Protein Science (1995), *4*:159–166. Cambridge University Press. Printed in the USA.

Rearranging the domains of pepsinogen

XINLI LIN, GERALD KOELSCH, JEFFREY A. LOY, AND JORDAN TANG

Protein Studies Program, Oklahoma Medical Research Foundation and the Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104

(RECEIVED September 27, 1994; ACCEPTED November 15, 1994)

Abstract

Most eukaryotic aspartic protease zymogens are synthesized as a single polypeptide chain that contains two distinct homologous lobes and a *pro* peptide, which is removed upon activation. In pepsinogen, the *pro* peptide precedes the N-terminal lobe (designated *pep*) and the C-terminal lobe (designated *sin*). Based on the three-dimensional structure of pepsinogen, we have designed a pepsinogen polypeptide with the internal rearrangement of domains from *pro-pep-sin* (native pepsinogen) to *sin-pro-pep*. The domain-rearranged zymogen also contains a 10-residue linker designed to connect *sin* and *pro* domains. Recombinant *sin-pro-pep* was synthesized in *Escherichia coli*, refolded from 8 M urea, and purified. Upon acidification, *sin-pro-pep* autoactivates to a two-chain enzyme. However, the emergence of activity is much slower than the conversion of the single-chain zymogen to a two-chain intermediate. In the activation of native pepsinogen and *sin-pro-pep*, the *pro* region is cleaved at two sites between residues 16P and 17P and 44P and 1 successively, and complete activation of *sin-pro-pep* requires an additional cleavage at a third site between residues 1P and 2P. In pepsinogen activation, the cleavage of the first site is rate limiting because the second site is cleaved more rapidly to generate activity. In the activation of *sin-pro-pep*, however, the second site is cleaved slower than the first, and cleavage of the third site is the rate limiting step. The reason for these differences is the result of the presence of activation intermediates bearing *pro* peptide 1P–16P, which is still covalently attached to the *sin* domain after the first and second cleavages. This peptide is known to have affinity to the enzyme moiety. Its presence apparently prevents the full expression of proteolytic activity, which catalyzes the cleavage of sites 2 and 3. A mechanism of intramolecular cleavage of site 1 is proposed that involves the local conformational change only near site 1 in the N-terminal region of the *pro* peptide.

Keywords: aspartic protease; pepsinogen; protein engineering; zymogen activation

Eukaryotic aspartic proteases are synthesized as single-chain zymogens. Upon activation, the N-terminal *pro* region of about 40–50 residues is removed. The resulting single-chain proteases have typically bilobal three-dimensional structures (Davies, 1990) with the substrate binding cleft and the catalytic aspartic acid residues situated between the lobes. The structures of the N- and C-terminal lobes are highly similar. This conformational homology is thought to have resulted from gene duplication and fusion in the evolution of aspartic proteases where the primordial aspartic protease would be a homodimer (Tang et al., 1978). Because the retroviral aspartic proteases are homodimers, considerable interest exists in their structure and function relationship with eukaryotic single-chain enzymes. We have demonstrated that the two lobes of pepsinogen expressed from separate genes can independently fold and recombine to form a viable two-chain zymogen, and that the N-terminal lobe can also form an active homodimer, as do the retroviral proteases (Lin et al., 1992). The independently folded recombinant lobes of pepsinogen were also used as probes to show that the N-terminal lobe of pepsin specifically denatures in neutral and alkaline solutions (Lin et al., 1993). Because of the innate capability of the lobes of pepsinogen to fold independently, it is possible to consider a rearrangement of their order by protein engineering. The functional perturbation of the resulting molecule may provide insights to the structure–function relationships of this family of enzymes and their precursors.

The domains in the peptide chain of pepsinogen are in the order of [pro]–[N-terminal lobe]–[C-terminal lobe] (*pro-pep-sin*). Using the crystal structure of pepsinogen (Hartsuck et al., 1992), we have designed a single-chain pepsinogen with the domain rearrangement of *sin-pro-pep*. Recombinant *sin-pro-pep* is a viable zymogen with drastically altered activation properties.

Reprint requests to: Jordan Tang, OMRF, 825 Northeast 13th Street, Oklahoma City, Oklahoma, 73104; e-mail: jordan-tang@omrf.uokhsc.edu.

Abbreviations: amu, atomic mass units; bis-Tris, bis(2-hydroxyethyl)-iminotris(hydroxymethyl)methane; FPLC, fast protein liquid chromatography; the following names are used for various pepsinogen domains and combinations thereof: *pro*, pro peptide; *pep*, the N-terminal lobe of pepsin; *sin*, the C-terminal lobe of pepsin; *pro-pep-sin*, native pepsinogen; *sin-pro-pep*, domain-rearranged pepsinogen; *pep·sin*, two-chain pepsin comprised of noncovalently associated *pep* and *sin*.

Results

Expression, refolding, and purification of sin-pro-pep

The gene of domain-rearranged pepsinogen, *sin-pro-pep*, is schematically summarized in Figure 1. The rearrangement involved placement of the cDNA region encoding *sin* upstream from that of *pro* and *pep*. Connection of the C-terminus of *sin* to the N-terminus of the *pro* domain required construction of a new connecting strand, preferably a random coil. The designed connecting strand has a sequence of -Gly-Pro-$(Gly)_6$-Pro-Gly-, which was chosen for minimal secondary structure and for minimal interaction with native pepsinogen. The length of this sequence was chosen based on analysis of the pepsinogen crystal structure to fit the distance between the C-terminus of *sin* (residue Ala^{326}) and the N-terminus of *pro* (residue Leu^{1P}) (Fig. 2A; Kinemage 1). The modeled *sin-pro-pep* in Figure 2B (Kinemage 1) shows the expected position of this new connecting strand and the separation of the *pep* and *sin* domains by the effective cutting between residues Ser^{172} and Ser^{173} (Lin et al., 1992). The expression of this rearranged gene in *Escherichia coli* resulted in the synthesis of recombinant *sin-pro-pep* as "inclusion bodies" and was recovered, refolded from 8 M urea, and purified by gel filtration on Sephacryl S-300 and ion-exchange FPLC on a Resource Q column. The chromatogram of the latter (Fig. 3A) shows that a single proteolytically active peak eluted at 0.37 M NaCl. The purity of active *sin-pro-pep* is confirmed by SDS-PAGE (Fig. 4; lane for time = 0).

Activation of sin-pro-pep

Sin-pro-pep was found to be proteolytically active in an acidic solution (^{14}C hemoglobin assay), which suggested that the domain-rearranged pepsinogen is capable of auto-activation. To detect structural changes accompanying activation, a *sin-pro-pep* solution was acidified to pH 2.0 for 10 s and then chromatographed on a Resource Q column using the same conditions as for the experiment shown in Figure 3A. Fractions from the resulting peaks (Fig. 3B) were analyzed for activity and for peptide chain composition on SDS-PAGE (data not shown). The second peak (eluted at 12 mL) was unactivated *sin-pro-pep*. The third peak (13 mL) was active proteolytically and revealed a change in mobility on SDS-gel electrophoresis to that of a two-chain pepsin. Peak 1 (10.5 mL) contained two-chain pepsin but was not enzymically active. This material, which was minor as compared to the active enzyme (peak 3), was likely inactivated two-chain pepsin because two-chain pepsin activated from *sin-pro-pep* denatures slowly at pH 6.0 (data not shown). Peaks 4 (14 mL) and 5 (15 mL) were relatively minor and revealed no activity. These may be the activation peptide fragments liberated during the activation.

The course of activation of *sin-pro-pep* at pH 2.6 was analyzed for changes in peptide chain composition and for the appearance of activity. The SDS-PAGE of samples of different activation durations established that the conversion of single-chain *sin-pro-pep* to two-chain pepsin proceeded rapidly and was near completion within 4 min (Fig. 4). A 26-kDa intermediate band was visible in samples of 1, 2, and 4 min incubation. The smaller 22-kDa band appeared immediately after activation and emerged as a closely positioned doublet at 8 min and thereafter. (This is shown only as a slight thickening of this band in Figure 4, but was seen clearly in 20% SDS-PAGE, data not shown.) However, the appearance of proteolytic activity during the activation was much slower than the peptide chain cleavages manifested as the appearance of the 26- and 22-kDa bands. Figure 5A shows the rate of the appearance of proteolytic activity under the same experimental conditions as the experiment in Figure 4. At 8 min, where all the peptide cleavages apparently had been completed, only about 10% of the activity had been generated. The full activity was attained only after 60 min (data not shown). The initial activation suggested the presence of a lag in activity following this cleavage. This was confirmed in a similar experiment with lower zymogen concentration as shown in Figure 5B, which also shows the complete course of activation.

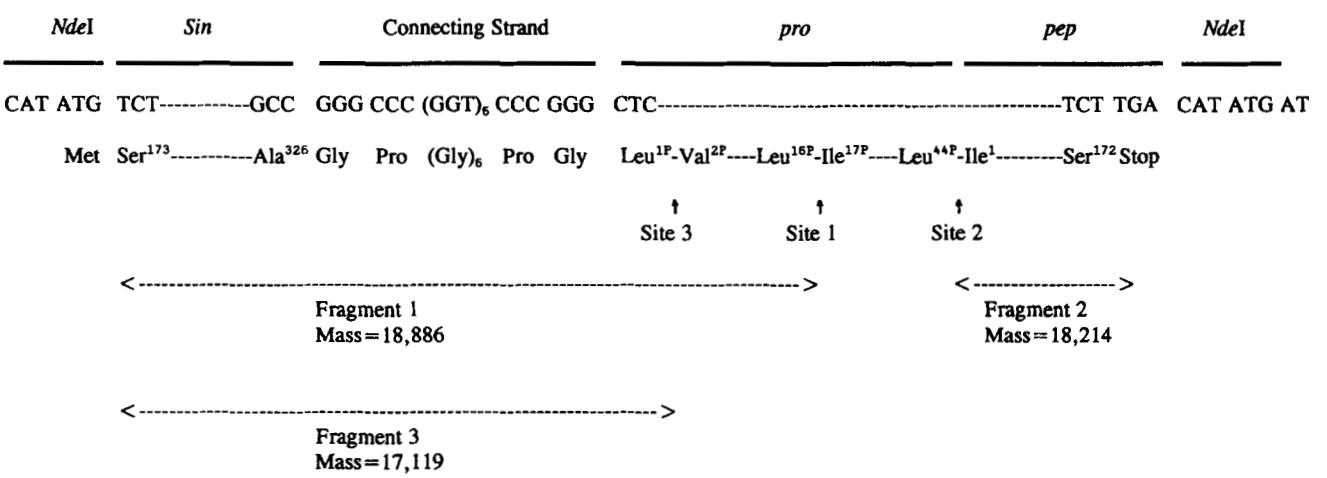

Fig. 1. Construction of the *sin-pro-pep* gene and rearrangement of the pepsinogen polypeptide. The first line represents the domains involved and their coding regions below. The gene consists of an initiation Met, a *sin* domain, a designed connecting strand, the *pro-pep* domains, and a stop codon. Relevant amino acid residues and their numbers are shown below the DNA sequence. The locations of sites cleaved during the activation of *sin-pro-pep* are shown by vertical arrows. The final fragments from the activation are shown by double-headed horizontal arrows.

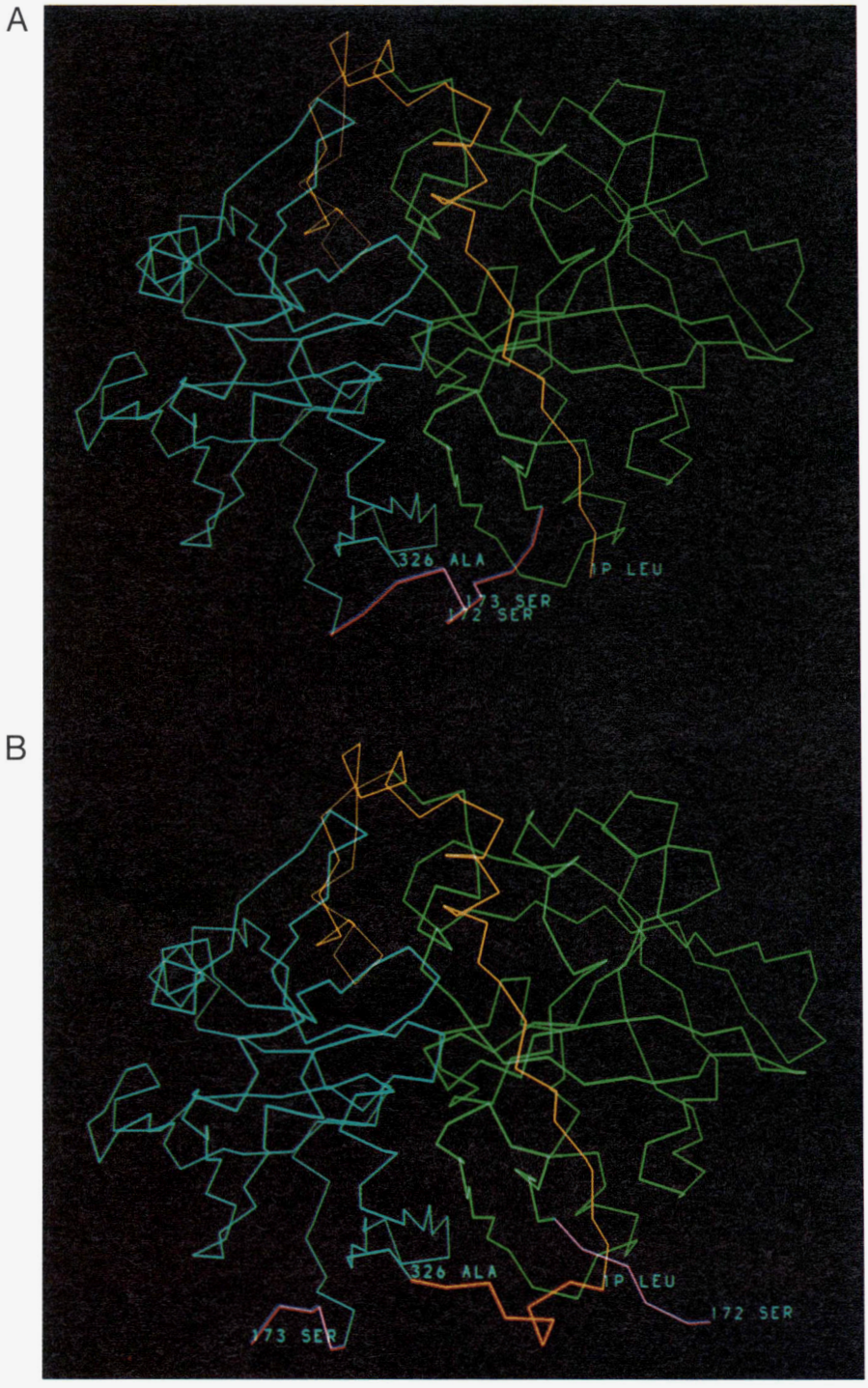

Fig. 2. Molecular models representing native pepsinogen and *sin-pro-pep*, the domain-rearranged pepsinogen. **A:** α-Carbon trace of pepsinogen. The yellow trace is the *pro* domain, green is the *pep* domain, and the light-blue trace is the *sin* domain. The violet trace represents the native random coil, which is the connecting strand between *pep* and *sin*. Termini are labeled 1P Leu (N-terminus) and 326 Ala (C-terminus). Serine residues 172 and 173 in the connecting strand are labeled. **B:** Graphic representation of *sin-pro-pep*. The red trace between 326 and 1P represents a modeled seven-residue new connecting strand between the C-terminus of *sin* and the N-terminus of *pro-pep*. The peptide bond between Ser^{172} and Ser^{173} is effectively cut in the connecting strand (violet color) of two lobes of pepsinogen. In this model, the cut strands are depicted as separated. These rearrangements create a new N-terminus at Ser^{173} and a new C-terminus at Ser^{172}. Remaining details are as in A.

To assess the proteolytic cleavage from the activation of *sin-pro-pep*, the short-time activation (10 s, pH 2.0) products were analyzed for mass by using mass spectrometry and their N-terminal sequences were determined. Purified, proteolytically active activation product (third peak in Fig. 3B) gave rise to two N-terminal sequences, which were assigned to be the N-terminal sequences of *sin* (observed: SYYTGSLNWVPVSVEGYWQIT LDS) and *pep* (observed: IGDEPLENYLDTEYFGTIGIGTPA). Electrospray mass spectrometric determination of the same activation product produced two masses of 18,925 and 18,235 amu. As a comparison, recombinant *sin* (Lin et al., 1992) produced masses of 16,359, 16,380, and 16,398 amu, which correspond, respectively, to *sin* (theoretical mass 16,362), and *sin* + either 1 or 2 ammonium counterions. Thus, the mass 18,235 was assigned to be *pep* (theoretical mass 18,214; Fig. 1, fragment 2) + 1 ammonium counterion, and the mass 18,925 was assigned to be a fragment consisting of *sin*, connecting strand, and the N-terminal 16 residues of *pro* (theoretical mass 18,886; Fig. 1,

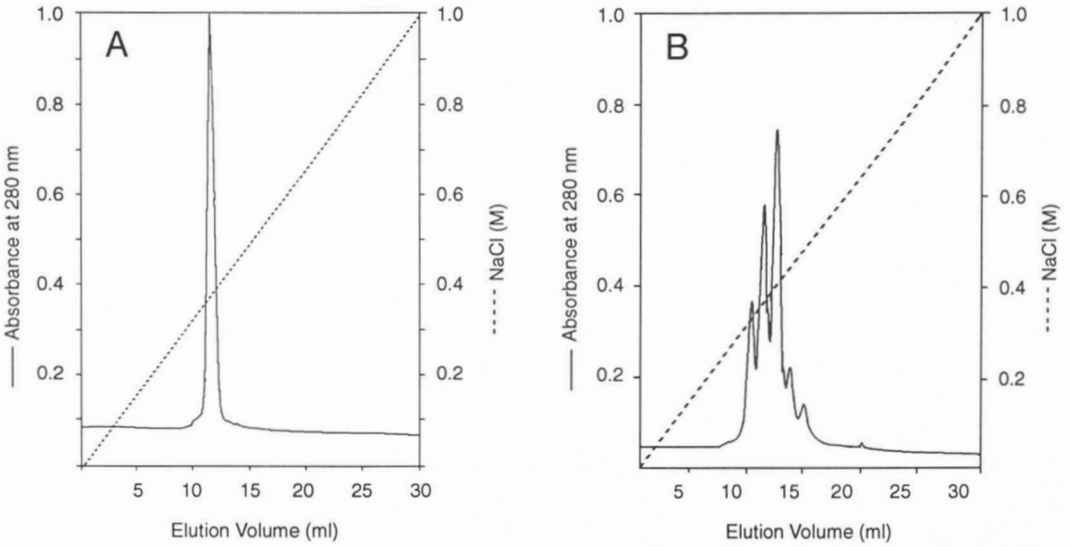

Fig. 3. FPLC separations on a Resource Q column of zymogen *sin-pro-pep* (**A**) and *sin-pro-pep* incubated in an acidic solution (**B**). The column was equilibrated in 20 mM bis-Tris-HCl, 0.8 M urea, pH 6.0. The elution was effected by a NaCl gradient of 0–1 M over 30 mL (dotted line).

fragment 1) + 2 ammonium counterions. The initial cleavage of *pro* between residues 16P and 17P (Fig. 1, site 1) was confirmed by an N-terminal sequence determination of a partially activated sample (pH 2.6, 2 min, 25 °C), which produced, in addition to the N-terminal sequences of *sin* and *pep*, a sequence of IKNGK identical to that of residues 17P–21P. Because the activation intermediate was clearly visible at 2 min of activation (Fig. 4), this observation suggested that the 26-kDa intermediate is the product of cleavage between residues 16P and 17P in the *pro* region.

Electrospray mass spectrometry of *sin-pro-pep* activated to generate full activity (20 h, pH 2.6 at 30 °C) indicated the presence of three species. Two masses were assignable to *pep* (18,208) and to the segment of protein sequence including *sin*, connecting strand, and residue 1P (17,119; Fig. 1, fragment 3). A very minor third species was assigned to the segment *sin*, connecting strand, and residues 1P–16P (18,880). In this experiment, the observed mass of each species exceeded that of its theoretical mass by about 6 amu, which we considered to be within the limits of experimental error.

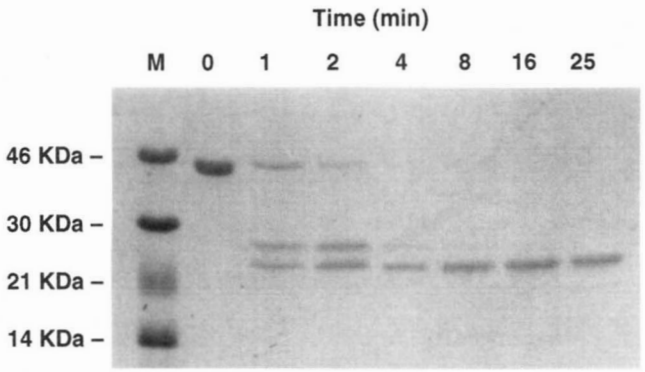

Fig. 4. Acid activation of zymogen *sin-pro-pep* as demonstrated on SDS-PAGE. Zymogen *sin-pro-pep*, 0.7 mg/mL, was activated at pH 2.6, 25 °C, for the different times noted. Lane M, molecular weight markers.

The pH dependence of the appearance of activity of *sin-pro-pep* activation (Fig. 6) was examined by measuring the substrate hydrolysis following a 15-min activation at 25 °C. A constant ionic strength buffer system ($I = 0.2$ M) was used to avoid a salt effect upon the rate of activation. The pH dependence in Figure 6 is similar to that of the first-order activation rate of pepsinogen (Al-Janabi et al., 1972), although there is an apparent break in pH dependence near pH 1.9, where the pepsinogen profile shows a more abrupt transition.

Kinetic parameters of fully activated sin-pro-pep

Fully activated *sin-pro-pep* is similar to two-chain pepsin, *pep·sin* (Lin et al., 1992) except it contains a linker and the N-terminal leucine of pepsinogen covalently linked to the *sin* subunit. So it was of interest to measure the kinetic parameters in order to determine the effects of these added structural elements. At pH 3.5, 30 °C, the K_m and k_{cat} were found to be 0.06 ± 0.01 mM and 170 ± 20 s^{-1}. The amount of enzyme in these studies was determined by active-site titration using pepstatin.

Discussion

Nearly all zymogens of proteolytic enzymes, including pepsinogen, have the *pro* peptides at their N-termini followed by the protease domains. This order is biologically important during the biosynthesis of these precursors in order to prevent unwanted proteolysis (Neurath et al., 1967). We have demonstrated in this work that a zymogen with a *pro* peptide in the middle and transposed N- and C-terminal domains can be folded and activated successfully. These results substantiate the observation that these domains are capable of independent folding (Lin et al., 1992) prior to assembling into topologically correct relationships for

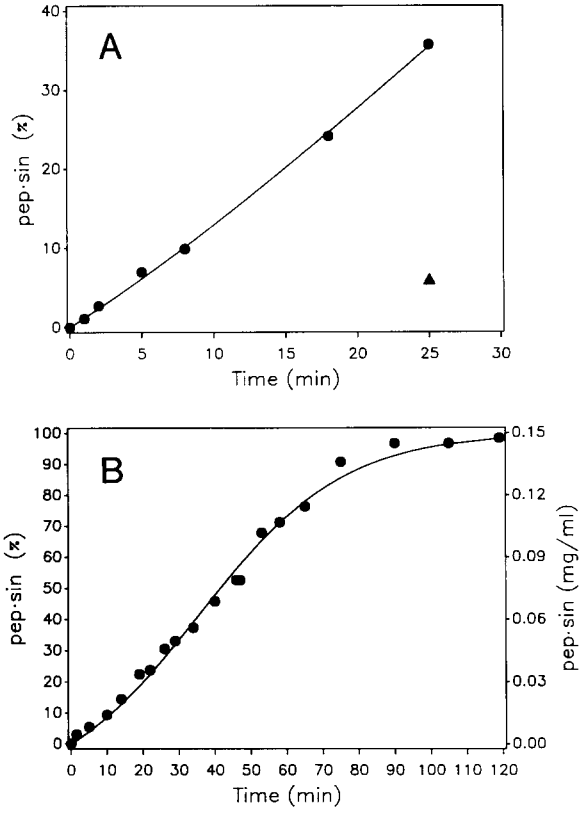

Fig. 5. Activation profiles for zymogen *sin-pro-pep*. **A:** Percentage of activation at various time intervals. Activation of *sin-pro-pep* was carried at 0.7 mg/mL, pH 2.6, 25 °C. Note that reaction conditions and sample time intervals correspond to the experiment analyzed in Figure 4, where apparent completion of activation within 8 min judged from SDS-PAGE corresponds to only a 10% fulfillment of activity. The ● symbol at each interval is the percentage activity acquired relative to a fully activated sample. The line passing through the data is a nonlinear fit to a mixed first- and second-order activation model (see text). The ▲ symbol represents the percentage activity for a 0.35-mg/mL sample activated under identical conditions. **B:** Percentage of activation at pH 2.0, 0.14 mg/mL *sin-pro-pep*. Data are presented identically to A.

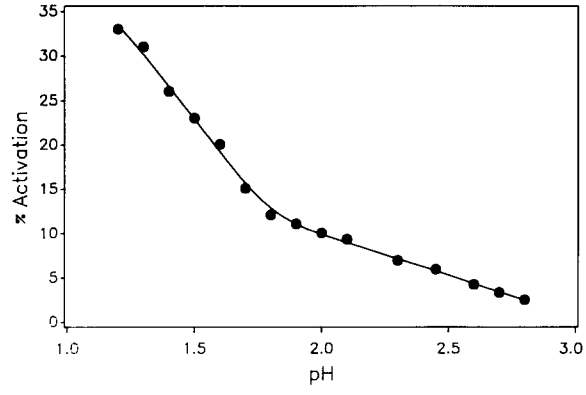

Fig. 6. Dependence of activation rate of *sin-pro-pep* on pH. Using a constant ionic strength buffer system ($I = 0.2$ M), the zymogen was activated at 25 °C, 0.038 mg/mL for 15 min at various pHs and assayed immediately using a turbidity assay at pH 5.3 to determine the percentage of activation.

potential activity. Therefore, in domain-rearranged pepsinogen, the spatial relationships of the domains should be nearly identical to that of the native pepsinogen, with the exception of the added connecting strand, which is expected to be on the surface of the three-dimensional structure (Fig. 2B; Kinemage 1) and should provide little perturbation to the tertiary structure.

Like pepsinogen, *sin-pro-pep* clearly manifests activity in acidic solutions with concomitant peptide cleavages in the *pro* region. However, the rate for the appearance of the activity relative to the rate of peptide bond cleavage in the *pro* region is very different in the activation of the two zymogens. Gel electrophoresis of acidified *sin-pro-pep* established that the single-chain is rapidly converted to a two-chain intermediate. Judging from the electrophoretic pattern (Fig. 4), the half-life of this first cleavage is less than 1 min, which is comparable to the half-life of native pepsinogen activated under similar conditions (Al-Janabi et al., 1972). This conclusion is also supported by the relative peak sizes of zymogen and enzyme in the chromatography of *sin-pro-pep* after activation at pH 2 for 10 s. But, unlike the activation of native pepsinogen, in which the chain cleavage proceeds at about the same rate as the manifestation of proteolytic activity (Sanny et al., 1975), the full activity appears in *sin-pro-pep* activation much slower than the apparent bond cleavages (Fig. 5A) by about 200-fold, at the conditions described.

The shape of the *sin-pro-pep* activation curve in Figure 5B suggests mixed first-order (intramolecular) and second-order (pepsin catalyzed) reactions, similar to the activation of pepsinogen at high pH and concentration, and the rate constants have been calculated using the previously developed model (Al-Janabi et al., 1972). The apparent first- and second-order constants k_1 and k_2 were determined to be 0.0072 ± 0.0007 (min^{-1}) and 0.30 ± 0.03 ($mL \cdot mg^{-1} \cdot min^{-1}$). The k_1 and k_2 for porcine pepsinogen (Al-Janabi et al., 1972) at pH 2.0 at 28 °C are 2.6 (min^{-1}) and 1.3 ($mL \cdot mg^{-1} \cdot min^{-1}$). However, the physical meaning of these constants with respect to *sin-pro-pep* is not clear, owing to the difference in the rate-limiting steps of activation discussed below.

The sites of cleavages in the *sin-pro-pep* activation were identified. Based on the sequencing and mass spectrometry data, there are three sites of activation cleavage (Fig. 1). Site 1 is a Leu–Ile bond between residues 16P and 17P of the *pro* region. This is the site rapidly cleaved in *sin-pro-pep* activation (Fig. 7) to produce an intermediate consisting of a 26-kDa band and a 22-kDa band in SDS-gel electrophoresis (Fig. 4). Site 2 is a Leu–Ile bond (residues 44P–1) between the *pro* and *pep* regions. This cut converts the 26-kDa fragment to another species of the 22-kDa band in the electrophoresis (which gave rise to a thickened band). The first two cleavages in the activation of *sin-pro-pep* accomplish the removal of residues 17P–44P of the *pro* peptide. However, we have established that the manifestation of full activity of *sin-pro-pep* requires a third cleavage to remove most of the *pro* peptide, 1P–16P, which after the first 2 cleavages is still attached to the C-terminus of the *sin* domain through the connecting strand (Fig. 7). This is clearly supported by the data, which demonstrate the first 2 cleavages have been completed in the initial minutes, yet activity is not completely expressed until 90 min of activation. We have also observed that the isolated two-cut intermediate possesses only low activity (data not shown). Finally, the position of cleavage site 3 was established by the mass determination of a fully activated sam-

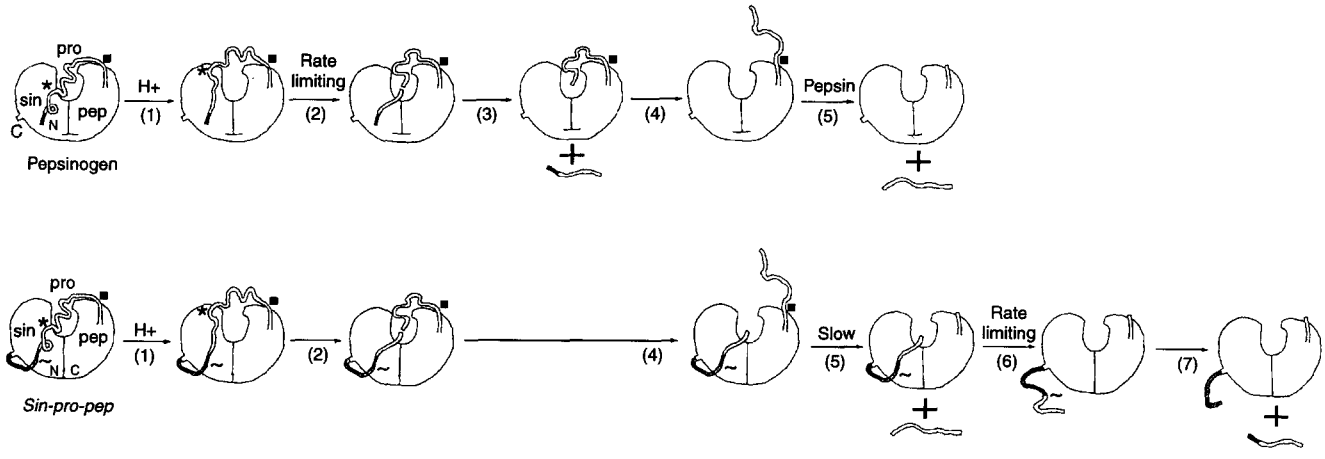

Fig. 7. Schematic presentation of the activation of native pepsinogen (upper line) and domain-rearranged pepsinogen, *sin-pro-pep* (lower line). The starting zymogens are marked for their N- and C-terminal positions and the domains *pro*, *pep*, and *sin*. The activation cleavage site 1 (*), site 2 (■), and site 3 (~) are marked whenever appropriate. The N-terminal nine residues of pepsinogen (residues 1P–9P) and their corresponding residues in *sin-pro-pep* are marked solid. These residues are hypothesized to remain bound during the initial steps of activation. The new connecting strand between *sin* and *pro* is shaded. The hypothesized mechanism involves the following: step (1), the initial conformational change of the region near site 1 in the *pro* peptide; step (2), the binding of site 1 to the active site and cleavage of site 1, the rate-limiting step for pepsinogen activation (see Discussion); step (3), the dissociation of peptide 1P–16P, a fast step in pepsinogen, but a very slow step in *sin-pro-pep* activation, which is actually shown as step (6); step (4), the "conformational dissociation" of peptide 17P–44P of the *pro* region; step (5), the active intermediate-catalyzed cleavage of site 2 at 44P–1; step (6), the slow "conformational dissociation" of *pro* peptide 1P–16P in *sin-pro-pep* activation, the rate-limiting step; step (7), the hydrolysis of site 3 and the dissociation of peptide 2P–16P.

ple to be between residues 1P and 2P, which implies the removal of 15 residues, 2P–16P, from the two-cut intermediate.

In the initial stage, *sin-pro-pep* and pepsinogen both activate with cleavages in the order of sites 1 and 2 (Dykes & Kay, 1976), suggesting that the mechanisms in the cleavage of these two bonds are similar in these two zymogens. Because the N-terminus of the *pro* peptide is covalently linked to *sin*, the initial cleavage at site 1 during the activation of *sin-pro-pep* likely does not involve the "conformational dissociation" (the dissociation of the N-terminal region of *pro* from the *pep·sin* moiety of the zymogen) of the N-terminal 9 residues (1P–9P). These residues form one of several strands in a β-sheet, and Nielsen and Foltmann (1993) have provided evidence that they are important for the interaction of the activation peptide to the pepsin moiety in the native pepsinogen molecule. The cleavage of site 1 in pepsinogen is an intramolecular event (for review, see Tang & Wong, 1987). The evidence discussed above supports a site 1 cleavage mechanism similar to that proposed previously for human pepsinogen by Kagayama et al. (1989). In this mechanism, the N- and C-terminal regions of the *pro* peptide are proposed to remain in their native conformations during the initial step of the activation (Fig. 7). The region near site 1 may be "locally denatured" and "conformationally dissociated" (Fig. 7, step (1)). This region may then bind back into the active site resulting in bond cleavage (Fig. 7, step (2)).

The rates of the activation cleavage steps in these two zymogens are very different and are also summarized in Figure 7. In the activation of native pepsinogen, the cleavage of site 1 is intramolecular (first-order reaction). The rate of the site 2 cleavage (measured as the appearance of the N-terminus of pepsin) is the same as that for the appearance of activity (Sanny et al., 1975), which is first-order. Thus, the rate of site 2 cleavage must be faster than the cleavage of site 1, and the rate-limiting step in pepsinogen activation is the first cleavage. In the activation of *sin-pro-pep*, the rate of site 1 cleavage, derived from gel electrophoresis of the short-term activation and products analyzed by N-terminal sequencing, is similar to that for pepsinogen activation. In Figure 4, this step is complete in about 3 min. However, the site 2 cleavage is slower; it is not complete until after 4 min. The cleavage of site 3 is much slower and is the rate-limiting step in the *sin-pro-pep* activation. It is complete near 90 min, as shown by the activity (Fig. 5) and by mass data. Thus, the presence of the N-terminal 16 residues of *pro* before the third cleavage strongly restricts the catalytic activity of the intermediates. In the single-cut intermediate of *sin-pro-pep* (Fig. 7, product of step (2)), the "conformational dissociation" of the 1P–16P peptide must be slower because it is covalently attached to *sin*. The slowness in this step, which is portrayed in Figure 7 as the missing step (3) in *sin-pro-pep* activation, then retards the cleavage of the second site between residues 44P and 1, proceeding to steps (6) and (7) in the figure.

Why does the presence of covalently linked peptide 1P–16P have such a strong effect on activity and why is the third cleavage so slow? The Leu^{1P}–Val^{2P} bond at the third cleavage site is quite similar to the Leu–Ile bonds cleaved at the other two sites. So, the bond specificity is not likely the cause of slowness. It is more likely that, even in an acidic activation solution, this peptide has strong affinity toward the *pep·sin* moiety in the activation intermediate. It is known that the peptide corresponding to amino acids 1P–16P of the pro region has strong affinity to the enzyme (Dunn et al., 1983; Nielsen & Foltmann, 1993). Also, the N-terminal nine residues of the *pro* region comprise one of several strands of a β-sheet in pepsinogen (Hartsuck et al., 1992) and may have considerable stability to resist "conformational dissociation." Our current view is that the covalently attached 1P–16P peptide stays mostly in its native conformation of pep-

sinogen, making the cleavage of site 2 slow and site 3 rate limiting. The concentration dependence of the overall apparent rate of activation of *sin-pro-pep* indicates that it is limited by a bimolecular cleavage event.

The kinetic parameters of fully activated *sin-pro-pep* indicate that its K_m is similar and k_{cat} is about half those for pepsin (Lin et al., 1992). The k_{cat} value is nearly three times higher than that of *pep·sin* obtained from independent refolding of the two lobes (Lin et al., 1992). The cause for this difference is not clear and it seems unlikely that this difference can be attributed to the presence of a linker in the activation product of *sin-pro-pep*. Other possible reasons are the different stabilities and refolding efficiencies to attain conformationally correct proteins.

Materials and methods

Materials

Oligonucleotide primers were synthesized locally at the Molecular Biology Resource Center, University of Oklahoma Health Sciences Center, using an Applied Biosystem DNA Synthesizer 380B. The TA Cloning System was purchased from Invitrogen. Other reagents were the highest purity obtained commercially.

Construction of sin-pro-pep gene

The *sin-pro-pep* gene was constructed by using a PCR-based synthesis using porcine pepsinogen cDNA as template (Lin et al., 1989). Primer 1 (Sin-N, 5′-TACATATGTCTTACTACA CAGG) and primer 2 (Sin-C, 5′-ACCACCACCACCACCAC CGGGCCCGGCCACGGGAGCCAGGCC) were used in PCR to synthesize the 480-bp *sin* fragment (containing pepsinogen nucleotides 731–1,192; Lin et al., 1989). The 660-bp *pro-pep* fragment (containing pepsinogen nucleotides 83–730) was also PCR synthesized using primer 3 (Propep-N, 5′-GGTGGTGGT GGTGGTGGTCCCGGGCTCGTCAAGGTCCCGCTG) and primer 4 (Propep-C, 5′-ATCATATGTCAAGAATCGATGCC). These two resulting fragments were isolated from agarose gel, annealed, and used as a template for the PCR synthesis of *sin-pro-pep* using primers 1 and 4. The resulting DNA was cloned into the TA Cloning System (Invitrogen), and selected clones were sequenced to confirm the construct.

Expression of sin-pro-pep in E. coli and the refolding and purification of the recombinant protein

Sin-pro-pep genes were cloned into the *Nde* I site of plasmid pET-11a and transfected into *E. coli* strain BL21(DE3) for expression (Studier et al., 1990). *Sin-pro-pep* is synthesized as inclusion bodies, which were recovered and washed as previously described (Lin et al., 1992). The inclusion bodies were dissolved in 20 mM CAPS (cyclohexylamino-1-propanesulfonic acid), pH 10.5, containing 8 M urea, 1 mM glycine, 1 mM EDTA, and 0.1 M β-mercaptoethanol and centrifuged at 36,000 × *g* to remove the insoluble material. The supernate was diluted with the same buffer until the A_{280} was near 2.0. The protein–urea solution was diluted with rigorous stirring into 10 volumes of 20 mM Tris-HCl, pH 9.0, and then kept at room temperature for 20 min. The pH of the solution was adjusted to 8.0 by the addition of 1 M bis-Tris, pH 6.0, and kept standing at 4 °C for 16 h. The pH of the solution was further adjusted to 6.0 with 1 M HCl and further incubated at 4 °C for 24 h. The stepwise pH changes were important for good yield of this refolded protein. The recombinant *sin-pro-pep* was purified as previously described for *pro-pep* (Lin et al., 1992) except that, in the final step, the protein was chromatographed on an anion-exchange column of Resource Q (Pharmacia) equilibrated in 20 mM bis-Tris, pH 6.0, with 0.8 M urea and eluted by linear gradient of 0–1 M NaCl over 30 mL at 4 mL/min.

Assay of proteolytic activity

Acid proteolytic activity was assayed using ^{14}C-labeled hemoglobin as substrate (Lin et al., 1989). For kinetic studies and whenever specified, proteolytic activity was also measured using a synthetic chromogenic substrate, KPAEFF(NO_2)AL as previously described (Fusek et al., 1990). The number of active sites in the assay of kinetic parameters to the chromogenic substrate was determined using aliquots of a 1 μM stock of pepstatin A in 10% methanol (Tomasselli et al., 1990).

Mass spectrometry and determination of molecular mass

Determination of molecular mass was done using a Sciex API III triple-quadrupole electrospray mass spectrometer at the Molecular Biology Resource Center, University of Oklahoma Health Sciences Center. Negative ions produced by a pneumatically assisted electrospray interface were sampled with an atmospheric pressure ion source with scanning every 0.2 amu. Polypropylene glycol ions of known mass were used for mass calibration. The average molecular mass was calculated using proprietary software.

Activation studies

Activation as a function of time was determined by incubation with 0.2 M citric acid, pH 2.0, 25 °C. The amount of active enzyme at various time intervals was assayed by monitoring the cleavage of the chromogenic substrate KPAEFF(NO_2)AL at pH 3.0. Activation during the assay was not evident. The formation of active enzyme as a function of time was analyzed using SAS 5.06 (SAS, 1993). The pH dependence of activation for purified *sin-pro-pep* was determined by incubation of 5 μL of 0.75 mg/mL zymogen solution with 5 μL of activation buffer at intervals spanning pH 1.2–2.8, using either 0.2 M KCl-HCl or 0.2 M citric acid–sodium citrate buffer. After 15 min, a 2-μL aliquot was assayed immediately at pH 5.2 using a milk clotting turbidity assay in which the amount of activated enzyme is inversely proportional to the milk clotting time (McPhie, 1976). Activation at pH 5.2 was determined to be negligible within the duration of the assay. The percentage of activation was quantified from the inverse clotting time of fully activated *pep·sin*.

Acknowledgments

We thank Dr. Jean Hartsuck for helpful discussions regarding the activation of *sin-pro-pep*. This work was supported by grant DK-01107.

References

Al-Janabi J, Hartsuck JA, Tang J. 1972. Kinetics and mechanism of pepsinogen activation. *J Biol Chem 247*:4628–4632.

Davies DR. 1990. The structure and function of the aspartic proteinases. *Annu Rev Biophys Biophys Chem 19*:189–215.

Dunn BM, Lewitt M, Pham C. 1983. Inhibition of pepsin by analogues of pepsinogen-(1-12)-peptide with substitutions on the 4-7 sequence region. *Biochem J 209*:355-362.
Dykes CW, Kay J. 1976. Conversion of pepsinogen into pepsin is not a one-step process. *Biochem J 153*:141-144.
Fusek M, Lin X, Tang J. 1990. Enzymic properties of thermopsin. *J Biol Chem 265*:1496-1501.
Hartsuck JA, Koelsch G, Remington SJ. 1992. The high-resolution crystal structure of porcine pepsinogen. *Proteins Struct Funct Genet 13*:1-25.
Kagayama T, Ichinose M, Miki K, Athauda SB, Tanji M, Takahashi K. 1989. Difference of activation processes and structure of activation peptides in human pepsinogen A and progastricsin. *J Biochem (Tokyo) 105*:15-22.
Lin X, Lin Yz, Koelsch G, Gustchina A, Wlodawer A, Tang J. 1992. Enzymic activities of two-chain pepsinogen, two-chain pepsin, and the amino-terminal lobe of pepsinogen. *J Biol Chem 267*:17257-17263.
Lin X, Loy JA, Sussman F, Tang J. 1993. Conformational instability of the N- and C-terminal lobes of porcine pepsin in neutral and alkaline solutions. *Protein Sci 2*:1383-1390.
Lin X, Wong RNS, Tang J. 1989. Synthesis, purification, and active site mutagenesis of recombinant porcine pepsinogen. *J Biol Chem 264*:4482-4489.
McPhie P. 1976. A turbidimetric milk-clotting assay for pepsin. *Anal Biochem 73*:258-261.
Neurath H, Walsh KA, Winter WP. 1967. Evolution of structure and function of proteases. *Science 158*:1638-1644.
Nielsen FS, Foltmann B. 1993. Activation of porcine pepsinogen A. *Eur J Biochem 217*:137-142.
Sanny CG, Hartsuck JA, Tang J. 1975. Conversion of pepsinogen to pepsin. *J Biol Chem 250*:2635-2639.
Studier WF, Rosenberg AH, Dunn JJ, Dubendorff JW. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol 185*: 60-89.
Tang J, James MNG, Hsu IN, Jenkins JA, Blundell TL. 1978. Structural evidence for gene duplication in the evolution of the acid proteases. *Nature 271*:618-621.
Tang J, Wong RNS. 1987. Evolution in the structure and function of aspartic proteases. *J Cell Biochem 33*:53-63.
Tomasselli AG, Olsen MK, Hui JO, Staples DJ, Sawyer TK, Heinrikson RL, Tomich CSC. 1990. Substrate analogue inhibition and active site titration of purified recombinant HIV-1 protease. *Biochemistry 29*:264-269.

Forthcoming Papers

Direct evidence of a heterotrimeric complex of human interleukin-4 with its receptors

R.C. Hoffman, B.J. Castner, M. Gerhart, M.G. Gibson, B.D. Rasmussen, C.J. March, J. Weatherbee, M. Tsang, A. Gustchina, C. Schalk-Hihi, L. Reshetnikova, and A. Wlodawer

Purification and preliminary crystallographic studies of penicillin G acylase from *Providencia rettgeri*

H.E. Klei, G.O. Daumy, and J.A. Kelly

Refined solution structure of human profilin I

W.J. Metzler, B.T. Farmer II, K.L. Constantine, M.S. Friedrichs, T. Lavoie, and L. Mueller

Structural basis of substrate specificity in the serine proteases

J.J. Perona and C.S. Craik

Transmembrane helices predicted at 95% accuracy

B. Rost, R. Casadio, P. Fariselli, and C. Sander

An automatic method involving cluster analysis of secondary structures for the identification of domains in proteins

R. Sowdhamini and T.L. Blundell

The structures of partridge egg-white lysozyme with and without tri-*N*-acetylchitotriose inhibitor at 1.9 Å resolution

M.A.Turner and P.L. Howell

Recombinant immunoglobulin variable domains generated from synthetic genes provide a system for in vitro characterization of light-chain amyloid proteins

P.W. Stevens, R. Raffen, D.K. Hanson, Y.-L. Deng, M. Berrios-Hammond, F.A. Westholm, C. Murphy, M. Eulitz, R. Wetzel, A. Solomon, M. Schiffer, and F.J. Stevens

Approaches to labeling and identification of active site residues in glycosidases

S.G. Withers and R. Aebersold