The Design of Substrates for Cathepsin X

Gopal Devanathan

A Thesis

In

The Department

Of

Chemistry and Biochemistry

Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Science at
Concordia University
Montreal, Quebec, Canada

April 2003

Your file Votre référence
ISBN: 0-612-83852-8
Our file Notre référence
ISBN: 0-612-83852-8

ABSTRACT

The Design of Substrates for Cathepsin X

Gopal Devanathan

Cysteine proteases constitute an important group of enzymes involved in a number of physiological processes. As these enzymes have been linked to diseases such as arthritis, Alzheimer's, and cancer, they are attractive targets for inhibitor design. Cathepsin X is a cysteine protease that was only recently discovered. The primary structure of cathepsin X contains several unique features that clearly distinguish it from the other human cysteine proteases. The aim of this study is to perform a systematic study on the S2, S1, and S1' subsites of the cathepsin X active site and to gain a detailed understanding of the enzyme's substrate specificity.

Three libraries of compounds have been synthesized based on the parent compound 2-Abz-Phe-Arg-Phe($4NO_2$). In each library, the 20 natural amino acids were substituted into P2, P1, and P1' sites respectively, while keeping the other positions fixed. In reference to the parent compound, P2 is occupied by Phe, P1 by Arg, and P1' by Phe(NO_2). All 59 compounds were purified and analyzed using RP-HPLC and MS. The catalytic efficiency values for P2 and P1 variants were assessed kinetically by analyzing progress curves using a fluorescent substrate analogue while P1' variants were analyzed by discontinuous assay using RP-HPLC. Molecular modeling studies were also performed by docking 2-Abz-Phe-Arg-Phe($4NO_2$) and analogues to the cathepsin X active site in order to gain a detailed understanding of factors underlying substrate

specificity. Knowledge obtained from this study will aid in the development of specific inhibitors useful as research tools to investigate the physiological and potential pathological roles of this novel enzyme.

Acknowledgements

I would like to sincerely thank Dr. Joanne Turnbull for the invaluable help and support as on campus supervisor at Concordia University. I would like to thank Dr. Traian Sulea for all help and useful discussions regarding the modeling aspect of this research, Dr. Edmund Ziomek for aiding me in understanding the principles behind the kinetic analyses used here, especially those involving RP-HPLC. I would like to acknowledge Dr. Enrico Purisima for directing this project at the Biotechnology Research Institute/National Research Council of Canada (BRI/NRCC).

Table of Contents

Glossary of symbols and abbreviations

2-Abz	*ortho*-aminobenzoic acid
βME	β-Mercaptoethanol
Boc	*tert*-Butyloxycarbonyl
Fmoc	9-Fluorenylmethoxycarbonyl
DIPCDI	Diisopropylcarbodiimide
DIPEA	*N,N*-Diisopropylethylamine
DMAP	Dimethylaminopyridine
DMF	Dimethyl Formamide
DMSO	Dimethyl Sulfoxide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
E-64	(L-3-carboxy-2,2-*trans*-epoxypropionyl-leucyl-amino(4-guanidino)butane
MeCN	Acetonitrile
MMTS	Methyl Methanethiosulfonate
MS	Mass Spectrometry
RP-HPLC	Reversed Phase High Performance Liquid Chromatography
TBTU	2-(H-Benzatriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate
TFA	Trifluoroacetic acid
TIPS	Triisopropylsilane

List of Figures

List of Tables

Chapter 1. Introduction

Papain-related enzymes are the most numerous among the cysteine protease family. Papain is isolated from a plant product (the latex from the melon tree, *Carica papaya*).[1] Homologs are found in a wide variety of life forms spanning fungi, plants, protozoans, invertebrates, fish, and mammals.[1,2] The main role of these enzymes is protein degradation via the hydrolysis of peptide bonds in polypeptide chains (Fig. 1).

All of the mammalian lysosomal cysteine proteases are known as cathepsins.[3] Along with lysosomal protein degradation these enzymes are involved in several well-ordered processes such as antigen processing, bone resorption, and apoptosis.[1,2,3] Besides these normal physiological roles, cathepsins have been linked to pathological processes such as tumor malignancy and progression, muscular dystrophy, arthritis, and Alzheimer's disease.[2] As detection and identification techniques become increasingly sophisticated and sensitive, more cathepsins are being characterized. Currently 11 human cathepsins are known at the amino acid sequence level (B, C, F, H, K, L, O, S, V, W, and X).[1] Some cathepsins (B, C, F, H, L, O, and X) are expressed in a wide variety of cells, whereas others (S, K, V, and W) are tissue-specific.[1] Structural information is important for understanding the subtleties with which these enzymes function and for the precise design of specific inhibitors with therapeutic benefits.

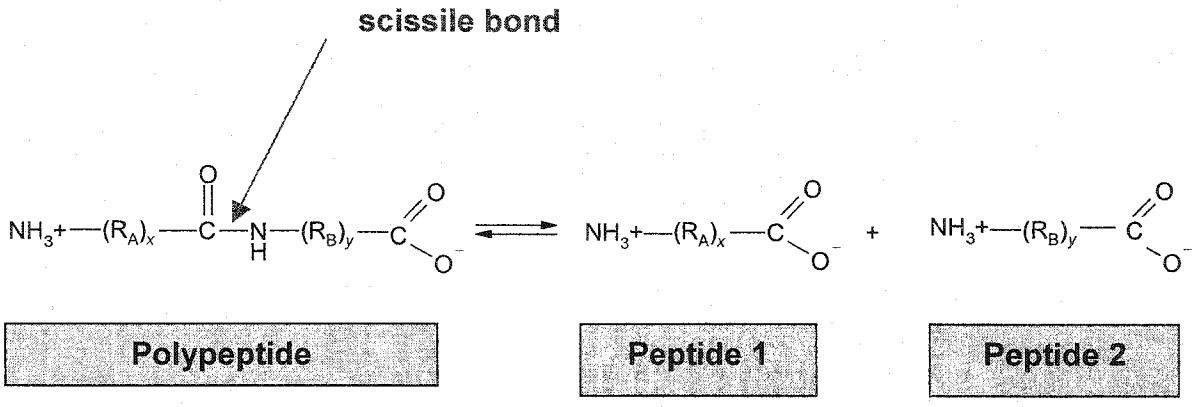


Figure 1: Proteases catalyze the hydrolysis of polypeptide chains yielding separate peptide moieties. The peptide bond that is cleaved is called the scissile bond.

1.1 Cysteine protease architecture

Papain was the first enzyme in the family of cysteine proteases to have its high-resolution structure determined by X-ray crystallographic methods.[4] The mature forms of papain-related cysteine proteases are monomeric globular proteins consisting of two domains of roughly equal size. The N-domain consists mainly of α helices, while the C-domain is made of β sheets.[4,5] The proteolytic machinery comprising the active site cysteine residue is housed in a cleft that separates the two domains. The active site of cysteine proteases is divided into a series of subsites (S3, S2, S1, S1', S2').[1,4] When a polypeptide enters the active site, amino acid side chains from the backbone of the substrate (denoted as P3, P2, P1, etc.) interact with the various subsites of the enzyme (S3, S2, S1, etc.) (Fig. 2). Site-directed mutagenesis studies[2] have shown that among the active site subsites, S2, S1, and S1' are the main determinants of substrate specificity.[1-5]

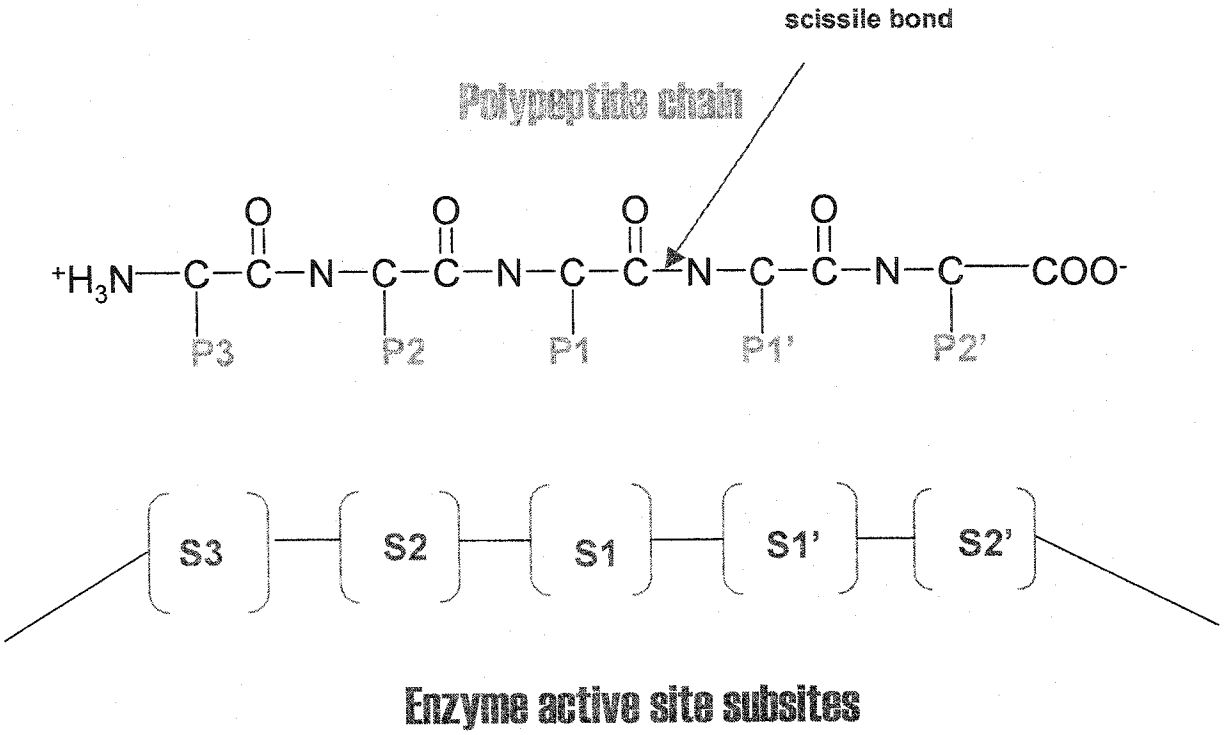


Figure 2: Annotation of active site subsites in papain-related cysteine proteases. Papain functions as an endopeptidase, cleaving in the middle of a polypeptide chain. The scissile bond of the polypeptide chain in the active site is shown.

1.2 Cysteine Protease Mechanism of Action

While peptide bond hydrolysis is well understood in the serine proteases, the same activity in cysteine proteases shares many analogous features but also relies on subtleties that are still being discovered.[6] A plausible mechanism for cysteine protease action is shown in Fig. 3. In the resting phase (pH 5.0), the active site cysteine residue and a proximal histidine residue form a thiolate-imidazolium ion pair in which the pKas of the two residues are perturbed by approximately 4 units (Cys to pH 4) and 1.5 units (His to pH 8.5).[6] The active site Cys residue is found at the N-terminus of an alpha helix. The net dipole of the helix is believed to stabilize the ion pair. The Cys and His residues participate directly in the cleavage of the scissile bond of the substrate. An asparagine residue in the active site region allows the formation of a charge relay network and completes the catalytic triad.[2,3]

Peptide hydrolysis is initialized by the nucleophilic attack of the carbonyl carbon atom of the scissile bond by the sulfur atom of Cys, resulting in the formation of a tetrahedral transition state. The negatively charged oxygen is stabilized by enzyme residues within a region called the "oxyanion hole". Concomitant with the collapse of the transition state, His acts as an acid catalyst and donates a proton to the peptide substrate. This renders it a good leaving group, resulting in cleavage of the peptide bond. A water molecule now performs the deacylation of the active site Cys residue. First, His extracts a proton from the water molecule.

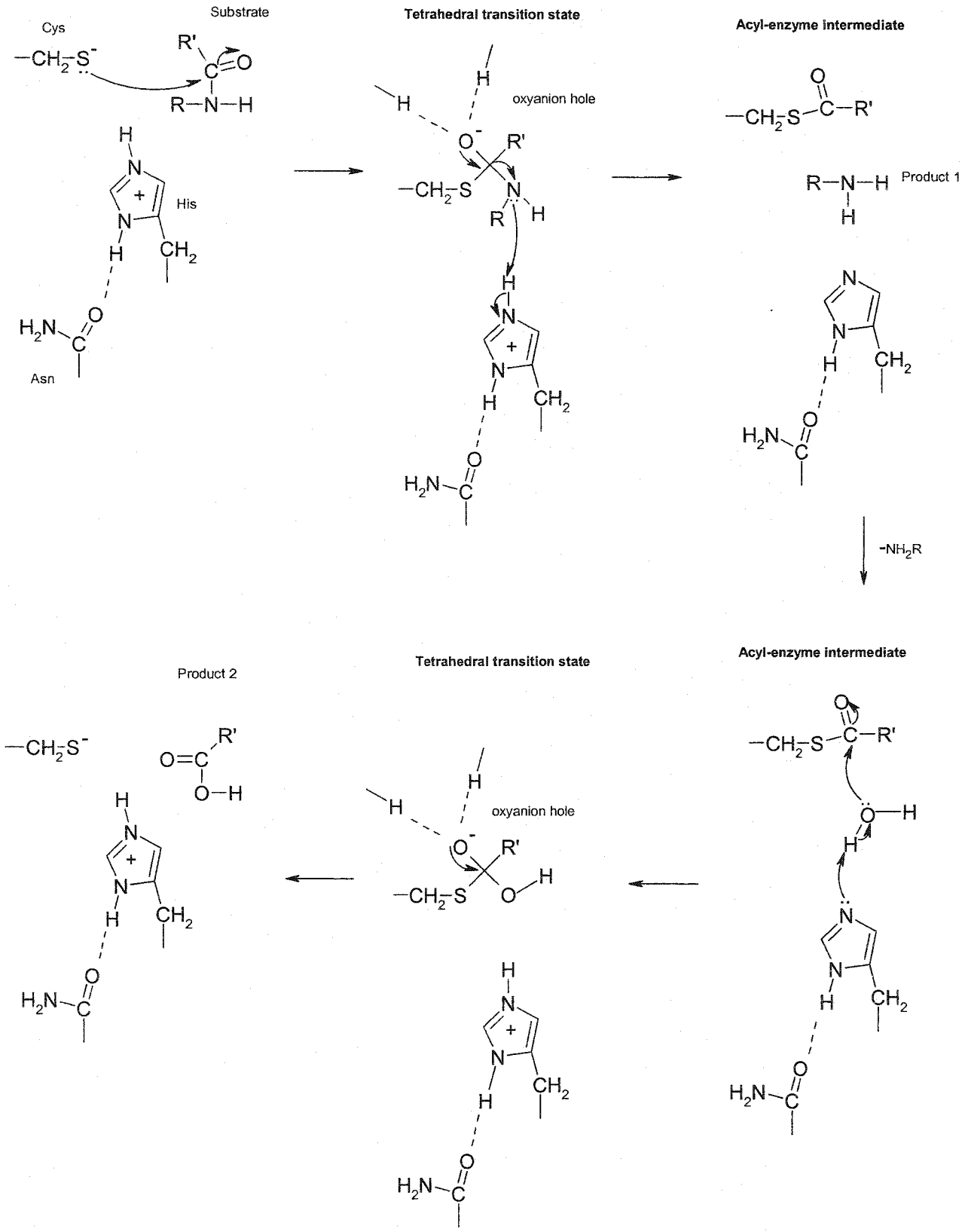


Figure 3: Reaction mechanism of peptide hydrolysis by a cysteine protease. Drawn using IsisDraw 2.4 (MDL Informations Systems, Inc.)

The resulting hydroxyl ion attacks the carbonyl carbon of the acyl-enzyme forming a tetrahedral transition state. The transition state collapses releasing the second peptide product. The enzyme's proteolytic machinery is then free to perform another round of catalysis.[2,3,6]

1.3 Cathepsin X

Cathepsin X is a novel cysteine protease that was identified from the EST (expressed sequence tags) database. The complete sequence of human cathepsin X was determined in 1998, when full-length cDNA clones were isolated independently by two groups.[7,8]

1.3.1 Structural Chemistry

The complementary DNA coding for human cathepsin X contains an open reading frame of 912 nucleotides which translates into a prepro-protein of 303 amino acid residues with a predicted M_r of 33,881.[7,8] Of these, 23 N-terminal residues correspond to the signal sequence, 38 residues to the proregion, and the remaining 242 residues to the mature enzyme. Cathepsin X clearly belongs to the papain family and all highly conserved regions in papain-like cysteine proteases,[7,8,9] including the catalytic residues, are present. Cathepsin X, like other members of the papain-related cysteine proteases, is a globular protein consisting of two domains: the N-domain which is composed of α helices and the C-domain consisting largely of β sheets.[10] The two domains are separated

by a V-shaped cleft which houses the enzyme's active site. The active site nucleophile is Cys31.[10] In the inactive precursor form (zymogen) of cathepsin X, the proregion which inhibits the enzyme's activity is very short (38 residues) compared to the proregions of all papain-like cysteine proteases which vary in length from 68 residues in procathepsin B to 251 residues in procathepsin F.[11] In particular, the proregion of cathepsin X does not share a significant degree of similarity to other cathepsins and contains no secondary structural elements. The X-ray crystallographic structures have been determined for the human procathepsin X at 1.7 Å resolution[10] (Fig. 4) and for the human mature cathepsin X at 2.67 Å resolution.[11] The procathepsin X crystal structure reveals that the short proregion is bound to the active site cysteine via a disulfide linkage (between Cys10 of the proregion and the active site Cys31), and makes fewer contacts with the mature enzyme as compared with other procathepsins. It is the first example of a zymogen in which the inhibition of enzyme's proteolytic activity is achieved through a reversible covalent modification of the active site nucleophile.[11,12]

A distinctive feature in the mature part of the enzyme is a three residue insertion[11] in a highly conserved region before the active site cysteine residue, Cys31. The consensus sequence for papain-related cysteine proteases is Gln-Gly-Xxx-Cys-Gly-Xxx-Cys (nucleophile).[10] Based on structural alignment, the corresponding sequence in cathepsin X is: Gln22-His-*Ile-Pro-Gln*-Tyr-Cys-Gly-Ser-Cys31, with the inserted residues shown in italics.[10] The insertion changes the backbone torsional angles of His23 and Tyr27, resulting in their side chains pointing directly toward the active site. The five residues (His23-Tyr27) form a loop structure, called the "mini-loop".

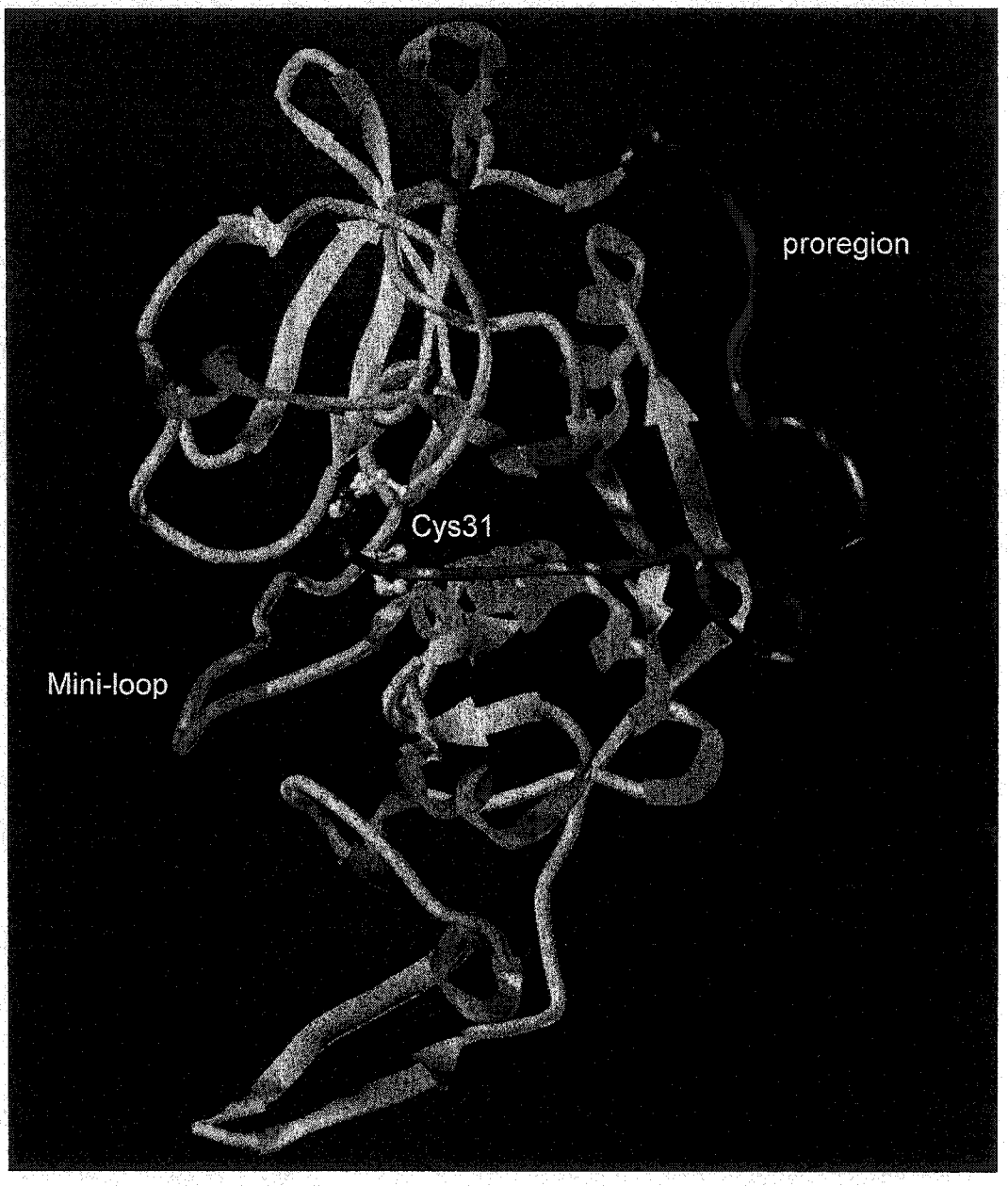


Figure 4: Ribbon diagram of procathepsin X based on the X-ray crystallographic structure determined for human procathepsin X at 1.7 Å resolution [Sivaraman J., et al. (2000) J. Mol. Biol. 295, 939-951]. α helices are in violet. β sheets are in yellow. The proregion (shown as a blue tube) renders the enzyme inactive via a disulfide bond with the sulfur atom of active site Cys31 residue (shown in yellow). The unique mini-loop structure of this enzyme is highlighted as a red tube.

1.3.2 Biological Aspects

Relatively little is known regarding the biology of human cathepsin X. The human cathepsin X gene maps to chromosome 20q13.[13] The gene comprises six exons and five introns and the cDNA contains an open reading frame of 912 nucleotides.[13] Human cathepsin X contains two potential *N*-glycosylation sites at positions Asn123 and Asn163 of the mature enzyme. The presence of a signal sequence and potential glycosylation sites indicates that the enzyme may be targeted to the endosomal/lysosomal compartment *via* the mannose 6-phosphate receptor pathway.[14]

Cathepsin X is ubiquitously expressed in human tissues, and is also highly expressed in several cancer cell lines.[7,8] Lysosomal localization suggests a role in the intracellular protein degradation system. However, a precise role in a physiological or pathological process has not been clearly established. It has been suggested that cathepsin X might be implicated in Alzheimer's disease through generation of β-amyloid peptide in cells.[14] However, the evidence is based on the use of compounds that inhibit production of β-amyloid peptide, a process believed to be due to inhibition of cathepsin X. This remains to be confirmed.

1.3.3 Specificity and Activity

Most cathepsins function as endopeptidases and are capable of cleaving a peptide bond in the middle of a polypeptide chain.[1,2,3] Cathepsin X is predominantly a carboxypeptidase (cleaving a peptide bond at the C-terminal end of a polypeptide chain) and displays very little endopeptidase activity.[7,15,16] The unique mini-loop structure seems

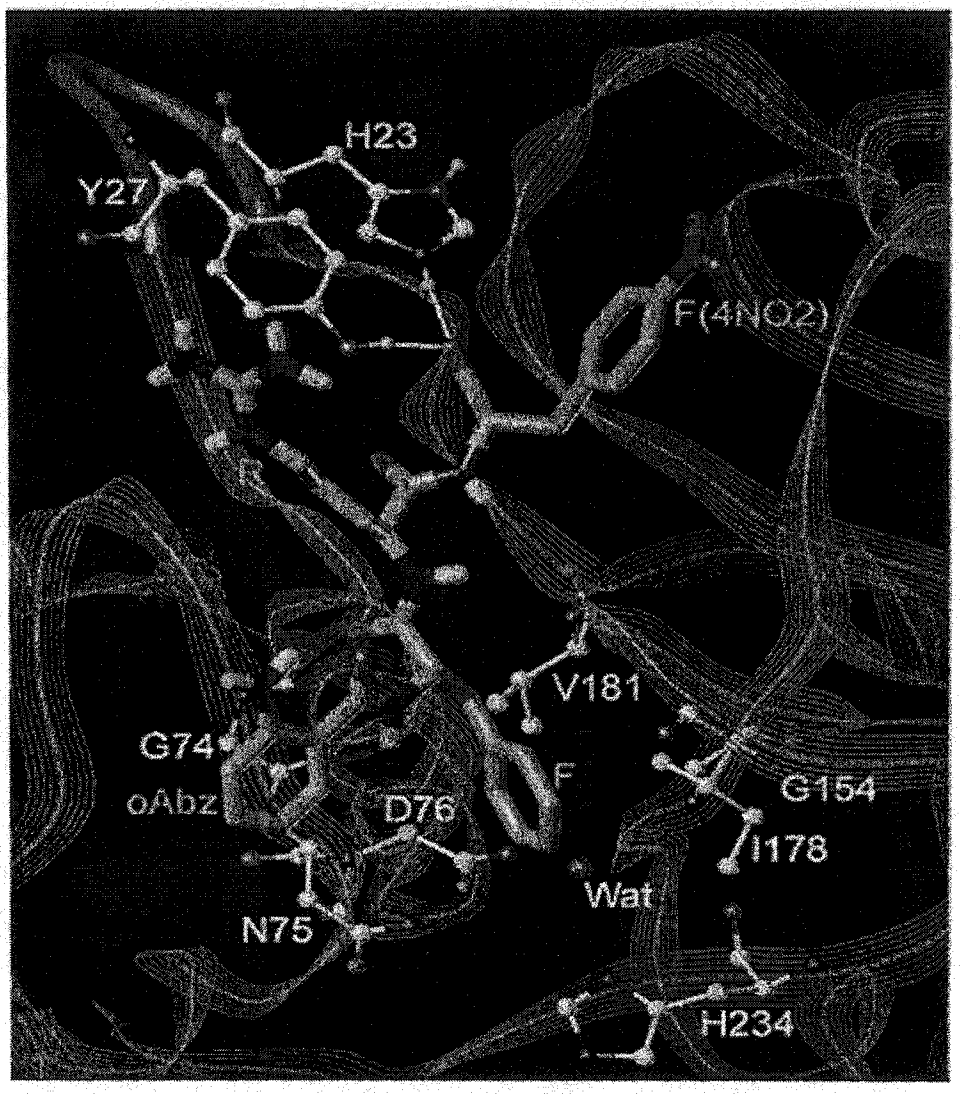


Figure 5: Model of a peptide substrate (in green) in the binding site of cathepsin X [Nagler, D, et al. (1999) Biochemistry 38,12648-12654]. Hydrogen bonds are shown in yellow. The mini-loop of the enzyme (His23-Tyr27) is highlighted as a red tube. The C-terminal carboxylate of the substrate makes hydrogen bond interactions with residues His23, and Tyr27. Note that the one-letter codes for amino acid residues are used in this figure (D=Asp, F=Phe, F($4NO_2$)=Phe($4NO_2$), G=Gly, H=His, I=Ile, N=Asn, oAbz=2-Abz, R=Arg, V=Val, Wat=water, Y=Tyr).

to be the main determinant of cathepsin X carboxy-monopeptidase activity by restricting accessibility to the S2' pocket of the binding site. Molecular docking of monopeptidyl carboxypeptidase substrates into the cathepsin X binding site predict specific hydrogen bonding interactions between the C-terminal carboxylate of the substrate and the side chains of His23 and Tyr27 mini-loop residues[7,11] (shown in Fig. 5).

Furthermore, these interactions cannot be formed by any other known cathepsin. His23 and Tyr27 occlude the S2' subsite so that binding of a peptide substrate with amino acid side chains spanning more than the S3-S1' region of the active site, is difficult to accommodate.[7,11,17] The docked substrate shown in Fig. 5, and used in previous studies is the fluorogenic substrate, 2-Abz-Phe-Arg-Phe($4NO_2$)[7]. The Abz group is a fluorophore[15] and the Phe($4NO_2$) moiety functions as a quencher.[15] In the intact tripeptide (Fig. 6), close proximity of the nitrophenylalanine decreases the fluorescence emission of the aminobenzoic acid.[18,19,20] The design of this substrate was based on the knowledge of the structures of cathepsins B, L, K, and S as well as homology modeling studies[7] since the crystal structure of cathepsin X was not determined until recently.

For cathepsins in general, a hydrophobic amino acid residue such as Phe is well accepted in the S2 subsite.[1] Arg, which possesses a fairly unbranched, positively charged side chain is well accepted in the S1 subsite.[1] The S1' subsite is the least studied of the three subsites and shows more variability within the cathepsin family.[1]

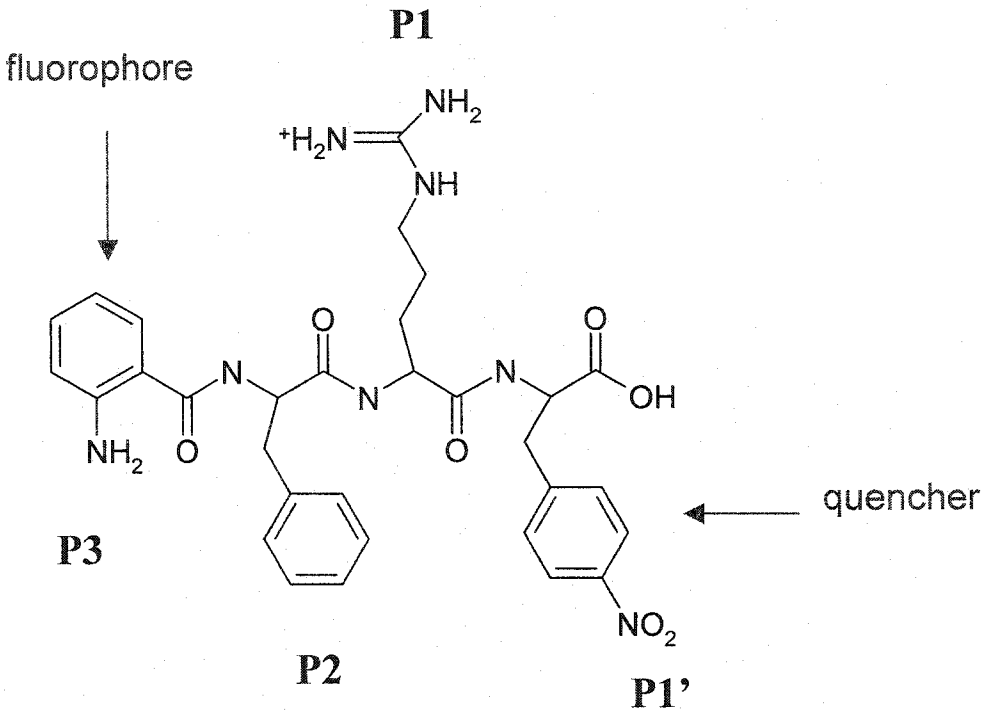


Figure 6: The fluorogenic substrate, 2-Abz-Phe-Arg-Phe(4NO_2). The Abz fluorophore is found in P3 and the Phe(4NO_2) quencher is found in P1'. Phe and Arg are found at the P2 and P1 positions, respectively, of the substrate.

1.3.4 Objectives of this Research

The purpose of this study was to gain a detailed understanding of cathepsin X substrate specificity. Probing of the S2, S1, and S1' subsites was conducted using peptide substrates (based on 2-Abz-Phe-Arg-Phe($4NO_2$)) that vary in their composition at the P2, P1, and P1' positions (shown in Fig. 7) and by determining the catalytic efficiency (k_{cat}/K_M) for each of these.

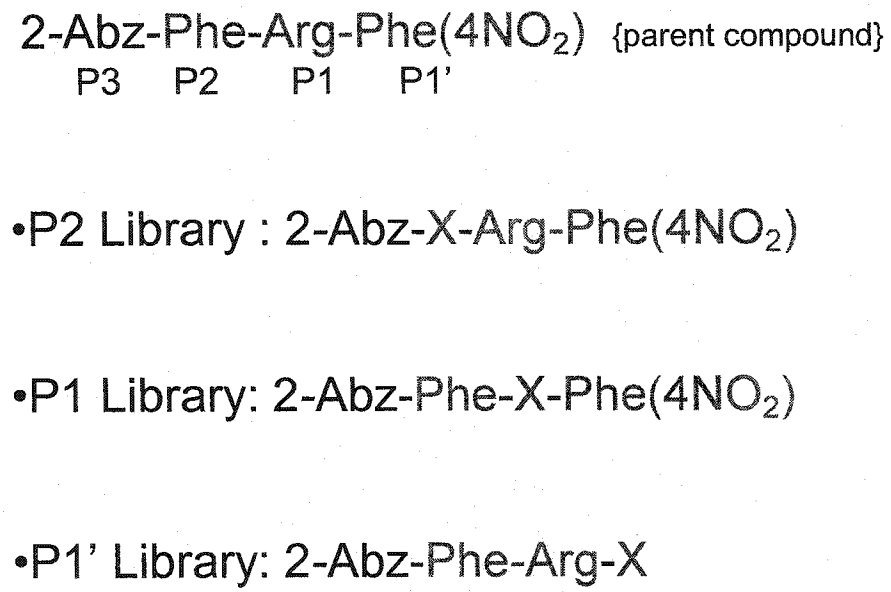


Figure 7: A summary of the three libraries of peptide substrates used in this study. X represents the 20 natural amino acids. The libraries contained a total of 59 substrates.

Chapter 2. Experimental Procedures

2.1 Materials

Fmoc-Phe(4NO_2)-OH, Wang Resin (100-200 mesh) 1.11 mmol/g, and all Fmoc protected natural amino acids pre-coupled to Wang resin were purchased from Novabiochem-Calbiochem (San Diego, Calif.). Boc-2-Abz-OH was purchased from Chem-Impex International (Wood Dale, Il). DMAP, E-64, βME, DTT, and all Fmoc protected natural amino acids were purchased from Sigma-Aldrich (St.Louis, MO). Sodium citrate, NaCl, acetic anhydride, and acetic acid were purchased from Anachemia (Rouses Point, NY). TBTU, DMF, DMSO, diethyl ether, piperidine, and EDTA were purchased from EM Science (Gibbstown, NJ). TFA, HPLC-grade MeCN, and HPLC-grade water were purchased from J. T. Baker (Philipsburg, NJ). DIPCDI was purchased from Fluka (TO, ONT). DIEA was purchased from Advanced Chemtech (Louisville KY). Purified recombinant human cathepsin X, stored at 4 °C in inhibited form by 100 μM of MMTS, was supplied by the enzymology group of Robert Menard at the Biotechnology Research Institute/NRCC (Montreal, QC).

2.2 Synthesis and Analysis of substrate libraries

A summary of the procedures used to obtain and analyze all peptide substrates used in this study is shown in Fig. 8.

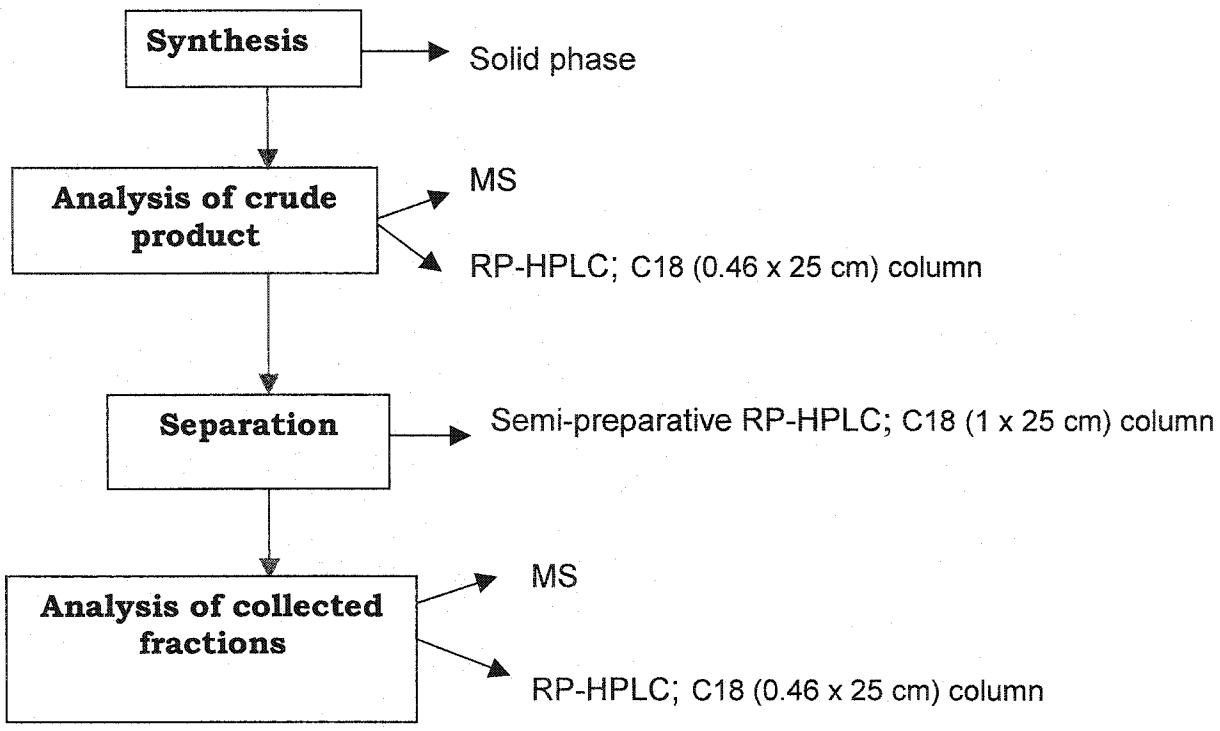


Figure 8: Scheme for synthesis and purification of peptide substrate libraries.

For the synthesis of P2 and P1 substrates, Wang resin was coupled for 24 h in a 60 mL reaction vessel with 3 equivalents of Fmoc-Phe($4NO_2$)-OH using DIPCDI (4 equiv.) as a coupling agent and DMAP (0.1 equiv.) as a catalyst, all in DMF.[19,20] The reaction mixture was drained and resin was washed 5 times with DMF. The coupling procedure was repeated. Any free hydroxy functions on the Fmoc-Phe($4NO_2$) coupled resin were blocked to prevent side reactions.[19,20] This was achieved by incubating the coupled resin for 1 h with acetic anhydride (6 equiv.) and DMAP (0.1 equiv.), all in DMF.[20,21] The reaction mixture was drained and resin was washed with DMF 10 times. Fmoc-Phe($4NO_2$) coupled resin was dried to facilitate separation into fractions to be used for the synthesis of peptide substrates.

Fmoc was removed from each fraction of Fmoc-Phe($4NO_2$) coupled resin using a 20% piperidine solution in DMF for 20 min. All peptide substrates were then synthesized by Fmoc solid-phase chemistry (Appendix I) using manual coupling of Fmoc-amino acid, 4 equiv; TBTU, 4 equiv; DIPEA, 8 equiv; all in DMF for 1 h. Completion of the reaction was verified by a ninhydrin assay (Kaiser test).[21] The N-terminus was then Fmoc deprotected, and each peptide was blocked with Boc-2-Abz by manual coupling using Boc-2-Abz-OH, 4 equiv; TBTU, 4 equiv; DIPEA, 8 equiv; all in DMF.[19,20] The peptides were cleaved from the resin by incubating the resin for 24 h with a cleavage cocktail composed of 92.5 % TFA and 2.5 % TIPS, in water.[21] Peptides were precipitated in a centrifuge tube containing cold diethyl ether, and then centrifuged. Samples were dissolved in a 10 % acetic acid solution in water and then analyzed using a triple quadropole SCIEX API III electrospray MS (PE SCIEX, Thornhill, ON) with an orifice potential of 80 V and a flow rate of 2 μL/min. Crude samples were then injected

into a Waters 600 (Waters Ltd., Mississauga, ON) RP-HPLC with a Vydac C18 (0.46 x 25 cm) analytical column using an 80-min linear gradient of 0-60 % MeCN (containing 0.1 % TFA) with a UV detector set at 215 nm and a flow rate of 1 mL/min.[22] All crude samples were subsequently separated using a Waters Delta Prep 4000 RP-HPLC with a semi preparative[22] Vydac C18 (1 x 25 cm) column using a 100-min linear gradient of 0-60 % MeCN (containing 0.1 % TFA) with a UV detector set at 215 nm and a flow rate of 13 mL/min.[22]. The molecular masses of fractions collected from semi preparative RP-HPLC were verified using electrospray MS by injecting RP-HPLC effluent into the MS (MS parameters identical to those described previously). Purity of the fractions was evaluated by RP-HPLC with a Vydac C18 (0.46 x 25 cm) analytical column using a 60-min linear gradient of 0-60% MeCN (containing 0.1 % TFA) with a UV detector set at 215 nm and a flow rate of 1 mL/min.[22] Samples were freeze-dried and obtained in solid form.

Identical procedures were used for the synthesis of P1' substrates except that commercially available Wang resin coupled with Fmoc-protected natural amino acids was used. The initial coupling step for substrates in this library was omitted.

2.3 Kinetic Analysis

2.3.1 Determining the concentration of active cathepsin X

The concentration of active cathepsin X was determined by titration with the tight binding irreversible inhibitor E-64.[23] A 2 mM stock solution of E-64 in water was obtained, along with 200 μM and 2 μM dilutions. 200 μL of cathepsin X (approximately 0.2 mg/mL) inhibited with MMTS, in sodium citrate buffer (100 mM sodium citrate, 0.4 M NaCl, 2 mM EDTA), pH 5.0, was activated by incubating the enzyme for 30 min with 1 μL of a 1 mM DTT solution in the same buffer. MMTS forms a disulfide bond with the active site Cys residue. DTT is a reducing agent that displaces MMTS by allowing the formation of a thiol group. A 100 μM stock solution of the substrate, 2-Abz-Phe-Arg-Phe($4NO_2$), in DMSO was also obtained.

From the stock solution of activated enzyme, 8 solutions were made such that each contained 20 μL of cathepsin X and increasing concentrations of E-64 (0, 0.10, 0.15, 0.20, 0.30, 0.40, 0.50, 0.60 μM). The volume of each solution was brought to 200 μL with sodium citrate buffer. Each solution containing enzyme and inhibitor was incubated for 60 min, then 100 μL aliquots were removed and placed in a 2 mL cuvette containing 2 μM 2-Abz-Phe-Arg-Phe($4NO_2$) (substrate) in the presence of sodium citrate buffer (50 mM sodium citrate, 0.2 M NaCl, 1 mM EDTA), 2 mM DTT, and 3 % DMSO. Hydrolysis of the substrate by cathepsin X while inhibited by varying E-64 concentrations was measured using a CARY spectrofluorometer with the excitation and emission wavelengths set at 320 nm and 420 nm, respectively. Excitation and emission

slit widths of 5 nm were used. All reactions were carried out at a pH of 5.0 and 25 °C. Progress curves were analysed and the initial reaction rates were assessed (as described in Results and Discussion). A plot of E-64 concentrations vs. initial reaction rates was generated in order to extrapolate the cathepsin X concentration present in the cathepsin X stock solution.[23]

2.3.2 Assays for the hydrolysis of P2 and P1 substrates

100 μM stock solutions were obtained for all thirty-nine substrates in DMSO. A calibration curve was generated in order to correlate relative fluorescence with the amount of product formed. Fluorescence intensity was measured for a series of samples containing various concentrations of both 2-Abz-Phe-Arg (product) and 2-Abz-Phe-Arg-Phe($4NO_2$) (parent substrate) in a 2 mL cuvette in the presence of sodium citrate buffer (50 mM sodium citrate, 0.2 M NaCl, 1 mM EDTA), 2 mM DTT, 3 % DMSO. The total concentration of both species in each sample was kept constant at 2 μM. Fluorescence readings were obtained using a CARY spectrofluorometer with the excitation and emission wavelengths set at 320 nm and 420 nm, respectively. Excitation and emission slit widths of 5 nm were used.

Cathepsin X, stored in inhibited form, was pre-activated by incubation with 2 mM DTT in sodium citrate buffer for 30 min.[23,24] Hydrolysis of P2 and P1 substrates was measured in a 2 mL cuvette in the presence of sodium citrate buffer (50 mM sodium citrate, 0.2 M NaCl, 1 mM EDTA), 2 mM DTT, and 3 % DMSO as an increase in

fluorescence intensity as a function of time on a CARY spectrofluorometer. (Spectrofluorometer parameters identical to those used for calibration).

All reactions were carried out at a pH of 5.0 and 25 °C at substrate concentrations of 1 μM, 2 μM, and 3 μM, that were much lower than K_M, and a cathepsin X concentration of 23 nM. Progress curves were analysed and the initial reaction rates were assessed (as described in Results and Discussion). Catalytic efficiency (k_{cat}/K_M) was calculated using GRAFIT 3.0 in units of M^{-1} s^{-1} by dividing the initial reaction rates by substrate and enzyme concentrations used in the assay.[7,16,17,18,24]

2.3.3 Assays for the hydrolysis of P1' substrates

100 μM stock solutions were obtained for all twenty substrates in DMSO. A calibration curve was generated to correlate integrated areas of RP-HPLC peaks with the amount of product formed. Various concentrations of commercially obtained 2-Abz-Phe-Arg (0.5, 2, 5, 10, 25 μM) were injected into a Waters 600 RP-HPLC with a Vydac C18 (0.46 x 25 cm) analytical column using an 80-min linear gradient of 0-60 % MeCN (containing 0.1 % TFA) and broad emission spectrum fluorescence detector. A flow rate of 1 mL/min was selected.

Cathepsin X, stored in inhibited form, was pre-activated by incubation with 2 mM βME in sodium citrate buffer for 30 min.[23,24] A progress curve was generated for 2-Abz-Phe-Arg-Phe (parent substrate) via a time-course experiment[24] by incubating 4 μM of the substrate with 23 nM of cathepsin X in a series of reaction vials containing sodium citrate buffer (50 mM sodium citrate, 0.2 M NaCl, 1 mM EDTA), 2 mM βME, and 3 % DMSO.

The reaction was quenched at various time points (0, 5, 10, 15, 20, 30, 40, 60, 180 min) using a 5 % TFA solution. Samples corresponding to each time point were injected into RP-HPLC with a Vydac C18 (0.46 x 25 cm) analytical column with RP-HPLC parameters identical to those used for calibration. The amount of product present in each was assessed (as described in Results and Discussion). The time-course progress curve was analyzed to deduce what concentrations of product were within the linear portion of the curve and allowed initial rate measurements of hydrolysis.

Next, a screening experiment was performed which involved incubating 4 μM of all P1' substrates with cathepsin X for 30 min using the same reaction conditions as described for the assay of 2-Abz-Phe-Arg-Phe. Each reaction was quenched using a 5 % TFA solution. Samples were then injected into RP-HPLC with a Vydac C18 (0.46 x 25 cm) analytical column as previously described. The amount of product formed after 30 min was assessed for each substrate. P1' substrates were then incubated with cathepsin X using incubation times ascertained from the screening experiment that allowed initial rate measurements of hydrolysis. Incubation reactions were quenched using a 5% TFA solution and injected into RP-HPLC with a Vydac C18 (0.46 x 25 cm) analytical column as previously described.

Hydrolysis of P1' substrates was measured using RP-HPLC in the presence of sodium citrate buffer (50 mM sodium citrate, 0.2 M NaCl, 1 mM EDTA), 2 mM βME, 3% DMSO. All reactions were carried out at a pH of 5.0 and 25 °C at substrate concentrations of 2 μM, 4 μM, 10 μM, 15 μM, and 20 μM, that were much lower than $K_{M,}$ and a cathepsin X concentration of 23 nM. Catalytic efficiency (k_{cat}/K_M) values were obtained as described in section 2.3.2.

The parent substrate, 2-Abz-Phe-Arg-Phe(4NO_2) was assayed using RP-HPLC for comparison with 2-Abz-Phe-Arg-Phe. Various concentrations (2, 4, 10, 15, and 20 μM) of 2-Abz-Phe-Arg-Phe(4NO_2) were incubated with 23 nM of cathepsin X for 10 min. Incubation reactions were quenched using a 5% TFA solution and injected into RP-HPLC with a Vydac C18 (0.46 x 25 cm) analytical column as previously described. Reaction conditions were identical to those of P1' assays and catalytic efficiency (k_{cat}/K_M) values were obtained as described in section 2.3.2.

2.4 Molecular Modeling

The crystal structure of human procathepsin X was used for substrate docking. The proregion segment of the enzyme and most of the water molecules were removed, and hydrogen atoms were added using SYBYL 6.6 (Tripos, Inc., St. Louis, MO). Crystallographic water molecules buried in the mature enzyme were retained. The His residues were protonated and the catalytic Cys was modeled in the thiolate form. The monopeptidyl carboxypeptidase substrate 2-Abz-Phe-Arg-Phe(4NO_2) was docked in the binding sites of human cathepsin X as a noncovalent Michaelis complex. Energy minimization was carried out in SYBYL 6.6 using AMBER all-atom force field[26] with the Powel minimizer,[26] a 4*R* distance dependent dielectric constant and an 8Å non-bonded cutoff up to an rms gradient of 0.01 kcal/mol·Å.[26,27] Refinement of substrate docking was carried out by a conformational search using a Monte Carlo with energy minimization procedure.[26,27] Starting conformations for minimization were generated by randomly perturbing one or more dihedral angles of the substrate side chains as well as crank-shaft

rotation of the substrate peptide units.[26,27] A total of 1000 Monte Carlo steps were carried out, each followed by an energy minimization in which the substrate and the protein were allowed to relax.[27] The enzyme-substrate complex was used for the docking of all substrates from the three libraries. This involved making mutations at the appropriate positions (P2, P1, and P1') and performing energy minimizations, this time solely around the modified residue and then allowing the protein to relax. Depending on the number of torsional angles present in the substituted amino acid side chains a total of 200, 500 or 1000 MCM cycles was carried out for each substrates.[27]

Chapter 3. Results and Discussion

3.1 Overview of Synthesis of P2, P1, and P1' Substrate libraries

The synthesis of the P2, P1, and P1' substrate libraries required Wang resin.[20,21,24] Wang resin was selected because it allows for the synthesis of peptide acids, with a free C-terminus.[21] The goal of this study was to assess cathepsin X carboxypeptidase activity, hence it was necessary to obtain the peptides with an unblocked C-terminal carboxyl group. Rink resins, which are commonly used in solid phase synthesis yield peptides that are blocked by an amide functional group (-$CONH_2$) at their C-terminus.[20,21]

3.2 Synthesis of P2 substrate library: 2-Abz-X-Arg-Phe($4NO_2$)

The synthesis of the P2 library was performed first. The initial coupling of the Fmoc-Phe($4NO_2$) to the resin (Appendix I) was time-consuming and as a result, was the rate limiting step of the synthesis.[20,21] Therefore, large amounts of Wang resin and an excess of Fmoc-Phe($4NO_2$) were used as starting materials, so that a sufficient amount of Fmoc-Phe($4NO_2$)-resin could be made for the synthesis of all P2 variant substrates. Each of the substrates in this library required Arg in the P1 position. Hence, Fmoc[22] was removed from Fmoc-Phe($4NO_2$)-resin, and sufficient Fmoc-Arg was coupled to it resulting in a "stock" of Fmoc-Arg-Phe($4NO_2$)-resin. The stock was then dried and divided into twenty fractions. Fmoc was cleaved from each of the fractions so that each of the twenty natural amino acids could be coupled to Arg as the P2 variant. The final

step involved blocking the N-terminus of each peptide substrate with Boc-2-Abz. Both Boc and Fmoc are protecting groups.[20] However, a Boc protected N-terminal blocking acid was selected because it can be removed using TFA which is also used for cleavage of peptide from the resin.[20] Fmoc must be removed using piperidine, after which requires extraction of the peptide from the piperidine solution using an organic solvent such as ethyl acetate.[19] The organic solvent is removed by rotoevaporation. The peptide is reconstituted using a 5-10% acetic acid solution in water and then freeze-dried to be obtained in solid form.[20,21] This multi-step procedure is laborious and can result in a greater chance of losing product. The use of Boc simplified the procedure.

3.3 Synthesis of P1 substrate library: 2-Abz-Phe-X-Phe($4NO_2$)

P1 substrates vary at the P1 position (where Arg is found in the parent compound). Hence, the synthetic procedure was modified slightly from that used to generate the P2 library. Instead of using a stock of Fmoc-Arg-Phe($4NO_2$)-resin, a stock of Fmoc-Phe($4NO_2$)-resin was synthesized and divided into nineteen fractions. One less peptide appears in this library because Arg already appears in the P1 position of the parent compound. Each of the nineteen fractions of Fmoc-Phe($4NO_2$)-resin had substitutions incorporated at the P1 position using Fmoc protected natural amino acids. Fmoc-Phe was then coupled to each P1 variant residue. Fmoc removal, N-terminal blockage, and cleavage of the peptides from resin were performed as described for the synthesis of the P2 library.

3.4 Synthesis of P1' substrate library: 2-Abz-Phe-Arg-X

In this series of substrates, the P1' position was occupied by each of the 20 natural amino acids so the coupling of Phe($4NO_2$) was omitted. However, since coupling each of the 20 amino acids individually with Wang resin separately would still prove labor-intensive, the synthesis was facilitated by purchasing Wang resin already coupled with each natural amino acid (blocked with Fmoc). After Fmoc removal from each P1' residue (variant), both Fmoc-Arg and Fmoc-Phe were used in consecutive coupling reactions resulting in Arg and Phe at the P1 and P2 positions, respectively. Fmoc removal, N-terminal blockage, and cleavage of the peptides from resin were performed as described for the synthesis of both P2 and P1 libraries.

3.5 Analysis of P2, P1, and P1' substrates

Yields for all fifty-nine purified substrates ranged from 60-100 mg. The RP-HPLC chromatograms and MS profiles for the parent compound 2-Abz-Phe-Arg-Phe($4NO_2$) (crude and RP-HPLC purified) are shown in Figs. 9a-e. They are representative of the data obtained and analyzed for all compounds. It is important to note that when using analytical RP-HPLC (Fig. 9e) to evaluate the purity of all fractions collected from the semi-preparative RP-HPLC, a steeper gradient of MeCN was used than when separating all crude samples (Fig. 9b).

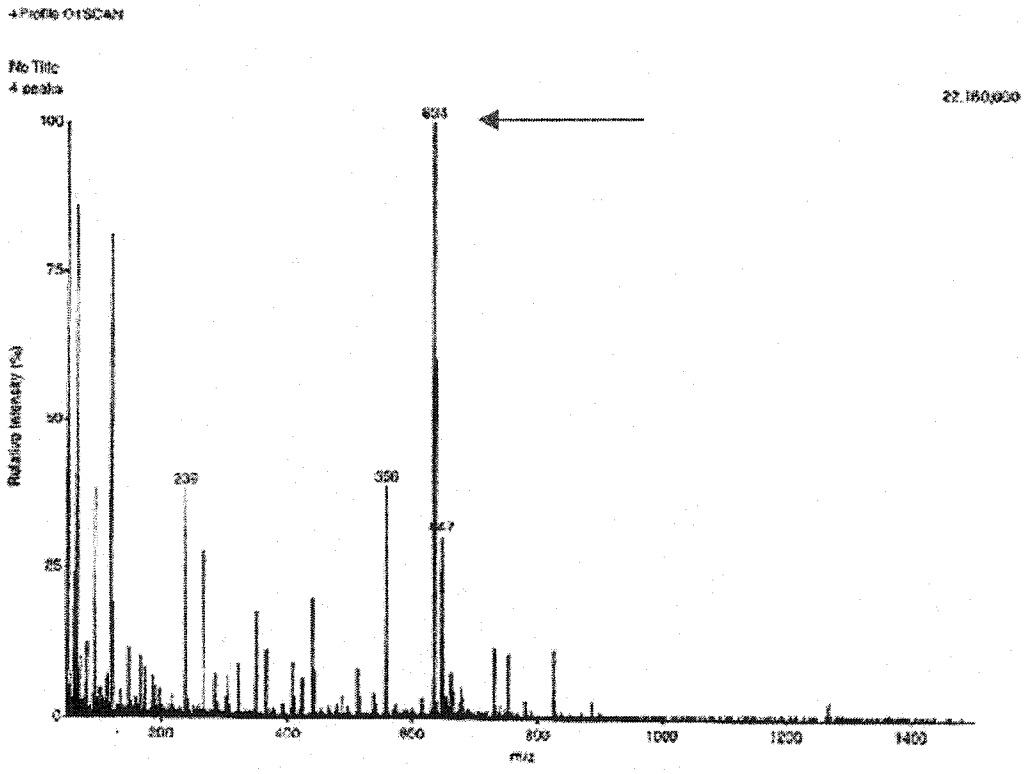


Figure 9a: MS profile of crude 2-Abz-Phe-Arg-Phe($4NO_2$). The arrow identifies the peak corresponding to the substrate. The m/z ratio of the substrate peak is shown. See section 2.2 for details.

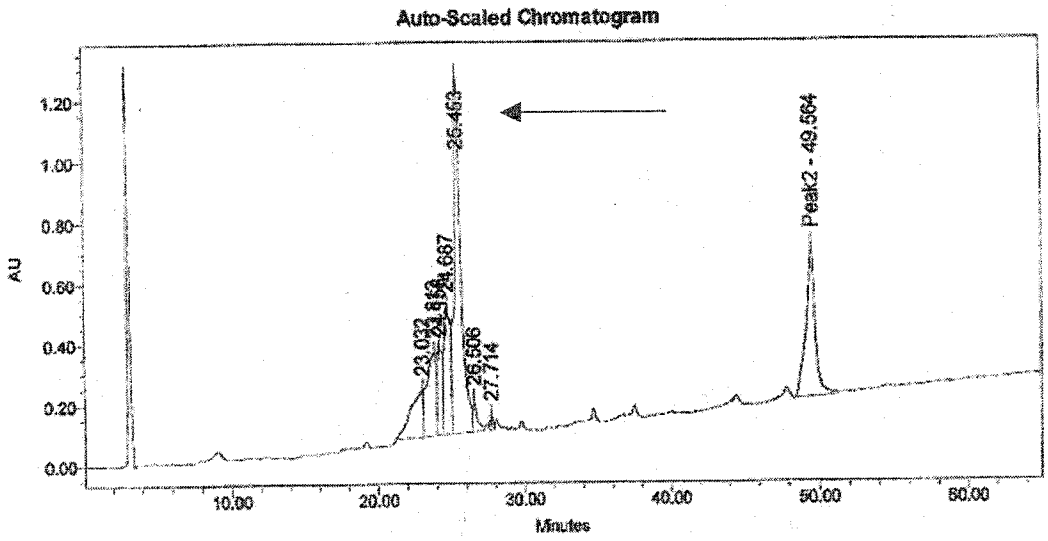


Figure 9b: RP-HPLC profile of crude 2-Abz-Phe-Arg-Phe($4NO_2$). The arrow indicates the peak corresponding to the substrate. The retention time is shown. See section 2.2 for details.

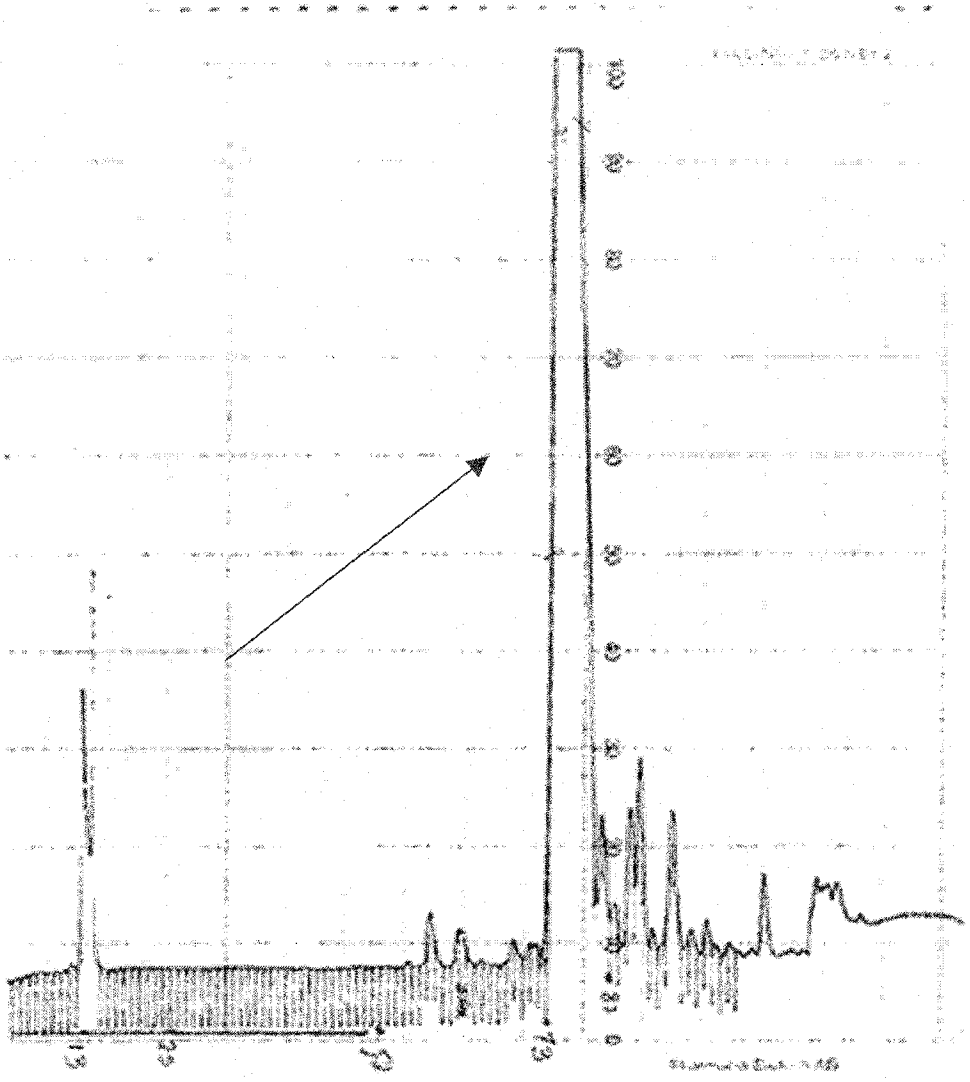


Figure 9c: RP-HPLC profile for the separation of crude 2-Abz-Phe-Arg-Phe(4NO_2). The arrow indicates the peak corresponding to the purified substrate. See section 2.2 for details.

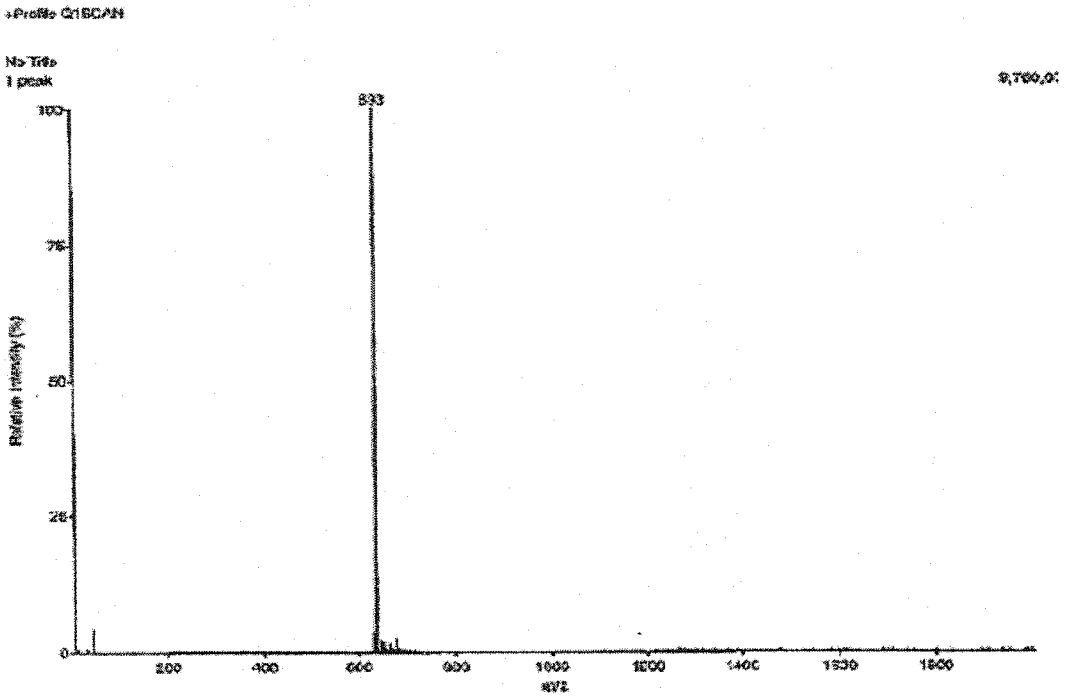


Figure 9d: MS profile of purified 2-Abz-Phe-Arg-Phe($4NO_2$). The peak corresponds to the substrate. The m/z ratio is shown. See section 2.2 for details.

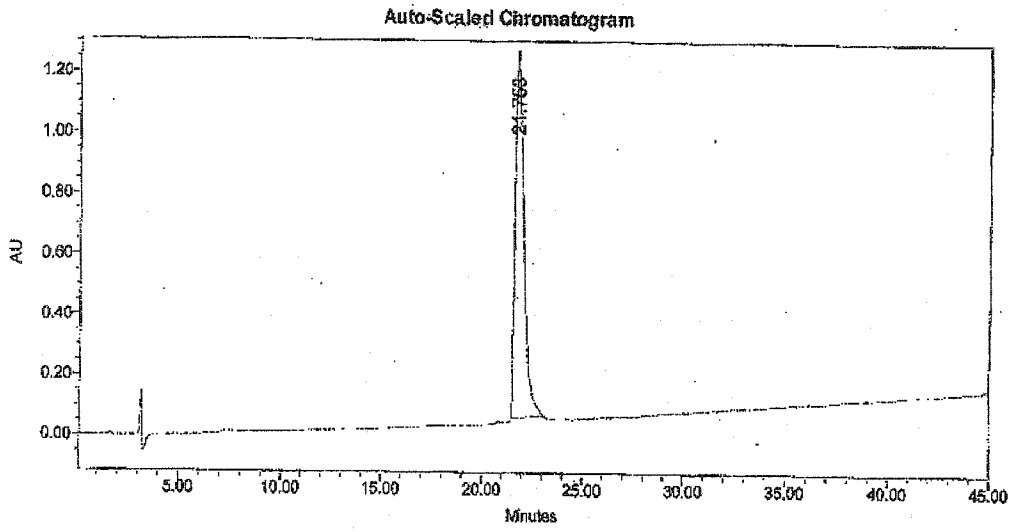


Figure 9e: RP-HPLC chromatogram of purified 2-Abz-Phe-Arg-Phe($4NO_2$). The retention time is shown. See section 2.2 for details.

Different gradients of MeCN were used when analyzing crude samples before and after separation using semi-preparative RP-HPLC. This was performed in order to save time and accounts for the variation in retention times between the peaks representing both crude and purified 2-Abz-Phe-Arg-Phe($4NO_2$) (Figs. 9b, 9e). It was not important if retention times on the chromatograms varied slightly as long as only one peak was present after purification and could be identified by MS.

3.6 Measurement of substrate hydrolysis by cathepsin X

Hydrolysis of 2-Abz-Phe-Arg-Phe($4NO_2$) and substrates within the P2 and P1 libraries by cathepsin X was monitored as a time dependent increase in Abz fluorescence.[7,17,18] Upon cleavage, the increase in Abz fluorescence could be detected using a spectrofluorometer (Fig. 10). Substrates within the P1' library, however, have the quenching agent replaced with one of 20 natural amino acids (Fig. 11). The emission of the intact substrates is already high. Upon enzymatic cleavage, there is no increase in Abz fluorescence. Hence, in this case, enzymatic hydrolysis was measured by quantifying the reaction product using RP-HPLC.[25]

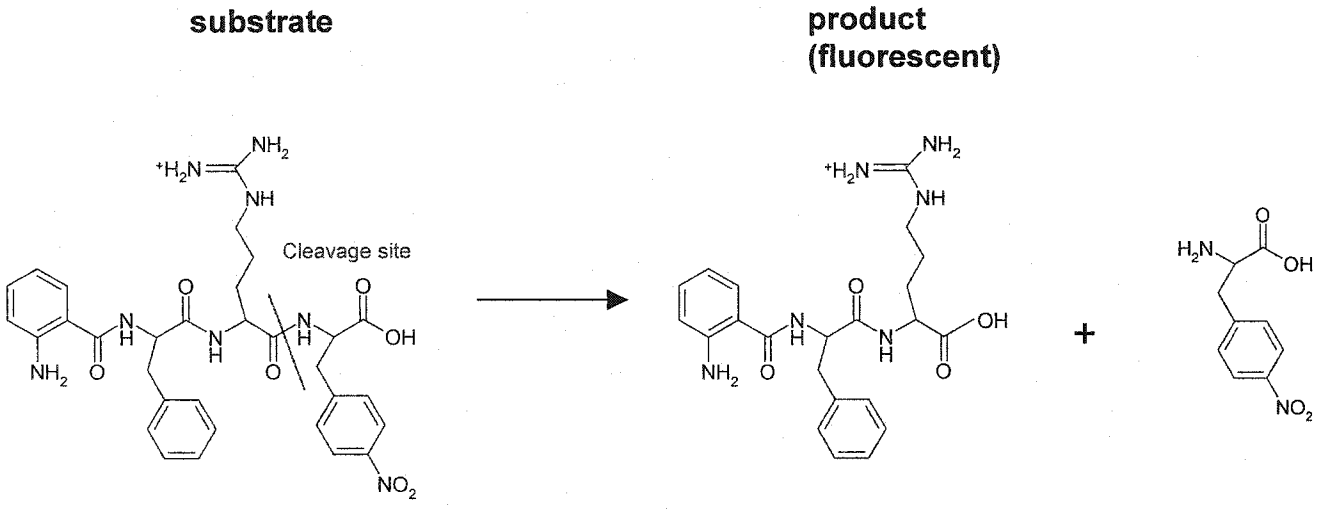
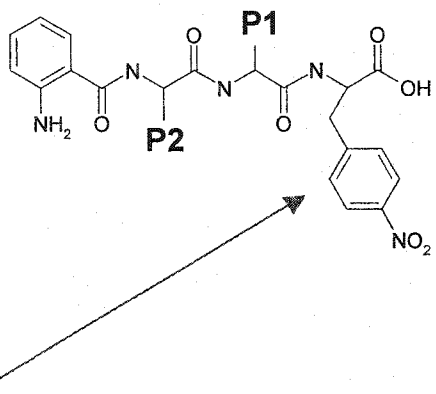
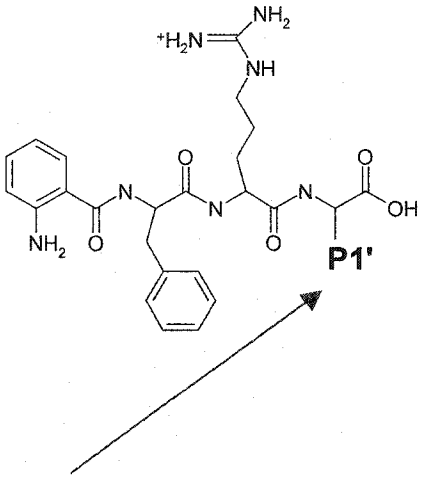


Figure 10: Basis for the fluorescence assay. After C-terminal cleavage of substrate by cathepsin X occurs, quenching by Phe($4NO_2$) decreases, resulting in an increase in Abz fluorescence



Quencher present for both P2 and P1 libraries

Spectrofluorometer analysis



Quencher absent in P1' library

RP-HPLC analysis

Figure 11: Rationale for choice of hydrolysis assay.

3.7 Determining the concentration of active cathepsin X

Before kinetic analysis of the substrate libraries, the concentration of active enzyme was determined by titrating cathepsin X with increasing amounts of the tight binding irreversible inhibitor E-64. This inhibitor combines stoichiometrically to cathepsin X.[25] The rates of enzymatic hydrolysis of the parent substrate, 2-Abz-Phe-Arg-Phe($4NO_2$), in the presence of E-64 were determined by analyzing progress curves. The reaction rates were expressed as a change of relative fluorescence per second. The enzyme's ability to hydrolyze the substrate diminished with increasing concentrations of inhibitor. The progress curve for the hydrolysis of 2 μM of 2-Abz-Phe-Arg-Phe($4NO_2$) by cathepsin X is shown in Fig. 12. A substrate concentration of 2 μM was used and was well below the K_M, which is greater than 50 μM. The reaction was terminated at approximately 10 min. Only the linear portion of the progress curve was analyzed in order to calculate the initial rate of the reaction (Fig. 13). The linear portion of the progress curve for the hydrolysis of 2 μM of 2-Abz-Phe-Arg-Phe($4NO_2$) by cathepsin X (inhibited by 0.2 μM) is overlaid in Fig. 13 for comparison. Notice that the slope of the reaction in the presence of E-64 is less than when no inhibitor is added to the reaction. Using a variety of E-64 concentrations, a plot of E-64 concentration vs. reaction rate (Fig. 14) was then generated.[23,25] The extrapolated value at the x-intercept represented the amount of E-64 required to completely inactivate the enzyme and to prevent hydrolysis of the substrate by cathepsin X.

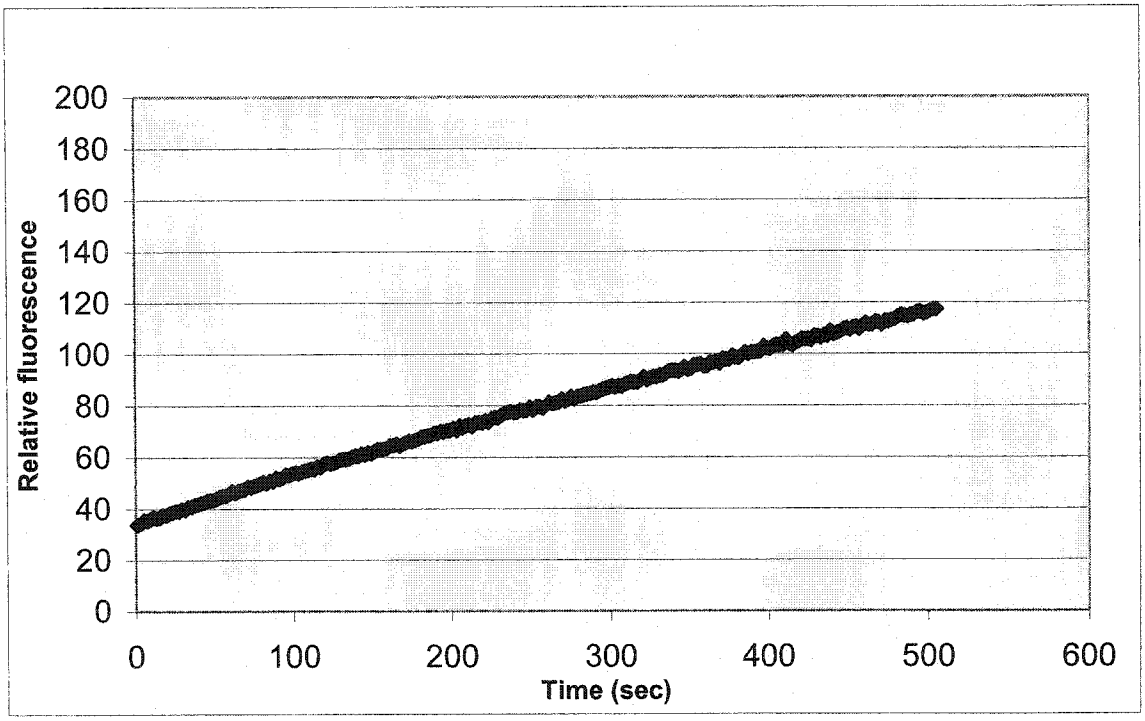


Figure 12: Progress curve of 2 μM 2-Abz-Phe-Arg-Phe(NO_2) hydrolyzed by cathepsin X. See section 2.3.1 for details.

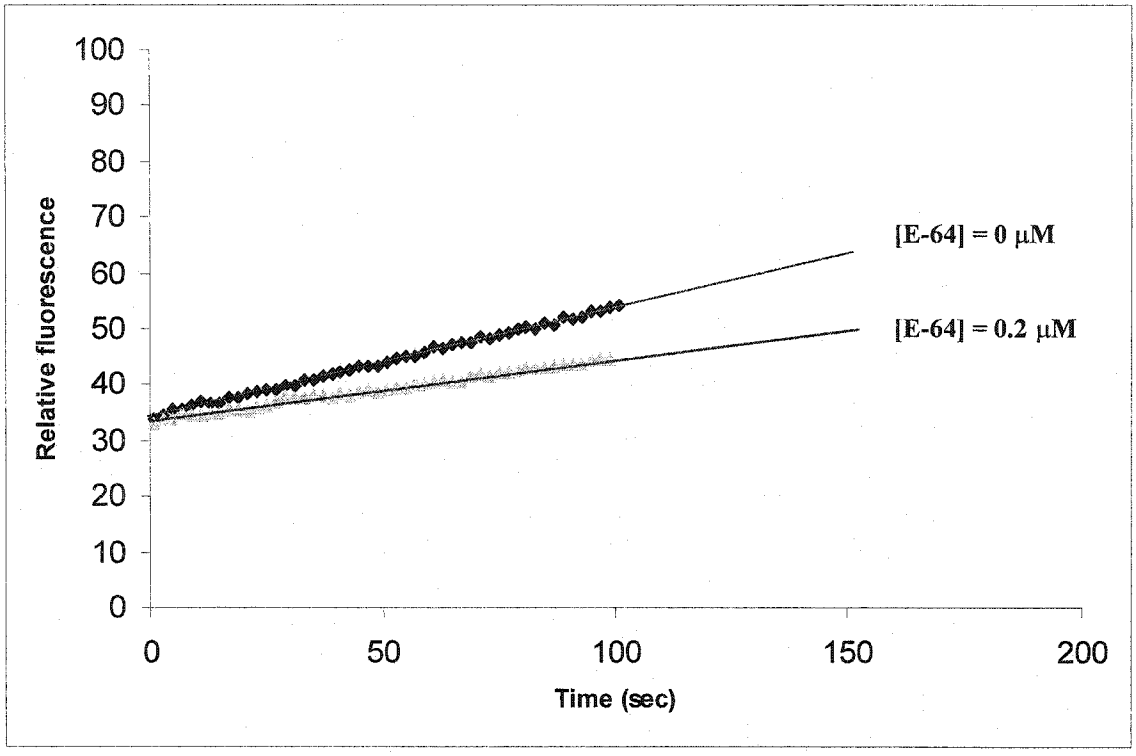


Figure 13: Linear portions of progress curves for the hydrolysis of 2 μM 2-Abz-Phe-Arg-Phe(NO_2) by cathepsin X with and without inhibitor added to the reaction mixture. Trendlines are shown in red. See section 2.3.1 for details.

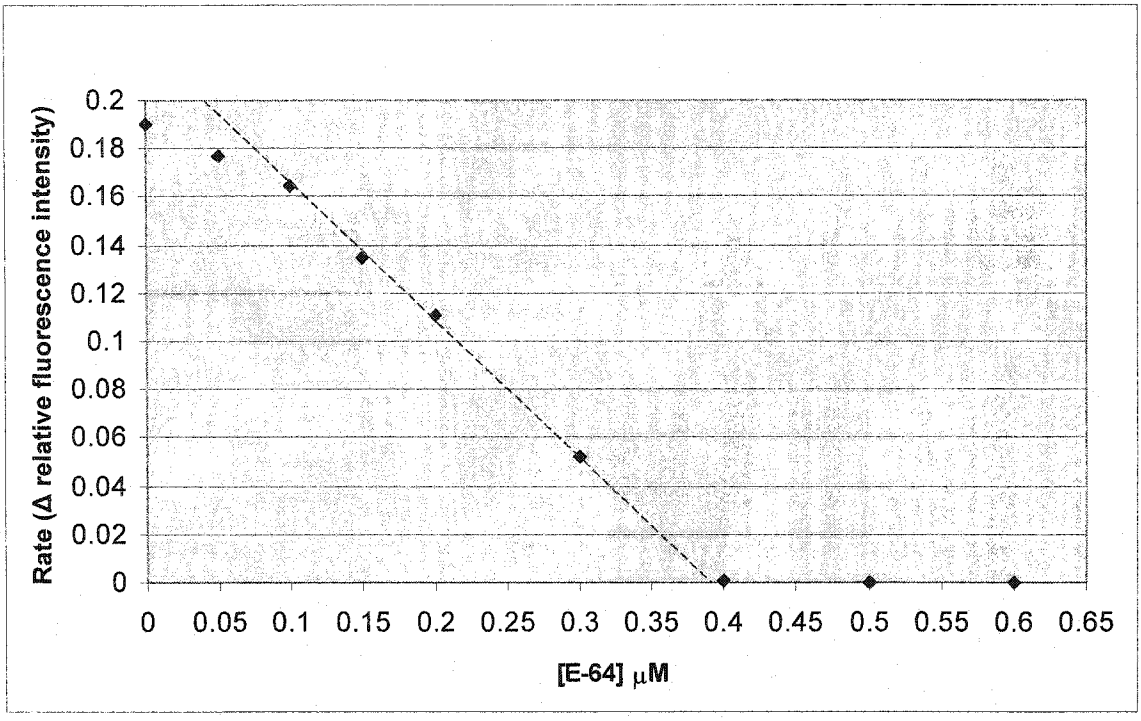


Figure 14: Plot of E-64 concentrations used to inhibit cathepsin X vs. initial rates of hydrolysis reactions. The red dashed line was extrapolated to the x-axis to yield the final concentration of cathepsin X in the reaction mixture. See section 2.3.1 for details.

This value (0.39 μM) was also equivalent to the concentration of cathepsin X in the reaction mixture. The enzyme was diluted ten fold in the reaction mixture, hence, the active concentration of cathepsin X in the stock solution was 3.9 μM.

3.8 Assays for the hydrolysis of P2 and P1 substrates

A calibration curve[26] was generated first in order to correlate the change in fluorescence with the amount of product formed (see Fig. 15 for details). Increasing concentrations of 2-Abz-Phe-Arg (product) with an appropriate amount of 2-Abz-Phe-Arg-Phe($4NO_2$) (substrate) were used such that the total concentration of product and substrate in the reaction mixture was 2 μM. A total concentration of 2 μM was used because it is well below K_M. Substrate was included in the assay for the calibration curve in order to account for possible quenching of the product by the substrate in the reaction mixture. The substrate, 2-Abz-Phe-Arg-Phe($4NO_2$), possesses intrinsic fluorescence that could mask the signal of the product, 2-Abz-Phe-Arg. The plot of product concentration vs. relative fluorescence yielded a straight line with a slope of 2.24×10^8 fluorescence units/M of product. This value was used for all enzymatic assays of P2 and P1 variants (39 substrates) in order to express reaction velocity in terms of moles of product formed per second.

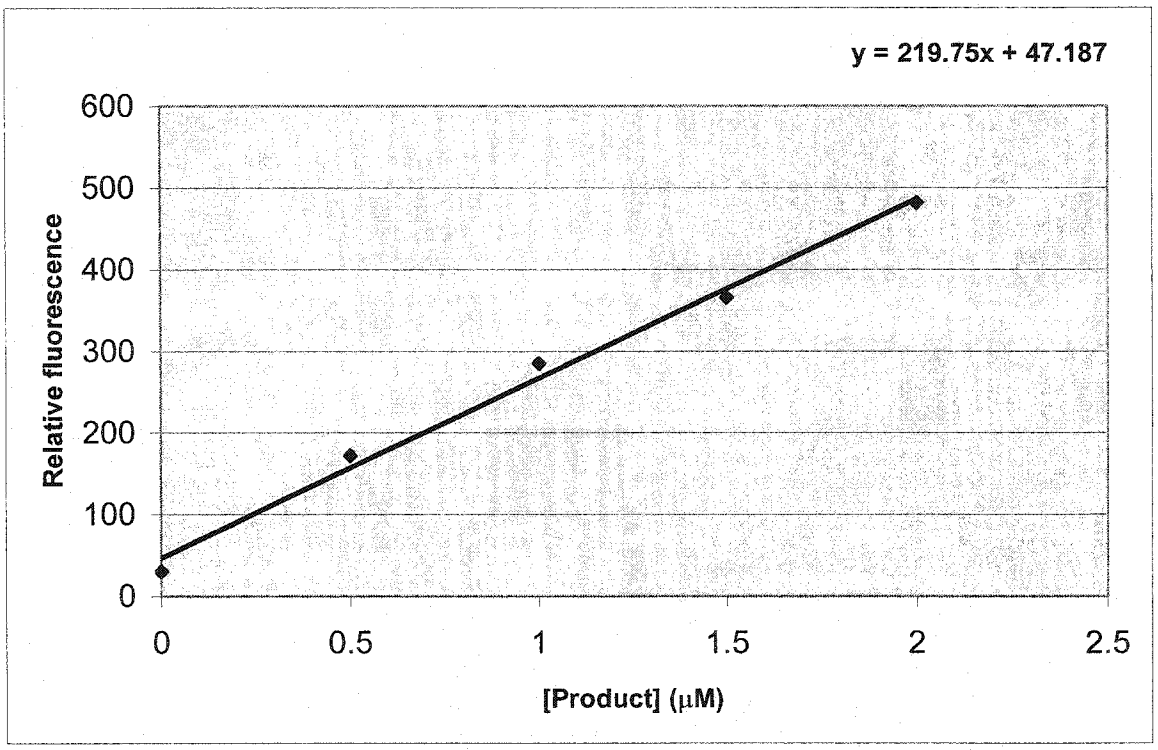


Figure 15: Calibration curve for the hydrolysis of P2 and P1 substrates (see section 2.3.2 for details). Each data point corresponds to a combination of 2-Abz-Phe-Arg (product) and 2-Abz-Phe-Arg-Phe($4NO_2$) (substrate) for a total concentration of 2 μM in each sample. (i.e. a sample containing 1.5 μM 2-Abz-Phe-Arg also contained 0.5 μM 2-Abz-Phe-Arg-Phe($4NO_2$)). The slope of this line correlates the change in relative fluorescence to the amount of product formed (2.24×10^8 fluorescence units /M of product).

The P2 and P1 substrates were incubated with cathepsin X and progress curves for each reaction were obtained. A typical progress curve for the reaction of 3 μM of 2-Abz-Phe-Arg-Phe($4NO_2$) with 23 nM of cathepsin X is shown in Fig. 16a and is representative of the data obtained for P2 and P1 substrates. A concentration of 23 nM of cathepsin X was used in all hydrolysis assays in order to remain in first order kinetics.[23,24] Only the linear portion of the curve (Fig. 16b) was analyzed to calculate the initial rate of hydrolysis.[23] The slope of the line represents the rate of the reaction as a change in relative fluorescence units per second. Using the conversion factor obtained from the calibration curve, the reaction rate was expressed as moles of product formed per second.[23] Three different substrate concentrations were used, all well below K_M in order to remain in first order kinetics. Catalytic efficiencies of P2 and P1 substrates were calculated as described in (2.3.2).

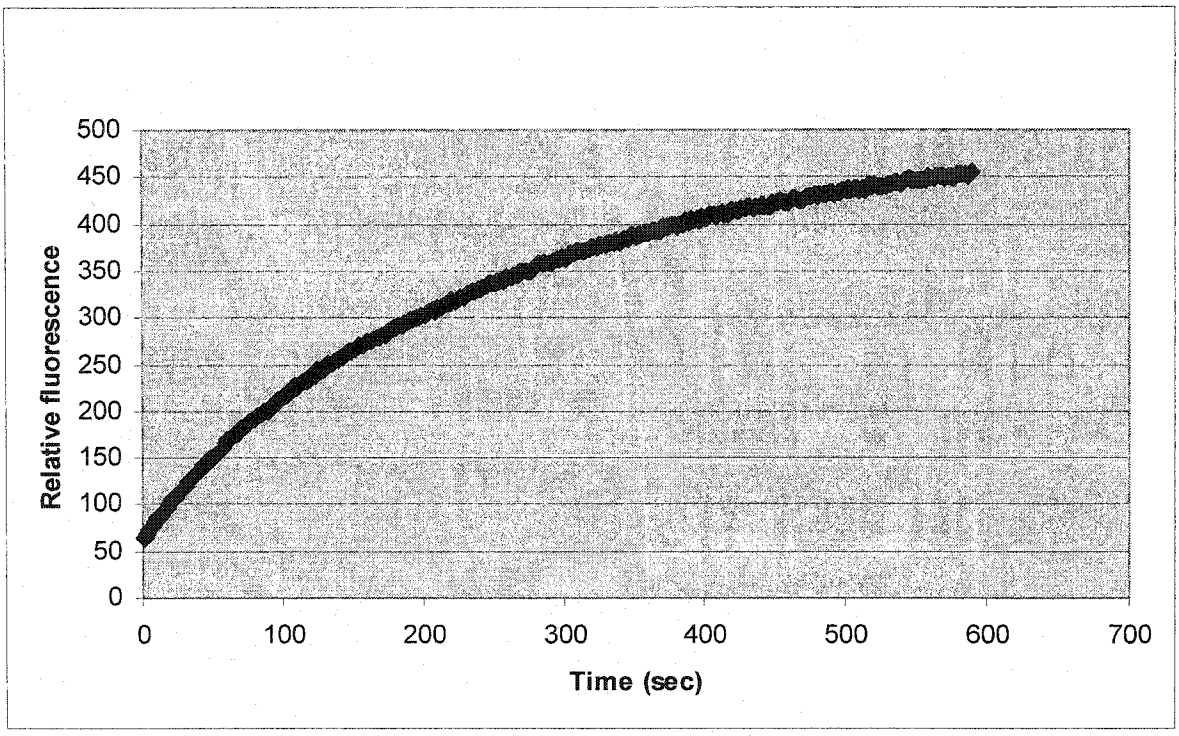


Figure 16a: Progress curve for the hydrolysis of 3 μM of 2-Abz-Phe-Arg-Phe(4NO_2) by 23 nM of cathepsin X. See section 2.3.2 for details.

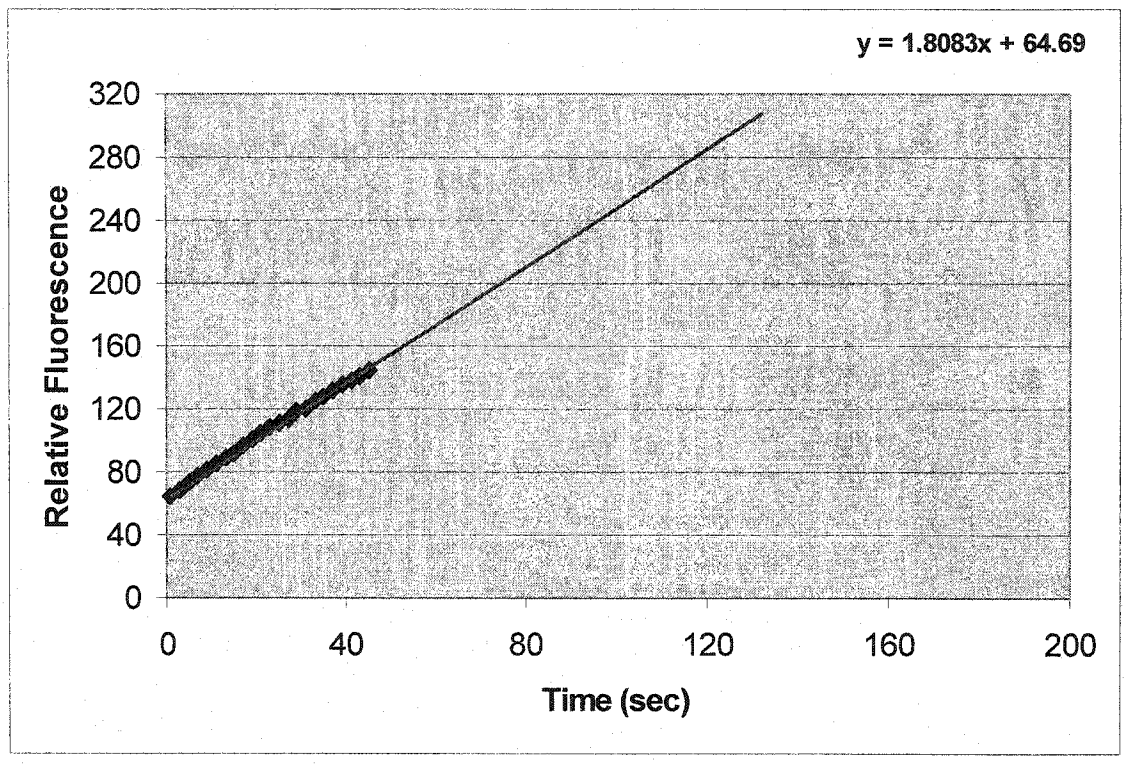


Figure 16b: Linear portion of progress curve for the hydrolysis of 3 μM of 2-Abz-Phe-Arg-Phe($4NO_2$) by 23 nM of cathepsin X used to calculate catalytic efficiency. The trendline is shown in red. See section 2.3.2 for details.

3.9 Assays for the hydrolysis of P1' substrates

Substrates in the P1'library differ with respect to each other at the C-terminal end of the tripeptide. Since cathepsin X is a carboxymonopeptidase, it is the varying amino acid residue in the substrate that is liberated upon hydrolysis, resulting in the formation of a product that will always have the sequence 2-Abz-Phe-Arg. It was necessary to first identify where the product peak would appear on the chromatogram. This was accomplished by injecting a small amount of commercially available 2-Abz-Phe-Arg into the RP-HPLC and determining its retention time (Fig. 17a). Each substrate appeared as a peak on the chromatogram with a unique retention time. The chromatogram for the parent substrate 2-Abz-Phe-Arg-Phe, which lacks the unnatural Phe($4NO_2$) in P1', is shown in Fig. 17b. When 4 μM of 2-Abz-Phe-Arg-Phe was incubated with 23 nM of cathepsin X for 30 min both the resulting product and unreacted substrate were resolved on the chromatogram (Fig. 17c) which is representative of the data collected and analyzed for the hydrolysis of all P1' substrates. The area under the product peak was integrated and could be converted into moles of product formed using a calibration curve. The calibration curve was generated by injecting increasing concentrations of commercially obtained 2-Abz-Phe-Arg (product) into RP-HPLC (illustrated in Fig. 18). The slope of the line represents 5×10^{12} integrated area units/M of product formed.

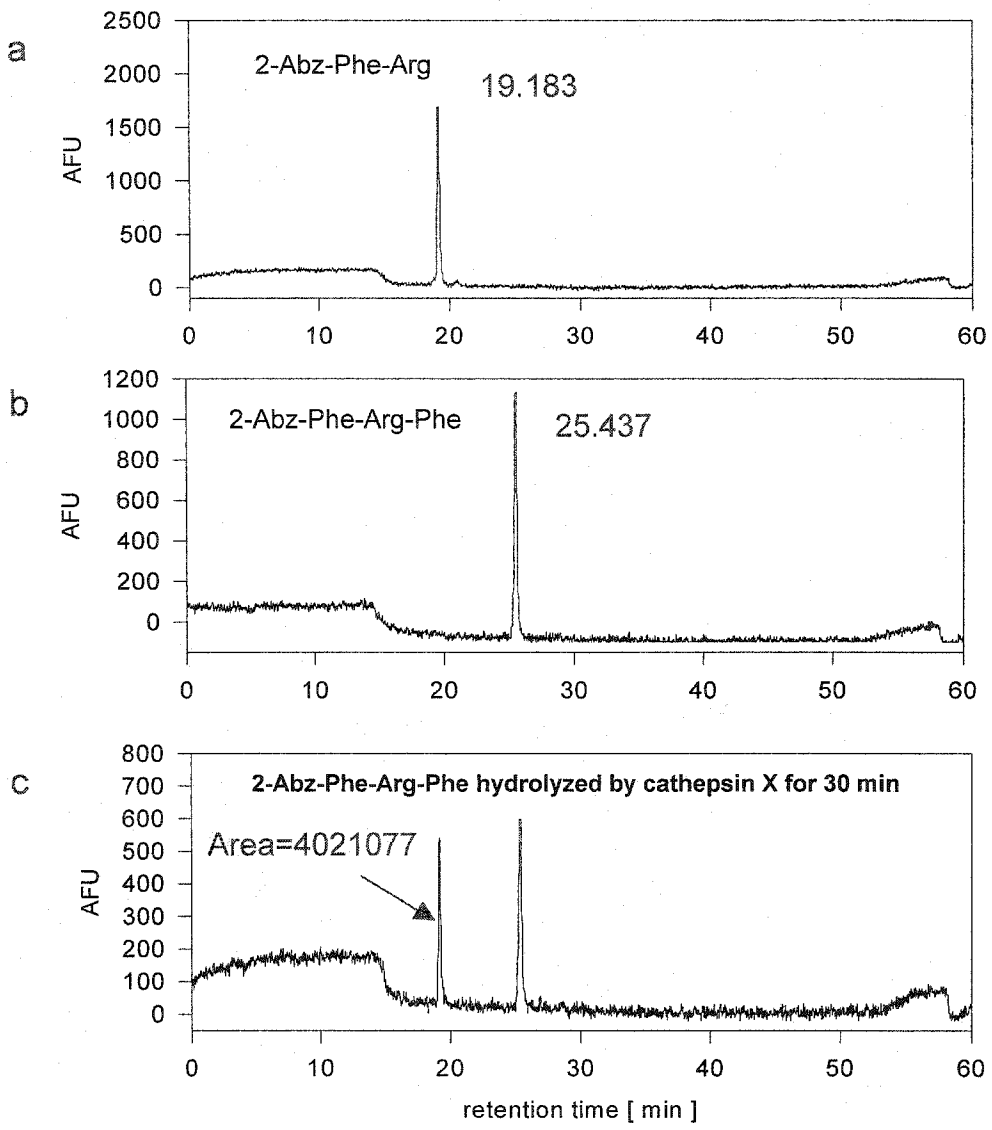


Figure 17: Chromatograms for the HPLC-based assay of P1' substrates.

a) chromatogram of the commercially obtained product. b) chromatogram of the parent substrate. c) chromatogram of the reaction of 4 μM of 2-Abz-Phe-Arg-Phe incubated with 23 nM of cathepsin X for 30 min. See section 2.3.3 for details. Integrated area of product peak and retention times of both product and substrate peaks are shown. Note: AFU stands for arbitrary fluorescence units.

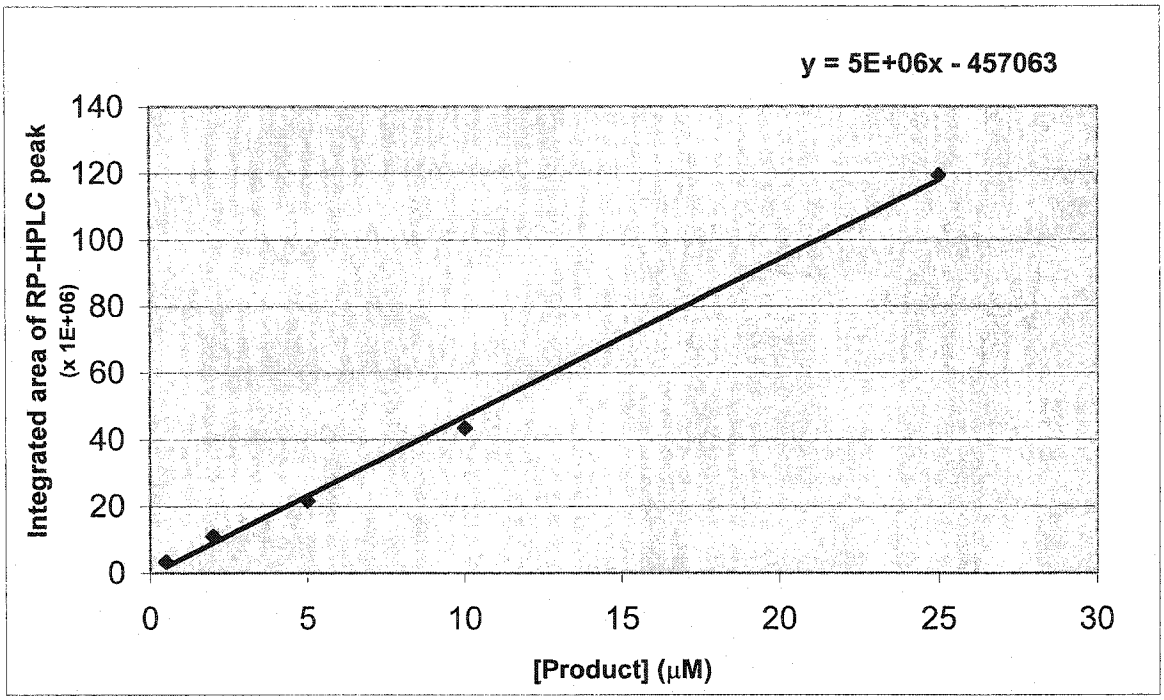


Figure 18: Calibration curve for P1' hydrolysis assays. Various concentrations (0.5, 2, 5, 10, 25 μM) of product, 2-Abz-Phe-Arg, were injected into a Waters RP-HPLC (see section 2.3.3 for details). The areas under the peaks were integrated. The slope of this line correlates the change in integrated area to the amount of product formed (5×10^{12} integrated area units/M of product formed).

The activation of cathepsin X and incubations of P1' substrates with cathepsin X were performed using βME rather than DTT. Any oxidized DTT[23] was resolved as a large peak on the chromatogram at a retention time that coincided with the resolution of the much smaller product peaks (data not shown). In contrast, βME eluted with the solvent and posed no problems.[23]

The reaction of each P1'substrate with cathepsin X was assessed with a time-course experiment.[23,24] This involved the incubation of 2-Abz-Phe-Arg-Phe with cathepsin X and quenching the reaction at various time points. The resulting samples were then analyzed by RP-HPLC. A progress curve of time vs. product formed was then plotted (Fig. 19). The plot remained linear until about 1.1 μM of product formed, or approximately 27% hydrolysis of the substrate initially present in the reaction mixture. Assuming that the other P1' substrates would behave similarly when incubated with cathepsin X, all reactions needed to be quenched when 25-30% of each substrate was hydrolyzed. In the case of 2-Abz-Phe-Arg-Phe, 27% hydrolysis corresponded to approximately 18 min. In order to screen all other substrates P1' library, 30 min incubations were used. As a result, P1' substrates that were more efficiently hydrolyzed than the parent substrate by cathepsin X could be identified from those that were poorer substrates (more slowly degraded). Incubation times were adjusted appropriately for each P1' substrate to ensure that 25-30% of the substrates were hydrolyzed. The RP-HPLC based assay was less sensitive than the assay involving a spectrofluorometer, hence, five different substrate concentrations were used that were all well below K_M in order to remain in first order kinetics. Catalytic efficiencies of P1' substrates were calculated as described in (2.3.2).

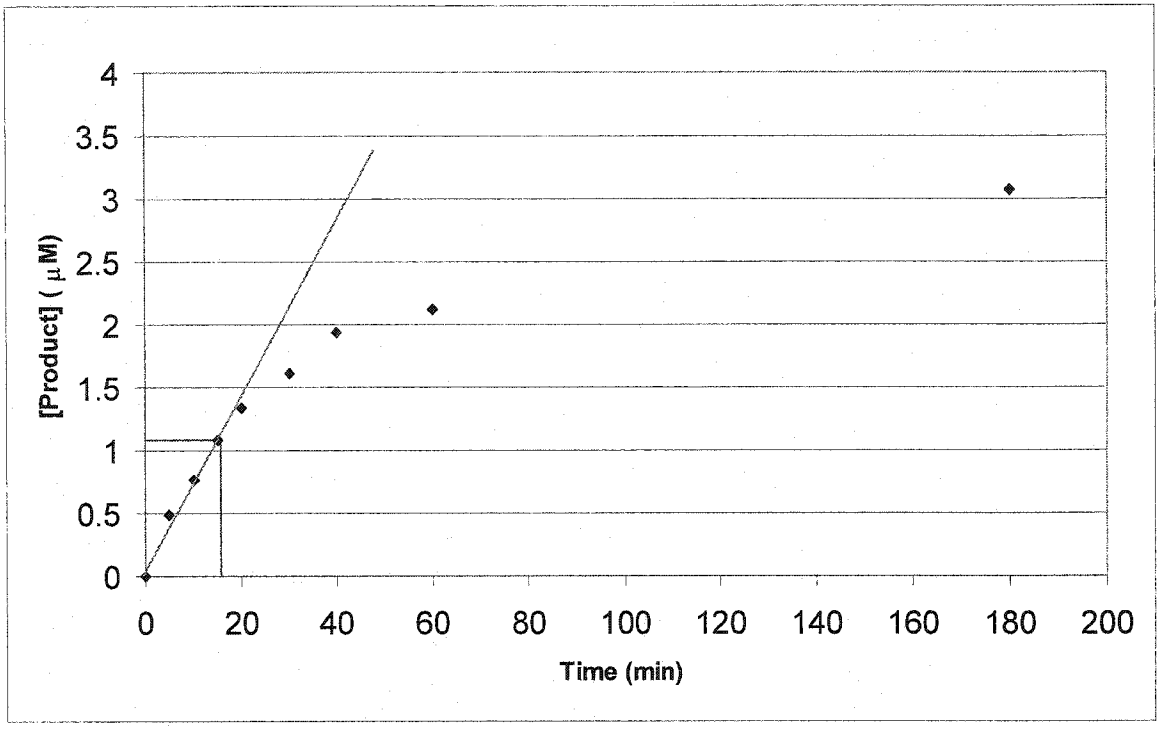


Figure 19: Time-course generated progress curve for the reaction of 4 μM of 2-Abz-Phe-Arg-Phe, parent substrate, incubated with 23 nM cathepsin X. The reaction was quenched using a 5% TFA solution at various time points (0, 5, 10, 15, 20, 30, 40, 60, 180) min. Samples were analyzed using a Waters RP-HPLC (see section 2.3.3. for details). Integrated areas of product peaks corresponding to each time point were converted to moles of product formed. The plot remained linear until approximately 1.1 μM of product was formed which was equal to approximately 27% hydrolysis of 2-Abz-Phe-Arg-Phe by cathepsin X.

3.10 P2 substrate specificity: 2-Abz-X-Arg-Phe(4NO_2)

The catalytic efficiency of the reactions of cathepsin X with each compound within the P2 substrate library is shown in Fig. 20. The data indicate that the S2 subsite exhibits moderately broad substrate specificity. The substrate containing Tyr in the P2 position yields the highest catalytic efficiency, and only three other substrates (Leu, Phe, and Gln in P2 position) possess values that are within one order of magnitude of the value for Tyr. Nevertheless, from the remaining substrates, fourteen have k_{cat}/K_M values within one log unit (4.5 - 3.5) of each other. One exception is the substrate housing Pro in P2. It is the poorest substrate for cathepsin X and differs from the substrate with Tyr in P2 by over 3 orders of magnitude.

The trend of the efficiency constants indicates that bulky hydrophobic side chains (Tyr, Leu, Phe) are more easily hydrolyzed than others. This could be due to strong interactions of these hydrophobic side chains with the largely hydrophobic S2 subsite of the enzyme.[11,12] Val181 and Ile178 are hydrophobic residues present in the S2 subsite of cathepsin X. These residues are capable of stabilizing a peptidyl P2 side chain via hydrophobic interactions.

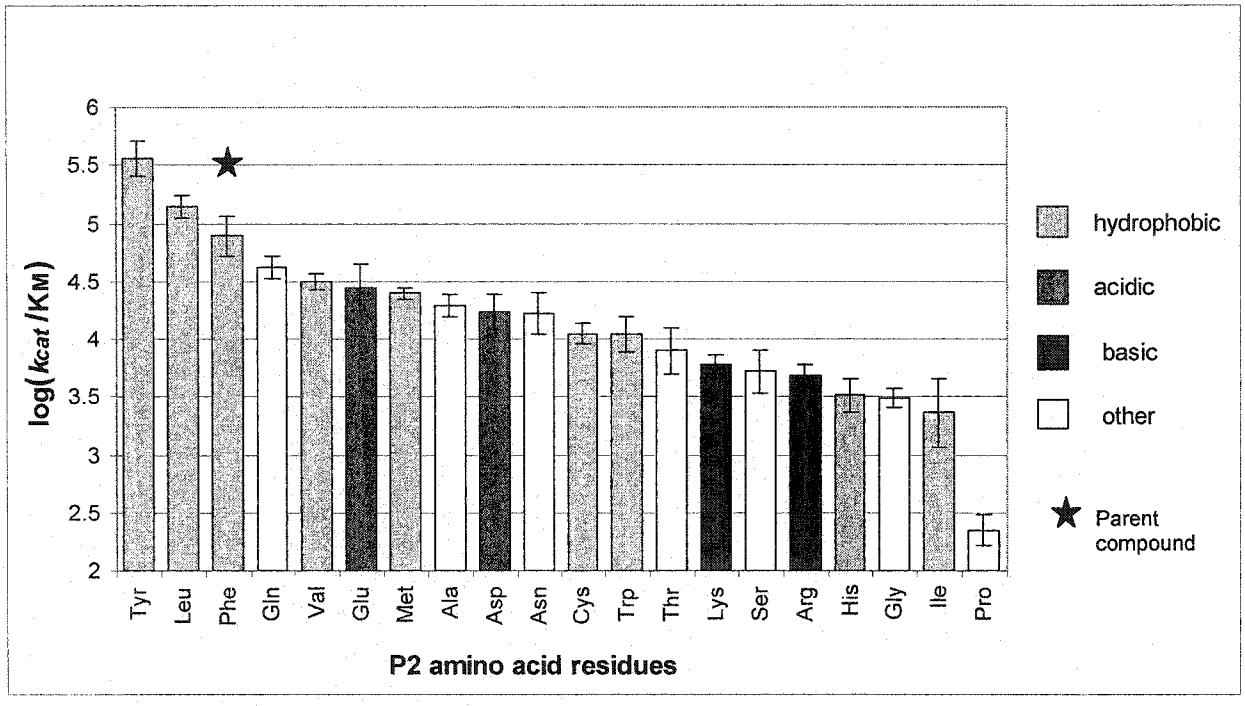


Figure 20: Catalytic efficiency of P2 variants. Values of k_{cat}/K_M have been expressed in logarithmic form. Each bar on the graph represents an average of catalytic efficiency values calculated at three different substrate concentrations. Error bars signify standard deviations from the mean (Microsoft Excel). The parent compound, 2-Abz-Phe-Arg-Phe($4NO_2$), is indicated by a star. All amino acid residues in P2 have been color-coded according to their degree of hydrophobicity based on their logP (water to octanol coefficients).[28-32] The bar for His is shown in purple since it can be protonated at pH 6.0.[30]

Modeling studies were performed as described in section 2.4 in order to correlate the experimental results with structural data. In the modeled complex, Tyr in the P2 position packs against Val181 on one side of the pocket (Fig. 21). Tyr's polar hydroxyl group points away from of the pocket and is solvent-exposed.

Substrates containing positively charged residues at the P2 position (Arg, Lys) are less efficiently hydrolyzed than those with negatively charged side chains (Glu, Asp). Furthermore, polar residues (Asn, Ser) yield comparable catalytic efficiency values to charged residues (Asp, Arg). However, a substrate with Gln in the P2 position is within one order of magnitude of Tyr and within 0.3 log units of Phe. Gln has a small polar side chain and differs from the bulky, hydrophobic side chains of Tyr, and Phe. Docking of the substrate with Gln in the P2 position (Fig. 22) shows that several favorable hydrogen bonds can be formed within the S2 subsite of cathepsin X. A range of (2.7-3.3) Å was used to determine whether a hydrogen bond could form between a hydrogen donor and acceptor. The side chain amide group of Gln can hydrogen bond to both the carbonyl oxygens associated with the backbone of Gly74 and the side chain of Asp76. Asp76 is residue specific to cathepsin X;[11,12] normally a hydrophobic residue is found at this position (Pro in cathepsin B[33-35] and L[36,37] ; Met in cathepsin K [38,39] and S[40,41]). Hence, the presence of Asp76 appears to stabilize the polar Gln in a manner unique to cathepsin X (see Appendix II). The Gln amide group also hydrogen bonds with a water molecule found in the S2 subsite. Asn, which is also polar, shows lower activity than Gln in the P2 position. This is most probably due to the fact that Asn is shorter by one methylene group,

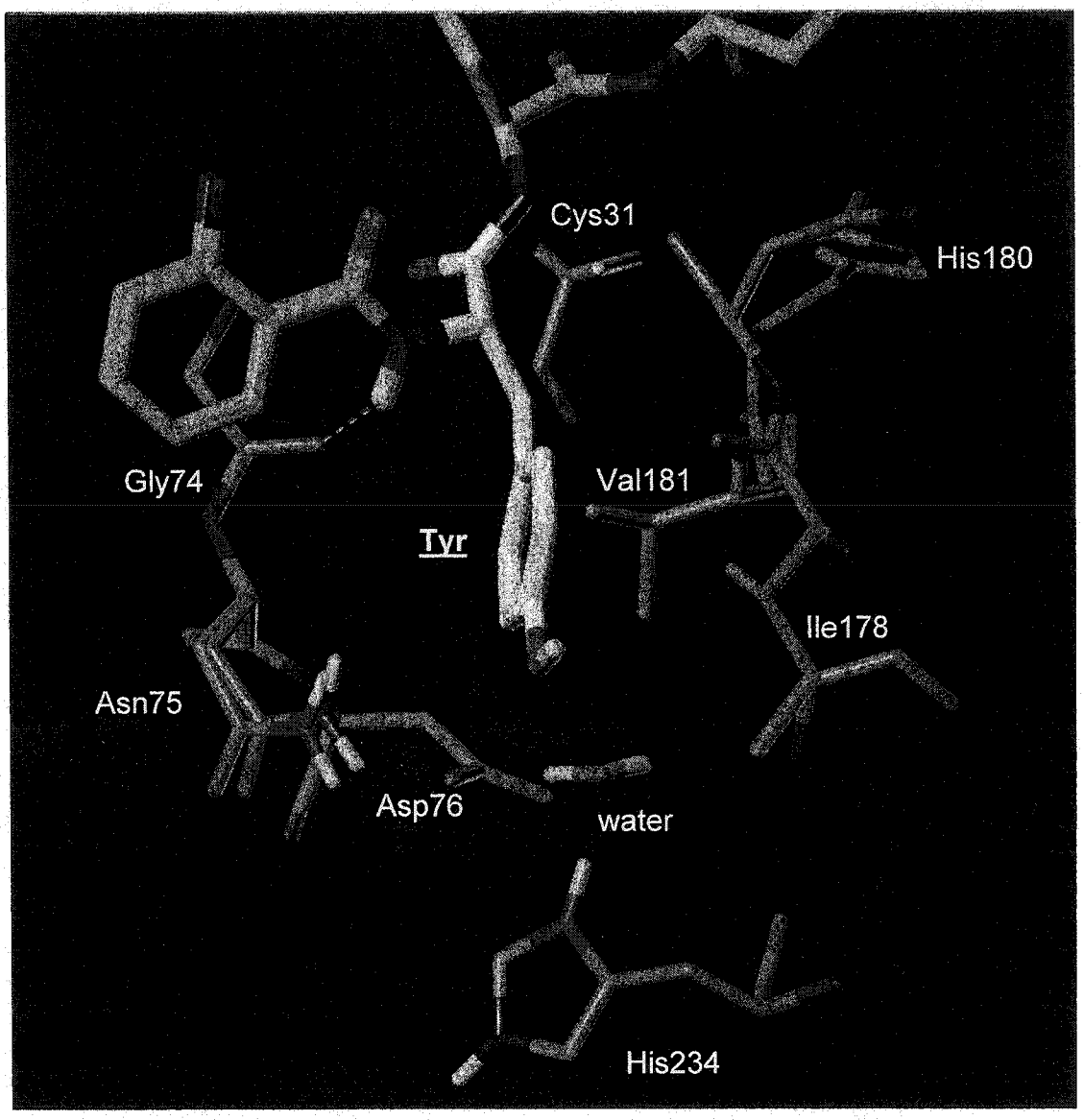


Figure 21: S2 subsite interactions with Tyr in the P2 position. Residues in purple belong to cathepsin X. The substrate is shown in green with the P2 side chain in white. Hydrogen bonds are shown using yellow dashed lines. This model was generated as described in section 2.4.

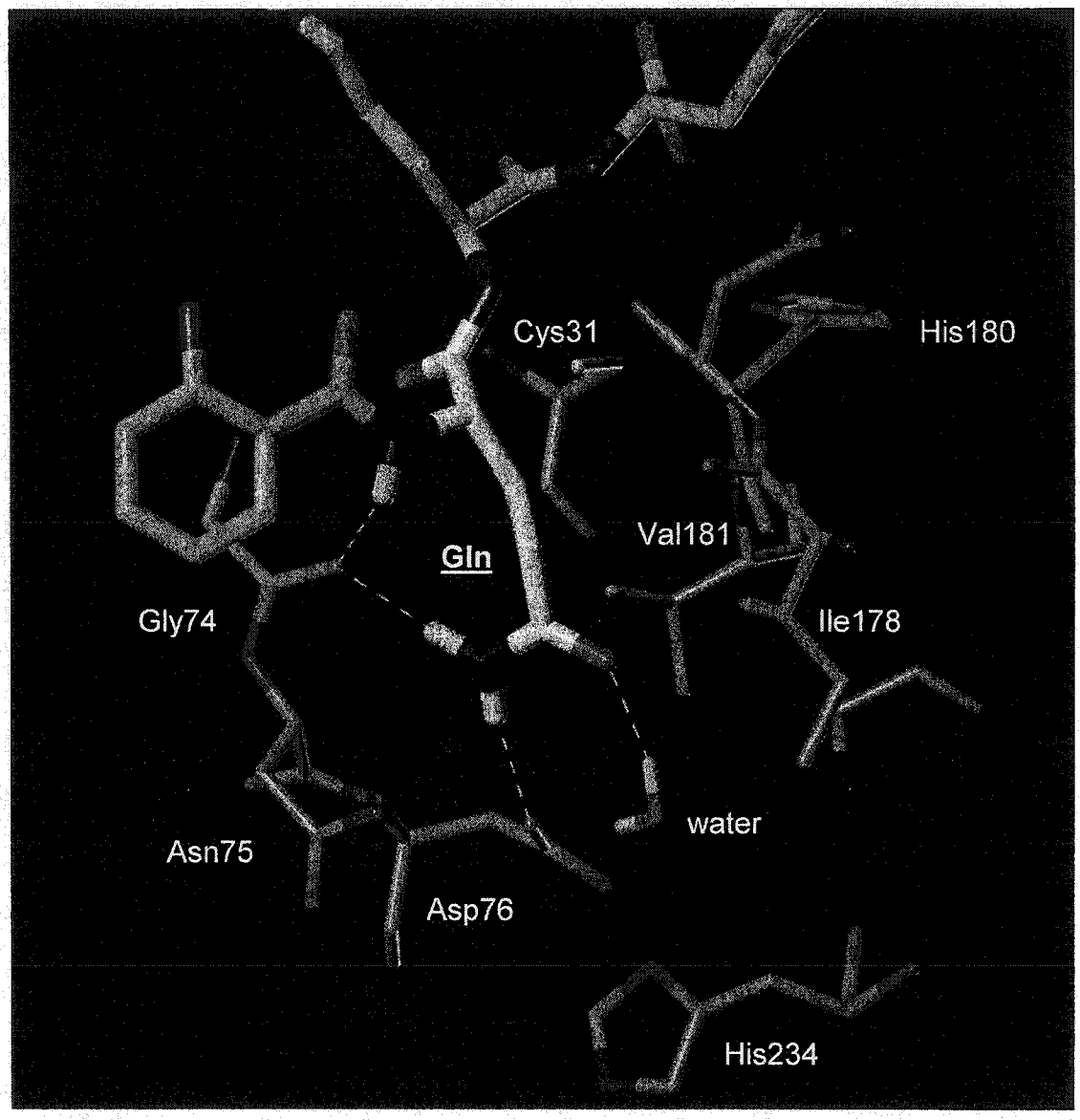


Figure 22: S2 subsite interactions with Gln in the P2 position. Residues in purple belong to cathepsin X. The substrate is shown in green with the P2 side chain in white. Hydrogen bonds are shown using yellow dashed lines. This model was generated as described in section 2.4.

preventing the proper positioning of its side chain amide to the side chain carboxylate of Asp76 as well as the Asn side chain carbonyl group to the water molecule in the S2 pocket.

Why is Lys, which is relatively isosteric with respect to Gln and possesses a positive charge, a poorer substrate? Comparison of the modeled bound conformation of the protonated side chain of Lys in the P2 position to the polar Gln (Fig. 23) illustrates that Lys establishes only one hydrogen bond with the enzyme whereas Gln side chain makes three hydrogen bonds with the S2 subsite of cathepsin X. For Lys, the hydrogen bond is between the positively charged ε amino group and the side chain carbonyl group of Asn75. The fact that Gln's side chain can form three hydrogen bonds could explain the higher catalytic efficiency of the substrate containing Gln at the P2 position. The increased k_{cat}/K_M value reflects an increase in binding interactions that may affect the turnover number of the enzyme.

It is not surprising that the substrate having Pro in the P2 position was cleaved with the lowest efficiency. Proline is unique because it lacks the main-chain NH group (Fig. 24) and thus the hydrogen bond with the backbone Gly74 carbonyl of the enzyme cannot be formed. In addition, since the Pro side chain is cyclized to the backbone, its bulkiness could sterically hinder the S2 subsite instead of extending further into the more spacious region of the pocket, as in the case of Tyr, Phe, or Leu side chains. Thus, these two factors may be responsible for the poor hydrolysis of a substrate containing Pro in the P2 position.

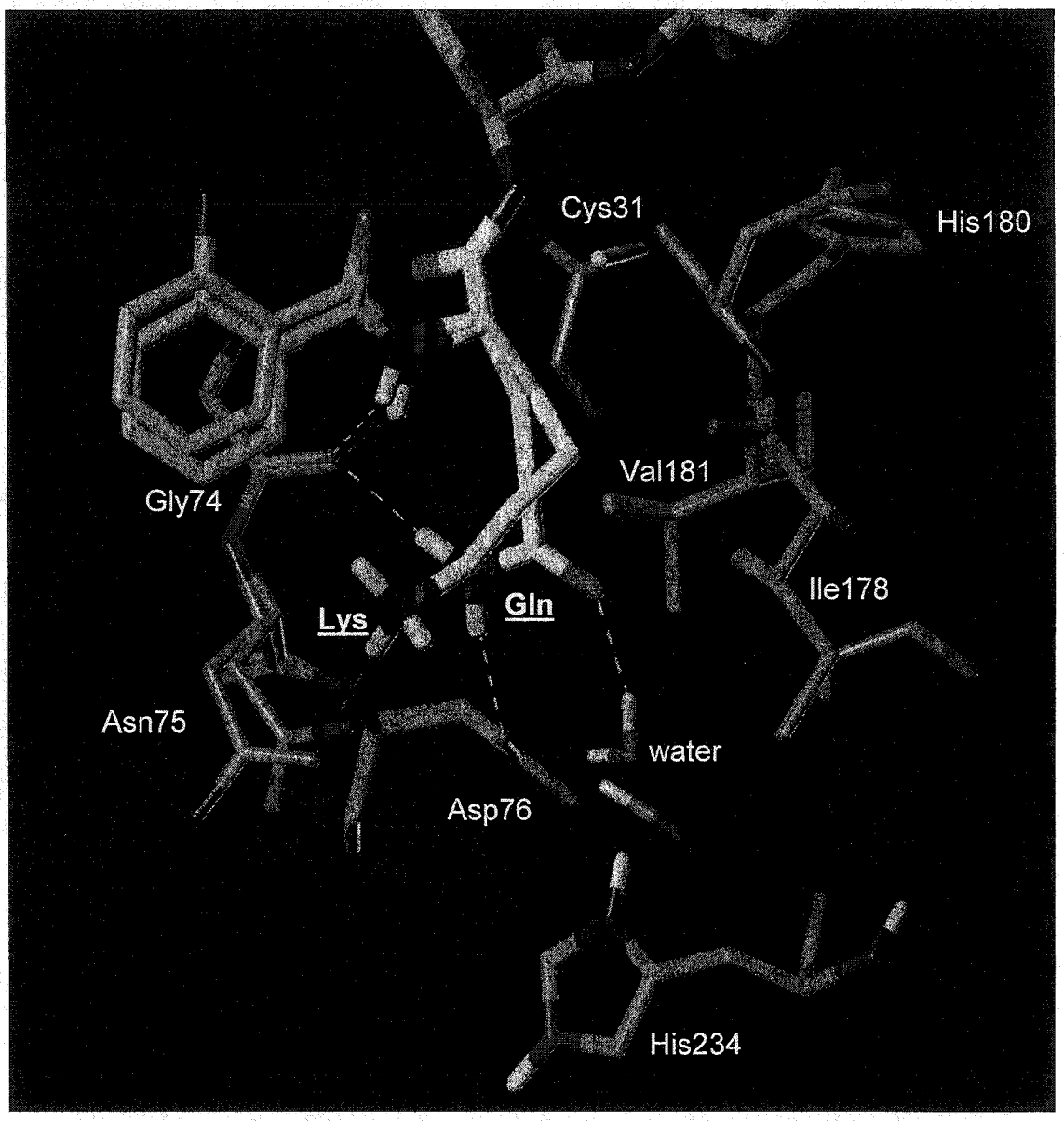


Figure 23: Comparison between Gln and Lys in the P2 position. Residues in purple belong to cathepsin X. The substrate is shown in green with the P2 side chains in white. Hydrogen bonds between Gln-containing substrate and cathepsin X are shown using yellow dashed lines. Hydrogen bonds between Lys-containing substrate and cathepsin X are shown using red dashed lines. This model was generated as described in section 2.4.

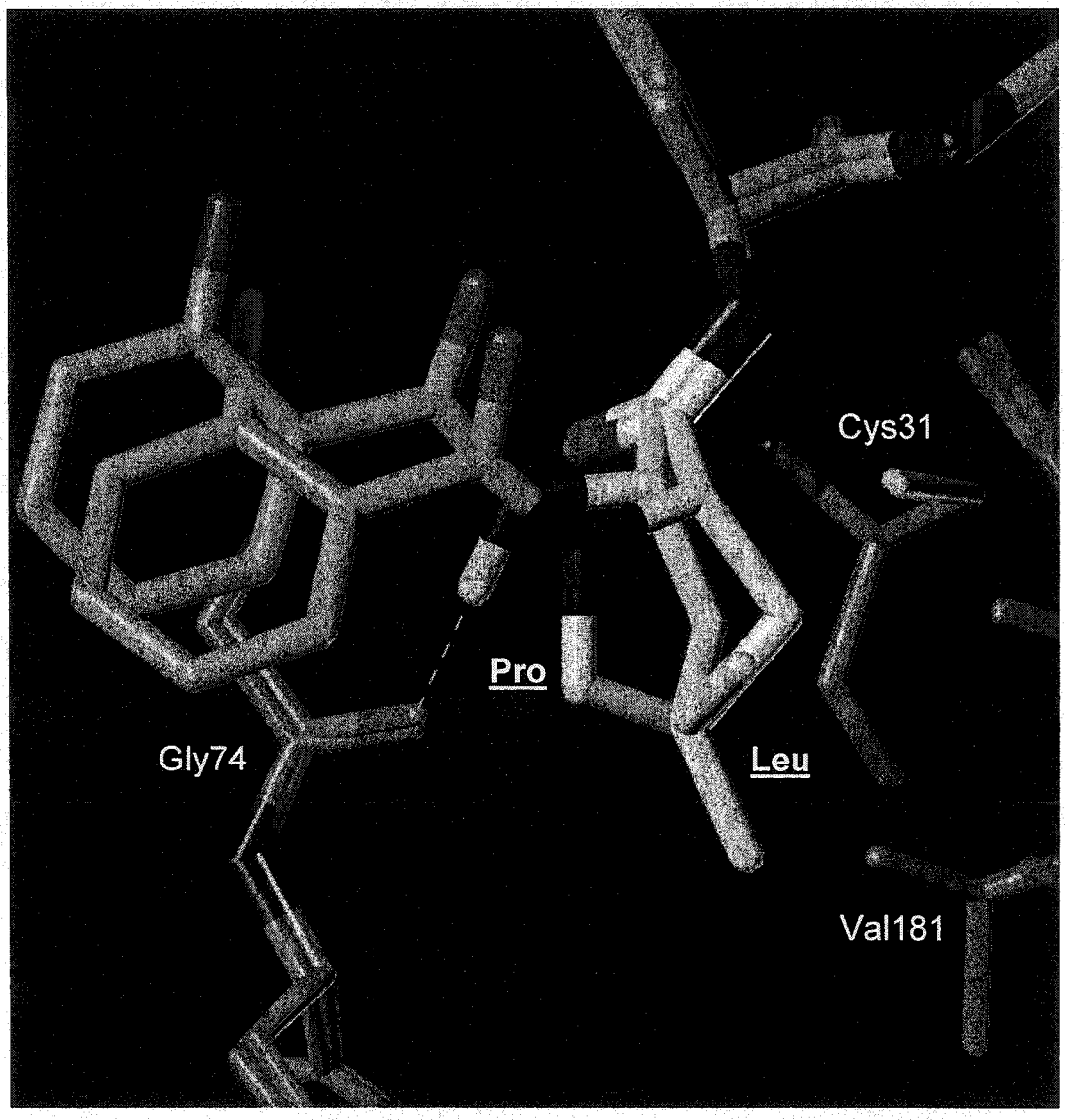


Figure 24: Comparing Pro in the P2 position with Leu to show its lack of hydrogen bonding. Residues in purple belong to cathepsin X. The substrate is shown in green with the P2 side chains in white. This model was generated as described in section 2.4.

In summary, the S2 subsite of cathepsin X prefers bulky, hydrophobic amino acid side chains in the P2 position of a peptide substrate, in accordance with other known cathepsins. A substrate with Gln in P2 is hydrolyzed efficiently due to the presence of the unique Asp76 of cathepsin X.

3.11 P1 substrate specificity: 2-Abz-Phe-X-Phe($4NO_2$)

Figure 25 shows the catalytic efficiency of substrates varied at the P1 position. The distribution is broader than that of the P2 substrate library (Fig. 20), indicating that the S1 subsite is slightly less specific than the S2 subsite. Ten substrates possessed k_{cat}/K_M values within one order of magnitude (5.2 - 4.2) of each other. Morever, values for nineteen out of the twenty substrates span only 2 orders of magnitude (5.2 - 3.2). Again, the hydrolysis of the Pro-containing substrate was poor, with a k_{cat}/K_M value almost three log units lower than that of the best substrate.

Substrates containing positively charged residues (Arg, Lys) in the P1 position are efficiently hydrolyzed by cathepsin X. This suggests that they are also well accommodated in the S1 subsite of the enzyme and is in accordance with previous findings using cathepsin X[16] and cathepsins B, L, K, and S as catalysts.[1] Furthermore, the Met- and Glu-containing substrates are both efficiently hydrolyzed. The former contains a hydrophobic side chain, whereas the latter is negatively charged. Ala has previously been shown to be accepted in the S1 subsite of cathepsins.[1] We note that all these side chains are unbranched. Substrates containing branched side chains in the P1 position (Ile, Val) are more poorly hydrolyzed. These data indicate that the steric contributions are more important than the hydrophobic and electrostatic requirements for binding to the S1 subsite.

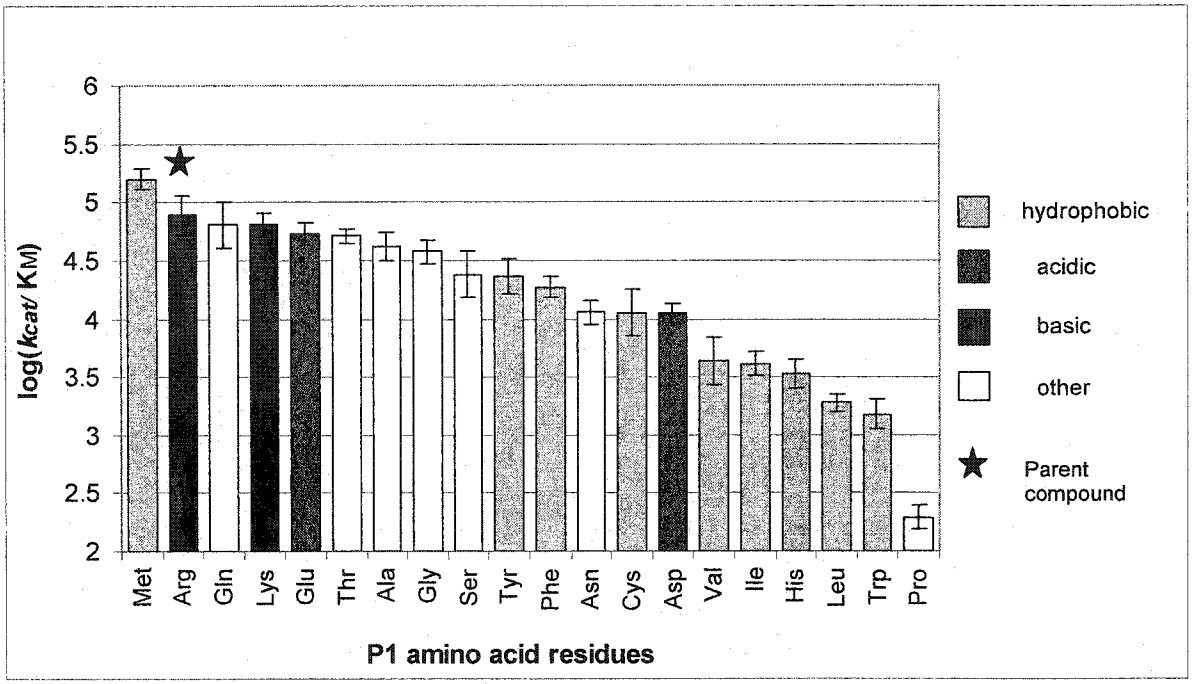


Figure 25: Catalytic efficiency of P1 variants. Values of k_{cat}/K_M have been expressed in logarithmic form. Each bar on the graph represents an average of catalytic efficiency values calculated at three different substrate concentrations. Error bars signify standard deviations from the mean (Microsoft Excel). The parent compound, 2-Abz-Phe-Arg-Phe($4NO_2$), is indicated by a star. All amino acid residues in P1 have been color-coded according to their degree of hydrophobicity based on their logP (water to octanol coefficients).[28-32] The bar for His is shown in purple since it can be protonated at pH 6.0.[30]

These results are in accordance with crystal structures of cathepsin X[11,12] which show that the S1 subsite is a shallow, long depression rather than a pocket as is the S2 subsite.[11,12] Thus, the enzyme would not tolerate bulky (Trp) or branched (Ile) P1 side chains but accommodates longer, unbranched side chains (Lys, Met, Glu) with greater ease.

The Met-containing substrate, is shown docked to cathepsin X in Fig. 26. It is apparent that the linear Met side chain is well packed along the S1 subsite of the enzyme. In Fig. 27 the γ-branched side chain of Leu in the P1 position is overlaid with the linear side chain of Met. Of the two methyl groups on the δ carbon of Leu, one points away from the S1 subsite, while the other appears abutted against the enzyme. Thus, although both Leu and Met contain hydrophobic side chains, P1 Leu is more poorly accommodated than Met in the S1 subsite. This is a plausible explanation for the difference in the hydrolysis efficiencies (about 2 orders of magnitude) of these two substrates.

The side chains of Arg and Lys are also, for the most part, linear and aliphatic. Their charged ε amino and guanidino groups are at the very end where branching seems to be tolerated. A comparison of Arg and Glu side chains in P1 is shown in Fig. 28.

As observed in the results for the P2 substrates, the peptide containing Pro in the P1 position was the least efficiently hydrolyzed by cathepsin X. This agrees with our docking studies, which show that Pro lacks a critical hydrogen bond which is normally formed between the backbone NH group of the P1 residue and the main-chain carbonyl group of Asn179 (Fig. 29).

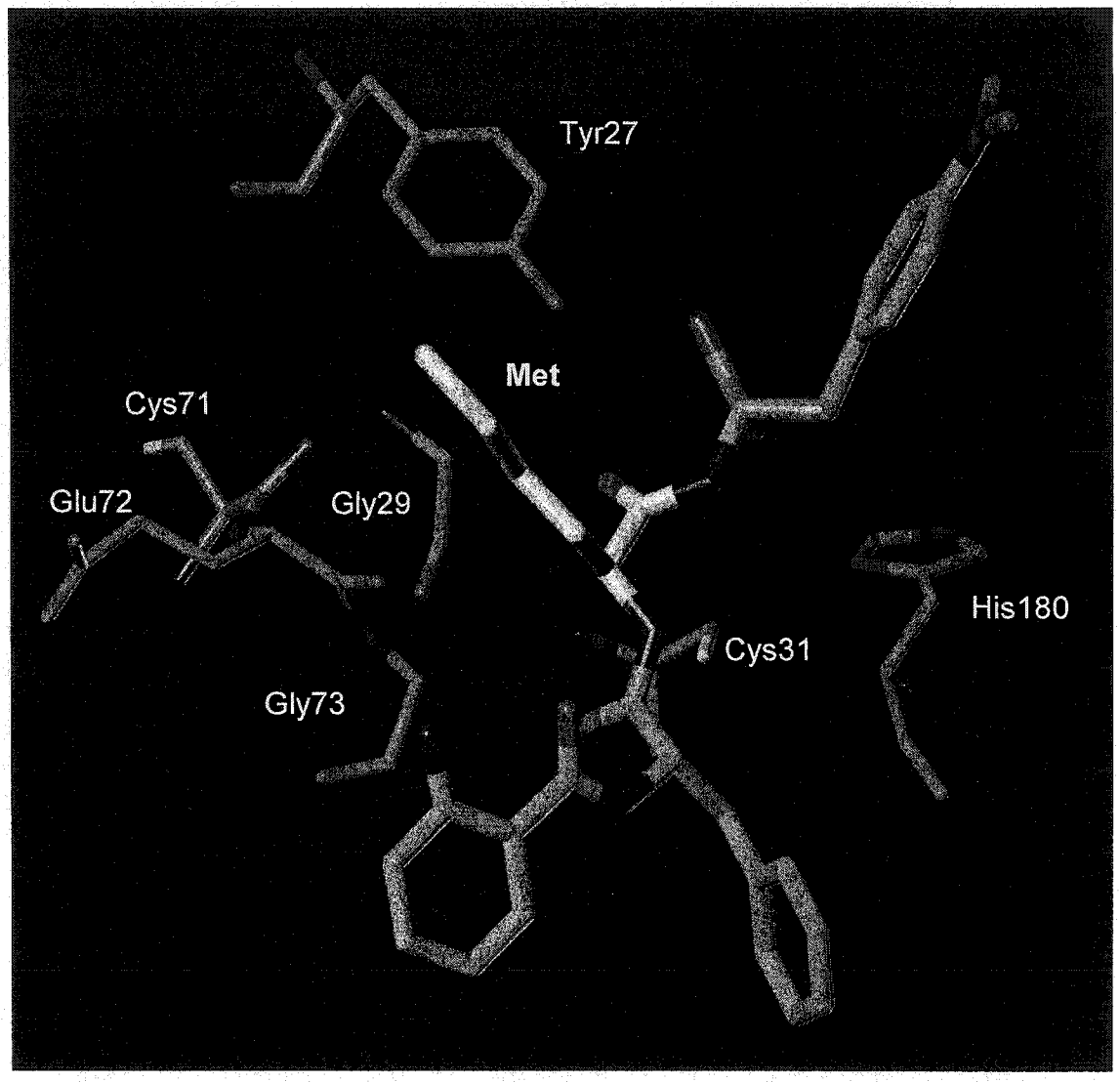


Figure 26: S1 subsite interactions with Met in the P1 position. Residues in purple belong to cathepsin X. The substrate is shown in green with the P1 side chain in white. This model was generated as described in section 2.4.

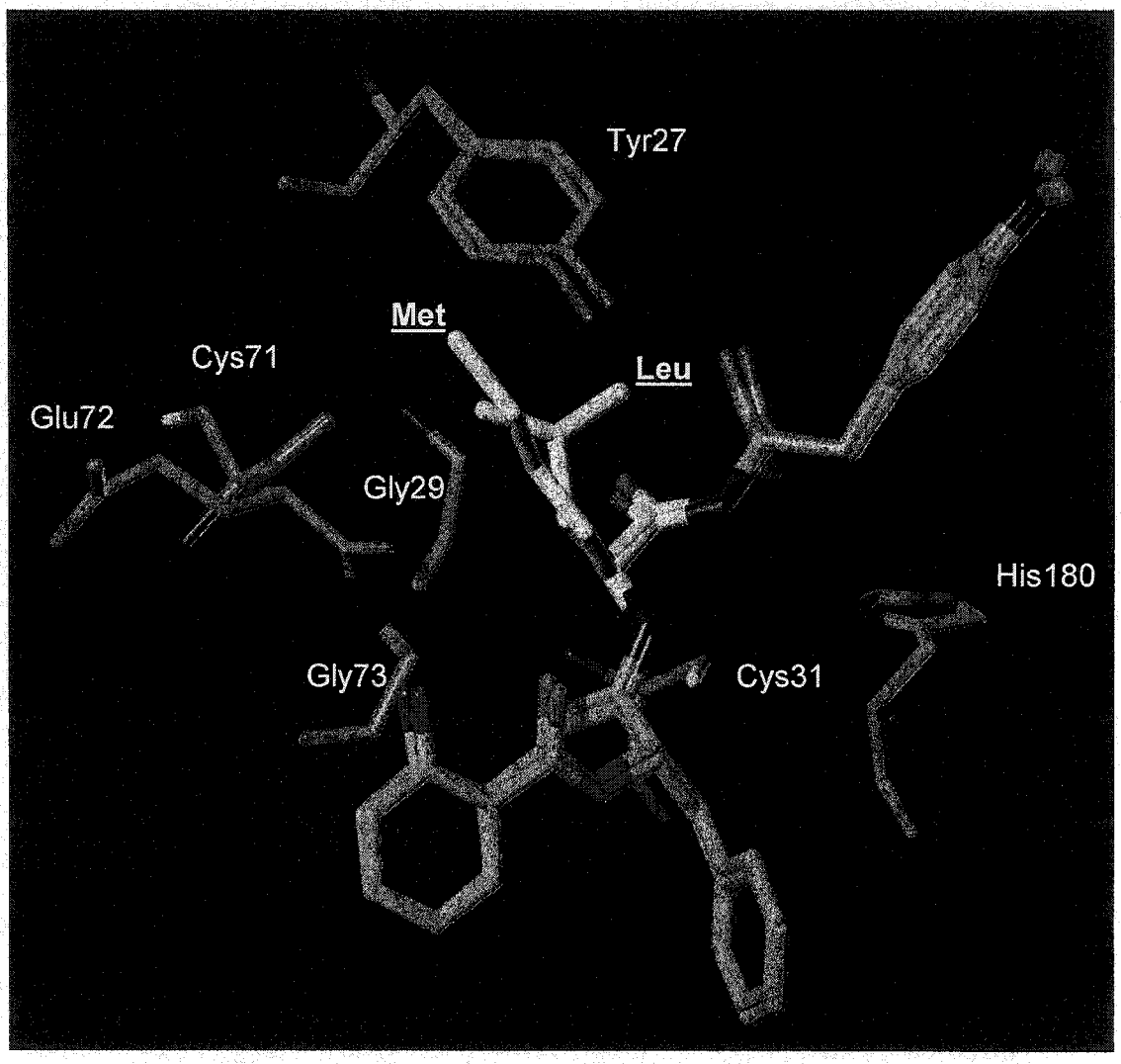


Figure 27: Comparison of Met and Leu in the P1 position. Residues in purple belong to cathepsin X. The substrate is shown in green with the P1 side chains in white. This model was generated as described in section 2.4.

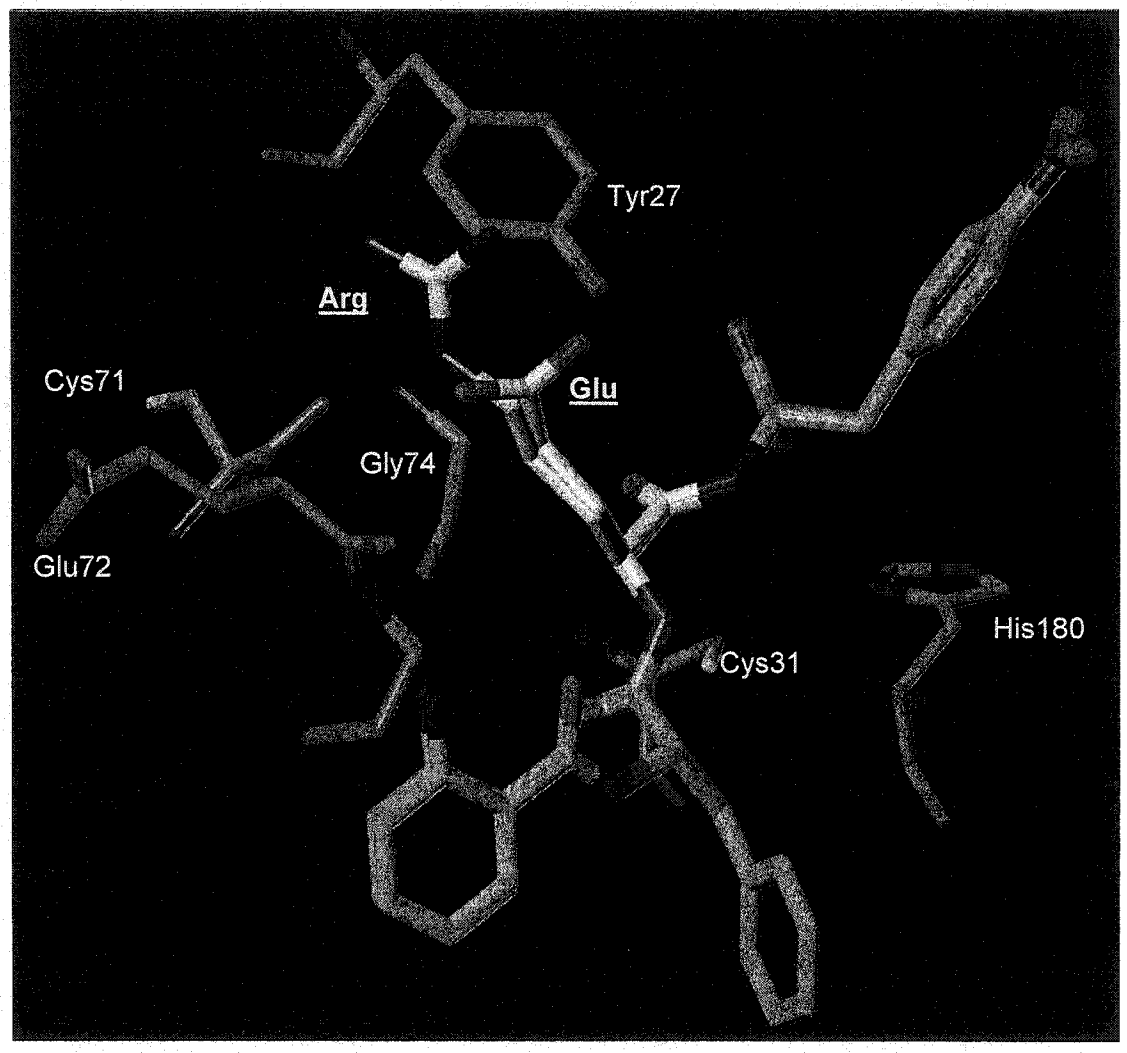


Figure 28: Comparison of Arg and Glu in the P1 position. Residues in purple belong to cathepsin X. The substrate is shown in green with the P1 side chains in white. This model was generated as described in section 2.4.

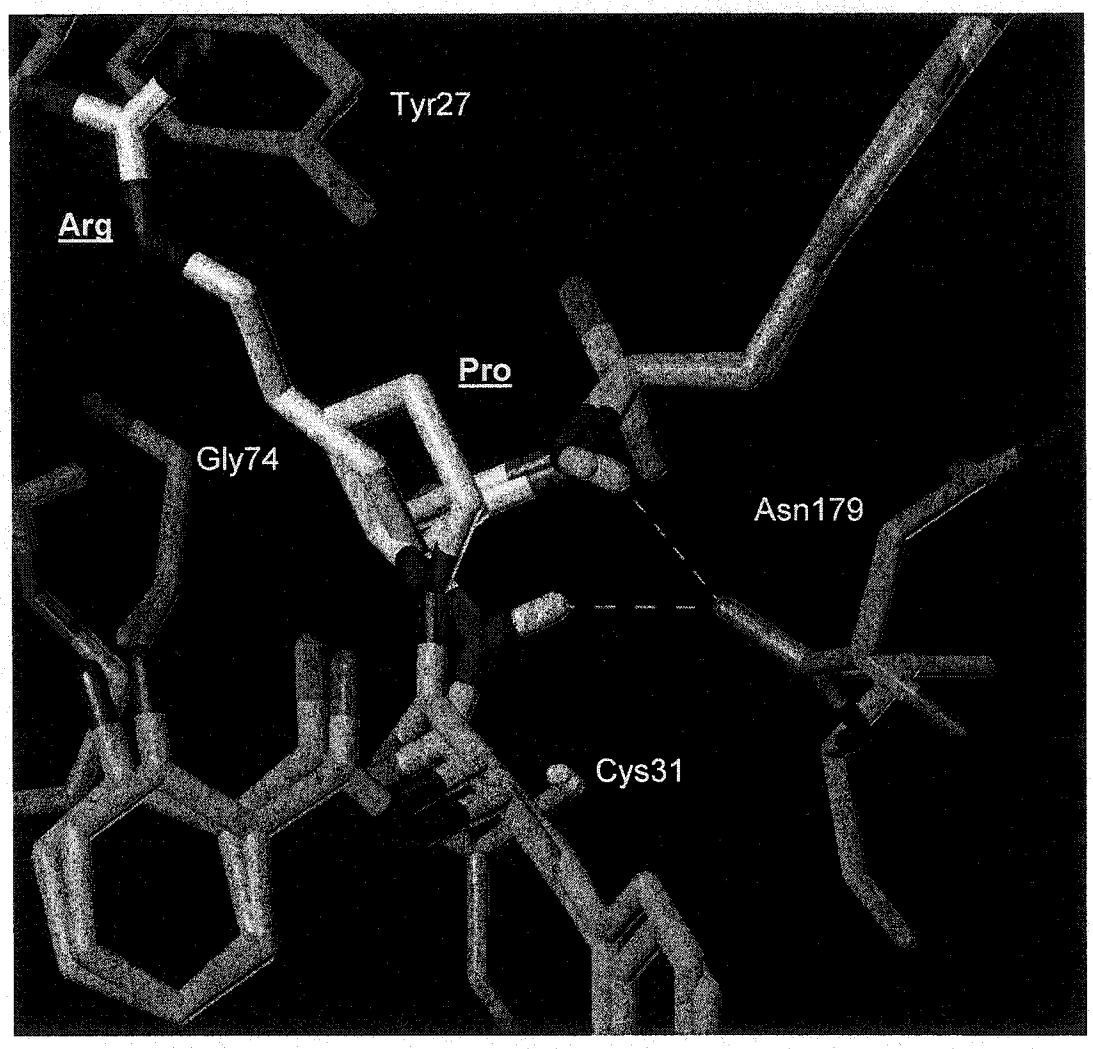


Figure 29: Comparison of Pro and Arg in the P1 position to illustrate Pro's lack of hydrogen bonding capability. Residues in purple belong to cathepsin X. The substrate is shown in green with the P1 side chains in white. Hydrogen bonds are shown using yellow dashed lines. This model was generated as described in section 2.4.

In summary, geometry appears to be the driving force for acceptance of a peptidyl side chain in the S1 subsite of cathepsin X. This is in accordance with higher catalytic efficiency values for unbranched side chains, and confirms that cathepsin X exhibits similar S1 subsite specificity as other cathepsins.

3.12 P1' substrate specificity results: 2-Abz-Phe-Arg-X

From the P1' library of substrates, cathepsin X is fairly selective for Cys and Ser (Fig. 30). The catalytic efficiencies for sixteen other substrates are within 0.5 log unit of each other (4.2 - 3.7), illustrating the enzyme's broad specificity for these substrates. The parent compound is not included in this set of results because Phe($4NO_2$) is not found in the P1' position of any of the 20 substrates assayed. However, to view the interactions at the S1' subsite, the previously docked Phe($4NO_2$)-containing parent compound is shown in Fig. 31. The S1' subsite is the least studied of all subsites. It is structurally quite shallow[1,42-44] and contains a Trp residue (Trp202) and an Ala residue (Ala157) which contribute to its hydrophobic character.[3-5] The aromatic phenyl ring of Phe($4NO_2$) is stabilized by packing against Trp202 and Ala157. The polar NO_2 group points away from the hydrophobic residues.

The lowest energy conformation obtained for Phe in P1' is somewhat different from that of Phe($4NO_2$). An overlay of these two substrates is shown in Fig. 32. The aromatic ring of Phe appears to be oriented toward Asn179 and Ala157. This positioning of Phe does not compromise important hydrogen bonds that are formed between the C-terminal free carboxylate of the substrate and the mini-loop residues (i.e., the imidazole εNH group of the protonated His23, the phenolic hydroxyl of Tyr27, the indole NH group of Trp202, and the side chain amide of Gln19).

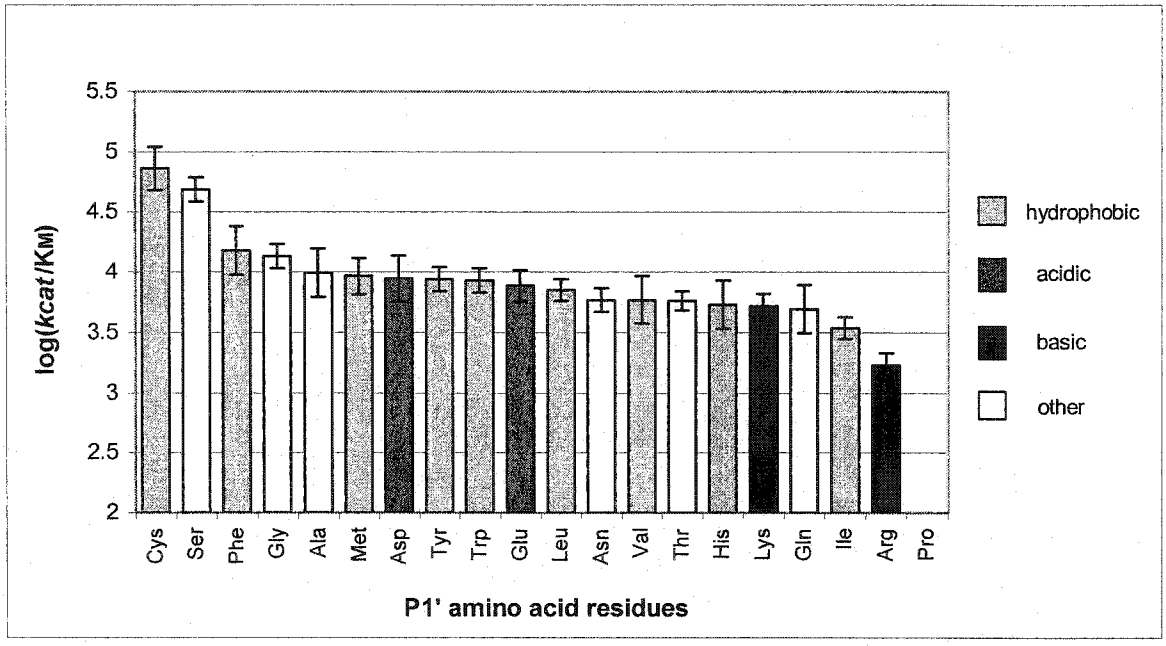


Figure 30: Catalytic efficiency of P1' variants. Values of k_{cat}/K_M have been expressed in logarithmic form. Each bar on the graph represents an average of catalytic efficiency values calculated at five different substrate concentrations. Error bars signify standard deviations from the mean (Microsoft Excel). The parent compound, 2-Abz-Phe-Arg-Phe($4NO_2$), is not present in this library. All amino acid residues in P1' have been color-coded according to their degree of hydrophobicity based on their logP (water to octanol coefficients).[28-32] The bar for His is shown in purple since it can be protonated at pH 6.0.[30]

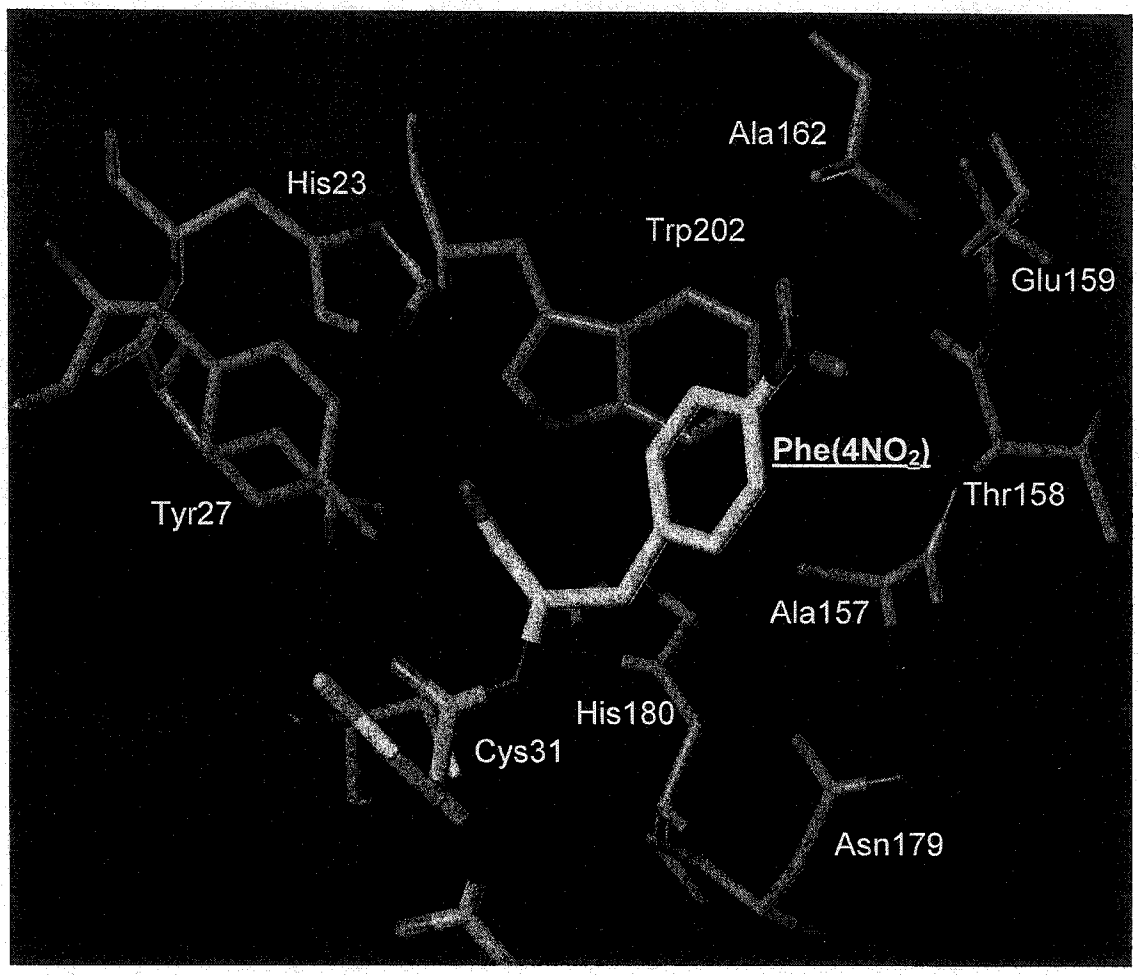


Figure 31: S1' subsite interactions with Phe($4NO_2$) in the P1' position. Residues in purple belong to cathepsin X. The substrate is shown in green with the P1' side chains in white. This model was generated as described in section 2.4.

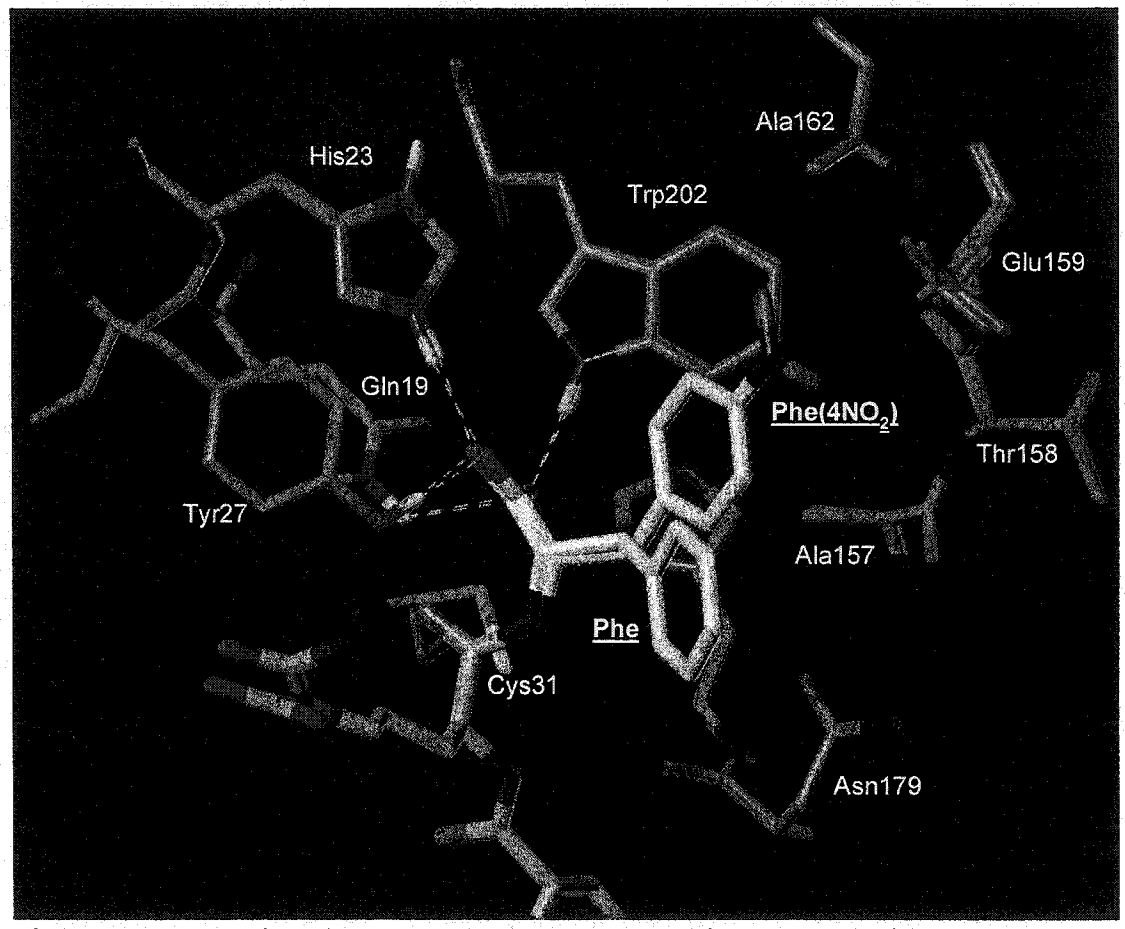


Figure 32: Comparison between Phe($4NO_2$) and Phe in the P1' position. Residues in purple belong to cathepsin X. The substrate is shown in green with the P1' side chains in white. Hydrogen bonds are shown using yellow dashed lines. This model was generated as described in section 2.4.

As shown in Fig. 30, the substrate containing Cys in the P1' position was hydrolyzed with the highest efficiency amongst all substrates in the P1' library. Cysteine is hydrophobic[28-32] as is the S1' subsite. However, the lowest energy conformation (Fig. 33, has the thiol group of the Cys side chain in close proximity to the Asn179 residue of cathepsin X rather than Ala157 or Trp202. The distance between the thiol group of Cys and the carbonyl oxygen of Asn179 is too long to allow formation of a hydrogen bond.

Serine is polar residue.[28-32] Nevertheless, the catalytic efficiency value for Ser is within 0.2 log units of Cys in the P1' position (Fig. 30). The side chain of the docked Ser has the same orientation as with the Cys side chain (Fig. 33). However, the polar hydroxyl group points away from the binding site and is solvent-exposed. It is difficult to speculate on the driving force for such a preferred geometry. One possible explanation is that the large desolvation cost of the polar hydroxyl group is not compensated by a strong electrostatic interaction upon its binding to the protein pocket, and the hydroxyl group prefers the solvent environment. In contrast, the decreased polarity of the thiol would favor the interaction with the protein.

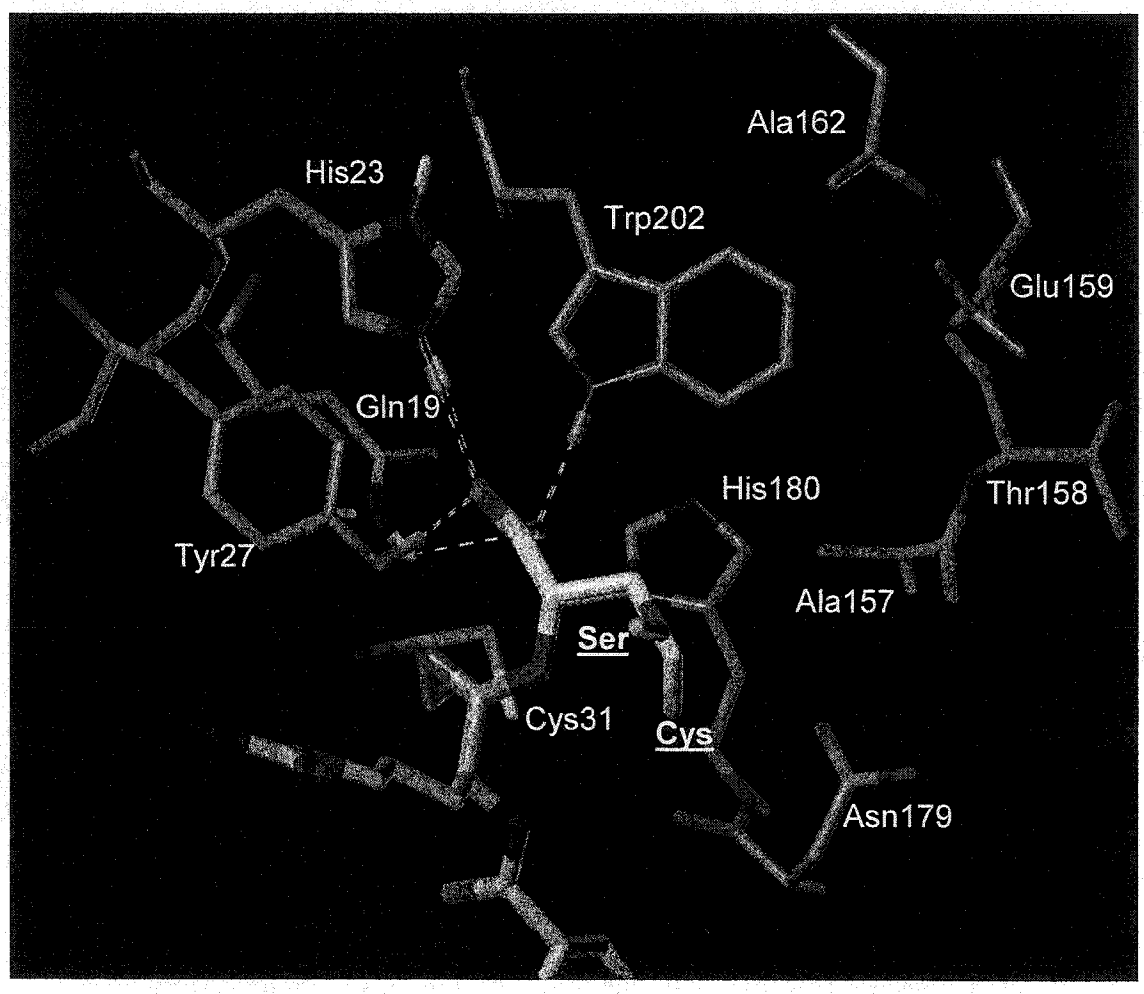


Figure 33: Comparison of Cys and Ser in the P1' position. Residues in purple belong to cathepsin X. The substrate is shown in green with the P1' side chains in white. Hydrogen bonds are shown using yellow dashed lines. This model was generated as described in section 2.4.

The catalytic efficiency of the Pro-containing substrate could not be quantified possibly due to the RP-HPLC-based assay not being sensitive enough. As shown in Fig. 34, Pro at the P1' position causes similar disruptions to substrate binding in the S1' subsite, as was shown previously for the S2 and S1 subsites. Comparison of Pro with Phe($4NO_2$) at the P1'position shows, one of two hydrogen bonds between the peptide backbone NH group and the carbonyl group of Asn179 is lost (Fig. 34). Regardless, all other hydrogen bonds are possible. These include Gln19, Trp202, His23, and Tyr27 with the C-terminal carboxylate of the substrate, and one from the peptide backbone NH group with the carbonyl of Asn179.

As mentioned previously, the parent compound 2-Abz-Phe-Arg-Phe($4NO_2$) was not present in the P1' library. Its catalytic efficiency in both the P2 and P1 libraries was approximately 4.8 ± 0.17 log units (Figs. 20, 25). By contrast, the substrate containing Phe in P1' of the substrate possessed a catalytic efficiency value of approximately 4.2 ± 0.2 log units (Fig. 30), a small but significant difference. In order to ascertain whether the NO_2 moiety contributed to the elevated catalytic efficiency of the parent compound or whether there was a discrepancy between the two assay methods, the hydrolysis of 2-Abz-Phe-Arg-Phe($4NO_2$) by cathepsin X was measured using RP-HPLC.[25] The catalytic efficiency for the parent compound was then calculated and the results are shown in Fig. 35. The k_{cat}/K_M value obtained was 4.6 ± 0.15 log units, which is not statistically different from the k_{cat}/K_M calculated after performing the fluorescence assay. The results suggest that it is the presence of the NO_2 moiety in the parent compound that accounts for its elevated catalytic efficiency with respect to 2-Abz-Phe-Arg-Phe.

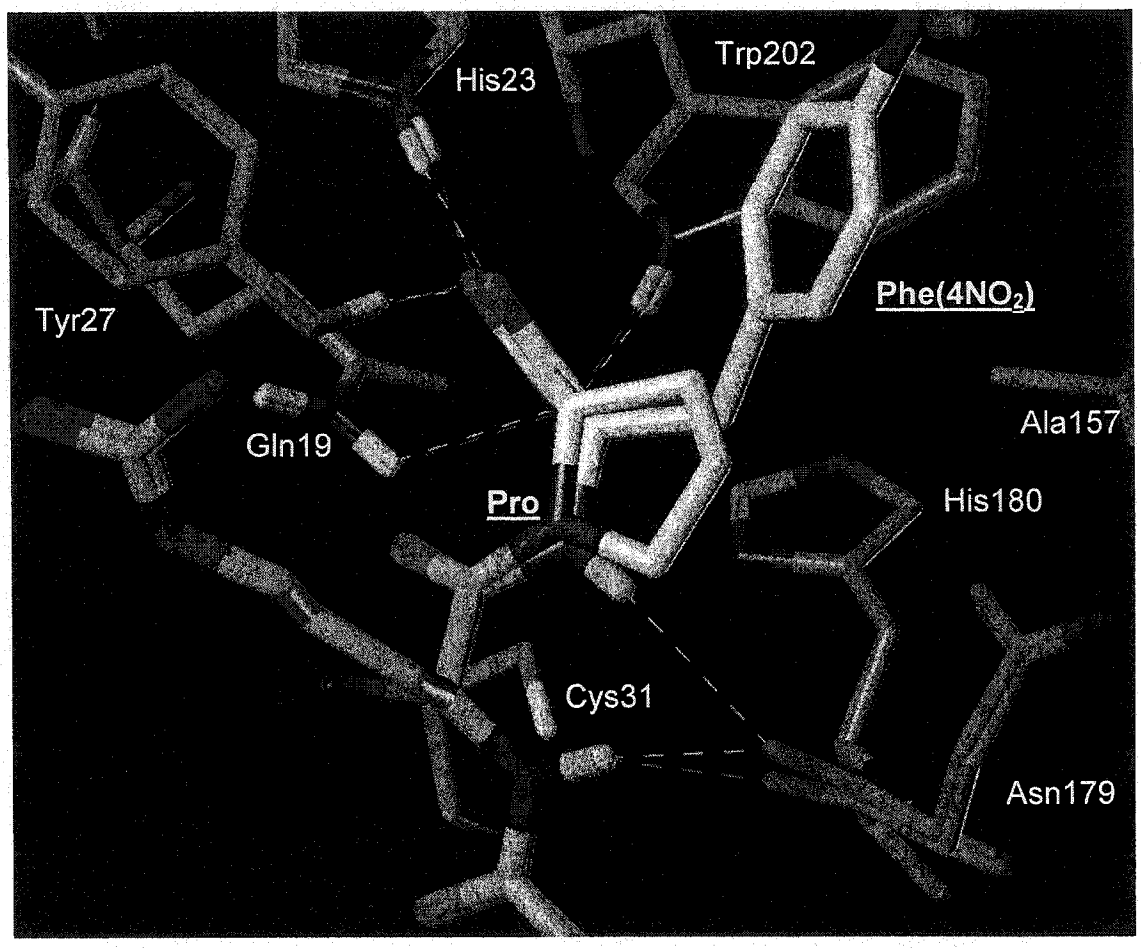


Figure 34: Comparison of Pro with Phe($4NO_2$) at the P1' position illustrating the loss of one hydrogen bond. Residues in purple belong to cathepsin X. The substrate is shown in green with the P1' side chains in white. Hydrogen bonds between Phe($4NO_2$)-containing substrate and cathepsin X are shown using yellow dashed lines. Hydrogen bonds between Pro-containing substrate and cathepsin X are shown using red dashed lines. This model was generated as described in section 2.4.

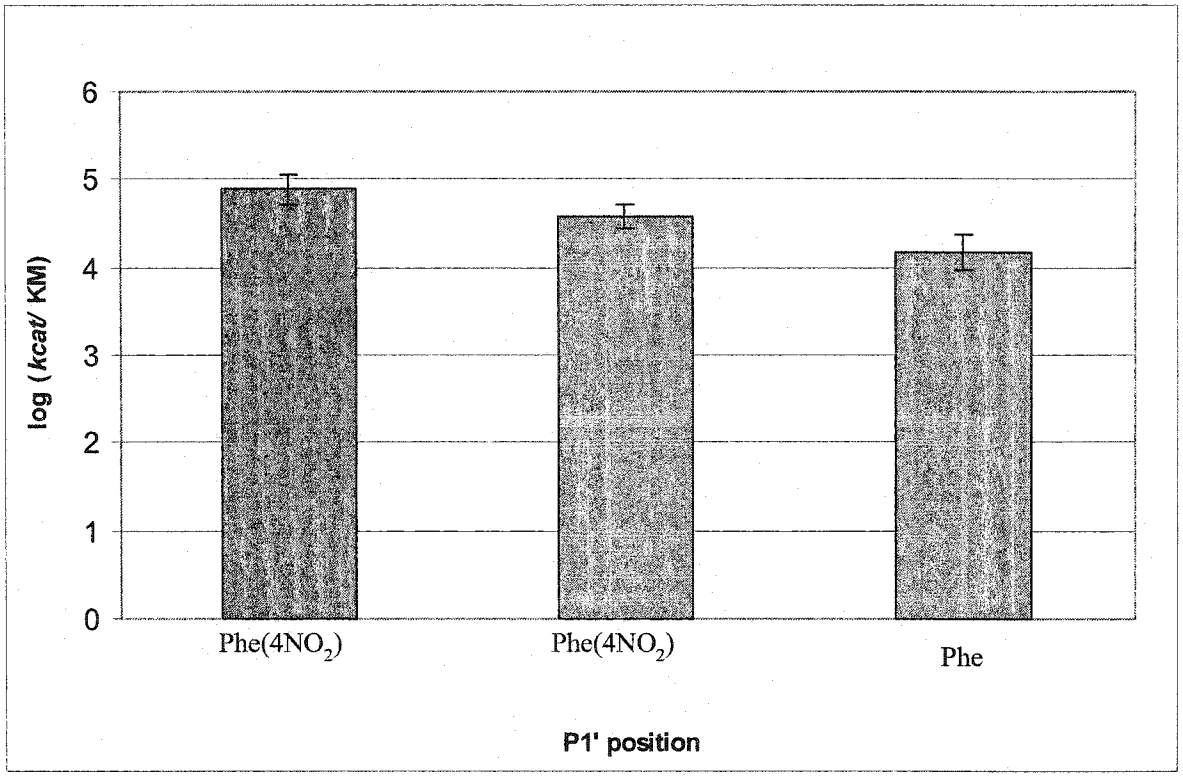
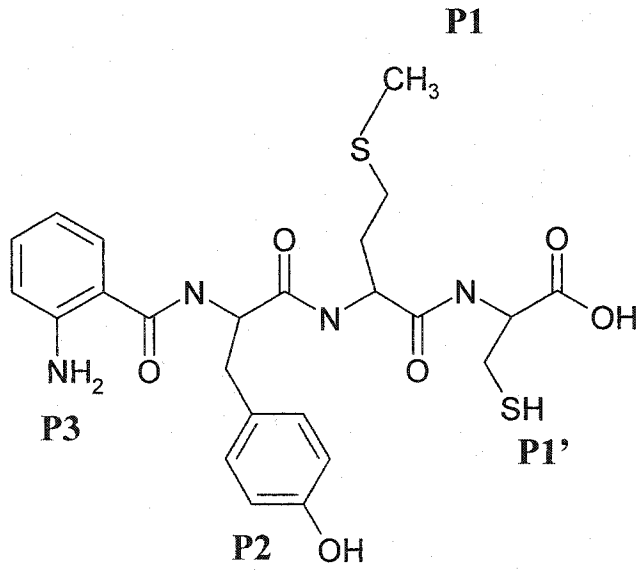


Figure 35: Comparison of catalytic efficiency values for the parent substrate, 2-Abz-Phe-Arg-Phe($4NO_2$), and 2-Abz-Phe-Arg-Phe after RP-HPLC analysis of both compounds (shown in purple). Catalytic efficiency of 2-Abz-Phe-Arg-Phe($4NO_2$) from fluorescence assay is shown in green for comparison.

In summary, the S1'subsite cathepsin X shows variability for preference of P1' peptidyl side chains in accordance with other cathepsins. Although selectivity is higher for Cys and Ser in P1' with respect to other P1' substrates, both differ in their physicochemical properties.

Conclusion

The results of this study demonstrate that cathepsin X exhibits a broad substrate specificity, thus being similar to the other studied human cathepsins (B, L, K, and S). [1,7,44] However, some amino acid substitutions in the tripeptide scaffold have been identified in this study yielding substrates that are more efficiently hydrolyzed. For instance, in future a substrate containing Tyr at the P2 position, Met at the P1 position, and Cys at the P1' position can be modeled for hydrolytic activity using cathepsin X.



Combinatorial chemistry can be applied to this research to generate thousands of compounds in a very short period of time, which can then be screened for enzymatic hydrolysis by cathepsin X. Substrates cleaved with high efficiency can serve as lead compounds applied in further investigations aimed at defining the role(s) of cathepsin X in both normal and pathological conditions.

References

1. McGrath, M.E. The lysosomal Cysteine Proteases. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 181-204.

2. Kirschke, H., and Barrett, A.J. Chemistry of lysosomal proteases. In *Lysosomes: Their role in protein breakdown.* (Glaumann H., Ballard F.J. eds) **1987**, pp 193-238, Academic Press, London.

3. Kirschke, H., Barrett, A.J., and Rawlings, N.D. Lysosomal cysteine proteases. *In Protein Profile* (Sheterline, P., ed.) **1995**, Vol.2, pp. 1581-1643. Academic Press, London.

4. Varughese, K.I., and Storer, A.C. Crystal structure of papain-E-64 complex. *Biochemistry* **1989**, *28*, 1330-1332.

5. Kim, M.J., and Kitamura, K. Crystal structure of papain-E-64-c complex. Binding diversity of E-64-c to papain S2 and S3 subsites. *Biochem J.* **1992**, *287*, 797-803.

6. Storer, A. C., Menard, R. Catalytic Mechanism in Papain family of Cysteine Peptidases. *Methods in Enzymology* **1994**, Vol 244, 486-500.

7. Nagler, D.K., and Menard, R. Human cathepsin X: a novel cysteine protease of the papain family with a very short proregion and unique insertions. *FEBS Lett.* **1998**, *434*, 135-139.

8. Santamaria, I., Velasco, G., Pendas, A.M., Fueyo, A., and Lopez-Otin, C. Cathepsin Z, a novel cysteine proteinase with a short propeptide domain and a unique chromosomal location. *J. Biol. Chem.* **1998b**, *273*, 16816-16823.

9. Pungercar, J., and Ivanovski, G. Identification and molecular cloning of cathepsin P, a novel human putative cysteine protease of the papain family. *Pflugers Arch.* **2000**, *439*, R116-R118.

10. Sivaraman, J., Nagler, D.K., Zhang, R., Menard, R., and Cygler, M. Crystal structure of human procathepsin X: A cysteine protease with the proregion covalently linked to the active site cysteine. *J. Mol. Biol.* **2000**, *295*, 939-951.

11. Guncar, G., Klemencic, I., Turk, B., Turk, V., Karaoglanovic-Carmona, A., Juliano, L., and Turk, D. Crystal structure of cathepsin X: a flip-flop of the ring of His23 allows carboxy-monopeptidase activity of the protease. *Structure* **2000**, *8*, 305-313.

12. Deussing, J., Tisljar, K., Papazoglou, A., and Peters, C. Mouse cathepsin F: cDNA cloning, genomic organization and chromosomal assignment of the gene. *Gene* **2000**, *251*, 165-173.

13. Appenzeller, C., Andersson, H., Kappeler, F., and Hauri, H-P. The lectin ERGIC-53 is a cargo transport receptor for glycoproteins. *Nat Cell Biol.* **1999**, *1*, 330-334.

14. Chapman, H.A., Riese, R.J., and Shi, G.P. Emerging roles for cysteine proteases in human biology. *Annu. Rev. Physiol.* **1997**, *59*, 63-88.

15. Klemencic, I., Carmona, A.K., Cezari, M.H., Juliano, M.A., Juliano, L., Guncar, G., Turk, D., Krizaj, I., Turk, V., and Turk, B. Biochemical characterization of human cathepsin X revealed that the enzyme is an exopeptidase, acting as carboxymonopeptidase or carboxydipeptidase. *Eur. J. Biochem.* **2000**, *267*, 5404-5412.

16. Therrien, C., Lachance, P., Sulea, T., Purisima, E.O., Qi, H., Ziomek, E., Alvarez-Hernandez, A.D., Rousch, W.R., and Menard, R. Cathepsins X and B can be differentiated by their respective mono- and dipeptidyl carboxypeptidase activities. *Biochemistry.* **2001**, *40*, 2702-2711.

17. Menard, R., Therrien, C., Lachance, P., Sulea, T., Qo, H., Alvarez-Hernandez, A.D., and Rousch, W.R. Cathepsins X and B display distinct activity profiles that can be exploited for inhibitor design. *Biol. Chem.* **2001**, *382*(*5*), 839-845.

18. Nagler, D.K., Zhang, R., Tam, W., Sulea, T., Purisima, E.O., and Menard, R. Human Cathepsin X: a Cysteine Protease with unique Carboxypeptidase Activity. *Biochemistry* **1999**, *38*, 12648-12654.

19. Novabiochem 2002/2003 Catalog. *Novabiochem-Calbiochem Corp.* San Diego, Calif. 92121.

20. Marshall, G.R., *Peptides: Chemistry and Biology.* (Proceedings of the Tenth American Peptide Symposium). ESCOM Science Publishers. Leiden, Netherlands **1988**.

21. Gross, E., and Meienhofer, J. *The Peptides:Analysis, Synthesis, Biology* Vol I. Major Methods of Peptide Bond Formation. Academic Press New York, NY. **1979**.

22. Vydac 1995 Second Ed. The Handbook of Analysis and Purification of Peptides and Proteins by Reversed-Phase HPLC. *Mandel Scientific.* Hesperia CA. 92345.

23. Segel I.H. *Biochemical Calculations 2nd Edition.* John Wiley & Sons, Inc. (Canada) **1976**.

24. Neuberger, A., and Van Deenen, L.L.M. *Modern Physical Methods in Biochemistry.* Elsevier Science Publishers B.V. (Biomedical Division). Amsterdam, Netherlands **1985**.

25. Bieth, J.G. Theoretical and practical aspects of proteinase inhibition kinetics. *Methods Enzymol.* **1995**, *248*, 59-84.

26. Cornell, A.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

27. Leach, A.R., *Molecular Modelling: Principles and Applications* Second Edition. Pearson Education Limited. Prentice Hall. Essex England **1996**.

28. Fauchere, J-L., and Pliska, V. Hydrophobic parameters (π) of amino-acid sidechains from the partitioning of N-acetyl-amino-acid amides. *Eur. J. Med.* **1983**, *No.4*. 369-375.

29. Fauchere, J-L., Charton, M., Lemont, K.B., Verloop, A., and Pliska, V. Amino acid side chain parameters for correlation studies in biology and pharmacology. *Int. J. Peptide Protein Res.* **1988**, *33*, 269-278.

30. Collantes, E.R., and Dunn, III W.J. Amino Acid Side Chain Descriptors for Quantitative Structure-Activity Relationship Studies of Peptide Analogues. *J. Med. Chem.* **1995**, *38*, 2705-2713.

31. Sotomatsu-Niwa, T., and Ogino, A. Evaluation of the hydrophobic parameters of amino acid side chains of peptides and their application in QSAR and conformational studies. *Journal of Molecular Structure* (*Theochem*). **1997**, *392*, 43-54.

32. Musil, D., Zucic, D., Turk, D., Engh, R.A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katanuma, N., and Bode, W. The refined 2.15 A X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *The EMBO Journal* **1991**, *vol 10*, 2321-2330.

33. Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A.C., and Mort, J.S. Structure of rat procathepsin B:model for inhibition of cysteine protease activity by the proregion. *Structure* **1996**, *vol 4*, 405-416.

34. Cezari, M.H., Puzer, L., Juliano, M.A., Carmona, A.K., and Juliano, L. Cathepsin B carboxypeptidase specificity analysis using internally quenched fluorescent peptides. *Biochem J.* **2002**, *368*, 365-369.

35. Fujushima, A., Imai, Y., Nomura, T., Fujisawa, Y., Yamamota, Y., and Sugawara, T. Crystal Structure of human cathepsin L complexed with E-64. *FEBS Lett.* **1997**, *407*, 47-50.

36. Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J.S., and Cygler, M. Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.* **1996,** *15*, 5492-5503.

37. McGrath, M.E., Klaus, J.L., Barnes, M.G., and Bromme, D. Crystal structure of human cathepsin K complexed with a potent inhibitor. *Nature Struct. Biol.* **1997**, *4*, 105-109.

38. Zhao, B., Janson, C.A., Amegadzie, B.Y., D'Allesio, K., Griffin, C., Hanning, C.R., Jones, C., Kurdyla, J., McQueney, M., Qui, X., Smith, W.W., and Abdel-Meguid, S.S. Crystal structure of human osteoclast cathepsin K complex with E-64. *Nature Struct. Biol.* **1997**, *4*, 109-111.

39. McGrath, M.E., Palmer, J.T., Bromme, D., and Somoza, J.R. Crystal structure of human cathepsin S. *Protein Science* **1998**, *7*, 1294-1302.

40. Bromme, D., and McGrath, M.E. High level expression and crystallization of recombinant human cathepsin S. *Protein Science* **1996**, *5*, 789-791.

41. Nagler, D.K., Tam, W., Storer, A.C., Krupa, J.C., Mort, J.S., and Menard, R. Interdependency of sequence and positional specificities for cysteine proteases of the papain family. *Biochemistry* **1999**, *38*, 4868-4874.

42. Rullmann, J.A.C., Bellido, M.N., and van Guijnen, P.T. The active site of papain: All-atom study of interactions with protein matrix and solvent. *J. Mol .Biol.* **1989**, *206*, 101-118.

43. Schechter, I., and Berger, A. On the size of the active site in proteases. I. Papain. *Biochem .Biophys. Res. Commun.* **1967**, *27*, 157-162.

44. Turk, B., Turk, D., and Turk, V. Lysosomal cysteine proteinases: more than scavengers. *Biochim. Biophys. Acta.* **2000**, *1477*, 98-111.

45. Tao, K., Sterns, N.A., Dong, J., Wu, Q., and Sahagian, G.G. The proregion of cathepsin L is required for proper folding, stability, and ER exit. *Arch. Biochem. Biophys.* **1994**, *311*, 19-27.

46. Bromme, D., Bonneau, P.R., Lachance, P., Wiederanders, B., Kirschke, H., Peters, C., Thomas, D.Y., Storer, A.C., and Vernet, T. Functional expression of human cathepsin S in *Saccaromyces cerevisiae. J .Biol. Chem.* **1993**, *268*, 4832-4838.

47. Kirschke, H., Wiederanders, B., Bromme, D., and Rinne, A. Cathepsin S from bovine spleen: purification, distribution, intracellular localization, and action on proteins. *Biochem J.* **1989**, *264*, 467-473.

48. Lalonde, J.M., Zhao, B., Janson, C.A., D'Alessio, K.J., McQueney, M.S., Orsini, M.J., Deboueck, C.M., and Smith, W.W. The crystal structure of human procathepsin K. *Biochemistry* **1999**, *38*, 862-869.

49. Sivaraman, J., Lalumiere, M., Menard, R., and Cygler, M. Crystal structure of wild-type human procathepsin K. *Protein Science* **1999**, *8*, 283-290.

50. Billington, C.J., Mason, P., and Magny, M.C. The slow-binding inhibition of cathepsin K by its propeptide. *Biochem. Biophys .Res. Commun.* **2000**, *273*, 924-929.

51. Hasnain, S., Hirama, T., Huber, C.P., Mason, P., and Mort, J.S. Characterization of cathepsin B specificity by site-directed mutagenesis. Importance of Glu245 in the S2-P2 specificity for arginine and its role in transition state stabilization. *J. Biol. Chem.* **1993**, *268*, 235-240.

52. Illy, C., Quarishi, O., Wang, J., Purisima, E., Vernet, T., and Mort J.S. Role of the occluding loop in cathepsin B activity. *J. Biol. Chem.* **1997**, *272,* 1197-1202.

53. Turk, B., Turk, V., and Turk, D. Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. *Biol. Chem.* **1997**, *378*, 141-150.

54. Baricos, H.W., Zhou, Y., Mason, R.W., and Barrett, A.J. Human kidney cathepsins B and L. Characterization and potential role in degradation of glomerular basement membrane. *Biochem. J.* **1988**, *252*, 301-30.

55. Kos, J., and Lah, T.T. Cysteine proteinases and their inhibitors: target proteins for prognosis, diagnosis and therapy in cancer. *Oncol. Reports.* **1998**, *5*, 1349-136.

56. Guay, J., Falgueyret, J.P., and Ducret, A. Potency and selectivity of inhibition of cathepsins K, L, and S by their respective propeptides. *Eur. J. Biochem* **2000**, *267*, 6311-6318.

Appendix A: Solid Phase Peptide Synthesis

Solid phase peptide synthesis is based on the sequential addition of α amino and side chain protected amino acid residues to an insoluble polymeric support[19] (Figs. 34, 35). The acid-labile Boc group or base-labile Fmoc group is used for N-α-protection (Fig. 36). The development of Fmoc solid phase synthesis arose out of concern that repetitive TFA acidolysis could lead to alteration of sensitive peptide bonds.[21] In Fmoc synthesis, the growing peptide is subjected to mild base treatment using piperidine during Fmoc-group deprotection and TFA is required only for the final cleavage and deprotection of peptidyl resin.[19,21] By contrast, cleavage and deprotection in Boc strategy requires the use of dangerous hydrofluoric acid and expensive laboratory equipment which is not always readily available to many researchers.[20] The purity of peptides made by Boc chemistry are comparable with that of the best Fmoc chemistry.[21] This suggests that, in skilled and experienced hands, either method can give good results. However, Fmoc-based synthesis in the hands of the "average user" can be far more accessible and more likely to provide the best avenue for routine synthesis of peptides.[19]

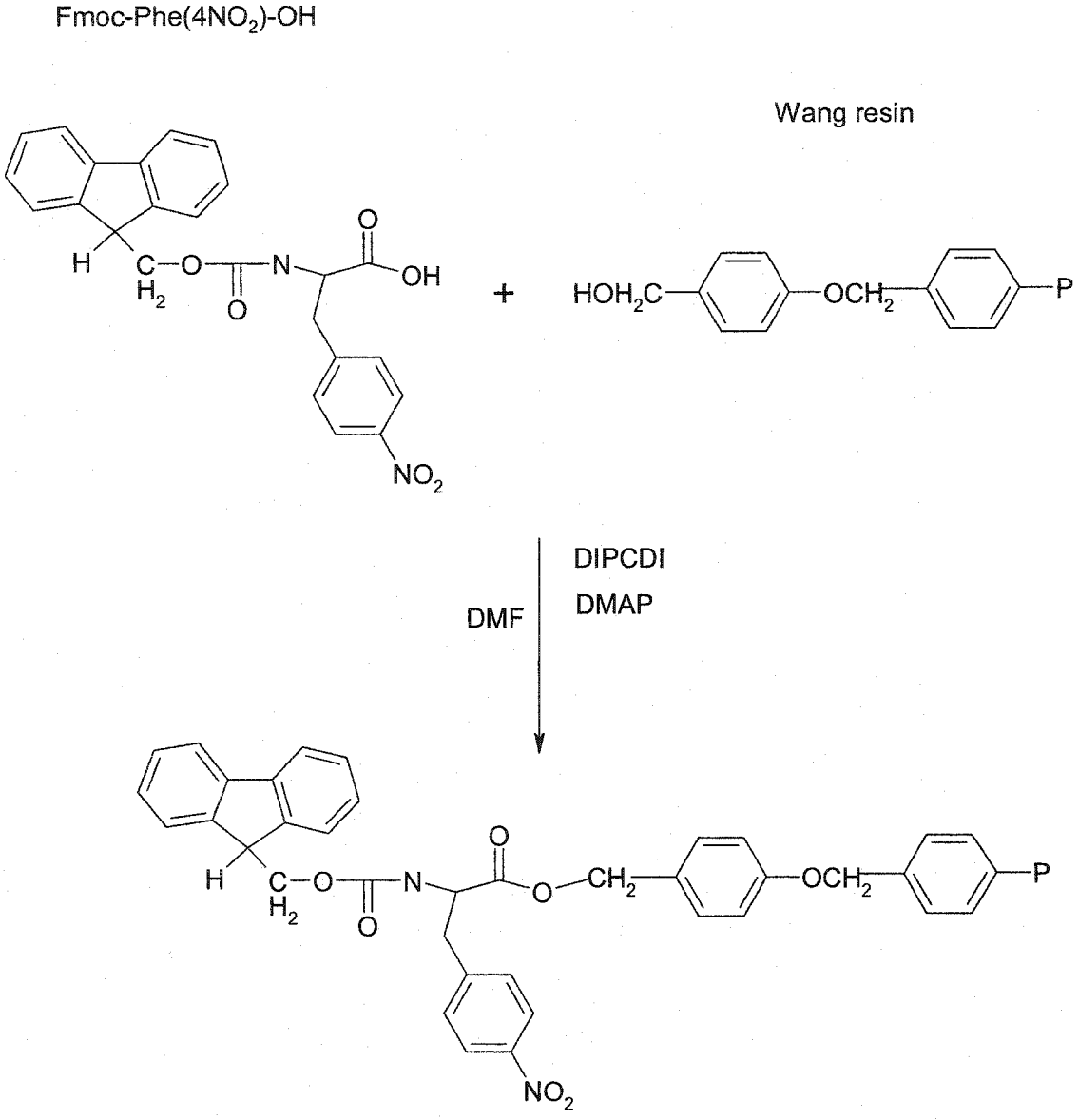


Figure 36: Coupling reaction of Fmoc-Phe(4NO_2) to Wang resin.

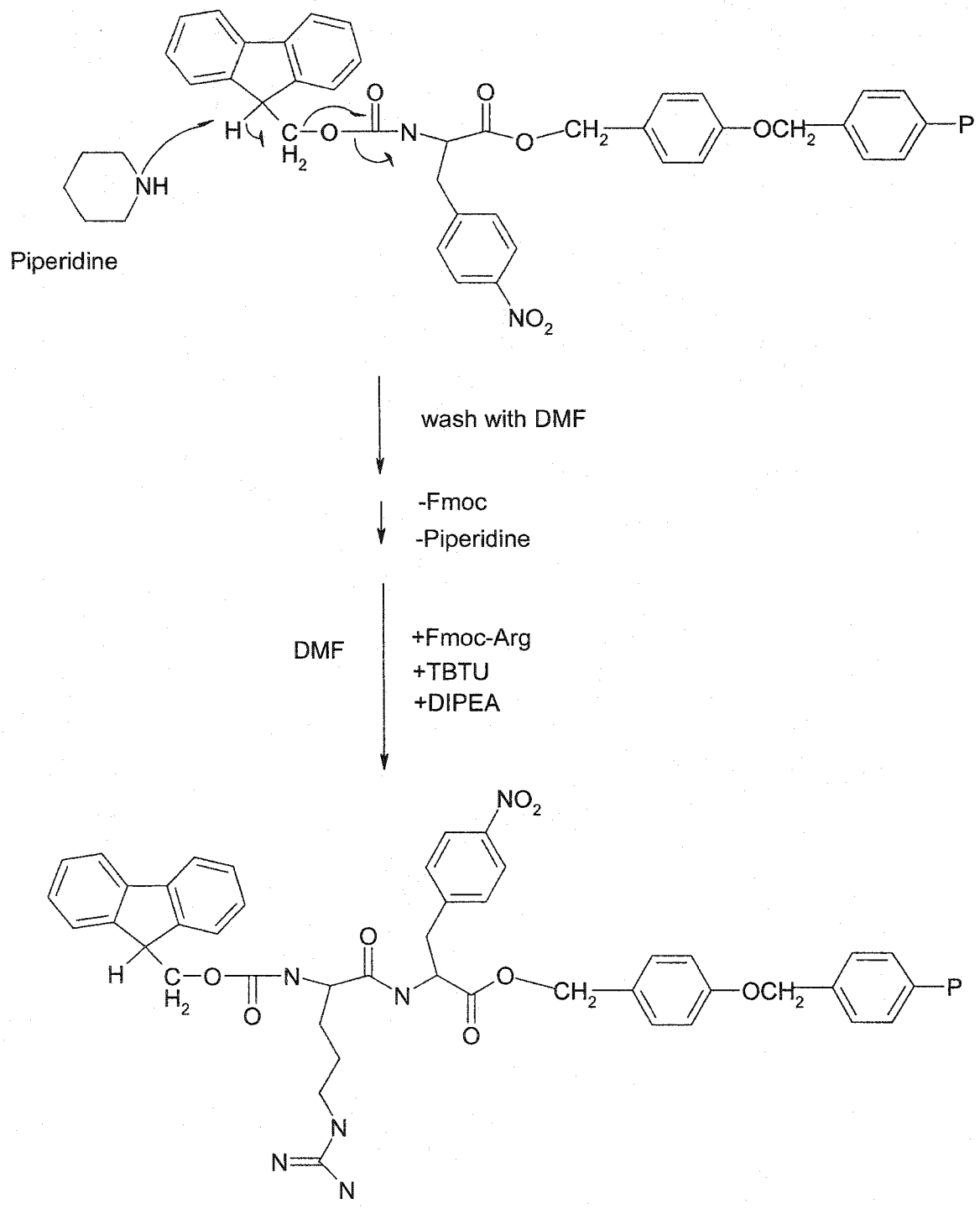


Figure 37: Coupling of Fmoc-Arg to Phe(4NO_2)-Wangresin after Fmoc removal. Deprotection and coupling can be performed n times to yield a polypeptide chain using solid phase chemistry.

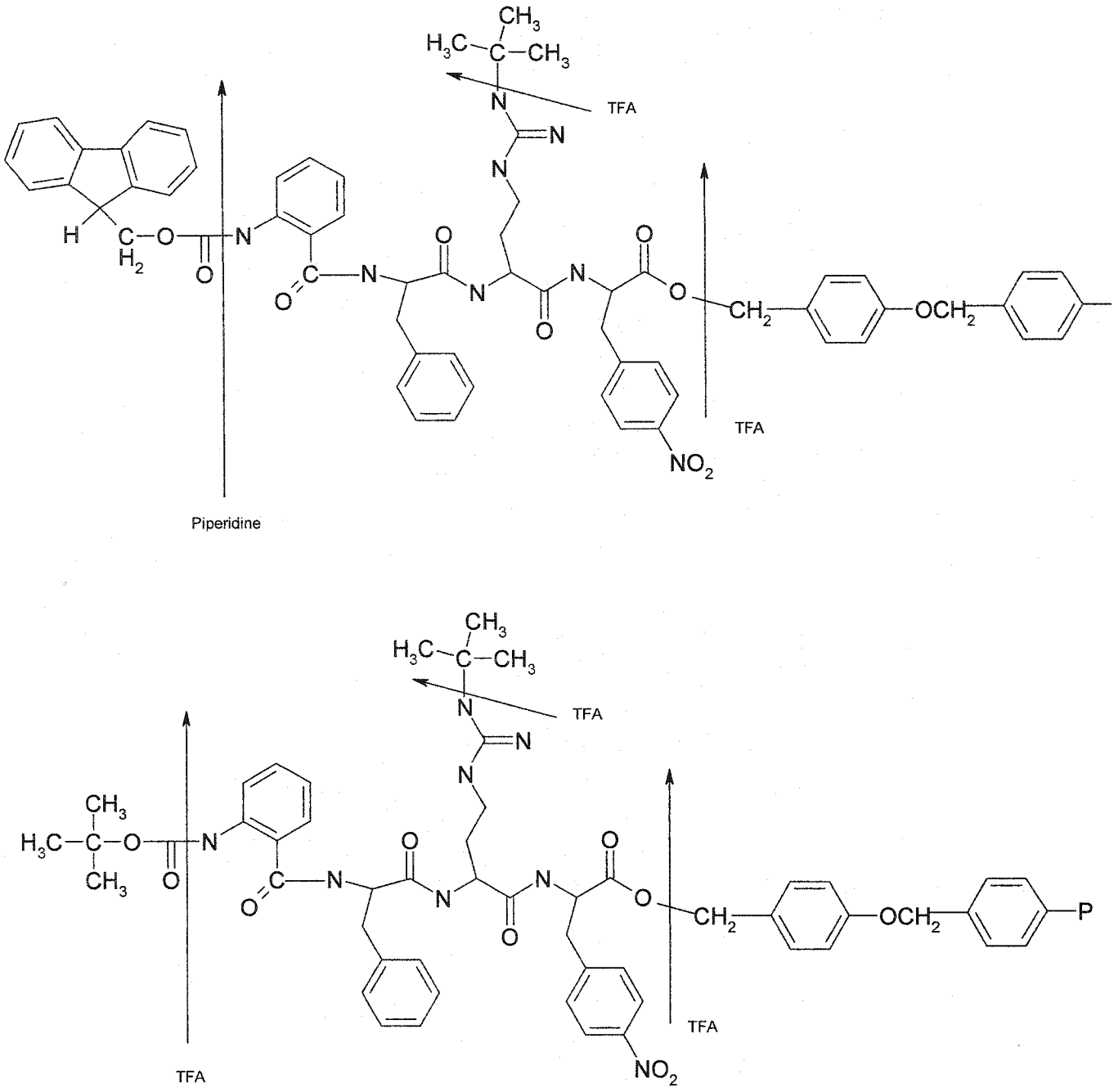


Figure 38: Protecting group strategy for solid phase synthesis. Fmoc removal requires piperidine whereas Boc removal can be accomplished using TFA, which is also used for deprotection of amino acid sidechains and peptide cleavage from the resin.

Appendix B: Comparison of cathepsin X with other cathepsins in the S2, S1, and S1' subsites

In cathepsin L, Ala-133 forms the floor of the S2 pocket and interacts with a P2 peptidyl chain, favoring the binding of aromatic residues.[36,43,45] A comparable location in the S2 pocket of cathepsin S contains a Phe205 that drives the specificity of cathepsin S toward branched hydrophobic side chains.[39,46,47] The presence of Gly160 and Gly133 in the S2 pocket of cathepsin S probably account for the binding site being larger than in the other cathepsins.[46,47] Leu at the P2 position of a substrate is preferred by cathepsin K. The presence of Leu205 at the floor of the S2 pocket renders it shallower than the S2 pockets of other cathepsins that can all accommodate a Phe residue in P2 of a substrate.[1,48-50] The S2 pocket of cathepsin B contains a Glu245 whose carboxylate is positioned to make a salt bridge with the guanidino group of an Arg side chain.[1,51,52] The acceptance of a charged residue in the S2 subsite of cathepsin B contrasts the preference of other cathepsins for a hydrophobic residue such as Phe in the P2 position of a substrate.[1,46-48] The presence of the charged Glu245 in cathepsin B is comparable to the non-hydrophobic His234 residue at the floor of the cathepsin X S2 subsite.[10,11]

No major differences in the S1 subsite exist between cathepsin X and the other studied cathepsins. This subsite is a fairly long, narrow groove in the active sites of cathepsins. [1,3,8,53]

The most notable structural difference between the active sites of cathepsin X and the other cathepsins is seen in the S1'region.[10,11,53-55] As mentioned previously, the unique "mini-loop"[7,11] of cathepsin X consists of a three residue insert (Ile24-Pro25-Gln26)

occupying the equivalent position of the S1' region of cathepsins L, K, and S. Similarly, cathepsin B has a structure analogous to the mini-loop of cathepsin X called the "occluding loop"[52]-an eighteen residue insertion spanning residues 106-124. X-ray analysis has shown that a selective inhibitor coupled with cathepsin B engages in hydrogen bonding with two His residues on the occluding loop, thereby stabilizing the inhibitor in the active site.[17,54,55] The occluding loop of cathepsin B therefore acts similarly to the mini-loop of cathepsin X but occludes the subsites beyond S2', i.e., that it is a flexible structure capable of accepting slightly longer substrates in the S' region of the active site. As a result, cathepsin B acts as a carboxy-dipeptidase.[54] The mini-loop is shorter and more constrained mini-loop of cathepsin X. [17,19,54] Nevertheless, cathepsin X is most similar to cathepsin B, the only other known papain-like carboxypeptidase. [7,17]

Cathepsins L, K, and S do not possess loop inserts and all function as endopeptidases cleaving within a polypeptide chain. [1,3,55,56]

Table 1: Putative determinants of Cathepsin X binding specificity with respect to other known cathepsins.

Subsite	Cathepsin X	Cathepsins B/L/K/S
S2	Asp76 His234	Pro/Pro/Met/Met Glu/Ala/Leu/Phe
S1	None	
S1'	Mini-Loop	Occluding Loop/-/-/-