases IV. Structure and refinement at 2.2 Å resolution of bovine chymosin. *J. Mol. Biol.* **221**, 1295–1309.
- Overington, J., Donnelly, D., Johnson, M.S., Sali, A., & Blundell, T.L. (1992). Environment-specific amino acid substitution tables: Tertiary templates and prediction of protein folds. *Protein Sci.* **1**, 216–226.
- Overington, J.P., Johnson, M.S., Sali, A., & Blundell, T.L. (1990). Tertiary structural constraints on protein evolutionary diversity. *Proc. R. Soc. Lond. Ser. B* **241**, 132–145.
- Pohl, J. & Dunn, B.M. (1988). Secondary enzyme-substrate interactions: Kinetic evidence for ionic interactions between substrate side chains and the pepsin active site. *Biochemistry* **27**, 4827–4834.
- Richards, A.D., Phylip, L.H., Farmerie, W.G., Scarborough, P.E., Alvarez, A.H., Dunn, B.M., Hirel, Ph.-H., Konvalinka, J., Strop, P., Pavlickova, L., Kostka, V., & Kay, J. (1990). Sensitive, soluble chromogenic substrates for HIV-1 proteinase. *J. Biol. Chem.* **265**, 7733–7736.
- Sali, A., Overington, J.P., Johnson, M.S., & Blundell, T.L. (1990). From comparisons of protein sequences and structures to protein modeling and design. *Trends Biochem. Sci.* **15**, 235–240.
- Scarborough, P.E., Richo, G.R., Kay, J., Conner, G.E., & Dunn, B.M. (1991). Comparison of kinetic properties of native and recombinant human cathepsin D. In *Structure and Function of the Aspartic Proteinases: Genetics, Structures and Mechanisms* (Dunn, B.M., Ed.), pp. 343–347. Plenum Press, New York.
- Sibanda, B.L., Blundell, T.L., Hobart, P.M., Fogliano, M., Bindra, J.S., Duminy, B.W., & Chirgwin, J.M. (1984). Computer graphics modeling of human renin. *FEBS Lett.* **174**, 102–111.
- Spyratos, F., Maudelonde, T., Brouillet, J.P., Brunet, M., Defrenne, A., Andrieu, C., Hacene, K., Desplaces, A., & Rochefort, H. (1989). Cathepsin D: An important marker predicting metastasis in primary breast cancer. *Lancet* **8672**, 1115–1118.
- Sutcliffe, M.J., Haneef, I., Carney, D., & Blundell, T.L. (1987a). Knowledge-based modeling of homologous proteins. Part 1: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Eng.* **1**, 377–384.
- Sutcliffe, M.J., Hayes, F.R.F., & Blundell, T.L. (1987b). Knowledge-based modeling of homologous proteins. Part 2: Rules for the conformations of substituted side chains. *Protein Eng.* **1**, 385–392.
- Tandon, A.K., Clark, G.M., Chamness, G.C., Chirgwin, J.M., & McGuire, W.L. (1990). Cathepsin D and prognosis in breast cancer. *N. Engl. J. Med.* **322**, 297–302.
- Thorpe, S.M., Rochefort, H., Garcia, M., Freiss, G., Christensen, I.J., Khalaf, S., Paolucci, F., Pau, B., Rasmussen, B.B., & Rose, C. (1989). Association between high concentrations of 52K cathepsin D and poor prognosis in primary breast cancer. *Cancer Res.* **49**, 6008–6014.
- Whitaker, J.N. & Seyer, J.M. (1979). The sequential limited degradation of bovine myelin basic protein by bovine brain cathepsin D. *J. Biol. Chem.* **254**, 6956–6963.
- Yankner, B.A., Duffy, L.K., & Kirschner, D.A. (1990). Neurotrophic and neurotoxic effects of amyloid β protein: Reversal by tachykinin neuropeptides. *Science* **250**, 279–282.
- Zhu, Z.-Y., Sali, A., & Blundell, T.L. (1992). A variable gap penalty function and feature weights for protein 3-D structure comparisons. *Protein Eng.* **5**, 43–51.