THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 267, No. 7, Issue of March 5, pp. 4713-4721, 1992
Printed in U.S.A.

Characterization of Recombinant Rat Cathepsin B and Nonglycosylated Mutants Expressed in Yeast

NEW INSIGHTS INTO THE pH DEPENDENCE OF CATHEPSIN B-CATALYZED HYDROLYSES*

(Received for publication, August 30, 1991)

Sadiq Hasnain‡ and Tomoko Hirama

From the Protein Structure and Design Section, Institute for Biological Sciences, National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada

Anna Tam and John S. Mort§

From the Joint Diseases Laboratory, Shriners Hospital for Crippled Children and the Department of Surgery, McGill University, Montreal, Quebec H3G 1A6, Canada

The cysteine proteinase rat cathepsin B was expressed in yeast in an active form and was found to be heterogeneously glycosylated at the consensus sequence for *N*-linked oligosaccharide substitution. Purified enzyme fractions containing the highest levels of glycosylation were shown to have reduced activity. A glycosylation minus mutant constructed by site-directed mutagenesis (by changing the Ser to Ala in the consensus sequence) was still secreted by the yeast and was shown to be functionally identical with purified rat liver cathepsin B. Recombinant cathepsin B was used to further characterize the pH dependence of cathepsin B-catalyzed hydrolyses using 7-amido-4-methylcoumarin (AMC) and *p*-nitroaniline (*p*NA) substrates with arginine as the P_1, and either arginine or phenylalanine as the P_2 residue. The AMC and *p*NA groups give insights into the leaving group binding site (P′) of cathepsin B. These studies show for the first time that at least seven dissociable groups are involved in substrate binding and hydrolysis in cathepsin B activity. Two of these groups, with pK_a values of 6.9 and 7.7 in the recombinant enzyme, are in the leaving group binding site and are most likely His^{110} and His^{111}. The same groups in rat liver cathepsin B have higher pK_a values than in recombinant cathepsin B, but have identical function in the two enzymes. Two other groups are probably the active site Cys^{29} and His^{199} with pK_a values of 3.6 and 8.6, respectively. A group with a pK_a of 5.1 interacts with substrates containing Arg at P_2, and the group is most likely Glu^{245}. The remaining two groups, one with a pK_a of about 4.9 and the other about 5.3, are most likely carboxyl residues possibly interacting with Arg at P_1 in the substrate. The possible candidates on the basis of the x-ray structure are Asp^{22}, Asp^{69}, Glu^{171}, and Glu^{122}, all found within a 13 Å radius from the active site thiol of Cys^{29}.

Cathepsin B is a major lysosomal cysteine proteinase which has been implicated in a number of diseases such as arthritis, muscular dystrophy, and tumor metastasis (1-7). The protein is synthesized as a preproenzyme (8, 9) and is targeted to the lysosomes following glycosylation. In the lysosome, procathepsin B is further processed, largely to a two-chain form consisting of a 47-residue light chain, containing the active site cysteine 29, and a 205-residue heavy chain, containing the active site histidine 199 (10, 11). Comparison of the cDNA-derived protein sequence with the primary structure of the two-chain form determined by Edman degradation revealed that processing involves removal of residues 48 and 49 (some single chain form also remains), as well as 6 residues at the COOH terminus (residues 255-260) (8, 9). The single chain form of the mature enzyme contains 254 residues and is the basis for the numbering used throughout this paper.

The mature form of rat liver cathepsin B contains one consensus sequence for *N*-linked oligosaccharide substitution, -Asn-Gly-Ser- (residues 113-115), which was suggested to be glycosylated on the basis of amino acid analysis (12). Takahashi *et al.* (13) found *N*-linked glycosylation at Asn^{113} in porcine spleen cathepsin B and showed by NMR analysis that the predominant form contained a single *N*-acetylglucosamine residue. A minor form contained 2 units of *N*-acetylglucosamine, 1 fucose, and 2 mannose residues. *In vitro* studies have demonstrated that glycosylation is essential for targeting cathepsin B to the lysosome (14, 15).

The mechanism of cathepsin B specificity and catalysis is still not well characterized. Cathepsin B has both endopeptidase and dipeptidyl carboxypeptidase activities (16, 17) and its pH activity profiles have suggested complex behavior involving several dissociable groups (17-19). Recently, a study on the pH-dependent hydrolysis of AMC[1] substrates for the investigation of S_2-P_2 interactions has been published for human liver cathepsin B (20). This study suggests that the thiolate-imidazolium ion pair functions in a similar manner in cathepsin B as in the better characterized cysteine proteinase, papain. The pK_a of the active site thiol (Cys^{29}) was shown to be about 3.4, and evidence was provided for the role of a group with a pK_a of about 5.4, suggested to be Glu^{245}, in the binding of substrate with Arg in the P_2 position. This was recently corroborated by the x-ray structure determination of human liver cathepsin B (21).[2] The structure suggests that Glu^{245} forms part of the S_2 subsite. Also of note in the x-ray

* This is National Research Council of Canada Publication 31970. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence and reprint requests should be addressed. Tel: 613-990-0899; Fax: 613-952-0583.

§ Supported by the Shriners of North America.

[1] The abbreviations used are: AMC, aminomethylcoumarin; *p*NA, *p*-nitroaniline; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; DTT, dithiothreitol.

[2] C. P. Huber, R. Campbell, T. Hirama, X. Lee, R. To, and S. Hasnain, manuscript in preparation.

structure is the position of a large 18-residue insertion in the cathepsin B structure compared to papain. This insertion forms a surface loop which occludes the top of the active site cleft (when the active site is oriented facing down). Two important histidine residues, His^{110} and His^{111}, are strategically located in the loop just before the turn and projecting into the active site cleft. His^{110} is in position to form a salt bridge with Asp^{22}, which is adjacent to Gln^{23}, which forms the oxyanion hole. Musil *et al.* (21) speculated that these histidines may play a role in the exopeptidase activity (dipeptidyl carboxypeptidase) of cathepsin B through an interaction of the protonated imidazoles with the terminal carboxylate of the substrate. They went further to suggest that His^{111} may have a pK_a of 5.5. They took the lead from previous suggestions that a group dissociating with a pK_a of 5.5 controlled the pH-dependent exopeptidase activity of cathepsin B.

Due to the considerable interest in cathepsin B structure-function relationships, several attempts have been made to express recombinant cathepsin B. Attempts to express active cathepsin B in *Escherichia coli* by Mort *et al.* (22) and Chan and Fong (23) were unsuccessful although cathepsin B was detected immunologically by both groups. Similar problems were experienced with the expression of other cysteine proteinases in *E. coli*. Papain has been expressed in *E. coli* (24, 25), but no active enzyme was obtained. Smith and Gottesman (26) expressed cathepsin L in *E. coli* as inclusion bodies, and some activity was recovered following denaturation and refolding. However, the enzymatically active form differed from native human liver cathepsin L both physically and functionally. More recently, active rat cathepsin B has been expressed in yeast by Mort *et al.*,[3] and the crystallization of the recombinant nonglycosylated mutant and preliminary x-ray data analysis have been published (27).

Here we discuss the effects of glycosylation on the activity of rat cathepsin B expressed in yeast and compare the kinetic properties of two nonglycosylated mutants, Ser^{115}-Ala and (Ser^{115}-Ala, Gln^{255}-Termination codon) (Ser^{115}-Ala, Gln^{255}-Term) with those of the enzyme purified from rat liver. In addition, by using substrates with different leaving groups (aminomethylcoumarin (AMC) and *p*-nitroaniline (*p*NA)) and different P_2 residues (Arg and Phe), we have made a comprehensive inventory of the dissociable groups in the enzyme active site involved in substrate binding and hydrolysis. We show for the first time that at least 7 dissociable groups play a role in cathepsin B activity and provide evidence for the possible involvement of 2 histidines in the leaving group binding subsite (S′) of cathepsin B.

EXPERIMENTAL PROCEDURES

Cloning and Expression—The cDNA for rat cathepsin B (8) was expressed in *Saccharomyces cerevisiae* as an α-factor fusion by Mort *et al.*[3] A mutant was constructed by site-directed mutagenesis to eliminate glycosylation in the mature protein by changing the serine at position 115 to alanine (Ser^{115}-Ala) to eliminate the consensus sequence -Asn-Gly-Ser-. Since initially we did not have evidence for the correct processing of the 6 extra residues in the COOH terminus coded for by the gene but not found in the mature cathepsin B in animal tissues, we decided to eliminate the COOH-terminal extension by constructing a double mutant incorporating a stop codon at position 255 along with the elimination of the glycosylation consensus sequence (Ser^{115}-Ala, Gln^{255}-Term).[4]

Yeast starter cultures were grown in synthetic medium (28) and then were grown in 4-liter shake flasks as reported before (27).

[3] J. S. Mort, A. Tam, L. Mach, P. Mason, A. D. Rowan, S. J. Chan, and D. F. Steiner, submitted for publication.

[4] A. D. Rowan, R. Feng, Y. Konishi, and J. S. Mort, submitted for publication.

Purification of Glycosylated Rat Cathepsin B Expressed in Yeast—The secreted, glycosylated recombinant enzyme was harvested when activity reached a maximum in the culture medium and concentrated using a Millipore Pellicon concentrator and then by an Amicon concentrator fitted with a YM-5 filter as described previously (27). The concentrated enzyme was purified by affinity chromatography using the ligand Gly-Phe-Gly-semicarbazone linked to Sepharose 4B (29). The ligand was purchased from AminoTech, Ottawa.

Purification of the Nonglycosylated Rat Cathepsin B Mutant (S115A) and the Double Mutant (S115A, Q255Term) Expressed in Yeast—The cathepsin B Ser^{115}-Ala and Ser^{115}-Ala, Gln^{255}-Term mutants expressed in yeast were harvested and concentrated as described above for the wild type form of the enzyme. However, only a two-step purification using DEAE-cellulose (DE52, Whatman) and then Sephadex G-75 (Pharmacia LKB Biotechnology Inc.) chromatography was required for the two mutants (27).

Purification of Rat Liver Cathepsin B—Livers were excised from male rats (250 g) sacrificed by inhalation of carbon dioxide, and the livers were frozen at −20 °C. 500 g of liver was thawed overnight at 4 °C, the tissue was chopped and homogenized, and the enzyme was partially purified by acetone precipitation as previously described by Barrett and co-workers (16, 30). The partially purified enzyme was resuspended in 50 mM sodium phosphate, 0.1% Brij-35, 1 mM EDTA, pH 6.0, and dialyzed extensively into the same buffer. After concentration of the enzyme solution using an Amicon YM-5 membrane, the enzyme was purified on a Sepharose 4B-Gly-Phe-Gly-semicarbazone affinity column as described above. The eluted fractions containing enzyme activity were pooled, concentrated, and chromatographed on Sephadex G-75. Fractions with similar specific activity were pooled and stored at 4 °C. Purity was confirmed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) as described before (27).

Active Enzyme Determination—Immediately after purification and before kinetic analyses were carried out, the concentration of active enzyme was determined by free thiol titration using the cysteine proteinase inhibitor E-64 (*N*-[*N*-(L-3-*trans*-carboxyoxiran-2-carbonyl)-L-leucyl]-agmatine) (Boehringer Mannheim Canada) as described before (16).

Enzyme Assays—For routine assays, cathepsin B activity was determined by continuous spectrophotometric assay using *N*-benzyloxycarbonyl-L-arginyl-L-arginine-*p*-nitroanilide (*Z*-Arg-Arg-*p*NA) (BACHEM, Philadelphia), and *p*-nitroaniline release was monitored at 410 nm at 24 °C in an SLM Aminco DW2000 spectrophotometer. The reaction buffer contained 100 mM sodium phosphate, 1 mM EDTA, 10 mM dithiothreitol (DTT), pH 6.0. The enzyme was activated in the DTT-containing reaction buffer for 5 min, and the reaction was started by the addition of substrate to a final concentration of 0.8 mM. Initial rates were determined from the linear portion of the optical density profile.

For the determination of kinetic constants, the enzyme was preactivated in 10 mM DTT for 1 h at room temperature and then placed on ice. The activated enzyme was determined to be stable for the duration of the experiment. Michaelis-Menten kinetic parameters were determined using a reaction buffer containing 25 mM sodium phosphate, 250 mM NaCl, 1 mM EDTA, 3% dimethyl sulfoxide, pH 6.0. Initial rates were measured from the linear portion of the optical density profile. At low substrate concentration, care was taken to avoid effects of substrate depletion for the determination of initial rates. The data were fitted by the nonlinear regression data analysis program, Enzfitter, of Leatherbarrow, supplied by Elsevier-Biosoft.

pH activity profiles of the enzymes were determined by measuring k_{cat}/K_m at 0.1 pH interval over the range of pH 3.2–8.4 at 24 °C using the relationship $v = [E][S]k_{cat}/K_m$ at $[S] << K_m$. In the range of pH 3.2–6.0, the reaction buffer consisted of 25 mM sodium citrate, 250 mM NaCl, 1 mM EDTA, and 3% dimethyl sulfoxide; in the pH range of 5.7–7.8, the reaction buffer consisted of 25 mM sodium phosphate, 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 borate, 250 mM NaCl, 1 mM EDTA, and 3% dimethyl sulfoxide. The citrate buffer consistently gave lower activity with the Arg-Arg-containing *p*NA and AMC substrate when compared with phosphate buffer at overlapping pH values. This effect was not observed with the Phe-Arg-containing substrates. Therefore, the data for Arg-Arg-containing substrates measured in citrate buffer were corrected by a factor determined from measurements in phosphate buffer at four separate pH intervals which overlapped with the citrate buffer. Samples were monitored to ensure that pH did not fluctuate over the course of the enzyme reaction. Data were fitted by nonlinear regression with the

Enzfitter program using appropriate equations, depending on whether the substrate had Phe or Arg in the P_2 position, and on the type of leaving group (*p*NA or AMC), as described under "Results" and "Discussion."

Kinetic measurements with the coumarin substrates, *Z*-Arg-Arg-AMC and *Z*-Phe-Arg-AMC (IAF Biochemicals International, Laval, Quebec), were performed by monitoring fluorescence of the AMC released (16) in an SLM Aminco spectrophotometer equipped with the Total Fluorescence Accessory. The reaction conditions were the same as for the *p*NA substrates.

Kinetic constants for acylation and deacylation for the cathepsin B-catalyzed hydrolyses of *p*NA and AMC substrates, in the presence of the nucleophile Gly-Gly, were determined according to the method of Bajkowski and Frankfater (31, 32) or as indicated.

RESULTS

The Effect of Glycosylation on Rat Cathepsin B Expressed in Yeast—The rat cathepsin B expressed in yeast, and secreted into the medium, was found to be extremely heterogeneous after purification by affinity chromatography. Active site titration of the affinity-purified enzyme with the cysteine proteinase inhibitor, E-64, confirmed that 100% of the protein was active enzyme when compared to total protein concentration determined by absorption at 280 nm. The affinity-purified sample chromatographed on Sephadex G-50 eluted with a broad profile as monitored by absorption at 280 nm (data not shown). Enzyme that was retarded showed higher specific activity when assayed with *Z*-Arg-Arg-*p*NA than the enzyme which eluted early from the Sephadex column and therefore had a larger molecular mass. The gel filtration fractions were pooled into three groups, A, B, and C. SDS-PAGE (Fig. 1*A*) of the three fractions and the original sample applied to the column shows that the enzyme in the original sample applied to the column, and in fractions A and B (Fig. 1*A*, *lanes 2, 3,* and *4*, respectively), ran as smears consistent with the behavior of highly glycosylated proteins on SDS gels. The migration of the protein in fraction A (Fig. 1*A*, *lane 3*) is retarded with respect to fraction B and C (Fig. 1*A*, *lanes 4* and *5*, respectively), suggesting enzyme molecules of different molecular mass in each of the fractions. The difference in mobility of fractions A, B, and C both on Sephadex G-50 and SDS-PAGE suggests heterogeneous glycosylation of recombinant cathepsin B. Analysis by SDS-PAGE of the recombinant rat cathepsin B following endoglycosidase H treatment gave a single discrete band[3] supporting the suggestion that the enzyme secreted by yeast is substituted with a *N*-linked oligosaccharide.

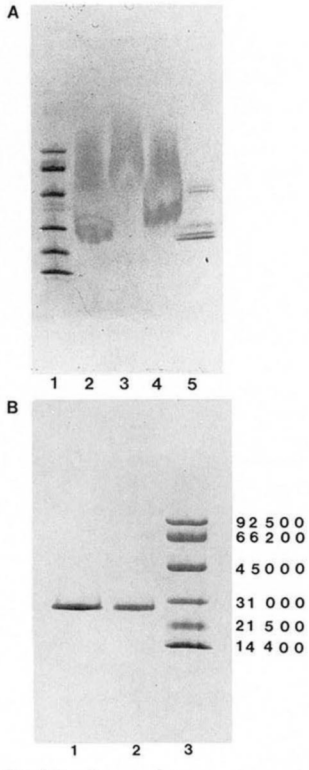


FIG. 1. *A*, SDS-PAGE of recombinant glycosylated rat cathepsin B fractionated on Sephadex G-50. Pooled fractions from Sephadex G-50 chromatography were subjected to SDS-polyacrylamide electrophoresis and stained with Coomassie Brilliant Blue. *Lane 1*, molecular weight markers. *Lane 2*, sample before gel filtration. *Lane 3*, fraction A. *Lane 4*, fraction B. *Lane 5*, fraction C. Molecular weight markers are the same as in *B*. *B*, purified recombinant glycosylation minus mutants of rat cathepsin B. *Lane 1*, (S115A, Q255Term) mutant. *Lane 2*, S115A mutant. *Lane 3*, molecular weight markers.

To investigate the effect of glycosylation on enzyme activity, the pooled fractions from the gel filtration experiment were subjected to kinetic analysis with two substrates, *Z*-Arg-Arg-*p*NA and *Z*-Phe-Arg-AMC. As shown in Table I, k_{cat} and k_{cat}/K_m for both substrates are significantly lower for fraction A than for fraction C (fraction A appears to be more highly glycosylated than fraction C on the basis of its mobility and SDS-polyacrylamide gels and Sephadex G-50 chromatography). The activity (k_{cat} and k_{cat}/K_m) in fraction C was 3 times higher than in fraction A, using the *p*NA substrate, while for the AMC substrate the ratio was about 6. K_m values did not appear to be affected.

Characterization of the Nonglycosylated Mutant of Rat Cathepsin B Expressed in Yeast—Since glycosylation of cathepsin B in yeast was observed to be heterogeneous and to affect activity, and since this posed a barrier to our primary goal of characterization enzyme function, we decided to eliminate *N*-linked glycosylation by site-directed mutagenesis. The Ser^{115} in the consensus sequence -Asn-Gly-Ser- was mutated to Ala. As well, a double mutant was constructed (Ser^{115}-Ala, Gln^{255}-Term) to eliminate a possible 6-residue COOH-terminal extension which is coded for by the gene but not found in the mature cathepsin B from animal and human tissues. The purified cathepsin B (Ser^{115}-Ala, Gln^{255}-Term) double mutant and the Ser^{115}-Ala single mutant ran as single bands with an M_r of about 29,000 on SDS-PAGE (Fig. 1*B*, *lanes 1* and *2*) showing no signs of the heterogeneity observed with the wild type gene product. The theoretical molecular mass of the unglycosylated, 254-amino acid mature rat cathepsin is 27,753 Da. With the additional 6-amino acid NH_2-terminal exten-

TABLE I

Kinetic constants for the hydrolysis of Z-Arg-Arg-pNA and Z-Phe-Arg-AMC by glycosylated recombinant rat cathepsin B pooled fractions from Sephadex G-50 chromatography (Fig. 1)

For *Z*-Arg-Arg-*p*NA, the assay buffer consisted of 50 mM phosphate, 1 mM EDTA, 1% acetonitrile, 10 mM DTT, pH 6.0. For *Z*-Phe-Arg-*p*NA, the assay buffer consisted of 50 mM phosphate, 1 mM EDTA, 10% dimethyl sulfoxide, 10% acetonitrile, 10 mM DTT, pH 6.0. For both substrates, enzyme was preincubated in the assay buffer for 5 min at 24 °C, and substrate was added to start the reaction.

Substrate	Pool	k_{cat}	K_m	k_{cat}/K_m
		s^{-1}	mM	$M^{-1}\ s^{-1}$
Z-Arg-Arg-*p*NA				
	A	2.71 ± 0.07	0.53 ± 0.03	5,400 ± 160
	B	5.57 ± 0.13	0.48 ± 0.03	11,800 ± 380
	C	7.08 ± 0.16	0.44 ± 0.03	16,800 ± 500
Z-Phe-Arg-AMC				
	A	1.38 ± 0.09	0.72 ± 0.10	1,860 ± 110
	B	8.22 ± 0.49	1.15 ± 0.12	7,020 ± 190
	C	9.96 ± 0.43	0.87 ± 0.07	11,100 ± 740

TABLE II

Kinetic constants for rat liver cathepsin B and for the glycosylation minus mutants expressed in yeast

Enzyme was activated by preincubation with 10 mM DTT as described. The pH of the reaction buffer was 6.0, $I = 0.3$, and dimethyl sulfoxide concentration was 3%. The kinetic data and standard deviations were calculated from three separate determinations except for the rat liver enzyme with Z-Phe-Arg-AMC, for which two determinations were made, and the averages of the two experiments are reported. The range of substrate concentration used was 0.100 to 2.30 mM for Z-Arg-Arg-pNA, 0.027 to 0.160 mM for Z-Phe-Arg-pNA, 0.012 to 0.100 mM for Z-Phe-Arg-AMC, and 0.200 to 1.000 mM for Z-Arg-Arg-AMC. Above 2.5 mM, substrate inhibition is observed for Z-Arg-Arg-pNA, while, above 0.18 mM, Z-Phe-Arg-pNA becomes insoluble under the assay conditions used. For the mutant S115A and the substrate Z-Phe-Arg-pNA, k_{cat}/K_m was calculated from the relationship $v = k_{cat}/K_m\,[S][E]$ when $[S] \ll K_m$.

	Rat liver	S115A	S115A, Q255Term
Z-Arg-Arg-pNA			
k_{cat} (s^{-1})	8.29 ± 0.73	8.55 ± 1.31	9.67 ± 1.31
K_m (mM)	1.01 ± 0.12	0.99 ± 0.03	1.19 ± 0.11
k_{cat}/K_m (M^{-1} s^{-1})	8,310 ± 1,010	8,600 ± 1,000	8,110 ± 445
Z-Phe-Arg-pNA			
k_{cat} (s^{-1})	35.2 ± 1.59		34.4 ± 3.90
K_m (mM)	0.62 ± 0.09		0.60 ± 0.08
k_{cat}/K_m (M^{-1} s^{-1})	57,100 ± 7,120	52,200 ± 173	57,700 ± 1,470
Z-Arg-Arg-AMC			
k_{cat} (s^{-1})	57.9 ± 3.83		51.4 ± 5.45
K_m (mM)	1.27 ± 0.23		1.11 ± 0.17
k_{cat}/K_m (M^{-1} s^{-1})	46,370 ± 5,405		46,400 ± 3,260
Z-Phe-Arg-AMC			
k_{cat} (s^{-1})	81.4		77.1 ± 9.49
K_m (mM)	0.28		0.23 ± 0.03
k_{cat}/K_m (M^{-1} s^{-1})	297,900		335,000 ± 4,930

sion,[3] and a possible 6-amino acid COOH terminus extension, the theoretical molecular mass of the 266-residue single mutant should be 29,297 Da based on the protein sequence, and for the double mutant 28,459 Da. These values are consistent with the value determined by SDS-PAGE (Fig. 1*B*, *lanes 1* and 2). Isoelectrofocusing of double and single mutants (data not shown) showed that the single mutant occurs as two major forms while the double mutant appears to be a single major form. When the single mutant was preactivated with DTT and then subjected to nondenaturing isoelectrofocusing, a single band equivalent to the double mutant was observed (data not shown). This suggests that the single mutant, Ser^{115}-Ala, possibly contains some if not all of the 6 extra COOH-terminal residues (255–260) which do not occur in the mature lysosomal form of the enzyme. Upon prolonged activation (1 h) of the enzyme with DTT, the extra COOH-terminal residues may be removed by autoprocessing resulting in a protein with the same charge as the double mutant, Ser^{115}-Ala, Gln^{255}-Term.

Kinetic Analysis of Recombinant and Rat Liver Cathepsin B—Kinetic analysis of the cathepsin B Ser^{115}-Ala mutant, the double mutant (Ser^{115}-Ala, Gln^{255}-Term), and native cathepsin B purified from rat liver was carried out with substrates containing different P_2 residues (Phe and Arg) and P′ leaving groups (AMC and pNA). The results show that the activities of the recombinant and rat liver cathepsin B are essentially identical at pH 6.0 for all kinetic parameters calculated for the substrates Z-Arg-Arg-pNA and Z-Phe-Arg-pNA and Z-Arg-Arg-AMC and Z-Phe-Arg-AMC (Table II). k_{cat}/K_m for Phe-Arg is approximately 7 times higher than for Arg-Arg irrespective of the leaving group, while k_{cat}/K_m is 5-fold higher for the AMC leaving group compared to pNA. The difference between Z-Arg-Arg-pNA and Z-Arg-Arg-AMC is primarily in k_{cat} (5-fold) as opposed to the difference between Z-Phe-Arg-pNA and Z-Phe-Arg-AMC where the difference is about 2-fold in k_{cat} and about 3-fold in K_m.

Comparison of the pH Activity Profiles for Recombinant and Rat Liver Cathepsin B Using the Substrates Z-Phe-Arg-pNA and Z-Arg-Arg-pNA—The data for k_{cat}/K_m *versus* pH for the substrates Z-Phe-Arg-pNA and Z-Arg-Arg-pNA for the native rat liver enzyme and the recombinant rat cathepsin B (Ser^{115}-Ala, Gln^{255}-Term) double mutant expressed in yeast are shown in Fig. 2, *A* and *B*, respectively. The calculated apparent pK_a values and the theoretical limit values of k_{cat}/K_m for each of the ionization events are listed in Table III. The pH profile data for the Z-Phe-Arg-pNA substrate is best fitted to an equation describing a model (Model 1) involving four dissociation events important for enzyme activity, three in the

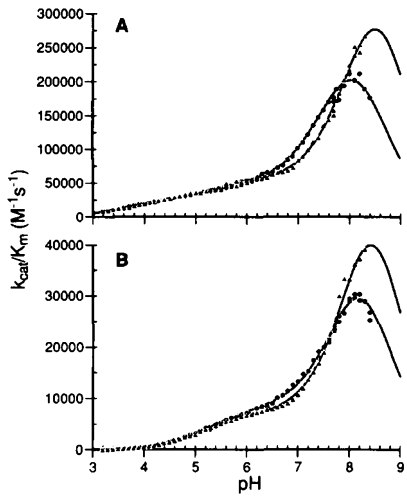


FIG. 2. **pH activity profiles of rat liver cathepsin B compared with the double mutant (S115A, Q255Term).** *A*, with the substrate Z-Phe-Arg-pNA. *B*, with the substrate Z-Arg-Arg-pNA. ●, double mutant (S115A, Q255Term). ▲, rat liver cathepsin B. The conditions for assays are as described under "Experimental Procedures." The *lines* through the data represent the best fit using equations described in the text.

TABLE III

Ionization constants and k_{cat}/K_m limit values of dissociable groups involved in substrate binding and catalysis calculated from pH activity profiles for the substrates Z-Phe-Arg-pNA and Z-Arg-Arg-pNA

The pK_a values and their standard deviations were calculated from three separate determinations. The k_{max} values are the theoretical limit values for k_{cat}/K_m (M^{-1} s^{-1}) for each of the ionization events. The value in parentheses was determined from the experiment with Z-Phe-Arg-pNA as substrate and used as a constant for fitting the rat liver cathepsin B pH profile with the Z-Arg-Arg-pNA substrate. For the rat liver enzyme and the Z-Arg-Arg-pNA substrate, the experiment was done twice, and the values reported are for one of these data sets along with the standard errors.

	Rat liver	S115A, Q255Term
Z-Phe-Arg-pNA		
pK_a1	3.59 ± 0.04	3.58 ± 0.04
pK_a2	5.29 ± 0.03	5.38 ± 0.09
pK_a3	8.10 ± 0.15	7.66 ± 0.04
pK_a4	8.76 ± 0.28	8.55 ± 0.02
$k_{1\,max}$	27,530 ± 1,224	25,260 ± 1,556
$k_{2\,max}$	51,100 ± 3,164	53,540 ± 1,346
$k_{3\,max}$	471,000 ± 94,620	314,300 ± 13,390
Z-Arg-Arg-pNA		
pK_a1	5.03 ± 0.02	5.13 ± 0.11
pK_a2	8.15 ± 0.01	7.75 ± 0.15
pK_a3	(8.76)	8.68 ± 0.16
$k_{1\,max}$	6,737 ± 66	6,908 ± 1,011
$k_{2\,max}$	76,040 ± 571	44,240 ± 7,796

ascending and one in the descending limb. EH4 and E are inactive forms of the enzyme.

$$\begin{array}{ccccccccc} & & \mathrm{p}K_1 & & \mathrm{p}K_2 & & \mathrm{p}K_3 & & \mathrm{p}K_4 \\ EH_4 & \rightleftarrows & EH_3 & \rightleftarrows & EH_2 & \rightleftarrows & EH & \rightleftarrows & E \\ & & \Downarrow & & \Downarrow & & \Downarrow & & \\ & & k_1 & & k_2 & & k_3 & & \end{array} \qquad \text{(Model 1)}$$

The equation derived to fit the data to this model is as follows:

$$\begin{aligned} k_{obs} = {} & k_{1\,limit}/([H^+]/K_1 + 1 + K_2/[H^+] + K_2K_3/[H^+]^2 + K_2K_3K_4/[H^+]^3) \\ & + k_{2\,limit}/([H^+]^2/K_1K_2 + [H^+]/K_2 + 1 + K_3/[H^+] + K_3K_4/[H^+]^2) \\ & + k_{3\,limit}/([H^+]^3/K_1K_2K_3 + [H^+]^2/K_2K_3 + [H^+]/K_3 + 1 + K_4/[H^+]) \end{aligned} \qquad (1)$$

k_{obs} is the experimentally observed k_{cat}/K_m, and $k_{1\,limit}$, $k_{2\,limit}$ etc. are the theoretical maximal k_{cat}/K_m values for the respective dissociation events.

The data for the Z-Arg-Arg-pNA was best fitted to an equation described by Model 2 which involves the enzyme undergoing three dissociation events, two in the ascending and one in the descending limb. EH3 and E are inactive forms of the enzyme.

$$\begin{array}{ccccccc} & & \mathrm{p}K_1 & & \mathrm{p}K_2 & & \mathrm{p}K_3 \\ EH_3 & \rightleftarrows & EH_2 & \rightleftarrows & EH & \rightleftarrows & E \\ & & \Downarrow & & \Downarrow & & \\ & & k_1 & & k_2 & & \end{array} \qquad \text{(Model 2)}$$

The equation derived to fit the data to this model is as follows:

$$\begin{aligned} k_{obs} = {} & k_{1\,limit}/([H^+]/K_1 + 1 + K_2/[H^+] + K_2K_3/[H^+]^2) \\ & + k_{2\,limit}/([H^+]^2/K_1K_2 + [H^+]/K_2 + 1 + K_3/[H]) \end{aligned} \qquad (2)$$

This model, for the Z-Arg-Arg-pNA substrate, ignores the effect of the deprotonation of the group presumed to be the active site Cys[29] (20) (pK_a 3.6, the first dissociation constant calculated from a pH activity profile with substrate Z-Phe-Arg-pNA) since a group with a $pK_a \approx 5.1$ has a very large effect (20-fold) on activity (Fig. 2B), and very little activity can be measured below pH 4.2.

While the shape of the pH profiles for the native rat liver enzyme and the recombinant mutants with the pNA substrates are essentially the same, part of the ascending limb, from about pH 6 onward, is shifted to the left for the recombinant protein compared with that of the rat liver enzyme. This shift is due to a change in the pK_a of the last ionizable group in the ascending limb, from a pK_a of 8.1 for the rat liver enzyme to 7.7 for the recombinant cathepsin B. The first and second pK_a values, 3.6 and 5.3, appear to be similar for the rat liver and recombinant enzymes. The pK_a values in the descending limb for Z-Phe-Arg-pNA and rat liver and recombinant enzymes, 8.76 ± 0.28 and 8.55 ± 0.02 (Table III), appear to be associated with the same group taking into consideration the larger experimental error for the rat liver enzyme. The descending limb pK_a of 8.68 ± 0.16 for the recombinant enzyme with Z-Arg-Arg-pNA (Table III) lends support to the possibility that the group associated with this pK has a similar environment in the recombinant and rat liver enzymes. Although the pH activity data only show the beginning of the descending limb, the suggestion that the dissociation of this group leads to an inactive enzyme (due to the deprotonation of the active site His[199]) was confirmed by activity measurements at higher pH at 4 °C (data not shown). The larger error in the calculation of the pK of His[199] for the rat liver enzyme is due to the fact that the maximal observed rate (the peak of the pH activity profile) is shifted to the right due to an increase in the final ascending limb pK_a. The denaturation of the enzyme above pH 8.4 at 24 °C made precise calculation of the pK of the final dissociating group in the rat liver enzyme difficult. As described by Alberty and Massey (33), when 2 pK values are close together, the shape of the activity profile changes depending on the proximity of the 2 pK values. The shape of the ascending limb also contains information on the descending limb pK and, therefore, with regression analysis using appropriate equations, both the ascending and descending limb pK values can be calculated with only half of the pH activity profile data. The accuracy of calculation depends on the extent and quality of the data. For example, the increased error for the Z-Arg-Arg-pNA substrate is due to reduced sensitivity of the assay compared to Z-Phe-Arg-pNA. The greater error for the Z-Arg-Arg-pNA substrate prevented the calculation of the descending limb pK with the rat liver enzyme, and, therefore, a constant of 8.76, calculated from the Z-Phe-Arg-pNA data (Table III), was used in the fitting.

Comparison of the pH Activity Profiles for Recombinant and Rat Liver Cathepsin B Using the Substrates Z-Phe-Arg-AMC and Z-Arg-Arg-AMC—pH activity profiles were also determined with Z-Phe-Arg-AMC and Z-Arg-Arg-AMC substrates (Fig. 3, A and B), and the calculated apparent pK values and the theoretical limit values of k_{cat}/K_m for each of the dissociation events are listed in Table IV. The data for the Z-Phe-Arg-AMC substrate can best be fitted to an equation describing a model (Model 3) considering five dissociation events involved in enzyme activity, three in the ascending limb and two in the descending limb.

$$\begin{array}{ccccccccccc} & & \mathrm{p}K_1 & & \mathrm{p}K_2 & & \mathrm{p}K_3 & & \mathrm{p}K_4 & & \mathrm{p}K_5 \\ EH_5 & \rightleftarrows & EH_4 & \rightleftarrows & EH_3 & \rightleftarrows & EH_2 & \rightleftarrows & EH & \rightleftarrows & E \\ & & \Downarrow & & \Downarrow & & \Downarrow & & \Downarrow & & \\ & & k_1 & & k_2 & & k_3 & & k_4 & & \end{array} \qquad \text{(Model 3)}$$

The equation derived to fit the data to this model is as follows:

$$\begin{aligned} k_{obs} = {} & k_{1\,limit}/([H^+]/K_1 + 1 + K_2/[H^+] + K_2K_3/[H^+]^2 + K_2K_3K_4/[H^+]^3 + K_2K_3K_4K_5/[H^+]^4) + \\ & k_{2\,limit}/([H^+]^2/K_1K_2 + [H^+]/K_2 + 1 + K_3/[H^+] + K_3K_4/[H^+]^2 + K_3K_4K_5/[H^+]^3) + \\ & k_{3\,limit}/([H^+]^3/K_1K_2K_3 + [H^+]^2/K_2K_3 + [H^+]/K_3 + 1 + K_4/[H^+] + K_4K_5/[H^+]^2) + \\ & k_{4\,limit}/([H^+]^4/K_1K_2K_3K_4 + [H^+]^3/K_2K_3K_4 + [H^+]^2/K_3K_4 + [H^+]/K_4 + 1 + K_5/[H^+] \end{aligned} \qquad (3)$$

The pH profile data for the Z-Arg-Arg-AMC substrate can be fitted to the equation described by Model 1 involving four dissociation events, except that there are two dissociation events in the ascending limb and two in the descending limb (Fig. 3B and Table IV) as compared to the Z-Phe-Arg-pNA substrate profile which can be fitted by the same equation but with three dissociation events in the ascending limb and only one in the descending limb.

Again, as for the pNA substrates, the overall shape of the AMC profiles are similar, and the same number of dissociable groups can be identified for the recombinant and rat liver enzymes. The first striking feature of the pH profiles with the AMC substrates is that there is an additional group which appears to influence activity compared to the pH profiles with pNA substrates. Secondly, with the AMC substrates, the difference between the rat liver and recombinant enzymes above pH 6 are more readily apparent due to a large shift in the pK_a of a group in the ascending limb as discussed below.

For the Z-Phe-Arg-AMC substrate, the first and second pK_a values, about 3.6 and 4.9, appear to be similar for the rat liver and recombinant enzymes. This group with the pK_a of 4.9, which is active in the Z-Phe-Arg-AMC hydrolysis, does not appear to be active in the Z-Phe-Arg-pNA hydrolysis. Instead, the second dissociable group in the Z-Phe-Arg-pNA

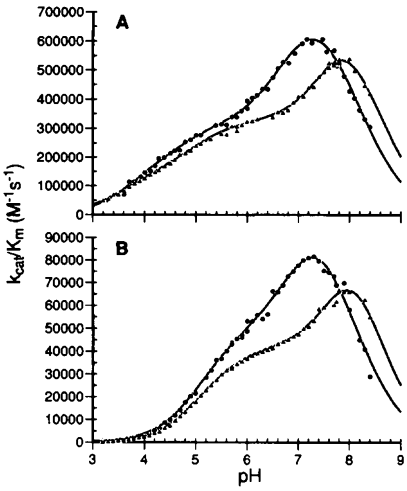


FIG. 3. **pH activity profiles of rat liver cathepsin B compared with the double mutant, (S115A, Q255Term).** *A*, with the substrate *Z*-Phe-Arg-AMC. *B*, with the substrate *Z*-Arg-Arg-AMC. ●, double mutant (S115A, Q255Term). ▲, rat liver cathepsin B. The conditions for assays are as described under "Experimental Procedures." The *lines* through the data represent the best fit using equations described in the text.

profile has a pK_a of 5.3.

With the *Z*-Arg-Arg-AMC substrate, as is the case with the *Z*-Arg-Arg-*p*NA, activity below pH 4 is too low to accurately calculate the pK_a of 3.6, the constant for the first dissociation event observed with Phe-Arg-containing substrates. As with the *Z*-Arg-Arg-*p*NA substrate, the first pK_a that can be calculated for the *Z*-Arg-Arg-AMC substrate is 5.09 ± 0.05 for the rat liver enzyme and 5.07 ± 0.03 for the recombinant enzyme Table IV (consistent with the value of 5.1 calculated from the *p*NA data (Table III)). For the last pK_a in the ascending limb, there is a major difference between the rat liver and recombinant enzymes, 7.58 ± 0.06 and 6.90 ± 0.15, respectively, for the *Z*-Phe-Arg-AMC and 7.81 ± 0.33 and 6.95 ± 0.09, respectively, for *Z*-Arg-Arg-AMC.

In contrast to the fitting of the *p*NA data, the AMC data were fitted according to a model describing two dissociation events in the descending limb of the pH activity profile. If only a single dissociation is considered for the descending limb of the AMC pH profile, the pK_a calculated is approximately 8.1 for the recombinant enzyme. Since the results from the *p*NA data show that the final dissociation leading to inactive enzyme has a pK_a of 8.6, a single dissociation event with a pK_a of 8.1 cannot correctly describe the descending limb of the AMC pH activity profile, and, therefore, this must be a composite of two pK values in the descending limb. The presence of two pK values in the descending limb was confirmed by activity measurements up to pH 9.4 at 4 °C (data not shown). For these reasons we used constants, calculated from the *p*NA data, for the final dissociation event leading to inactive enzyme in the fitting of the AMC pH activity profiles.

The first pK_a in the descending limb of the *Z*-Phe-Arg-AMC substrate pH activity profiles was calculated to be 7.82 ± 0.24 for rat liver enzyme and 7.70 ± 0.24 for the recombinant enzyme (Table IV). The group associated with this pK_a appears to be the same as the one associated with the last ascending limb pK in the *p*NA profiles, 8.10 ± 0.15 for rat liver and 7.66 ± 0.04 for the recombinant enzyme (Table III).

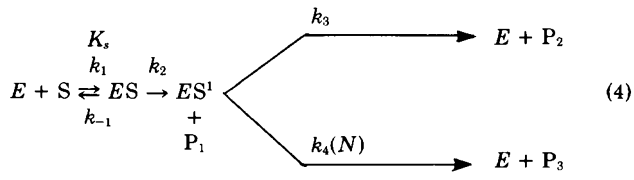
Determination of Acylation and Deacylation Rate Constants for a Better Understanding of the Function of the Group with a pK_a of 7.7—To understand why with the *p*NA substrates, dissociation of a group (pK_a of 7.7 for the recombinant enzyme) leads to an increase in k_{cat}/K_m while, with the AMC substrates, dissociation of the same group leads to a decrease in activity, we conducted nucleophile competition experiments at pH 7.0 and 8.2 to directly measure changes in k_2 and K_s (since $k_{cat}/K_m = k_2/K_s$). Bajkowski and Frankfater (31, 32) presented the following relationships to describe the influence of competing nucleophiles other than water in the deacylation reaction in the hydrolysis of substrates by cathepsin B (Equations 4 to 6).

TABLE IV

Ionization constants and k_{cat}/K_m limit values of dissociable groups involved in substrate binding and catalysis calculated from pH activity profiles for the substrates Z-Phe-Arg-AMC and Z-Arg-Arg-AMC

With the *Z*-Phe-Arg-AMC substrate, three separate determinations were made for the recombinant enzyme and the averages and standard deviations are listed while, for rat liver cathepsin B, the experiment was done twice and the data and standard errors of one of the experiments are reported. The values of 8.55 and 8.76 in parentheses were determined from experiments with the recombinant and rat liver enzymes respectively with the *Z*-Phe-Arg-*p*NA substrate. For the *Z*-Arg-Arg-AMC substrate, the value in parentheses (8.55) was determined from the experiment with the recombinant enzyme and *Z*-Phe-Arg-*p*NA as substrate, and the values 8.1 and 8.76 in parentheses were determined from experiments with rat liver enzyme and *Z*-Phe-Arg-*p*NA as substrate (Table III). These values in parentheese were used as constants for fitting the pH profile data with AMC-containing substrates. For the *Z*-Arg-Arg-AMC substrate, three separate experiments were conducted for the rat liver cathepsin B and the values listed are the averages and the standard deviations, while for the recombinant enzyme the experiment was done twice, and the data and standard errors for one of these are reported. The conditions are the same as in Table III.

	Rat liver	S115A
Z-Phe-Arg-AMC		
pK_a1	3.53 ± 0.07	3.70 ± 0.05
pK_a2	4.93 ± 0.02	4.85 ± 0.12
pK_a3	7.58 ± 0.06	6.90 ± 0.15
pK_a4	7.82 ± 0.24	7.70 ± 0.18
pK_a5	(8.76)	(8.55)
$k_{1\,max}$	149,500 ± 2,827	183,900 ± 11,910
$k_{2\,max}$	323,400 ± 2,902	334,300 ± 12,250
$k_{3\,max}$	700,600 ± 28,912	806,000 ± 73,270
$k_{4\,max}$	543,200 ± 20,940	372,600 ± 29,250
Z-Arg-Arg-AMC		
pK_a1	5.09 ± 0.05	5.07 ± 0.03
pK_a2	7.81 ± 0.33	6.95 ± 0.09
pK_a3	(8.10)	7.71 ± 0.12
pK_a4	(8.76)	(8.55)
$k_{1\,max}$	44,730 ± 3,030	49,560 ± 1,352
$k_{2\,max}$	115,900 ± 41,560	110,600 ± 5,728
$k_{3\,max}$	53,700 ± 31,270	46,560 ± 4,794



(4)

$$K_m = \frac{(k_3 + k_4(N))K_s}{k_2 + k_3 + k_4(N)} \quad (5)$$

$$V_{max} = \frac{k_2(k_3 + k_4(N))(E)_0}{k_2 + k_3 + k_4(N)} \quad (6)$$

They have shown that, for cathepsin B-catalyzed hydrolysis of *p*-nitrophenyl esters at pH 6.5, deacylation is rate-limiting (*i.e.* $k_{cat} = k_3$), while the hydrolysis of *p*-nitroanilide substrates is acylation rate-limiting (*i.e.* $k_{cat} = k_2$) (31, 32). Table V lists the results of the determination of the acylation rate constant k_2, the deacylation rate constant k_3, K_s, and k_{cat} from the nucleophile competition experiments at pH 7.0 and 8.2 for

the recombinant double mutant Ser[115]-Ala, Gln[255]-Term. The acylation and deacylation rate constants for substrate *Z*-Phe-Arg-AMC were calculated directly from nucleophile competition experiments using Gly-Gly as nucleophile, and k_{cat} was calculated from the following relationship.

$$k_{cat} = \frac{k_2 k_3}{k_2 + k_3} \quad (7)$$

K_s was calculated from the relationship:

$$K_m = \frac{K_s k_3}{k_2 + k_3} \quad (8)$$

K_m was calculated from Michaelis-Menten kinetics. Since the deacylation rate constants for both *Z*-Phe-Arg-AMC and *Z*-Phe-Arg-*p*NA should be identical, the values of k_2 and K_s for the *p*NA substrate (Table V) were calculated using the k_{cat} and K_m values from Michaelis-Menten experiments using k_3 values determined from the nucleophile competition experiments with *Z*-Phe-Arg-AMC (Equations 7 and 8). This was necessary since hydrolysis of *p*NA substrates, as reported by Bajkowski and Frankfater (31, 32) and confirmed by the present study, is mainly acylation rate-limiting below pH 7.0 (k_3 is 7 times larger than k_2 at pH 7.0 (Table V)), and, therefore, nucleophile competition cannot be used to determine k_2. At pH 8.2 for *Z*-Phe-Arg-*p*NA, as shown in Table V, $k_{cat} = k_2 k_3/(k_2 + k_3)$ since k_2 and k_3 are almost equal. This was confirmed by conducting nucleophile competition experiments with Gly-Gly as the nucleophile and *Z*-Phe-Arg-*p*NA as substrate at pH 8.2. The value of k_3 calculated from experiments with the *Z*-Phe-Arg-*p*NA substrate and Gly-Gly nucleophile was consistent with the value of k_3 determined from the *Z*-Phe-Arg-AMC data (78 and 60, respectively) (Table V, see legend). For comparison, Table V lists the values of k_{cat} and K_m calculated from Michaelis-Menten kinetics by varying substrate concentrations using *Z*-Phe-Arg-AMC and *Z*-Phe-Arg-*p*NA. It is clear that the values calculated for k_{cat} using the two different methods agree very well within the limits of the experimental error.

It is apparent from the data in Table V that, for the recombinant enzyme with the *Z*-Phe-Arg-AMC substrate, k_{cat} is dependent on both k_2 and k_3 (Equation 7) at pH 7.0 as well as 8.2. As the group with a pK_a of 7.7 dissociates on raising the pH from 7.0 to 8.2, k_2 drops from 111 ± 3.61 to 37.6 ± 0.61. Since there is only a small decrease in K_s, the drop in k_2 leads to a decrease in k_{cat}/K_m as observed in the pH activity profiles with the AMC substrates. With the *p*NA substrate, there is a small increase in k_2 over the same pH range and a small decrease in K_s leading to a 2-fold increase in k_{cat}/K_m as observed in the pH activity profiles.

DISCUSSION

Kinetic Characterization of Glycosylation Minus Mutants—The removal of the single glycosylation consensus sequence in the mature region by site-directed mutagenesis eliminated *N*-linked oligosaccharide substitution, and we have shown that glycosylation minus mutants expressed in yeast are functionally identical at pH 6.0 with cathepsin B purified from rat liver on the basis of the kinetic parameters k_{cat} and K_m for 4 substrates (Table II).

pH Activity Profiles and Evidence for 2 Histidines in the Leaving Group Binding Site—pH activity profiles can serve as sensitive measures of functional and structural integrity when comparing recombinant with native enzymes, as well as providing information on possible groups involved in binding and catalysis. The pH dependence of k_{cat}/K_m can be used to calculate pK_a values of groups on the free enzyme that affect activity upon dissociation, as discussed by Fersht (34). Small changes in the environment of the dissociable groups on recombinant proteins and their mutants may be reflected in altered pK values of these groups. The advantage of pH activity profiles using k_{cat}/K_m, as opposed to k_{cat} or K_m, is that since $k_{cat}/K_m = k_2/K_s$ (when $k_{-1} >> k_2$), changes in the rate-limiting step (Equation 4) do not affect the calculation of pK values. The pK values determined by this method reflect the ionization state of groups on the free enzyme (and free substrate) as opposed to the enzyme substrate complex, $E \cdot S$, or the acylenzyme, $E \cdot S'$. The models used to account for dissociable groups in the cathepsin B active site are minimal models with only the minimum number of dissociation events being considered in the equation to achieve a good fit for the

TABLE V

Effect of pH on kinetic constants for the S115A, Q255Term double mutant with Z-Phe-Arg-AMC and Z-Phe-Arg-pNA substrates

For the *Z*-Phe-Arg-AMC substrate, k_2 and k_3 were calculated from nucleophile competition experiments using Gly-Gly as the competing nucleophile. k_2 and k_3 were also calculated independently at pH 8.2 for *Z*-Phe-Arg-*p*NA in the presence of Gly-Gly, and the average of two experiments was calculated to be 32.8 s^{-1} and 78.1 s^{-1}, respectively, in reasonable agreement with the values calculated with the AMC substrate considering the larger errors for the *p*NA substrate. No effect of nucleophile was observed at pH 7.0 at any nucleophile concentration with *Z*-Phe-Arg-*p*NA, which is consistent with the fact that k_3 is 7 times greater than k_2. Nucleophile concentration ranged from 50 mM to 600 mM. The reaction buffers consisted of 25 mM sodium phosphate, 1 mM EDTA, 250 mM NaCl, pH 7.0, and 25 mM sodium borate, 1 mM EDTA, 250 mM NaCl, pH 8.2. The ionic strength was $I = 0.3$. The values in parentheses for k_{cat} and K_m were determined using Michaelis-Menten kinetics by varying the substrate concentration (see Table II for conditions). The data reported are the averages of three experiments except for the value of k_2 with *Z*-Phe-Arg-*p*NA which is an average of two determinations.

pH	k_2	k_3	k_{cat}	K_s	K_m	k_4
	s^{-1}	s^{-1}	s^{-1}	*mM*	*mM*	$mM^{-1}\ s^{-1}$
Z-Phe-Arg-AMC						
7.0	111 ± 3.61	244 ± 45.2	75.9 ± 5.63 (80.8)	0.193 ± 0.010	(0.132)	3.57 ± 1.12
8.2	37.6 ± 0.61	59.6 ± 1.32	23.1 ± 0.36 (24.2 ± 1.57)	0.109 ± 0.001	(0.067 ± 0.007)	3.79 ± 0.85
Z-Phe-Arg-*p*NA						
7.0	36.3	244 ± 45.2[a]	(31.6)	0.345	(0.300)	
8.2	45.8 ± 10.5	59.6 ± 1.32[a]	(23.1 ± 0.36)	(0.117 ± 0.011)	0.208 ± 0.020	

[a] For the *Z*-Phe-Arg-*p*NA substrate, the deacylation rate constants at pH 7.0 and 8.2 are from the experiments with the *Z*-Phe-Arg-AMC substrate and Gly-Gly nucleophile.

data. pK_a values of functional groups on the enzyme can be calculated from these profiles as long as there are not more pK_a values in the pH profile than can be reliably handled by regression analysis.

A comparison of the pH profiles of the rat liver and recombinant enzymes shows that while the enzymes are kinetically similar (based on the constants calculated from the Michaelis equation and from the theoretical limit values for k_{cat}/K_m calculated from the pH activity profiles), there is a difference in the pK_a values of two groups involved in activity. A group with a pK_a of 8.1 in the rat liver enzyme, which appears in the ascending limb of the *p*NA substrate pH profiles and in the descending limb of AMC profiles, has a pK_a of 7.7 in the recombinant enzyme. Similarly, a group which has a pK_a of about 7.6 in the rat liver enzyme and is found in the ascending limb of the AMC profile has a pK_a of about 6.9 in the recombinant enzyme. In summary, considering just the recombinant enzyme, the dissociation of a group with a pK_a of 6.9 results in an increase in k_{cat}/K_m for the AMC substrates but does not appear to influence *p*NA hydrolysis. Furthermore, dissociation of a group with pK_a of 7.7 results in an increase in k_{cat}/K_m for *p*NA substrates but a decrease in activity for AMC substrates. Since a different response is seen depending on whether substrates contain *p*NA or AMC, it appears that these groups with p*K* values of 6.9 and 7.7 are part of the recognition site for the leaving group (P′). Since the pK_a values are in the range possible for histidine imidazole moieties, we suggest that they are associated with His^{110} and His^{111}. These histidine residues, which are found in a large loop (an insert of 18 residues relative to papain) occluding the top of the active site cleft, are about 8 Å from the active site Cys^{29} (an appropriate distance and location for interacting with leaving groups) according to the crystal structure (21). This leads us to support the suggestion by Musil *et al.* (21) that His^{111} may play an important role in the exopeptidase activity (dipeptidyl carboxypeptidase) of cathepsin B by stabilization of the COOH-terminal carboxylate of the substrate through a salt bridge or hydrogen bond. From the x-ray structure, His^{110} appears to form a salt bridge with Asp^{22}, but it is still possible that the imidazole nitrogen not involved in the salt bridge could hydrogen-bond with the substrate carboxylate.

Musil *et al.* (21) have speculated that His^{111} may be the group associated with the pK_a of about 5.5 which was suggested to control the exopeptidase activity of cathepsin B (deprotonation of a group with a pK_a of 5.5 was reported to decrease exopeptidase activity) (19). Our present study indicates that His^{111} probably has a pK_a of either 6.9 or 7.7 in the recombinant enzyme (7.6 and 8.1, respectively, in the rat liver enzyme). While we cannot definitively say at this point which of the p*K* values is associated with His^{111}, the identity of the group with the pK_a of 5.5 controlling the exopeptidase activity still remains uncertain.

The difference in the p*K* values observed for His^{110} and His^{111}, for the rat liver and recombinant enzymes, may be due to the proximity of the *N*-linked oligosaccharide substitution at Asn^{113} in the rat liver enzyme. The carbohydrate may directly influence the p*K* values of these histidines through H-bond interactions or through a subtle change in the structure of the loop. Other possible but less likely explanations for the differences observed in the recombinant enzyme are, firstly, some influence of the additional negative charges (a Glu and an Asp) in the 6-residue NH_2-terminal extension,[3] and, secondly, an influence of the additional positive charge (Arg^{49}) in the Gly-Arg dipeptide sequence which is clipped out of cathepsin B in rat tissues during processing to the two-chain form. However, irrespective of the change in the p*K* values of these groups in the recombinant enzyme, their function remains the same as in the rat liver enzyme. This decrease in the 2 p*K* values in the recombinant enzyme is in fact serendipitous since it increases the separation of the p*K* values allowing us to characterize more accurately the role of these groups in the active site of this pH-sensitive enzyme.

The Role and Action of the Histidines in the Leaving Group Binding Subsite—On the basis of Michaelis-Menten kinetics, pH activity profiles, and the calculation of acylation and deacylation rate constants from nucleophile competition experiments, we have determined that, with respect to the synthetic endopeptidase substrates, the groups in the recombinant enzyme with pK_a values of 6.9 and 7.7 (most probably His^{110} and His^{111}) play an important role in substrate binding and catalysis. The data suggest that the group with a pK_a of 6.9 is involved in binding of AMC substrates, but not *p*NA substrates. The group with a pK_a of 7.7 has a more complicated role. For the *p*NA substrate, deprotonation of the group with a pK_a of 7.7 leads to an increase in k_{cat}/K_m (Fig. 2) which results from a small decrease in K_s and a small increase in k_2 as measured between pH 7.0 and 8.2 (Table V). With the AMC substrate, this group, upon deprotonation, leads to a drop in k_{cat}/K_m (Fig. 3) which results from a large drop in k_2, the acylation rate constant (Table V). This suggests a role for the imidazolium cation in increasing the binding energy of the enzyme-transition state complex when a substrate with the AMC leaving group binds in the active site.

The deprotonation of the group with a pK_a of 7.7 also leads to a drop in k_3, the deacylation rate constant, as determined for the *Z*-Phe-Arg-acylenzyme complex. In the deacylation reaction, the imidazolium cation cannot possibly interact directly with the substrate due to distance constraints, and, therefore, the group with a pK_a of 7.7 influences catalysis and the transition state through an electrostatic effect at a distance. The implication of the dissociation of the group with a pK_a of 7.7 being associated with the drop in k_3 must be treated with caution since these p*K* values were calculated for the free enzyme and may be different in the acylenzyme. However, since the group with the pK_a of 7.7 (His^{110} or His^{111}) is in the leaving group binding site, it is possible that its pK_a is the same in the free enzyme and the acylenzyme. Verification of the above possible mechanisms, as well as the possible role for His^{110} and His^{111} in the dipeptidyl carboxypeptidase activity, will have to await detailed characterization of the site-directed mutants of His^{110} and His^{111}, work which is in progress.

Evidence for Similarities in the Active Site Cleft of Rat Liver and Recombinant Cathepsin B—Notwithstanding the possible alteration of the loop on which His^{110} and His^{111} are located, the most critical indicator of structural integrity of the active site is the highly perturbed pK_a of 3.6 for the thiol of the active site Cys^{29}. As for papain, the low pK_a observed for the active site cysteine is probably due to the stabilization of the thiolate anion by the positive charge of active site His^{199} (35, 36) as well as the positive dipole moment of the α-helix which starts at Cys^{29}. This perturbed pK_a should be highly sensitive to alterations in the structure around the active site. The pK_a of 3.59 ± 0.04 for the rat liver enzyme and 3.58 ± 0.04 for the double mutant indicate a high degree of similarity of the active sites of the two enzymes. Four other groups, apart from His^{110} and His^{111}, have also been shown to be similar in the rat liver and recombinant enzymes, within the accuracy of the data. Therefore, we conclude from the comparison of the kinetic constants, the apparent pK_a values, and the theoretical limit values of k_{cat}/K_m, for each of the dissociation events, that the catalytic site and much of the substrate binding sites of the recombinant cathepsin B are structurally and function-

ally identical with the native rat liver enzyme.

Determination of the pK_a of the Active Site His[199]—The pH activity profiles of recombinant cathepsin B with the *p*NA substrates allowed the calculation of a pK_a of 8.6 associated with the final dissociation event leading to an inactive enzyme. This pK_a 8.6 is comparable to the pK_a of about 8.3 calculated for the active site His[159] of papain (20), and, therefore, we suggest that the pK_a of 8.6 in cathepsin B is probably associated with the active site His[199].

Further Evidence for the Role of a Carboxyl Group in the S2 Subsite—The apparent anomalous behavior of cathepsin B, whereby a group with a pK_a of 5.1 appears to be essential for activity when a substrate has an Arg as opposed to Phe in the P_2 position, has been the source of considerable speculation (20). The role of the group with a pK_a of 5.1 has been discussed in detail recently with respect to cathepsin B purified from human liver (20), and the suggestion has been made that this group is Glu[245] in the S_2 subsite. The pH activity profiles can be explained simply on the basis of a direct interaction of Glu[245] with the P_2 Arg, which is supported by the x-ray structure[2] (21). The difference in the free energy of binding of the Arg-Arg substrate with the neutral or ionized carboxyl, about 1.8 kcal/mol (equivalent to a 20-fold difference in k_{cat}/K_m), is reasonable for the difference in interactions between the uncharged carboxyl-charged guanidino and charged carboxylate-charged guanidino groups.

Considering the fact that there is a 7-fold preference for substrates containing Phe as opposed to Arg in the P_2 position (equivalent to approximately 1.15 kcal/mol difference in the free energy), the possibility exists that the P_2 Phe side chain of the substrate does not occupy the S_2 subsite in the same orientation as the P_2 Arg.

Evidence for Group(s) in the S_1 Subsite of Cathepsin B—In the pH activity profile of *Z*-Phe-Arg-*p*NA substrate, the pK_a of 5.3 (the second dissociation event) cannot be due to the same group as that associated with the pK_a of 5.1 (probably Glu[245]) determined with the *Z*-Arg-Arg-*p*NA substrate. The 2-fold increase in activity for the *Z*-Phe-Arg-*p*NA substrate resulting from the dissociation of the group with a pK_a of 5.3 is probably not due to an interaction of the P_2 Phe with what appears to be a carboxyl group. Since the substrate contains Arg in the P_1 position, one possibility is that the group with a pK_a of 5.3 is in the S_1 subsite. The possible candidates on the basis of their proximity to the active site are Glu[171], Asp[69], Glu[122], and Asp[22], all within a 13-Å radius of the active site thiol of Cys[29].[2] For the *Z*-Phe-Arg-AMC substrate pH activity profile, the second dissociation event is associated with a pK_a of 4.9 instead of a pK_a of 5.3 observed for the *p*NA substrate. This, together with the evidence that the AMC and *p*NA groups bind differently in the leaving group (P') subsite (see above), suggests that the binding of these substrates may differ in the S_1 subsite as well if the pK values of 4.9 and 5.3 are associated with groups in the S_1 subsite.

In summary, we have shown that the recombinant cathepsin B is functionally similar to rat liver enzyme. Furthermore, we have shown that the active site of cathepsin B is highly complex and that seven dissociable groups influence cathepsin B activity toward synthetic substrates. These groups may be important in the binding and hydrolysis of physiological substrates, and some may also influence the different pH optima observed for the exo- and endopeptidase activities of cathepsin B. Further clarification of the role of groups in the cathepsin B active site, such as Glu[245], His[110], His[111], Asp[22], Glu[171], Asp[69], and Asp[224], should be achieved through a systematic analysis using the technique of site-directed mutagenesis.

Acknowledgments—S. H. would like to acknowledge Drs. Peter Tonge and Jon Durkin of the Institute for Biological Sciences, National Research Council of Canada, for helpful suggestions on the manuscript.

REFERENCES

1. Mort, J. S., Recklies, A. D., and Poole, A. R. (1984) *Arthritis Rheum.* **27,** 509-515
2. van Noorden, C. J. F., Smith, R. E., and Rasnick, D. (1989) *J. Rheumatol.* **15,** 1525-1535
3. Gopalan, P., Dufresne, M. J., and Warner, A. H. (1987) *Can. J. Physiol. Pharmacol.* **65,** 124-129
4. Katunuma, N., and Kominami, E. (1987) *Rev. Physiol. Biochem. Pharmacol.* **108,** 2-20
5. Lah, T. T., Buck, M. R., Honn, K. V., Crissman, J. D., Rao, N. C., Liotta, L. A., and Sloane, B. F. (1989) *Clin. Exp. Metastasis* **7,** 461-468
6. Sloane, B. F. (1990) *Semin. Cancer Biol.* **1,** 137-152
7. Sloane, B. F., Moin, K., Krepela, E., and Rozhin, J. (1990) *Cancer Metastasis Rev.* **9,** 333-352
8. San Segundo, B., Chan, S. J., and Steiner, D. F. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 2320-2324
9. Chan, S. J., San Segundo, B., McCormick, M. B., and Steiner, D. F. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 7721-7725
10. Hara, K., Kominami, E., and Katunuma, N. (1988) *FEBS Lett.* **231,** 229-231
11. Kominami, E., Tsukahara, T., Hara, K., and Katunuma, N. (1988) *FEBS Lett.* **231,** 225-228
12. Ritonja, A., Popovi, T., Turk, V., Widenman, K., and Machleidt, W. (1985) *FEBS Lett.* **181,** 169-171
13. Takahashi, T., Schmidt, P. C., and Tang, J. (1984) *J. Biol. Chem.* **259,** 6059-6062
14. Nishimura, Y., Furuno, K., and Kato, K. (1988) *Arch. Biochem. Biophys.* **263,** 107-116
15. Nishimura, Y., Amano, J., Sato, H., Tsuji, H., and Kato, K. (1988) *Arch. Biochem. Biophys.* **262,** 159-170
16. Barrett, A. J., and Kirschke, H. (1981) *Methods Enzymol.* **80,** 535-561
17. Pohl, J., Davinic, S., Bláha, I., Trop, P., and Kostka, V. (1987) *Anal. Biochem.* **165,** 96-101
18. Willenbrock, F., and Brocklehurst, K. (1985) *Biochem. J.* **227,** 521-528
19. Polgar, L., and Csoma, C. (1987) *J. Biol. Chem.* **262,** 14448-14453
20. Khouri, H. E., Plouffe, C., Hasnain, S., Hirama, T., Storer, A. C., and Ménard, R. (1991) *Biochem. J.* **275,** 751-757
21. Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, V., Turk, V., Towatari, T., Katanuma, N., and Bode, W. (1991) *EMBO J.* **10,** 2321-2330
22. Mort, J. S., Tam, A., Steiner, D. F., and Chan, S. J. (1988) *Biol. Chem. Hoppe-Seyler* **369** (suppl.), 163-167
23. Chan, M. M.-Y., and Fong, D. (1988) *FEBS Lett.* **239,** 219-222
24. Vernet, T., Tessier, D. C., Laliberte, F., Dignard, D., and Thomas, D. Y. (1989) *Gene (Amst.)* **77,** 229-236
25. Cohen, L. W., Fluharty, C., and Dihel, L. C. (1990) *Gene (Amst.)* **88,** 263-267
26. Smith, S. M., and Gottesman, M. M. (1989) *J. Biol. Chem.* **264,** 20487-20495
27. Lee, X., Ahmed, F. R., Hirama, T., Huber, C. P., Rose, D. R., To, R., Hasnain, S., Tam, A., and Mort, J. S. (1990) *J. Biol. Chem.* **265,** 5950-5951
28. Ernst, J. F. (1986) *DNA (N. Y.)* **5,** 483-491
29. Rich, D. H., Brown, M. A., and Barrett, A. J. (1986) *Biochem. J.* **235,** 731-734
30. Barrett, A. J. (1973) *Biochem. J.* **131,** 809-822
31. Bajkowski, A. S., and Frankfater, A. (1983) *J. Biol. Chem.* **258,** 1645-1649
32. Bajkowski, A. S., and Frankfater, A. (1983) *J. Biol. Chem.* **258,** 1650-1655
33. Alberty, R. A., and Massey, V. (1954) *Biochim. Biophys. Acta* **13,** 347-353
34. Fersht, A. R. (1985) *Enzyme Structure and Mechanism*, pp. 155-175, W. H. Freeman and Co., New York
35. Lewis, S. D., Johnson, F. A., and Shafer, J. A. (1976) *Biochemistry* **15,** 5009-5017
36. Lewis, S. D., Johnson, F. A., and Shafer, J. A. (1981) *Biochemistry* **20,** 48-51