Characterization of the $S_3$ Subsite Specificity of Cathepsin B

(Received for publication, May 3, 1985, and in revised form, May 16, 1985)

Alpay Taralp, Harvey Kaplan, Iou-Iou Sytwu, Isidoros Vlattas, Regine Bohacek, Anna K. Knop, Tomoko Hirama, Carol P. Huber, and Sadiq Hasnain

From the Department of Chemistry, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada, and the Research Department of the Pharmaceuticals Division, Ciba-Geigy Corporation, Summit, New Jersey 07901, and the Institute for Biological Sciences, National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada

Five synthetic substrates containing different amino acid residues at the $P_3$ position (acetyl-X-Arg-Arg-AMC, where X is Gly, Glu, Arg, Val, and Tyr) and where AMC represents 7-amido-4-methylcoumarin) were used to investigate the $S_3$ subsite specificity of cathepsin B. At pH 6.0, the specificity constant, $k_{cat}/K_m$, for tripeptide substrate hydrolysis was observed to increase in the order Glu $<$ Gly $<$ Arg $<$ Val $<$ Tyr. Molecular modeling studies of substrates containing a P$_3$ Glu, Arg, or Tyr covalently bound as the tetrahedral intermediate to the enzyme suggest that the specificity for a P$_3$ Tyr is because of a favorable aromatic-aromatic interaction with Tyr$^\beta_6$ on the enzyme, as well as a possible H bond between the P$_3$ Tyr hydroxyl and the side chain carboxyl of Asp$^\alpha$.

Lysoosomal cathepsin B (EC 3.4.22.1), a cysteine proteinase normally active in protein turnover, has been implicated in several pathological conditions, including arthritis, muscular dystrophy, and tumor metastasis (1-7). It is unique among the cysteine proteinases in that it has both endopeptidase and exopeptidase activities, can accept basic residues in a substrate at both the P$_1$ and P$_2$ positions, and has complex pH dependences (8-11). The detailed analysis of the pH dependence of cathepsin B-catalyzed hydrolyses has revealed that at least seven groups that dissociate in the pH range of 3-9 can affect substrate binding and/or turnover. Reactions with N-terminal protected monopentidyldipeptide and dipeptidyl substrates or various irreversibly inhibitors demonstrated that the $S_3$ subsite prefers positively charged or straight chained aliphatic amino acids but exhibits a very poor affinity for negatively charged or bulky aromatic and branched aliphatic amino acids (11, 13-16). The $S_3$ subsite was shown to exhibit a preference for phenylalanine and arginine, although the specificity constant, $k_{cat}/K_m$, is 7-fold higher for the former (12). Interpretation of pH activity profile data, obtained with substrates containing an arginine at P$_3$ and either an arginine or phenylalanine at P$_3$, suggested the presence of carboxyl groups in the $S_3$ and $S_2$ subsites with $pK_a$ values of 5.4 and 5.1, respectively (12). The identity of the group with the $pK_a$ of 5.1 was confirmed to be Glu$^{249}$ by the kinetic characterization of the site-directed mutants Glu$^{249}$ $\rightarrow$ Glu$^{249}$ and Glu$^{249}$ $\rightarrow$ Ala$^{245}$ (17).

In contrast to the understanding of the $S_1$ and $S_2$ subsite specificities, $S_3$ specificity remains largely uncharacterized. Brönne et al. (18) have shown that the binding of a series of N-terminal succinylated alanine-containing peptide substrates could be improved by increasing the length of the peptide chain. When comparing tripeptide and dipeptide alanine substrates, the former demonstrated a 2-fold increase in the specificity constant $k_{cat}/K_m$. In addition, the $S_3$ subsite was shown to accept proline, unlike subsites $S_1$ or $S_2$ (18). Brönne et al. (19) also examined $S_3$ and $S_2$ specificity using six tripeptide and tetrapeptide coumaryl substrates. For two of these substrates, with valine and phenylalanine as the $P_3$ residue, they found no significant difference in binding energy. However, since the N-terminal blocking groups were not the same for the substrates and the $P_3$ residues selected did not have a sufficiently wide range of physicochemical properties, the optimal specificity of the $S_3$ subsite remained unresolved. In a different approach, Koga et al. (20) digested soluble denatured proteins with cathepsin B. Fragments arising from endopeptidase activity were analyzed, and interpretation of the results suggested a preference for the amino acids glycine, tryptophan, alanine, lysine, isoleucine, and proline at the $P_3$ position of the substrate. Although only qualitative, the results, excluding the observation of lysine at $P_3$, supported the possibility of a hydrophobic pocket at the $S_3$ subsite.

In the present paper we describe the synthesis and kinetic analysis of substrates of the type acetyl-X-Arg-Arg-AMC (where X is Gly, Glu, Arg, Val, or Tyr). These substrates were designed to define specificity at the $S_3$ subsite. In addition, the change in free energy of binding ($\Delta^\circ G$) values for these substrates was determined to compare the relative strengths of binding of the different $P_3$ side chains in the $S_3$ subsite. Finally, molecular modeling studies were carried out for the tripeptide substrates containing either Glu (in its protonated and deprotonated states), Arg, or Tyr in the $P_3$ position to gain further insights into enzyme-substrate interactions.

**EXPERIMENTAL PROCEDURES**

**Enzyme Expression and Purification**—The cDNA for rat cathepsin B was expressed in Saccharomyces cerevisiae as an n-factor fusion construct (21). Yeast starter cultures were grown in synthetic medium and then were grown in 4-liter shake flasks as reported before (22, 23). The recombinant enzyme was harvested and purified as described previously (21).

**Enzyme Assays for the Determination of Kinetic Constants**—Prior to

---

*This work was supported in part by a Natural Sciences and Engineering Research Council of Canada operating grant (to H.K.). This is National Research Council of Canada Publication 37414. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Recipient of a Natural Sciences and Engineering Research Council of Canada graduate scholarship.

‡‡ Present address: Biotechnology Research Institute, National Research Council of Canada, M5L 054, Ottawa ON K1A 0R6, Canada.

†† Present address: Marketing Services Office, Corporate Services, National Research Council of Canada, Ottawa ON K1A 0G6, Canada. To whom correspondence and reprint requests should be addressed: Tel.: 613-993-6040; Fax: 613-952-4660.
kinetic analysis, the enzyme concentration was determined by active site thiol titration using the cysteine proteinase inhibitor E-64 (N-N-[2-(4-carboxy-2-methyl-1,4-naphthylene)-2-carbonyl]-L-cysteinyl-L-cystatin) (Boehringer Mannheim). The enzyme was isolated according to the method of Barret and Kirschke (11). Just before conducting experiments for the determination of kinetic constants, the enzyme, in 50 mM sodium phosphate buffer, pH 6.0, 1 mM EDTA, was activated for 1 h by the addition of dithiothreitol to a final concentration of 10 mM and then incubated in pyridine. Due to the enzyme concentration was shown to be unchanged for the duration of the experiments. Michaelis-Menten kinetic parameters for the corresponding substrates were determined at 24 °C by continuous fluorescence spectrophotometry, using an SLM Aminco DW2000 spectrophotometer equipped with the Total Fluorecence accessory (12). Assays were analyzed by the nonlinear regression data analysis program Enzfitter, developed by Elsevier-Biosoft.

Assays were performed in 25 mM sodium phosphate buffer, 250 mM NaCl, 1 mM EDTA, 3% dimethyl sulfoxide, pH 6.0 or 7.7, and in 25 mM citrate buffer, 250 mM NaCl, 1 mM EDTA, 3% dimethyl sulfoxide, pH 4.0. The pH-activity profiles were determined for two substrates, acetyl-Val-Arg-Arg-AMP and acetyl-Arg-Arg-AMP. These pH-activity profiles were determined over the pH range 3.2–6.4 by measuring Kₐₑ₅₀/Kₚ at 0.1 pH unit intervals, using the relationship Kₑ₅₀ = ε[(S)[E][S]] when [S] = Kₑ₅₀. Assay samples were monitored to ensure that pH was stable over the course of the reaction. Over the pH range 3.2–6.0, the reaction buffer consisted of 25 mM sodium citrate, 250 mM NaCl, 1 mM EDTA, and 3% dimethyl sulfoxide, and in the pH range of 7.7–8.4, the reaction buffer was 25 mM sodium phosphate, 250 mM NaCl, 1 mM EDTA, and 3% dimethyl sulfoxide.

The citrate buffer consistently afforded lower activity measurements when compared with the phosphate buffer through the pH range 5.7–6.0, and for this reason there was a break in the continuity of the data of the pH-activity profile. Similar results were observed by Hassain et al. (12) for the substrates benzoylformylarginyl-Arg-Val-Arg-Val (where X is AMP or γ,δ-polyribose). The data in citrate buffer were corrected by a factor determined from measurements in phosphate buffer at four separate pH intervals (5.7–6.0) where citrate and phosphate buffers overlapped, as described by Hassain et al. (12). For the tripeptide substrates in this investigation, a similar effect was evident at pH 7.7, where borate and phosphate buffers overlapped, with greater activity in borate buffer. It was determined previously (12) that pH 7.7 was the only value in the pH range where both phosphate and borate buffers could be made to overlap under the ionic strength conditions used in the experiment. Therefore, the following strategy was used to correct for this buffer effect. The enzyme activity measurements at pH 7.7 for the substrate acetyl-Val-Arg-Arg-AMP in each buffer system were carefully repeated in triplicate and were shown to be highly reproducible. By averaging the enzyme activities in each buffer system at pH 7.7 and by calculating the ratio of these averaged activities for both buffer systems, a normalizing factor with small standard deviation was determined. Graphical analysis of the corrected data for two independent experiments with acetyl-Val-Arg-Arg-AMP was performed using the substrate acetyl-Arg-Arg-AMP demonstrated that the profiles gave a best fit to the four-proton ionization model of Hasain et al. (12), where dipeptide AMP substrates with an arginyl residue in the position were employed. The previous work supported the validity of this correction for the borate buffer.

Substrate Synthesis—The peptides acetyl-Arg-Arg-Arg and acetyl-Val-Arg-Arg were synthesized on a Saccarin resin solid support by the employment of standard Fmoc methodology. The side chain functionalization of Fmoc-Arg-Arg-Arg was blocked with the 2,5,7,8-pentamethoxychroman-6-sulfonyl group. Amino acids were coupled using HBTU in N-methylpyrrolidine. After being cleaved from the resin with 1% trifluoroacetic acid in dichloromethane, the peptides were coupled to AMMC with EDC in dimethylformamide and dichloromethane again in the presence of 1-hydroxybenzotriazole. Deblocking the completed substrate was effected by treatment with 95% trifluoroacetic acid and 5% water.

The substrates acetyl-Arg-Arg-AMP, acetyl-gluf-Arg-Arg-AMP, acetyl-Glu-Arg-Arg-AMP, and acetyl-Arg-Arg-AMP were prepared in solution by the employment of N₅,N₆,N₇,N₈-tetraacyclohexylcarbodiimide. The side chain functionalities of Arg, Tyr, and Glu were protected as p-toluenesulfonyl, 2-bromobenzylxycarbonyl, and benzyl derivatives, respectively. Peptide coupling reactions with dichlohexylcarbodiimide were generally achieved in a double solvent system. Typically, dichloromethane was used to dissolve the t-Boc amino acid and coupling agent. To this was combined the growing AMC peptide, previously dissolved in N,N-dimethylacetamide, which was then mixed with 1-hydroxybenzotriazole. Sequential removal of the t-Boc group required treatment with 50% trifluoroacetic acid in dichloromethane at room temperature. Substrates were subsequently acetylated with acetic anhydride in 1:10 of were8mentation of the side groups was affected by treatment with 10% v/v) mixture of hydrofluoric acid/anhidrole.

All substrates were purified by reverse-phase high performance liquid chromatography using varied linear gradients between 0 and 60% acetonitrile, in 0.1% (v/v) trifluoroacetic acid. Substrate concentration and purity were verified by amino acid, mass spectral, and 1H NMR spectral analyses.

Evidence Supporting Exclusive Hydrolysis of Substrates at the AMC Peptide Bond—Exclusive hydrolysis of the AMC peptide bond was established by using high voltage paper electrophoresis to separate the products after incubation of substrates with cathepsin B. Since AMC is ninhydrin-negative, the basic premise of this strategy was that cleavage at any site, other than the AMC peptide bond, would generate a free α-aminogroup, which would yield a ninhydrin-positive peptide. Substrates were incubated with cathepsin B in pH 6.5 buffer of the composition acetic acid/pyridine/water (3:100:1,800 v/v). Reaction completion was verified spectrophotometrically, whereupon the digests were applied to 3MM Whatman paper and subjected to electrophoresis in pH 5.6 at a voltage gradient of 250 volts/cm for 30 min. All substrate digestions yielded products that tested ninhydrin-negative, demonstrating that cleavage occurs only at the AMC peptide bond. After spraying the chromatogram with the modified Sakaguchi reagent (14), arginine residues were identified as a single red band, indicating that there was only one product containing arginine. The final position of this fragment and its lack of fluorescence under ultraviolet light were consistent with the expected behavior of each substrate after cleavage at the AMC peptide bond. The absence of other Sakaguchi-positive material also ruled out other possible cleavage reactions. Standards of unwanted cleavage products were employed at predetermined concentrations and demonstrated that undesirable digest fragments would have been detected at the 1 mol % level.

Molecular Modeling of Enzyme-Substrate Interactions—Modeling studies were carried out for the tripeptide substrates containing a P₃ Tyr, Arg, or Glu, the last in the protomoted and deprotonated forms. Each substrate was constructed using, as a starting point, the x-ray structure of the inhibitor, Pro-Phe-Arg-CMK bound to cathepsin B (8). The MACROMODEL molecular modeling package (25) was used for this operation. The substrate was covalently attached to the enzyme simulating a tetrahedral intermediate by forming a covalent bond between the active-site thiolate of Cys²⁹ and the P₁ carbonyl carbon of the substrate.

To refine these enzyme-substrate complexes, low energy conformers were generated using simulate annealing followed by energy minimization. Each complex was subjected to 20 simulating annealing experiments. In each experiment, the complex was subjected to dynamic simulation during which the temperature varied from 300 to 30 °C over a period of 3 ps. The time step was 3 fs. These procedures were carried out using the QXP program developed at Ciba-Geyco with parameters from the AMBER force field (28). Comparison of the positions of the enzyme atoms from the x-ray structures of cathepsin B complexed with several different inhibitors showed that the positions of most of the enzyme atoms did not change significantly with binding of different inhibitors (27). Therefore, most of the enzyme atoms were held stationary during energy minimization. Residues that were allowed to move were Glu²⁵, Cys²⁹, Arg¹², Glu¹³⁹, and Glu⁴⁴⁰.

RESULTS AND DISCUSSION

Michaelis-Menten Constants

Table I lists the substrates and the kinetic constants for their hydrolysis by cathepsin B. At pH 6.0, Kₐₑ₅₀/Kₚ for the substrate acetyl-Arg-Arg-AMP was 4-fold lower than benzoylcarbonyl-Arg-Arg-AMP. The difference appears largely due to an increase in Kₑ₅₀, which can be attributed to the substitution of the benzoylcarbonyl group by the acetyl group. This is consistent with the expected steric effects of the benzoylcarbonyl group, which is bulkier than the acetyl group.
TABLE I

Kinetic constants for recombinant cathepsin B expressed in yeast

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$k_{cat}$</th>
<th>$K_m$</th>
<th>$k_{cat}/K_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH 6.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z-Arg-Arg-AMC$^a$</td>
<td>51.4 ± 5.45</td>
<td>1.11 ± 0.17</td>
<td>46,400 ± 3,260</td>
</tr>
<tr>
<td>Acetyl-Arg-Arg-AMC</td>
<td>98.9 ± 14.6</td>
<td>8.67 ± 1.63</td>
<td>11,500 ± 1,080</td>
</tr>
<tr>
<td>Acetyl-Glu-Arg-Arg-AMC</td>
<td>&gt;10</td>
<td>8,730 ± 660</td>
<td></td>
</tr>
<tr>
<td>Acetyl-Gly-Arg-Arg-AMC</td>
<td>69.2 ± 4.84</td>
<td>4.60 ± 0.75</td>
<td>15,200 ± 1,600</td>
</tr>
<tr>
<td>Acetyl-Arg-Arg-AMC</td>
<td>120 ± 23.7</td>
<td>2.00 ± 0.35</td>
<td>59,600 ± 4,360</td>
</tr>
<tr>
<td>Acetyl-Val-Arg-Arg-AMC</td>
<td>108 ± 6.20</td>
<td>1.35 ± 0.05</td>
<td>80,500 ± 4,680</td>
</tr>
<tr>
<td>Acetyl-Tyr-Arg-Arg-AMC</td>
<td>175 ± 12.3</td>
<td>0.92 ± 0.09</td>
<td>190,000 ± 6,800</td>
</tr>
<tr>
<td>pH 4.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acetyl-Glu-Arg-Arg-AMC</td>
<td>15.0 ± 2.20</td>
<td>2.15 ± 0.21</td>
<td>7,000 ± 360</td>
</tr>
<tr>
<td>Acetyl-Arg-Arg-AMC</td>
<td>3.94 ± 2.45</td>
<td>1.24 ± 0.83</td>
<td>3,260 ± 230</td>
</tr>
<tr>
<td>Acetyl-Val-Arg-Arg-AMC</td>
<td>10.1 ± 0.81</td>
<td>1.94 ± 0.41</td>
<td>5,280 ± 730</td>
</tr>
<tr>
<td>Acetyl-Tyr-Arg-Arg-AMC$^a$</td>
<td>(17.7 ± 3.80)</td>
<td>(1.56 ± 0.40)</td>
<td>(11,500 ± 570)</td>
</tr>
<tr>
<td>pH 7.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acetyl-Arg-Arg-Arg-AMC$^a$</td>
<td>123 ± 14.1</td>
<td>1.65 ± 0.09</td>
<td>74,700 ± 4,500</td>
</tr>
<tr>
<td>Acetyl-Val-Arg-Arg-AMC</td>
<td>86.8 ± 5.00</td>
<td>0.89 ± 0.07</td>
<td>126,000 ± 10,400</td>
</tr>
</tbody>
</table>

$^a$ Data are from Hassain et al. (12).
$^b$ Data are from a single qualitative experiment, shown with standard errors.
$^c$ Data are from a duplicate run, with the averages of the two runs reported.

with the fact that the x-ray structure of the inhibitor, benzyl-oxidocarbonyl-Arg-Ser-(O-Bz)-CMK, in a cofactor complex with cathepsin B, reveals a favorable interaction between the N-terminal benzyl group and the phenyl ring of Tyr$^{76}$ (27). This x-ray structure also suggests that if the binding of the acetyl was analogous to the position of the oxycarbonyl moiety in the inhibitor's benzoxycarbonyl group, the N-terminal acetyl group, in acetyl-Arg-Arg-AMC, would not make any significant contacts with the enzyme.

For the tripeptide substrates at pH 6.0, specificity increased in the order Glu < Gly < Arg < Val < Tyr, with a 21-fold difference between the glutamate- and tyrosine-containing substrates. For the two substrates containing either arginine or valine at P$_3$, kinetic analysis at pH 7.7 showed that their second-order rate constants increased only slightly when compared with their values at pH 6.0. However, at pH 4.0, the second-order rate constants were significantly lower than the respective values at pH 6.0. The $k_{cat}/K_m$ values of the tripeptide substrates at pH 4.0 were generally 15–20-fold lower than at pH 6.0, the only exception being the result for acetyl-Glu-Arg-Arg-AMC. This major dependence of $k_{cat}/K_m$ on pH has been observed previously for dipeptide substrates of cathepsin B which contained a P$_3$ arginine (12, 17). The poor specificity observed for the substrate acetyl-Glu-Arg-Arg-AMC at pH 6.0 is best illustrated by a $K_m$ value that is too large (estimated at greater than 10 mm) to measure. At pH 4.0, however, under which condition the side chain carboxyl group of the P$_3$ glutamyl could be more than 50% protonated (assuming a $pK_a$ of 4.5) (28), there was a marked improvement in binding affinity as suggested by the drop in $K_m$ compared with the value at pH 6.0. As a result $k_{cat}/K_m$ values for this substrate were very similar at pH 4.0 and 6.0. For all other tripeptide substrates in this investigation (Table I), as well as for the substrate Z-Arg-Arg-AMC (12), the approximately 20-fold decrease in the value of $k_{cat}/K_m$ at pH 4.0, when compared with pH 6.0, is, to a larger extent, due to the protonation state of the Glu$^{445}$ side chain in the S$_3$ subsite of cathepsin B (17). It was concluded from the

TABLE II

Ionization constants and $k_{cat}/K_m$ limit values of dissociable groups that are important for substrate binding and catalysis obtained from pH activity profiles using the substrates acetyl-Arg-Arg-Arg-AMC and acetyl-Val-Arg-Arg-AMC

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$pK_a$</td>
<td>5.07 ± 0.03</td>
<td>5.06 ± 0.02</td>
<td>4.98 ± 0.02</td>
</tr>
<tr>
<td>$pK_b$</td>
<td>6.95 ± 0.09</td>
<td>6.90 ± 0.10</td>
<td>6.93 ± 0.04</td>
</tr>
<tr>
<td>$pK_c$</td>
<td>7.11 ± 0.12</td>
<td>7.09 ± 0.17</td>
<td>7.10 ± 0.08</td>
</tr>
<tr>
<td>$pK_d$</td>
<td>8.55 ± 0.02</td>
<td>8.53 ± 0.02</td>
<td>(8.55 ± 0.02)</td>
</tr>
<tr>
<td>$k_1$ (limit)</td>
<td>49,600 ± 1,350</td>
<td>46,700 ± 1,270</td>
<td>65,900 ± 1,100</td>
</tr>
<tr>
<td>$k_2$ (limit)</td>
<td>111,000 ± 6,750</td>
<td>111,000 ± 3,970</td>
<td>160,000 ± 3,600</td>
</tr>
<tr>
<td>$k_3$ (limit)</td>
<td>46,600 ± 4,790</td>
<td>78,300 ± 5,010</td>
<td>63,000 ± 7,500</td>
</tr>
</tbody>
</table>

$^a$ Data were taken from Hassain et al. (12).
investigation of site-directed mutants that the side chain carboxylate of the S$_2$ Glu$^{245}$ forms a salt bridge with the guanidinium cation of a P$_2$ arginine in dipeptide substrates, such as Z-Arg-Arg-p-nitroanilide. The strength of this interaction, which appears to stabilize the transition state complex, was shown to be weakened by 1.2 kcal/mol upon protonation of the Glu$^{245}$ side chain carboxylate. Any drop in $k_{cat}/K_m$ which may have been expected for acetyl-Glu-Arg-Arg-AMC at pH 4.0 relative to pH 6.0, due to the P$_2$-S$_2$ interaction, appears to have been compensated for by an increase in $k_{cat}/K_m$ because of an improved P$_2$-S$_3$ interaction. Indeed, if the data are normalized to account for the 20-fold drop normally observed at pH 4.0 relative to pH 6.0 for substrates with a P$_2$ Arg and a neutral P$_3$, there is an approximate 20-fold increase in affinity for the protonated side chain carboxyl of the P$_3$ Glu when compared with the deprotonated form.

For the tripeptide substrates in this investigation, with the exception of acetyl-Glu-Arg-Arg-AMC, the observed drop in the value of $k_{cat}$ and $k_{cat}/K_m$ at pH 4.0, compared with pH 6.0 (Table I), may similarly be explained by a weakening of the P$_2$-S$_2$ interaction upon the protonation of the Glu$^{245}$ side chain carboxylate. Any drop in $k_{cat}/K_m$ which may have been expected for acetyl-Glu-Arg-Arg-AMC at pH 4.0 relative to pH 6.0, due to the P$_2$-S$_2$ interaction, appears to have been compensated for by an increase in $k_{cat}/K_m$ because of an improved P$_2$-S$_3$ interaction. Indeed, if the data are normalized to account for the 20-fold drop normally observed at pH 4.0 relative to pH 6.0 for substrates with a P$_2$ Arg and a neutral P$_3$, there is an approximate 20-fold increase in affinity for the protonated side chain carboxyl of the P$_3$ Glu when compared with the deprotonated form.

### Table III

Change of free energy of binding ($\Delta G$) between (i) for the substrates Z-Arg-Arg-AMC and acetyl-Arg-Arg-AMC and (ii) and for tripeptide substrates with amino acid substitutions at P$_2$ and the reference substrate acetyl-Gly-Arg-Arg-AMC

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$k_{cat}/K_m$ (reference)</th>
<th>$\Delta G$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetyl-Arg-Arg-AMC</td>
<td>0.246</td>
<td>0.62</td>
</tr>
<tr>
<td>Acetyl-Glu-Arg-Arg-AMC</td>
<td>0.574</td>
<td>0.83</td>
</tr>
<tr>
<td>Acetyl-Val-Arg-Arg-AMC</td>
<td>3.92</td>
<td>0.81</td>
</tr>
<tr>
<td>Acetyl-Tyr-Arg-Arg-AMC</td>
<td>5.28</td>
<td>-0.96</td>
</tr>
<tr>
<td>Acetyl-Tyr-Arg-Arg-AMC</td>
<td>12.5</td>
<td>-1.50</td>
</tr>
</tbody>
</table>

Data were fitted to this model according to the equation of Hasnain et al. (12).

The pH dependence of cathepsin B-catalyzed hydrolysis of the substrates acetyl-Arg-Arg-Arg-AMC and acetyl-Val-Arg-Arg-AMC (Fig. 1) was very similar to that obtained previously for the substrate Z-Arg-Arg-AMC (12). The $pK_a$ values for the former tripeptide substrates were in close agreement with the values for the dipeptide substrate (Table I). The greater uncertainty in the values for $pK_a$ and $pK_b$ were due, in part, to the difficulty of obtaining accurate initial rates above pH 7.9.

---

**Analysis of pH-Activity Profiles for the Substrates Acetyl-Arg-Arg-Arg-AMC and Acetyl-Val-Arg-Arg-AMC**

The data for the pH-activity profiles (Fig. 1) were best fitted to a model involving two dissociation events in the ascending limb and two dissociation events in the descending limb as described before for the dipeptide substrate Z-Arg-Arg-AMC (12). $E_1$ and $E$ are inactive forms of the enzyme.

\[
\begin{align*}
E_1 & \underset{k_1}{\rightleftharpoons} E_H \overset{k_2}{\rightleftharpoons} E_2 \\
E_H & \overset{k_3}{\rightleftharpoons} E_2 \overset{k_4}{\rightleftharpoons} E
\end{align*}
\]

**Model 1**

---

**Fig. 2.** Possible binding mode of the substrate acetyl-Tyr-Arg-Arg-AMC. A possible binding mode of the substrate acetyl-Tyr-Arg-Arg-AMC to cathepsin B was determined by docking the substrate into a computer model of the active site obtained from the crystal structure of cathepsin B. The substrate was energy minimized as described under "Experimental Procedures," using the Amber force field as implemented by the MACROMODEL software package.
caused by the pH instability of cathepsin B in that pH range. These data suggest that the interaction of the P₃ Arg in the S₃ subsite is not to any significant degree a result of a charge-charge interaction, although a weak interaction cannot be ruled out.

**Nature of the S₈-P₃ Interaction**

Table III reports the change in apparent binding energies among the dipeptide and tripeptide substrates at pH 6.0. Results are summarized as follows.

(i) Apparent Binding Energies for the Dipeptide Substrates Acetyl-Arg-Arg-AMC and Z-Arg-Arg-AMC—The change in apparent binding energy for the substrate acetyl-Arg-Arg-AMC relative to Z-Arg-Arg-AMC was calculated to be 0.82 kcal/mol, suggesting that the benzyloxy carbonyl group interacts with the enzyme as a pseudo-P₃ residue. As mentioned previously, this finding is supported by the x-ray structure of a cathepsin B-inhibitor complex (27). In this complex, the benzylic ring of the N-terminal benzyloxy carbonyl group of the inhibitor, Z-Arg-Ser-(O-Bzl)-CMK, makes a direct contact with the enzyme, forming a vertical aromatic-aromatic interaction with Tyr⁷⁵, with the shortest distance of about 3.71 Å.

(ii) Binding Energies of Various P₃ Side Chains—The glycol-containing substrate served as the reference for the calculation of apparent binding energies of the P₃ side chains of the other tripeptide substrates (Table III). At pH 6.0, the only P₃ side chain that demonstrated a weaker binding energy relative to glycol was glutamyl, with a ΔΔG of 0.33 kcal/mol. The other three tripeptide substrates, with either Arg, Val, or Tyr at P₃, exhibited an improved binding in that order. Although the substrate containing a P₃ arginy1 has an increased binding energy compared with glycol, by about 0.81 kcal/mol, this enhancement cannot be to any large degree due to a charge-charge interaction at the S₈ subsite. First, a 4-fold effect arising from an ionic interaction between the P₃ guanidinium group and a negatively charged group in the S₈ subsite should have been manifested in the pH-activity profile. As shown in Fig. 1 and Table II, the pH-activity profile data for the tripeptide acetyl-Arg-Arg-Arg-AMC could be fitted by the same equation used to fit the data for the dipeptide Z-Arg-Arg-AMC, and the substrate with a neutral P₃, acetyl-Val-Arg-Arg-AMC. Furthermore, the dissociation constants derived from these data were virtually identical. In fact, the ratios of kₐ/d/Kₐ of acetyl-Arg-Arg-Arg-AMC and acetyl-Val-Arg-Arg-AMC are similar at pH 6.0 and 4.0. This is also true for the ratios of kₐ/d/Kₐ of acetyl-Arg-Arg-Arg-AMC and the tripeptide substrates with either glycyl or tyrosyl side chains at P₃, which suggests that the P₃ arginy1 side chain binding is not to any significant extent stabilized by an ionic interaction. The x-ray structure of the complex of cathepsin B and the inhibitor Z-Arg-Ser-(O-Bzl)-CMK suggests a relatively hydrophobic S₈ pocket defined by Tyr⁷⁵, Ca of Gly⁷⁹, Cβ of Asn⁷₂, and Ca of Asp⁶⁹ (27). Therefore one may suggest that the improved binding of arginine over glycine at P₃ may be largely due to hydrophobic interactions of the methylene groups in the arginine side chain with the relatively hydrophobic S₈ subsite. An interaction between the Asp⁶⁹ side chain carboxylate and the P₃ guanidinium cannot be entirely ruled out. However, if it occurs, this interaction does not contribute very significantly to substrate binding (see below for further discussion).

At pH 6.0, a comparison of the specificity constants for the substrates acetyl-Arg-Arg-AMC and acetyl-Gly-Arg-Arg-AMC (Table I) shows virtually no difference. It appears, therefore, that the acetyl methyl and the atoms forming the adjacent amide bond in the tripeptide substrate make no significant contribution to binding. In addition, the x-ray structure of the
The cathepsin B-inhibitor complex reveals that the oxyanion hole between the pseudo-P₂ N-terminal benzyl group and the P₃ residue does not make any contact with the enzyme (27). As such, the values for the change in free energy of binding (Table III) for different tripeptide substrates also serve to define the incremental increase in binding energy of the P₃ residue when compared with the dipeptide substrate, acetyl-Arg-Arg-AMC.

(iii) Evidence Supporting a Predominantly Hydrophobic Pocket at S₄—From the specificity constants reported in Table I and changes in binding energy in Table III, the nature of the S₄ subsite of cathepsin B can be deduced. The S₄ subsite shows a general preference for both aromatic and aliphatic groups. At pH 6.0, the substrate with the P₃ glutamyl is least favored, whereas at pH 4.0 it is preferred over the arginyl and valyl side chains. Therefore, it appears that a negative charge at P₃ seems to disrupt the binding at the S₄ subsite. This may be due to a charge-charge repulsion involving Asp₆⁰.

If the Asp₆⁰ carboxylate is responsible for destabilizing the binding of a P₃ negative charge, it is surprising that the 4-fold increase in specificity of the arginyl side chain compared with glysyl does not appear to result from a charge-charge attraction with Asp₆⁰. A possible explanation for the 4-fold increase in specificity of the Arg side chain compared with Gly is the possible van der Waals interaction of one or more of the arginyl side chain methylene carbons with Tyr₇₅.

The 2-fold increase in specificity of the P₂ tyrosyl side chain, relative to the P₃ valyl, and the 4-fold increase in specificity of the P₂ tyrosyl, relative to the benzyl ring of the substrate Z-Arg-Arg-AMC, may result from an improved interaction of the phenyl ring of the substrate tyrosyl and Tyr₇₅ in the S₂ subsite of the enzyme. In the x-ray structure of the cathepsin B-inhibitor complex, the benzyl ring of the benzoyloxycarbonyl moiety is about 3.7 Å from the Tyr₇₅ phenyl ring. The kinetic data suggest that both the valyl and tyrosyl side chains in the respective substrates make energetically more favorable contacts with Tyr₇₅ than the benzyl ring of the benzoyloxycarbonyl moiety.

Modeling Tripeptide Substrates Containing Tyr, Glu, or Arg at P₃

The modeling study with the substrate containing a P₃ Tyr (Fig. 2) revealed that the phenyl ring of the substrate is partly stacked on the phenyl of Tyr₇₅ of the enzyme, with the shortest distance between Ce of the substrate Tyr and Ce of Tyr₇₅ of 3.4 Å. In the model, the substrate tyrosyl hydroxyl forms a hydrogen bond with Asp₆⁰. Another potential contact in the S₄ pocket is between the substrate Tyr Ce and Ce of Gly₇₃ (3.6 Å). The improved contact of the P₂ tyrosyl in the S₄ subsite, as suggested by the model, relative to the position of the pseudo-P₂ benzyl group of the dipeptide inhibitor observed in the x-ray structure (27), is consistent with the kinetic data showing a 4-fold increase in specificity for the P₃ Tyr compared with the N-terminal benzoyloxycarbonyl group. A study by Serrano et al. (39) revealed that aromatic-aromatic interactions in protein structures can contribute between 0.5 and 1.3 kcal/mol to protein stability. This effect is due to a quadrupole-quadrupole interaction between the aromatic rings, for which there is an associated potential energy that varies as 1/r² (where r is the quadrupole separation distance) (30). Therefore, any binding energy that is contributed by aromatic side chains of substrates or inhibitors interacting with aromatics on enzymes would be very sensitive to the distance between the groups.

The P₃ Arg modeling study (Fig. 3) reveals that the methyl groups of the arginyl side chain are somewhat further from Gly₇₃ (4.3 Å) and Tyr₇₅ (4.4 Å) than the other modeled substrates, suggesting that a van der Waals interaction is unlikely. The guanidinium group of the P₃ Arg appears to form hydrogen...
bonds with the hydroxyl of Tyr$^{75}$ (3.0 Å) and the carboxyl oxygen of Asp$^{69}$ (2.7 Å) and, possibly, a weak hydrogen bond with the backbone carbonyl oxygen of Asn$^{72}$ (3.4 Å). As discussed before, the kinetic data suggest that if there is an interaction between the guanidinium moiety of the P$_3$ Arg and the carboxylate of Asp$^{69}$, there is no significant net gain in binding energy. Considering the fact that there would be a significant energy cost involved in desolvating the guanidinium cation and that this interaction would be completely solvent-exposed, it may be reasonable to expect that there may not be any significant gain in binding energy for this charge-charge interaction. Therefore, the 4-fold increase in specificity observed for a P$_3$ Arg, relative to Gly, may be due instead to the charge-neutral hydrogen bonds involving the substrate guanidinium and the hydroxyl of Tyr$^{75}$ and possibly the carboxyl of Asn$^{72}$.

The modeling study for the substrates containing either a protonated or deprotonated Glu (Fig. 4 and 5) shows that the side chain methylene carbons of Glu make van der Waals contacts with the Ca of Gly$^{72}$ (3.2 Å) and the phenyl ring of Tyr$^{75}$ (3.6 Å). The kinetic data show that $k_{cat}/K_m$ drops 20-fold when the P$_3$ Glu deprotonates. One possible explanation of this significant destabilization of binding may be a charge-charge repulsion between Asp$^{69}$ and the P$_3$ side chain carboxylate, which are about 4.4 Å apart. Since the potential energy of charge-charge interactions varies as $1/r$ (where $r$ is the charge separation), they can be manifested over relatively large distances. In fact, the modeling study with the deprotonated Glu shows the Asp$^{69}$ side chain moving away from the P$_3$ carboxylate relative to the Asp$^{69}$ side chain position in the modeling studies with the other substrates. Another possible explanation, however, may be that there is a repulsion between the π-electron cloud of Tyr$^{75}$ and the negative charge on the substrate side chain (30).

To summarize, the substrate binding in the S$_3$ subsite of cathepsin B is largely due to contacts with the phenyl ring of Tyr$^{75}$ and to a lesser extent may involve the Ca of Gly$^{72}$. Additionally, hydrogen bonds involving the Tyr$^{75}$ hydroxyl and the main chain carbonyl oxygen of Asn$^{72}$ may also make a contribution to substrate stabilization. The best substrate among the tripeptides tested has tyrosine at P$_3$, which in comparison with the reference substrate containing glycine, has an increase in binding energy of 1.5 kcal/mol. Although these findings are significant for inhibitor design strategies, it may be useful to examine further the S$_3$ specificity of cathepsin B with substrates containing residues such as tryptophan, phenylalanine, proline, methionine, leucine, and isoleucine at the P$_3$ position. The study of these additional residues would provide a more complete understanding of the S$_3$ specificity pocket. Extension of such a study to S$_4$ interactions would contribute additional information for the design of more specific inhibitors.

Acknowledgments—We are indebted to A. Szabo for the use of his laboratory facilities and to A. Wigg for assistance.

REFERENCES
Cathepsin B Specificity


