THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 11, Issue of March 16, pp. 8118–8124, 2001
Printed in U.S.A.

Identification of Internal Autoproteolytic Cleavage Sites within the Prosegments of Recombinant Procathepsin B and Procathepsin S

CONTRIBUTION OF A PLAUSIBLE UNIMOLECULAR AUTOPROTEOLYTIC EVENT FOR THE PROCESSING OF ZYMOGENS BELONGING TO THE PAPAIN FAMILY*

Received for publication, July 5, 2000, and in revised form, December 4, 2000
Published, JBC Papers in Press, December 13, 2001, DOI 10.1074/jbc.M005851200

Omar Quraishi‡§ and Andrew C. Storer‡¶‖

From the ‡Protein Engineering Network of Centres of Excellence and Department of Biochemistry, McGill University, Montreal, Quebec H3G 1Y6, and the ¶Pharmaceutical Biotechnology Sector, Biotechnology Research Institute, National Research Council Canada, Montreal, Quebec H4P 2R2, Canada

The steps involved in the maturation of proenzymes belonging to the papain family of cysteine proteases have been difficult to characterize. Intermolecular processing at or near the pro/mature junction, due either to the catalytic activity of active enzyme or to exogeneous proteases, has been well documented for this family of proenzymes. In addition, kinetic studies are suggestive of a slow unimolecular mechanism of autoactivation which is independent of proenzyme concentration. However, inspection of the recently determined x-ray crystal structures does not support this evidence. This is due primarily to the extensive distances between the catalytic thiolate-imidazolium ion pair and the putative site of proteolysis near the pro/mature junction required to form mature protein. Furthermore, the prosegments for this family of precursors have been shown to bind through the substrate binding clefts in a direction opposite to that expected for natural substrates. We report, using cystatin C- and N-terminal sequencing, the identification of autoproteolytic intermediates of processing *in vitro* for purified recombinant procathepsin B and procathepsin S. Inspection of the x-ray crystal structures reported to date indicates that these reactions occur within a segment of the proregion which binds through the substrate binding clefts of the enzymes, thus suggesting that these reactions are occurring as unimolecular processes.

Prior to being shuttled to the mature lysosome, cysteine proteases of the papain family are first synthesized as latent precursors of higher molecular weight. Zymogens of papain-like enzymes are composed of polypeptide extensions of various lengths at the N terminus of the mature enzyme domain which act as potent inhibitors toward the cognate enzyme (1–3). Because most precursors of the papain superfamily are susceptible to autoactivation upon their exposure to acidic pH environments (4–9), the stability of the prosegment[1]-enzyme complex is believed to rely mainly on electrostatic interactions. The crystal structures of rat (10) and human (11, 12) procathepsin B, human procathepsin L (13), procaricain (14), human procathepsin K (15, 16), and procathepsin X (17) have been reported recently. With the exception of procathepsin X (17), each of these structures reveals that the enzyme is inhibited by a small segment of the proregion which binds through the substrate binding cleft in a direction opposite to that expected for natural substrates. To protect cells from unregulated digestion, this reverse configuration is believed to help ensure proenzyme stability at neutral pH as the zymogen is passed from the endoplasmic reticulum to its final destination, *i.e.* the acidified lysosomal compartment of the cell.

In general, zymogens of all families of proteolytic enzymes may undergo maturation using either intermolecular or intramolecular processing pathways. For instance, autoproteolytic cleavage of prosubtilisin E, a serine protease precursor, has been suggested to occur at the same site near the pro/mature junction in either an intermolecular or intramolecular manner, and the mechanism that predominates is dependent mainly on the starting concentration of the proenzyme (18, 19). The conformational rearrangements involved in the unimolecular mechanism of prosubtilisin E, however, have yet to be elucidated. Similarly, kinetic studies that monitor the conversion of propapain (4, 7), procathepsin B (5), and procathepsin L (6) to active enzyme have revealed both an intermolecular and intramolecular component to processing. Significantly, the kinetic studies (5–7) have also revealed that the molecularity, *i.e.* the concentration of proenzyme at which the rate of the intermolecular events equals that of the unimolecular processes, is in the range of only 10^{-9} M. Thus it is only at very low concentrations of the proenzymes that the unimolecular mechanism plays a significant role in processing.

Interestingly, the proposal of a unimolecular step of maturation is inconsistent to what is observed in the three-dimensional structures for this family of precursors (10–16). For

* 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.This work was funded in part by the Government of Canada Network of Centres of Excellence Program supported by the Medical Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada through the Protein Engineering Network of Centres of Excellence, Inc. This is National Research Council of Canada Publication 42997. 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.

§ Present address: Dept. of Biochemistry and Molecular Biology, Merck Frosst Centre for Therapeutic Research, 16711 Trans-Canada Highway, Kirkland, Quebec H9H 3L1, Canada.

‖ To whom correspondence should be addressed: Pharmaceutical Biotechnology Sector, Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Ave., Montreal, Quebec H4P 2R2, Canada. Tel.: 514-496-6256; Fax: 514-496-1629; E-mail: Andrew.Storer@nrc.ca.

[1] In the text, the words *prosegment*, *proregion*, *prosequence*, and *prodomain* refer to the polypeptide stretch located N-terminal to the mature enzyme in the proenzyme; the word *propeptide* refers to the chemically synthesized polypeptide corresponding to the proregion sequence but without the mature enzyme.

example, the crystal structure of procathepsin B reveals that the pro/mature junction, *i.e.* the final site for proteolytic processing, is ~28 Å from the catalytic nucleophile. In addition, direct comparison of the crystal structures of procathepsin B (10–12) with that of mature cathepsin B (20) reveals no evidence for major N-terminal rearrangement within the mature segment following the maturation process as is found in zymogens belonging to other families (21–23). Hence, the crystal structures reported to date have not independently provided any clues to a plausible unimolecular step. Recently, a nonhomology knowledge-based strategy predicted that a unimolecular proteolytic step in propapain processing may involve the adjustment of a single β-turn which rearranges the first 12 residues[2] of the enzyme domain and allows the mature N terminus to reach the active site in a cleavable direction (24). Within this segment of the mature N terminus are found Asp-6 and Arg-8, which are both highly conserved residues among cysteine proteases of the papain family (25). Interestingly, all x-ray crystal structures of papain-like enzymes (pro- and mature) reported to date reveal that the side chains of Asp-6 and Arg-8 contribute to the formation of a conserved salt bridge (10–17, 20, 26) that contributes to the structural integrity of the mature enzyme's N terminus.

Here, we attempt to identify novel intermediates of processing for the precursors of cathepsin B and cathepsin S *in vitro*. This has been achieved by monitoring using SDS-PAGE[3] the processing of either purified procathepsin B or procathepsin S in the presence of the endogenous inhibitor, cystatin C. Cystatin C has been shown to be a substrate binding cleft-directed protein inhibitor of papain-like enzymes with K_i values in the subnanomolar range (27–29). Because the formation of a tight binding complex with cystatin C requires that the substrate binding cleft of papain-like enzymes be unobstructed (29), for example free from the prodomain, it may be reasonably assumed that the affinity of cystatin C for the mature enzyme would be superior to that for either the full-length proenzyme or any intermediates generated during autoproteolysis, *i.e.* hierarchy of cystatin C affinity for mature enzyme > intermediates > proenzyme. In its excess, therefore, cystatin C may provide the desired effect of decreasing the rate of intermolecular proteolytic processing, which is mainly the result of the activity of mature (processed) enzyme, and thereby favoring the detection of other processing events. Furthermore, we investigate a role for the N terminus of the mature enzyme domain in the processing of precursors belonging to the papain family. This has been achieved by performing site-directed mutagenesis of Arg-8 to an alanine residue in propapain, *i.e.* causing the destruction of the conserved Asp-6 → Arg-8 salt bridge, and monitoring the overall effect of this mutation on the ability of propapain to autoactivate.[4]

EXPERIMENTAL PROCEDURES

Materials—Human wild-type cystatin C were a generous gift from Dr. Irena Ekiel (Biotechnology Research Institute (BRI)/National Research Council) and monoclonal antibody to papain was provided by Daniel Tessier and Dr. David Y. Thomas (BRI/National Research Council). Recombinant human wild-type procathepsin B was expressed and purified as described previously (30). The pPIC9 vector and *Pichia pastoris* strain GS115 were purchased from Invitrogen (San Diego). Butyl-Sepharose resin was purchased from Amersham Pharmacia Biotech. The substrate benzyloxycarbonyl-L-phenylalanyl-L-arginine 4-methylcoumarinyl-7-amide hydrochloride (Z-Phe-Arg-MCA) and the irreversible inhibitor E-64 (L-*trans*-epoxysuccinyl)-L-leucyl-amino(4-guanidino)butane) were purchased from IAF Biochem International Inc. (Laval, Canada). Polyvinylidene difluoride (PVDF) membranes were purchased from Applied Biosystems.

Expression of Procathepsin B and Procathepsin S—A cDNA construct consisting of human wild-type procathepsin B or procathepsin S as a fusion with the preproregion of yeast α-factor was digested with *Xho*I and *Not*I, and the proenzyme fragment was subcloned into the pPIC9 vector (Invitrogen). For integration into the *Pichia* genome, the pPIC9-based constructs were linearized by cleavage with *Bgl*II and purified. The *P. pastoris* host strain GS115 (Invitrogen) was then transformed with the linearized constructs by electroporation. Positive transformants were grown for 2 days at 30 °C in medium containing glycerol as the carbon source followed by incubation in the presence of methanol for a further 3 days to induce expression of recombinant protein. The consensus sequence for oligosaccharide substitution located on the occluding loop within the mature enzyme domain of cathepsin B had been removed by the substitution S115A. All other sites for oligosaccharide substitution within procathepsin B and procathepsin S were left unaltered. Protein secreted into the culture supernatants was analyzed by SDS-polyacrylamide gel electrophoresis (12% gels).

Purification of Procathepsin B and Procathepsin S—The proenzymes were purified from the culture supernatant using a hydrophobic resin under nonacidic conditions. The culture supernatant (250 ml) was concentrated to 50 ml using an Amicon stirred cell (YM-10 membrane). During concentration, the supernatant was exchanged to 50 mM Tris (pH 7.4) containing 1.6 M $(NH_4)_2SO_4$. Concentrated recombinant proenzyme was then purified on a fast protein liquid chromatography system (Amersham Pharmacia Biotech) using a butyl-Sepharose fast flow column. Proenzyme fractions eluted from the column by applying a linear gradient of decreasing ammonium sulfate concentration. Glycosylated procathepsin B and procathepsin S eluted at 0.6–0.8 M and 0.3–0.5 M $(NH_4)_2SO_4$, respectively, and samples were stored at 4 °C.

In Vitro Processing of Procathepsin B and Procathepsin S—Purified procathepsin B (20 μM) and procathepsin S (4 μM) samples were dialyzed against 50 mM sodium acetate (pH 5.0), 1 mM dithiothreitol at 4 °C overnight in the presence (or absence) of 100 μM human wild-type cystatin C. Each sample was then treated with excess E-64 followed by the addition of reducing buffer and denaturation in a boiling water bath. Protein samples were then applied to SDS-PAGE (12% gels) and stained with Coomassie Brilliant Blue R-250 (Bio-Rad) or $AgNO_3$ (see Fig. 1).

Expression and Processing of Wild-type and R8A Propapain—Propapain was produced as described previously (4, 7). Briefly, the *Saccharomyces cerevisiae* strain BJ3501 was transformed with the expression vector derived from pVT100-U in which the propapain gene is under the control of the α-factor promoter. Yeast cells were first grown under selective conditions to ensure plasmid maintenance and then transferred into a rich medium. The cells were lysed using a French pressure cell, and the soluble fraction of the crude lysate included propapain. Propapain was then partially purified using butyl-Sepharose resin (Amersham Pharmacia Biotech) as discussed previously for procathepsins B and S. Complete processing *in cis* was achieved by incubating propapain at 65 °C in 50 mM sodium acetate (pH 3.8), 20 mM cysteine for 30 min. Samples were then analyzed by Western blot following separation of the proteins using SDS-PAGE.

N-Terminal Identification of Protein Bands—Following SDS-PAGE, protein bands were blotted onto hydrophobic PVDF membranes. The membranes were then stained with *Coomassie Brilliant Blue R-250*, and each protein band of interest was subjected to a minimum of five cycles of automated Edman degradation using the method described previously (31).

Fluorogenic Assay for Monitoring Proenzyme Processing—Processing of human wild-type procathepsin B and procathepsin S was followed in a continuous manner by carrying out the reactions in a 3-ml quartz cuvette in the presence of the substrate Z-Phe-Arg-MCA (10 μM) and measuring fluorescence as a function of time. The conversion of procathepsins B and S to active enzyme leads to hydrolysis of the substrate, and fluorescence of the MCA product was monitored using excitation and emission wavelengths of 380 and 440 nm, respectively. Processing was initiated by lowering the pH from 7.4 (pH of the stock

[2] In the text, residue numbering relates to that of cathepsin B for recombinant human cathepsin B, and to that of cathepsin S for recombinant human cathepsin S. Residues in the proregion are identified with the suffix p.

[3] The abbreviations used are: PAGE, polyacrylamide gel electrophoresis; E-64, *trans*-epoxysuccinyl-L-leucyl-amido-(4-guanidino)butane; Z-Phe-Arg-MCA, benzyloxycarbonyl-L-phenylalanyl-L-arginine 4-methylcoumarinyl-7-amide; PVDF, polyvinylidene difluoride.

[4] In the text, the terms *autoprocessing*, *autoactivation*, and *maturation* relate to the ability of zymogen to convert to mature protein at acidic pH due to autocatalytic cleavage at or near the pro/mature junction.

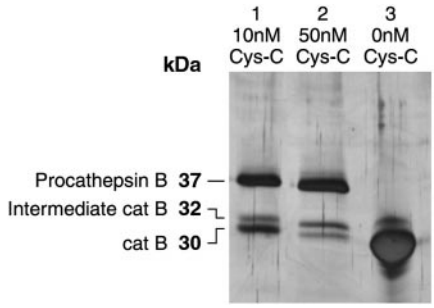

FIG. 1. **SDS-PAGE stained with AgNO3.** Procathepsin B (37 kDa) at low concentrations (2 nM) was exposed to 50 mM acetate buffer (pH 5.0) and 1 mM dithiothreitol for 12 h in the presence of 10 nM (*lane 1*), 50 nM (*lane 2*), and 0 nM (*lane 3*) human cystatin C. Note that the proportion of processing intermediate (32 kDa) to mature cathepsin B (30 kDa) is improved with increasing concentration of cystatin C.

solution of procathepsins B and S) to 5.0 (pH of the assay). Reactions were carried out at 25 °C in the presence of 50 mM sodium acetate buffer, 0.2 M NaCl, 2 mM EDTA, 2 mM dithiothreitol, and 3% dimethyl sulfoxide. The reaction mixture was stirred continuously in the cuvette during the reaction. The product *versus* time curves were fitted to the following equation (32–36).

$$[P] = v_E t - ((v_E - v_{PE})(1 - \exp(-k_{obs}t)))/k_{obs} \quad \text{(Eq. 1)}$$

where P is the MCA product formed, v_{PE} represents the initial rate of product release (which should reflect activity of the proenzyme, if any), v_E corresponds to the rate for mature (active) enzyme, and k_{obs} is a first-order rate constant.

RESULTS

Identification of Intermediate Cleavage Sites in Procathepsin B Using the Cystatin C Assay—On addition of cystatin C to a reaction mixture with which the processing of procathepsin B is being monitored, an intermediate species of cathepsin B is observed. This intermediate is observable even at nanomolar concentrations of proenzyme (as determined by $AgNO_3$-stained SDS-PAGE; Fig. 1) as well as at micromolar concentrations (Fig. 2*A*; PVDF membranes, Coomassie Blue staining). The ability to immobilize on PVDF membranes, and consequently purify, protein bands that correspond to presumably unstable processing intermediates ensures that the residual prosegment of these species remains resistant to further degradation. Direct N-terminal sequencing of the intermediate band of cathepsin B processing (Fig. 2*A*) indicated a mixture of intermediate species with autocatalytic cleavage having taken place at Cys-42p ↑ Gly-43p (70% of total signal) and Arg-40p ↑ Leu-41p (30% of total signal) (Fig. 3, cathepsin B prosegment numbering). After several weeks of incubation in the presence of cystatin C, both full-length procathepsin B and the processing intermediates disappear, and only the protein band corresponding to mature cathepsin B is observed. Therefore, it may be concluded that these are true intermediates of processing and not dead-end (side product) reactions. Because complete maturation of cathepsin B has been shown to also include autocatalytic trimming of six residues at the C-terminal end of the enzyme (37), it should be noted that the 30- and 32-kDa protein bands in Fig. 2*A*, *lane 2*, most likely correspond to cathepsin B composed of a mature and semimature N terminus, respectively, and an unprocessed C terminus.

Identification of Intermediate Cleavage Sites in Procathepsin S Using the Cystatin C Assay—In the case of procathepsin S, the presence of cystatin C is not strictly required to observe, using SDS-PAGE, the accumulation of an intermediate processing band that is ~2 kDa heavier than the mature enzyme (data not shown). To ensure the accumulation of sufficient quantities of this species for N-terminal identification, cystatin C was added to procathepsin S (Fig. 2*B*) under processing

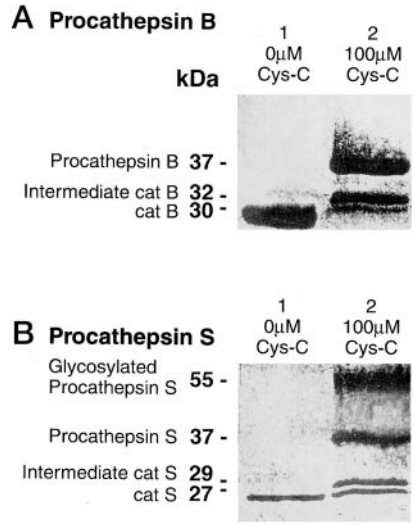

FIG. 2. **PVDF membranes (Applied Biosystems, Problott ™ membranes) stained with Coomassie Brilliant Blue R-250 (Bio-Rad) containing immobilized procathepsin B (*panel A*) and a mixture of glycosylated and deglycosylated procathepsin S (*panel B*).** *Lanes 1* and *2* of each membrane represent the conversion of proenzyme to mature enzyme in the absence and presence of 100 μM cystatin C (M_r = 12.5 kDa), respectively. The processing intermediates migrate at 32 and 29 kDa for procathepsin B and procathepsin S, respectively. The 29-kDa species for procathepsin S is detectable in low quantities using SDS-PAGE in the absence of cystatin C (data not shown).

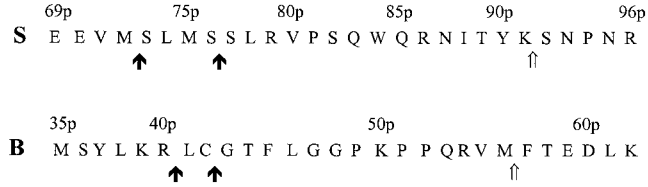

FIG. 3. **Structure-based sequence homology of the C-terminal ends of the prosegments of cathepsin B and cathepsin S.** Cathepsin B prosegment numbering was used for cathepsin B, and cathepsin S prosegment numbering was used for cathepsin S. The established cleavage sites to form mature protein near the pro/mature junction are represented as ⇑. The sites of autoproteolytic processing detected using the cystatin C assay are denoted as ↑.

conditions to inhibit bimolecular reactions as has been discussed previously for procathepsin B. Based on their migration on SDS-PAGE, it is assumed that the intermediate species formed in the presence of cystatin C are identical to those in the absence of the inhibitor. This protein band obtained for the sample treated with cystatin C was found to be a mixture of species corresponding to cleavage at Ser-76p ↑ Ser-77p (50% of total signal) and Met-72p ↑ Ser-73p (50% of total signal) (Fig. 3; cathepsin S prosegment numbering).

Continuous Monitoring of Wild-type Procathepsin B and Procathepsin S Autocatalytic Processing at pH 5.0—A continuous assay based on the hydrolysis of the substrate Z-Phe-Arg-MCA by the active enzyme generated in the autocatalytic process was used. The rate of substrate hydrolysis increases with time due to time-dependent release of active enzyme from the precursor until a constant rate is obtained which corresponds to the activity of fully processed enzyme. The curves can be fitted to a model that assumes a first-order increase in rate from an initial rate v_{PE}, corresponding to the activity of the precursors (if any), to a final rate v_E, corresponding to the activity of fully processed enzyme. Based on the results of nonlinear regression analysis, no significant activity of the precursors against the Z-Phe-Arg-MCA substrate could be detected, and the first-order rate for autocatalytic processing, k_{obs}, increases linearly

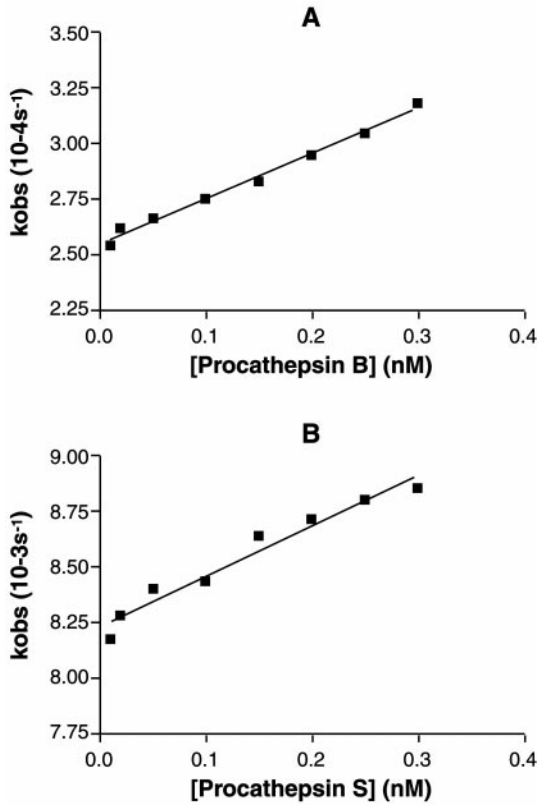

FIG. 4. **Continuous assay for the autocatalytic processing of wild-type procathepsin B (*panel A*) and procathepsin S (*panel B*).** Shown are plots of the first-order rates of processing (k_{obs}) obtained by nonlinear regression of the data discussed in Equation 1 under "Experimental Procedures" as a function of precursor concentration (determined by active site titration with the E-64 inhibitor). The data are in agreement with a first-order rate of processing. For both procathepsins B and S, the rate of processing, k_{obs}, increases linearly with proenzyme concentration, indicative of a bimolecular reaction that most likely corresponds to intermolecular processing of proenzyme by mature or semimature protein. In addition, there is a corresponding rate constant of precursor activation as the concentration of proenzyme is extrapolated to zero (2.5×10^{-4} s^{-1} for procathepsin B and 8.2×10^{-3} s^{-1} for procathepsin S), indicative of an activation event that is independent of the concentration of precursor.

with proenzyme concentration (Fig. 4). The direct link between the rates of autoprocessing and precursor concentration confirms the occurrence of a bimolecular reaction, *i.e.* intermolecular processing of proenzyme by fully or partially processed (active) enzyme. In support of the postulated unimolecular autoproteolytic reactions discovered using the cystatin C assay (discussed above), the extrapolated rate constants were calculated to be 2.5×10^{-4} s^{-1} and 8.2×10^{-3} s^{-1} for procathepsin B and procathepsin S, respectively, as the concentration of these precursors approach zero. This kinetic data support the involvement of an intramolecular event in the processing of these precursors which is independent of the concentration of proenzyme.

Role of the N Terminus of the Mature Segment in the Autocatalytic Processing of Precursors Belonging to the Papain Family—Site-directed mutagenesis was performed to test the hypothesis that a rearrangement of the N terminus of the mature domain of cysteine proteases is involved in an intramolecular processing mechanism. Specifically, to remove the salt bridge that is formed between the highly conserved residues Asp-6 and Arg-8 an R8A mutant of propapain was produced. From the data in Fig. 5 we conclude that R8A propapain remains competent to autoactivate to form mature protein.

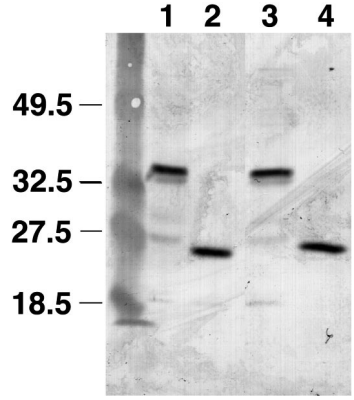

FIG. 5. **Monitoring the autocatalytic processing of wild-type propapain (*lane 1*) and R8A propapain (*lane 3*) using Western blot analysis following their incubation at 65 °C in 50 mM acetate buffer (pH 3.8) and 20 mM cysteine for 30 min (*lane 2*, wild-type mature papain; *lane 4*, R8A mature papain).** Molecular masses of the papain precursor and mature form are 37 and 25 kDa, respectively. The R8A variant of propapain was observed to autoprocess as efficiently as the wild-type precursor.

DISCUSSION

The crystal structure of procathepsin B (10–12) indicates that the Leu-41p → Gly-47p segment of the cathepsin B proregion binds through the active site cleft of the enzyme in the reverse substrate binding mode and in an extended conformation. These structures also reveal that the carbonyl carbon of Cys-42p is in closest proximity to the catalytic residue (Fig. 6) (note that the coordinates illustrated in Fig. 6 are those for C29S human procathepsin B at pH 5.7 (11, 12) and not the wild-type precursor). Inspection of this region of the cathepsin B precursor indicates that the carbonyl carbon of Cys-42p is located ~4.3Å from the catalytic nucleophile and the potential bond angle between the catalytic nucleophile and the carbonyl oxygen of Cys42p is 131°, *i.e.* conducive to forming a tetrahedral intermediate (Fig. 6). For the Arg-40p ↑ Leu-41p cleavage site, the distance and potential bond angle between the carbonyl group of Arg-40p and the catalytic nucleophile are ~6.9 Å and 45°, respectively. Hence, the ability of the carbonyl carbon of Arg-40p to reach the catalytic center of the enzyme for hydrolysis would suggest that significant conformational mobility exists within the segment composed of residues Asp-34p → Leu-41p which interact with the occluding loop crevice (10) as well as the primed subsites of the substrate binding cleft of cathepsin B.

Given the proximity of these carbonyl carbons (those of Cys-42p and Arg-40p) to the catalytic center, it is tempting to speculate that these reactions occur in an intramolecular manner. As discussed previously, kinetic studies are suggestive of a unimolecular step among members of the papain family of precursors whose molecularity is unusually low (10^{-9} M) yet is still much faster than noncatalyzed (spontaneous) peptide hydrolysis. The low molecularity may be accounted for by the reverse binding mode adopted by the prodomain in its interaction with the active site cleft of the enzyme. As a consequence of the reverse substrate binding mode, the formation of a tetrahedral intermediate at the carbonyl carbon of Cys-42p or Arg-40p would not be stabilized by the oxyanion hole formed by Gln-23. This structural incompatibility has, therefore, led to the suggestion that it would not be possible for the enzyme to perform such reactions (12). Previous work with oxyanion hole mutants of papain (38) and cathepsin B,[5] however, indicates

[5] D. K. Nägler, A. C. Storer, and R. Ménard, manuscript in preparation.

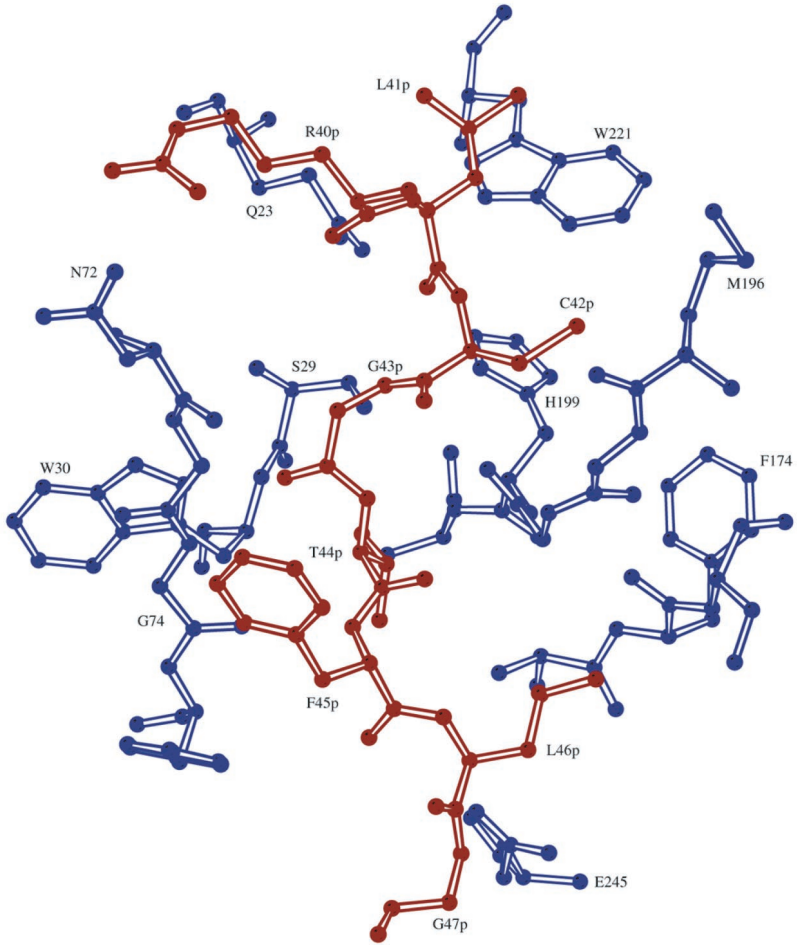

FIG. 6. **Active site cleft in human C29S procathepsin B at pH 5.7 (11, 12).** The cathepsin B prosegment (*red* residues) binds through the substrate binding cleft of cathepsin B (*blue* residues) in the reverse N → C direction to that taken by natural substrates. The carbonyl carbon of Cys-42p is in closest proximity to the catalytic nucleophile (4.3 Å), and the potential bond angle between the catalytic nucleophile and the carbonyl oxygen of Cys42p is 131°.

that the oxyanion hole is not an absolute requirement to hydrolyze small synthetic substrates but rather is a feature that may improve the catalytic efficiency of these enzymes by only 10–100-fold depending on the substrate under study. Alternatively, it may be proposed that the slow rate of these unimolecular reactions may be because the reverse complementarity between the bound prosegment and the enzyme's substrate binding cleft causes the distance between the δN of the catalytic His-199 and the backbone amide (leaving) group of either Gly-43p or Leu-41p to be larger than would be the case for natural substrates (39, 40). Hence, protonation from the δN of the catalytic histidine to the prosegment bound in the reverse mode is likely to be less efficient to that for substrates bound in the usual substrate binding mode, thus resulting in a reversible nucleophilic reaction that has difficulty going to completion. Because the catalytic thiol among cysteine proteases is more nucleophilic and constitutes a better leaving group than the catalytic oxygen found among serine proteases, it has been postulated that proton transfers effectuated by the catalytic histidines found among cysteine proteases would need to be more efficient than those found among serine proteases (41), *i.e.* the catalytic histidine found in cysteine proteases must compensate for the lower pK_a of the catalytic thiol group.

The ability to detect cleavage at the carbonyl carbon of Arg-40p indicates the existence of a significant degree of conformational mobility for the proregion within the prosegment-substrate binding cleft interface. This mobility may be accounted for by the pH-dependent stability of the enzyme's occluding loop, which consequently defines the pH-dependence of propeptide binding as well as the overall rate of procathepsin B processing (30). Competition between the prosegment and the occluding loop for the surface of the enzyme termed the *occluding loop crevice* (10) was shown to be regulated by the formation of a critical salt bridge between His-110 of the occluding loop and Asp-22 located within the primed subsites of the enzyme's active site cleft. Site-directed mutagenesis of either one of His-110 or Asp-22 to an alanine residue produces a variant of procathepsin B which is stable and incapable of autoactivation. Remaining elusive from these studies (30), however, was whether these mutations caused the perturbation of a unimolecular event involving proteolysis of the prosegment. As procathepsin B is exposed to acidic pH conditions, it is possible that salt bridge formation between His-110 and Asp-22 promotes competition between the occluding loop and the N-terminal α-helical cap of the proregion for the occluding loop crevice. From this competition, it follows that the remaining C-terminal residues of the proregion, *i.e.* prosegment residues that stretch from the substrate binding cleft to the pro/mature junction, would have reduced affinity for the surface of the enzyme and increased conformational mobility. In agreement with this proposal is the consistent lack of secondary or tertiary structure found within the C-terminal end of papain-like prosegments when bound to the cognate enzyme (10–17). Furthermore, truncated propeptides composed only of these C-terminal residues display significantly weaker affinity for the enzyme than the full-length propeptide (42–43). Additional evidence for mobility within the prosegment has been shown for the propeptide of cathepsin L which loses most of its tertiary structure yet almost none of its secondary structure at low pH (44). It is believed the high B-factors corresponding to the C-terminal end of papain-like prosegments facilitate the autocatalytic conversion of these zymogens upon their exposure to acidic pH. That the conformational mobility within the C-terminal end of papain-like prosegments is important for autoprocessing is ev-

idenced by the results that have been obtained recently for procathepsin H.[6] Because of the predicted preformation of a disulfide bond linking the C-terminal end of the cathepsin H prosegment to the main body of the enzyme, the pro/mature junction within the cathepsin H precursor was found to be highly resistant to proteolysis either *in cis* or *in trans*.[6]

Based upon the sequence homology between procathepsin S and other cysteine protease proforms for which structural information is available (10–16), with the exception of procathepsin X (17), it has been possible to model the structure of the relevant portion of the procathepsin S proregion. Based on this model the predicted structure-based sequence homology between the prosegments of procathepsin B and procathepsin S is presented in Fig. 3, and it reveals that cleavage at the carbonyl carbon of residues in position 42p of procathepsin B and 76p of procathepsin S are well aligned. This conservation is observed despite the difference in length for these two prodomains, *i.e.* 62 residues for procathepsin B and >90 residues for procathepsin S, which is a member of the cathepsin L subfamily. It follows, therefore, that Ser-76p is predicted to bind through the substrate binding cleft of cathepsin S in the reverse binding mode and that its carbonyl carbon is located closest to the catalytic residue, Cys-25, as has been documented for Cys-42p in procathepsin B.

Cathepsin S prefers to cleave at internal sites within its prosegment where serine residues are located in the S_1' position, namely at the putative Lys-91p ↑ Ser-92p site near the pro/mature junction needed to form mature cathepsin S as well as at the Met-72p ↑ Ser-73p and Ser-76p ↑ Ser-77p sites identified using the cystatin C assay (Fig. 3). It is interesting to note that two consecutive serine residues are located within the prosequence of cathepsin S, namely Ser-76p and Ser-77p. Curiously, proteolysis was only observed at the carbonyl carbon of Ser-76p and not at the carbonyl carbon of Met-75p. This result indicates that cleavage at the carbonyl carbon of Ser-76p may be selective. Similar to what was observed for procathepsin B, cleavage at the carbonyl carbon of Met-72p suggests that a significant amount of conformational mobility exists for prosegment residues that bind through the substrate binding cleft of cathepsin S. It is interesting to note that the conversion of procathepsin H to its mature form has been proposed to involve cleavage at the carbonyl carbon of Ser-77p (26) (cathepsin H prosegment numbering), which is located adjacent to the cleavage sites identified in this study for procathepsin B and procathepsin S using the cystatin C assay. This cleavage site has been proposed to be a prelude to the formation of an *N*-linked glycosylated octapeptide of prosegment residues composed of Glu-78p → Thr-85p (cathepsin H prosegment numbering), termed the *mini-chain*, which remains attached to mature cathepsin H via a disulfide bond (26).

Is it possible that the processes identified by the cystatin C assay are the result of intermolecular reactions caused by catalytic (undetectable) amounts of mature enzyme or activated proenzyme? If processing were assumed to be solely the result of intermolecular processes, then it would be expected that the conversion of full-length proenzyme to processing intermediates would be inhibited as efficiently by cystatin C as the conversion from processing intermediate to mature enzyme or the direct maturation of full-length proenzyme to mature protein. In the presence of cystatin C, however, this is not what is observed, but rather the time-dependent accumulation of intermediate protein bands is detected using SDS-PAGE.

Previous to this work, a nonhomology knowledge-based prediction of propapain activation proposed that an intramolecular proteolytic event may involve the N terminus of the mature enzyme domain moving toward the active site cleft, thus facilitating the release of the prosegment (24). Using the structure of mature papain as a template, *i.e.* the effect of the prosegment was not considered, the adjustment of a single β-turn was postulated to permit the extension of the first 12 residues at the N terminus of the enzyme and predicted to allow the pro/mature junction to reach the active site in the cleavable direction, *i.e.* the substrate binding mode. For this rearrangement to be made possible, it would be expected that the integrity of the salt bridge formed by Asp-6 and Arg-8 found in all structures of papain-like enzymes reported to date (10–17, 20, 26, 45), *i.e.* residues that contribute to the β-turn, would influence the overall pH-triggering mechanism of propapain processing. However, as reported above, the removal of the salt bridge in the papain mutant R8A does not influence the ability of propapain to autoactivate to form mature protein (Fig. 5). These results collaborate with the x-ray crystal structures of papain-like enzymes (pro- and mature enzymes) which demonstrate high resolution among residue side chains located at the N terminus of the mature segment, thus corresponding to a region of the molecule which is conformationally constrained (low B-factors). The N terminus of the mature segment within the precursors (10–17) is in a conformation that is essentially identical to that found in the crystal structures of mature enzyme (20, 26, 45), thus suggesting that no major N-terminal rearrangement is observed during precursor activation. Furthermore, the overall assumption that the putative site of proteolysis to form mature protein near the pro/mature junction is the only possible cleavage site, *i.e.* as has been proposed for prosubtilisin E (18, 19), remains speculative as an unidentified processing intermediate was observed for propapain at 30 kDa (7).

Nature of the Steps Involved in the Autocatalytic Processing of Procathepsin B and Procathepsin S—Using site-directed mutagenesis, previous studies have established that the reactivity of the catalytic cysteine residue found within the precursors of papain-like enzymes is responsible for the maturation of this family of zymogens (4–9). Hence, autoactivation of zymogens belonging to the papain family requires that the precursors be composed of a preformed and functional catalytic center and substrate binding cleft. In this study, we have described the identification of novel processing intermediates for procathepsin B and procathepsin S. The intermediates identified for cathepsin B are observable only in the presence of cystatin C, whereas for cathepsin S, the intermediates are weakly observable on SDS-PAGE in the absence of cystatin C (data not shown), and their identification is facilitated only by the addition of cystatin C. That these novel cleavage products in procathepsin B and procathepsin S are observed at all starting concentrations of proenzyme, including very low concentrations, suggests that these reactions are occurring as unimolecular processes and that they may be important. The crystal structures of precursors of the papain family (10–16) demonstrate that these cleavage reactions are taking place within a segment of the proregion which binds through the active site cleft of the enzyme in the reverse substrate binding mode. Thus in effect the intramolecular processing of the precursors is analogous to the cleavage of a polypeptide chain (substrate) when bound in the reverse mode to the active site. To the authors' knowledge this is the first reported example of the observation of a protease cleaving a peptide bond in the reverse direction. It is interesting to note that although the rate of cleavage of the reversed amide bond is much slower than that of an optimally oriented bond, it is nonetheless several orders

[6] O. Quraishi, R. Ménard, J. S. Mort, and A. C. Storer, manuscript in preparation.

of magnitude faster than noncatalyzed hydrolysis.

Following the completion of the intramolecular proteolytic step, more than 20 residues derived from the C-terminal end of the prosegment continue to remain covalently attached to the mature segment via the pro/mature junction. Intuitively, these intermediate species would be as catalytically competent as the mature enzyme because propeptides composed of amino acid sequences corresponding to the C-terminal end of papain-like prosegments possess low inhibitory activity compared with that of the full-length propeptide (2, 42, 43).

Given the kinetic and structural data presented here, it is tempting to speculate that these reactions are occurring as slow unimolecular steps that may be necessary for triggering the intermolecular proteolytic cascade, *i.e.* the first step may involve the slow intramolecular cleavage reactions presented here, followed by the rapid intermolecular proteolytic cascade performed by the catalytic activity of mature or semimature species whose quantities accumulate with time. From this study, it may also be concluded that the N terminus of the mature domain does not participate through a conformational rearrangement in the pH-dependent autoprocessing mechanisms of zymogens belonging to the papain family.

It has been proposed that deregulated secretion of papain-like enzymes to the extracellular matrix may serve as the catalyst for propagating several disease states. For example, cathepsin B has been implicated in tumor metastasis (46) as well as rheumatoid arthritis (47). In chronic inflammatory disease, a degradative phenotype of monocyte-derived macrophages have been shown to secrete enzymically active forms of cathepsins B, L, and S into the extracellular milieu (48). Furthermore, deregulated cathepsin K activity has been linked to osteoporosis (49). Because the prodomains are known to act as intramolecular chaperones (4), it may be reasonably assumed that these enzymes are targeted to the extracellular matrix in their precursor form. Despite the neutral pH environment generally associated to the extracellular matrix, it has been postulated that microenvironments of acidic pH may allow low concentrations of zymogens of papain-like enzymes to be converted to their mature forms. The discovery of a unimolecular mechanism of processing for this family of precursors may help to explain how lysosomal enzymes (even at low concentrations) have been implicated in a number of degradative and invasive pathological conditions extracellularly. Although inhibitors with sufficient potency are available for this class of enzymes, they often lack the required selectivity needed for therapeutic applications. In addition to the traditional approach of designing substrate binding cleft-directed inhibitors, an improved understanding into the molecular basis of autoprocessing for zymogens of the papain family may lead to novel therapeutic strategies in which the conversion of proenzyme is intervened. This would have significant consequences given that the prosegments of papain-like cysteine proteases are intrinsic appendices that are potent inhibitors of the enzyme from which they originate.

Acknowledgments—We thank France Dumas for the N-terminal sequence determinations and Dr. Irena Ekiel for her generous gift of human wild-type cystatin C. We also thank Daniel Tessier and Dr. Dave Thomas for providing the constructs of wild-type and R8A propapain, and Dr. Jayaraman Sivaraman for help in preparing Fig. 6 depicting the active site cleft of procathepsin B.

REFERENCES

1. Fox, T., de Miguel, E., Mort, J. S., and Storer, A. C. (1992) *Biochemistry* **31,** 12571–12576
2. Carmona, E., Dufour, É., Plouffe, C., Takebe, S., Mason, P., Mort, J. S., and Ménard, R. (1996) *Biochemistry* **35,** 8149–8157
3. Maubach, G., Schilling, K., Rommerskirch, W., Wenz, I., Schultz, J. E., Weber, E., and Wiederanders, B. (1997) *Eur. J. Biochem.* **250,** 745–750
4. Vernet, T., Khouri, H. E., Laflamme, P., Tessier, D. C., Musil, R., Gour-Salin, B. J., Storer, A. C., and Thomas, D. Y. (1991) *J. Biol. Chem.* **266,** 21451–21457
5. Mach, L., Mort, J. S., and Glössl, J. (1994) *J. Biol. Chem.* **269,** 13030–13035
6. Ménard, R., Carmona, E., Takebe, S., Dufour, É., Plouffe, C., Mason, P., and Mort, J. S. (1998) *J. Biol. Chem.* **273,** 4478–4484
7. Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Ménard, R., Magny, M.-C., Storer, A. C., and Thomas, D. Y. (1995) *J. Biol. Chem.* **270,** 10838–10846
8. Mach, L., Schwihla, H., Stüwe, K., Rowan, A. D., Mort, J. S., and Glössl, J. (1993) *Biochem. J.* **293,** 437–442
9. Rowan, A. D., Mason, P., Mach, L., and Mort, J. S. (1992) *J. Biol. Chem.* **267,** 15993–15999
10. Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A. C., and Mort, J. S. (1996) *Structure* **4,** 405–416
11. Turk, D., Podobnik, M., Kuhelj, R., Dolinar, M., and Turk, V. (1996) *FEBS Lett.* **384,** 211–214
12. Podobnik, M., Kuhelj, R., Turk, V., and Turk, D. (1997) *J. Mol. Biol.* **271,** 774–788
13. Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J. S., and Cygler, M. (1996) *EMBO J.* **15,** 5492–5503
14. Groves, M. R., Taylor, M. A., Scott, M., Cummings, N. J., Pickersgill, R. W., and Jenkins, J. A. (1996) *Structure* **4,** 1193–1203
15. LaLonde, J. M., Zhao, B., Janson, C. A., D'Alessio, K. J., McQueney, M. S., Orsini, M. J., Debouck, C. M., and Smith, W. W. (1999) *Biochemistry* **38,** 862–869
16. Sivaraman, J., Lalumière, M., Ménard, R., and Cygler, M. (1999) *Protein Sci.* **8,** 283–290
17. Sivaraman, J., Nägler, D. K., Zhang, R., Ménard, R., and Cygler, M. (2000) *J. Mol. Biol.* **295,** 939–951
18. Li, Y., and Inouye, M. (1996) *J. Mol. Biol.* **262,** 591–594
19. Volkov, A., and Jordan, F. (1996) *J. Mol. Biol.* **262,** 595–599
20. Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J., and Turk, V. (1988) *EMBO J.* **7,** 2593–2599
21. James, M. N. G., and Sielecki, A. R. (1986) *Nature* **319,** 33–38
22. Gallagher, T., Gilliland, G., Wang, L., and Bryan, P. (1995) *Structure* **3,** 907–914
23. Becker, J. W., Marcy, A. I., Rokosz, L. L., Axel, M. G., Burbaum, J. J., Fitzgerald, P. M. D., Cameron, P. M., Esser, C. K., Hagmann, W. K., Hermes, J. D., and Springer, J. P. (1995) *Protein Sci.* **4,** 1966–1976
24. Jaqueline, Padilla-Zúñiga, A., and Rojo-Domínguez, A. (1998) *Folding Design* **3,** 271–284
25. Berti, P. L., and Storer, A. C. (1995) *J. Mol. Biol.* **246,** 273–283
26. Guncar, G., Podobnik, M., Pungercar, J., Strukelj, B., Turk, V., and Turk, D. (1998) *Structure* **6,** 51–61
27. Abrahamson, M., Barrett, A. J., Salvesen, G., and Grubb, A. (1986) *J. Biol. Chem.* **261,** 11282–11289
28. Hall, A., Dalbøge, H., Grubb, A., and Abrahamson, M. (1993) *Biochem. J.* **291,** 123–129
29. Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T., and Mort, J. S. (1997) *J. Biol. Chem.* **272,** 1197–1202
30. Quraishi, O., Nägler, D. K., Fox, T., Sivaraman, J., Cygler, M., Mort, J. S., and Storer, A. C. (1999) *Biochemistry* **38,** 5017–5023
31. Matsudaira, P. (1987) *J. Biol. Chem.* **262,** 10035–10038
32. Cha, S. (1975) *Biochem. Pharmacol.* **24,** 2177–2185
33. Morrison, J. F. (1969) *Biochim. Biophys. Acta* **185,** 269–286
34. Morrison, J. F. (1982) *Trends Biochem. Sci.* **7,** 102–105
35. Morrison, J. F., and Stone, S. R. (1985) *Comments Mol. Cell. Biophys.* **2,** 347–368
36. Morrison, J. F., and Walsh, C. T. (1988) *Adv. Enzymol. Relat. Areas Mol. Biol.* **61,** 201–301
37. Rowan, A. D., Feng, R., Konishi, Y., and Mort, J. S. (1993) *Biochem. J.* **294,** 923–927
38. Ménard, R., Carrière, J., Laflamme, P., Plouffe, C., Khouri, H. E., Vernet, T., Tessier, D. C., Thomas, D. Y., and Storer, A. C. (1991) *Biochemistry* **30,** 8924–8928
39. Arad, D., Langridge, R., and Kollman, P. A. (1990) *J. Am. Chem. Soc.* **112,** 491–502
40. Winkler, F. K., and Dunitz, J. D. (1971) *J. Mol. Biol.* **59,** 169–182
41. Higaki, J. N., Evnin, L. B., and Craik, C. S. (1989) *Biochemistry* **28,** 9256–9263
42. Chen, Y., Plouffe, C., Ménard, R., and Storer, A. C. (1996) *FEBS Lett.* **393,** 24–26
43. Chagas, J. R., Ferrer-Di Martino, M., Gauthier, F., and Lalmanach, G. (1996) *FEBS Lett.* **392,** 233–236
44. Jerala, R., Zerovnik, E., Kidric, J., and Turk, V. (1998) *J. Biol. Chem.* **273,** 11498–11504
45. Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M., and Wolthers, B. G. (1968) *Nature* **218,** 929–932
46. Sloane, B. F. (1990) *Semin. Cancer Biol.* **1,** 137–152
47. Mort, J. S., Recklies, A. D., and Poole, A. R. (1984) *Arthritis Rheum.* **27,** 509–515
48. Reddy, V. Y., Zhang, Q.-Y., and Weiss, S. J. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 3849–3853
49. Bossard, M. J., Tomaszek, T. A., Thompson, S. K., Amegadzie, B. Y., Hanning, C. R., Jones, C., Kurdyla, J. T., McNulty, D. E., Drake, F. H., Gowen, M., Levy, and M. A. (1996) *J. Biol. Chem.* **271,** 12517–12524

Identification of Internal Autoproteolytic Cleavage Sites within the Prosegments of Recombinant Procathepsin B and Procathepsin S: CONTRIBUTION OF A PLAUSIBLE UNIMOLECULAR AUTOPROTEOLYTIC EVENT FOR THE PROCESSING OF ZYMOGENS BELONGING TO THE PAPAIN FAMILY

Omar Quraishi and Andrew C. Storer

J. Biol. Chem. 2001, 276:8118-8124.
doi: 10.1074/jbc.M005851200 originally published online December 13, 2000
