Covalent Structure of a Group-Specific Protease from Rat Small Intestine†

Richard G. Woodbury,*,‡ Nobuhiko Katunuma,§ Keiko Kobayashi,§ Koiti Titani, and Hans Neurath

Appendix: Crystallographic Data for a Group Specific Protease from Rat Intestine‖

Wayne F. Anderson,*,‖ Brian W. Matthews,*,‖ and Richard G. Woodbury#

ABSTRACT: "Group-specific"protease (GSP) is a serine protease, obtained from rat small intestine, which preferentially inactivates the apo forms of certain pyridoxal phosphate requiring enzymes. The enzyme contains 224 amino acid residues in a single polypeptide chain and three disulfide bonds. In the present work the covalent structure has been determined and its homologous relationship to those of chymotrypsin, trypsin, and elastase has been established (approximately 33% identity with each). The residues forming the "charge-relay" system of the active site of chymotrypsin (His-57, Asp-102, and Ser-195) are found in corresponding regions in GSP, whereas an alanyl residue at position 176 of GSP corresponds to a residue which participates in the primary substrate binding site in serine proteases (Asp-177 in trypsin; Ser-189 in chymotrypsin). Three disulfide bonds in GSP occur in similar positions in chymotrypsin, trypsin, and elastase. However, GSP lacks a disulfide bond which is present in all known serine proteases (linking Cys-191 to Cys-220 in chymotrypsin). In view of the close proximity of this bond to both the primary and the antiparallel binding sites of various serine proteases, it is likely that its absence in GSP is related to the substrate specificity of this enzyme. It is concluded that GSP diverged from a common ancestor preceding chymotrypsin but following trypsin.

The steady-state concentration of proteins in cells and tissues is controlled by the rates of their synthesis and degradation (Schimke, 1970, 1973). Although lysosomal proteases play a major role in intracellular protein turnover (Coffey & de Duve, 1968), there is considerable evidence that other proteases may be of equal importance. Goldberg & St. John (1976) discussed the likelihood of several alternate intracellular proteolytic pathways, particularly in muscle which undergoes autolysis at least as fast in the alkaline range (pH 8–9) as it does in the acid range (pH 4–5). They also noted that certain cells, e.g., reticulocytes, exhibit moderate rates of protein turnover, but contain few lysosomes.

In recent years, numerous hypotheses have been proposed suggesting that specific limited proteolysis may initiate the degradation of certain intracellular proteins or groups of similar proteins. Support for the existence of degradative pathways of this type has been derived from the investigations of Katunuma (1973) who has isolated from various rat tissues several proteases exhibiting restricted specificities. These proteases, which exhibit optimal activities at pH 8–9, have several properties in common including the capability to inactivate, presumably by a process of limited proteolysis, the apo forms of certain intracellular pyridoxal phosphate requiring enzymes. The holo forms of these enzymes and the apo forms of several enzymes requiring other cofactors were not inactivated by these proteases (Katunuma et al., 1975).

These so-called "group specific" proteases (GSP) have been shown to possess chymotrypsin-like specificity toward small substrates. Further, they are inactivated by DFP[1] and their catalytic sites are thought to be structurally related to those of the pancreatic serine proteases (Katunuma et al., 1975).

In view of the current interest in the mechanisms of intracellular protein degradation in eukaryotes, the present study was initiated in order to determine the amino acid sequence of the group-specific protease obtained from rat small intestine. The covalent structure described in this paper reveals that the enzyme is homologous to the pancreatic serine proteases, but also possesses unique structural features.

Experimental Section

Materials. Group-specific protease was isolated from thoroughly rinsed rat small intestine by methods previously described (Katunuma et al., 1975).

Trypsin (treated with TPCK) was obtained from Worthington Biochemical Corp. Protease from *Staphylococcus aureus* was a gift from Dr. G. R. Drapeau. Carboxypeptidase Y was a gift from Dr. M. Ottesen.

† From the Department of Biochemistry, University of Washington, Seattle, Washington 98195. *Received October 10, 1977.* This work was supported by research grants from the National Institutes of Health (GM-15731) and the American Cancer Society (BC91-P), and by a Grant-in-Aid for Scientific Research (No. 137017) and Cancer Research (No. 101008) from the Ministry of Education, Science and Culture, Japan.

‡ R.G.W. is a recipient of a postdoctoral award from the National Institutes of Health (AM 05144).

§ Department of Enzyme Chemistry, Tokushima University, School of Medicine, Tokushima 770, Japan.

¶ This work was supported in part by U.S. Public Health Service Grants GM-20066 and GM-15731 and by National Institutes of Health postdoctoral award GM 05972 to W.F.A.

‖ Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403.

Department of Biochemistry, University of Washington, Seattle, Washington 98195.

[1] Abbreviations used: GSP, group-specific protease; DFP, diisopropyl phosphofluoridate; DIP, diisopropyl phosphoryl; TPCK, tosylphenylalanyl chloromethyl ketone; Ptc, phenylthiocarbamoyl; Pth, phenylthiohydantoin; HPLC, high-pressure liquid chromatography.

0006-2960/78/0417-0811$01.00/0

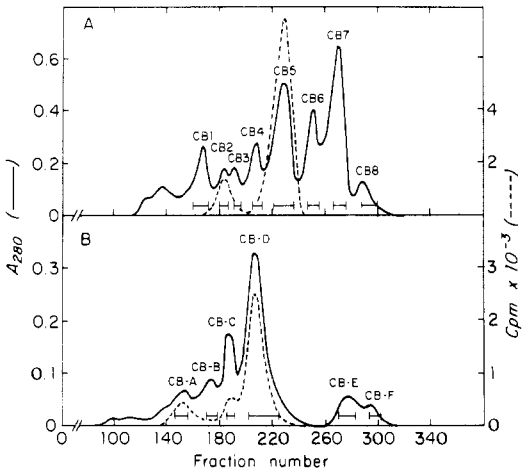

FIGURE 1: Gel filtration of fragments, generated by cleavage of GSP with CNBr, on a 2.5 × 100 cm column of Sephadex G-50 Super Fine equilibrated with 9% formic acid. Approximately 1.5-mL fractions were collected at a rate of 23 mL per h. Radioactivity, due to [^{14}C]DIP, was estimated in 50-μL aliquots of each fraction. Fractions were pooled as indicated by bars (|—|) under the peaks. (A) Elution profile of fragments from the digestion of approximately 2 μmol of pyridylethylated GSP. (B) Elution profile of fragments from the digestion of approximately 0.5 μmol of GSP with intact disulfide bonds.

[^{14}C]Diisopropyl phosphorofluoridate (DFP) was purchased from New England Nuclear. Methanesulfonic acid (4 N) containing 0.2% 2-aminoethylindole was obtained from Pierce Chemical Co. 4-Vinylpyridine (Aldrich Chemical Co.) was further purified by distillation. All other reagents were of the highest grade available from commercial sources.

Methods. In most experiments, a portion of GSP was inactivated by [^{14}C]DFP as described by Robinson (1971). The presence of the labeled DIP group in approximately 5% of the enzyme provided a marker for locating fragments containing the putative seryl residue of the active site during subsequent purification steps. Reduction and pyridylethylation of the protease were carried out as described by Hermodson et al. (1973).

Large fragments of the protease molecule were produced by adding solid cyanogen bromide (200 mg) to solutions of GSP (1–2 μmol) in 5 mL of 70% formic acid. The cleavage reaction was allowed to proceed for 18 to 24 h at room temperature. The fragments were separated by gel filtration on a 2.5 × 100 cm column of Sephadex G-50 Super Fine equilibrated with 9% formic acid. Two fragments of the enzyme (CB5A and CB5B) were further purified by ion-exchange chromatography on a 1 × 30 cm column of SP-Sephadex C-25 equilibrated with 7 M urea and 0.1 M sodium formate (pH 2.9). The fragments were eluted from the column by applying a linear gradient (0.1 M to 0.6 M) of sodium formate, pH 2.9. Salt and urea were removed from the desired pools of fractions by gel filtration on Sephadex G-25 equilibrated with 9% formic acid.

Amino acid analyses were performed on a Durrum Model D-500 amino acid analyzer. Duplicate samples were hydrolyzed in 6 N HCl at 110 °C for 24, 48, 72, and 96 h. Serine and threonine concentrations were determined by extrapolating to zero hydrolysis time. Tryptophan was determined after hydrolysis of samples in 4 N methanesulfonic acid (Liu & Chang, 1971). Half-cystine was measured as *S*-pyridylethylcysteine (Hermodson et al., 1973), or as cysteic acid following performic acid oxidation of the protein by the method of Hirs (1967).

Automated sequence analysis was performed on the Beckman sequencer Model 890B by the method of Edman & Begg (1967) as modified by Hermodson et al. (1972). Frequently, a peptide program of Crewther & Inglis (1973) was adapted to the dimethylbenzylamine system. Occasionally, special precautionary steps were taken. Thus, when a fragment contained a glutaminyl residue, which cyclizes rapidly in acid, the time allowed for the preceding acid cleavage step was decreased to 30 s. When a fragment contained a prolyl–aliphatic residue bond, which lowers the stepwise yield, the temperature was increased by 1.5 °C to 59 °C in order to increase the efficiency of cleavage of the Ptc-prolyl residue (Brandt et al., 1976).

Solid-phase sequence analysis was carried out on a Sequemat, Model 12, essentially as described by Laursen (1975). Several of the CNBr fragments were covalently attached to an insoluble matrix through their carboxyl-terminal homoseryl residues according to the method of Horn & Laursen (1973). Peptides from enzymatic digests of CNBr fragments were selectively attached to resin through the homoseryl residue without prior purification.

Pth-amino acids were identified by gas- and high-pressure liquid chromatography (HPLC). For analysis by HPLC, an aliquot of sample, dissolved in methanol, was injected into a μ-Bondapack C-18 column (Waters Associates) and fractionated by applying a linear methanol gradient (14% to 56%) in aqueous buffer (pH 4.18) pumped by a Waters Associates System Model 6000A at a rate of 2.2 mL/min (Bridgen et al., 1976).

The location of the [^{14}C]DIP group in one of the fragments was determined during sequence analysis by monitoring approximately 5% of each 1-chlorobutane extract for radioactivity in a Packard Model 3003 liquid scintillation spectrometer.

The amino acid sequence of GSP was optimally aligned with those of other serine proteases using an 8K, PDP-12 computer with an oscilloscope for editing purposes (De Haën et al., 1975).

Cyanogen bromide fragments were subdigested for 6 to 8 h either with TPCK-treated trypsin (enzyme to substrate ratio, 1/100 w/w) in 0.1 M ammonium bicarbonate (pH 7.8) at 37 °C, or with Staphylococcal protease (enzyme to substrate ratio, 1/24 w/w) in 0.1 M pyridine–acetate (pH 4.0) at 37 °C. Fresh portions of enzyme were added periodically to the digests to ensure that the desired peptides were generated in high yield. Succinylation of fragments prior to digestion by enzymes was carried out as described by Hermodson et al. (1973).

Digestion of pyridylethylated enzyme with carboxypeptidase Y (in 6 M urea), or of the CNBr fragment originating from the carboxyl-terminal region of the enzyme, was carried out in 0.1 M pyridine–acetate (pH 6.0) at 37 °C, using an enzyme to substrate ratio of 1/50 (w/w). Norleucine was included in the digest as an internal standard.

Results

Purification of Fragments from the CNBr Digest of GSP. Amino acid analysis of GSP reveals 5 methionyl residues per molecule (Table I). Hence, 6 major fragments should be generated by digestion of pyridylethylated GSP with CNBr. These fragments were fractionated by gel filtration on Sephadex G-50, equilibrated with 9% formic acid, as shown in Figure 1A. Four of the pooled fractions (CB1, CB6, CB7, and CB8) contained a single polypeptide as judged by sequenator analyses. Approximately 80% of the total radioactivity was recovered in pool CB5 and 5% to 10% in pool CB2.

Pool CB5 was further fractionated on a column of SP-

TABLE I: Amino Acid Compositions[a] of the Major Fragments Generated from the Digestion of Pyridylethylated GSP by CNBr.

Amino acid	Fragment CB1 AAA	CB1 SEQ	CB5A AAA	CB5A SEQ	CB5B AAA	CB5B SEQ	CB6 AAA	CB6 SEQ	CB7 AAA	CB7 SEQ	CB8 AAA	CB8 SEQ	Total AAA	Total SEQ	GSP
Asp	5.0	5	2.0	2	4.0	4	2.0	2	1.0	1	0.0	0	14	14	14.0
Thr[b]	4.0	4	0.8	1	3.0	3	2.0	2	2.8	3	0.0	0	13	13	12.5
Ser[b]	4.1	4	2.1	2	3.3	3	1.1	1	0.9	1	1.8	2	13	13	13.0
Glu	8.8	9	2.3	2	0.2	0	3.0	3	1.9	2	1.1	1	17	17	17.2
Pro	1.1	1	5.0	5	5.0	5	1.2	1	1.1	1	2.1	2	15	15	14.8
Gly	5.4	5	0.9	1	5.8	6	1.3	1	2.8	3	2.1	2	18	18	18.4
Ala	4.2	4	2.0	2	5.0	5	3.0	3	1.8	2	0.0	0	16	16	16.0
Val[c]	6.7	7	4.3	4	4.8	5	3.0	3	1.8	2	1.1	1	22	22	21.6
Ile[c]	7.9	9	0.8	1	3.7	4	0.1	0	1.2	1	2.8	3	17	18	16.8
Leu	6.1	6	5.0	5	2.3	2	1.2	1	1.8	2	0.0	0	16	16	15.6
Tyr	1.1	1	0.0	0	1.8	2	3.7	4	1.2	1	0.9	1	9	9	8.8
Phe	2.0	2	0.9	1	1.1	1	1.8	2	0.0	0	0.0	0	6	6	6.1
His	4.8	5	0.9	1	1.9	2	0.0	0	0.0	0	0.9	1	9	9	8.8
Lys	5.9	6	2.7	3	1.1	1	2.1	2	1.0	1	0.0	0	13	13	12.7
Arg	4.9	5	0.0	0	0.8	1	2.0	2	2.7	3	1.1	1	12	12	12.0
Trp[d]	0.0	0	0.0	0	0.0	0	0.0	0	1.6	2	0.0	0	2	2	1.7
Cys/2[e]	1.6	2	0.0	0	0.8	1	1.7	2	0.8	1	0.0	0	6	6	5.8
Homoser[f]/ Met	0.5	1	0.5	1	0.0	0	0.6	1	0.5	1	0.5	1	5	5	4.7
Total residues		76		31		45		30		27		15	223	224	
Mol wt		8570		3323		4516		3533		3078		1635		24 655	

[a] Duplicate samples were analyzed after 24, 48, 72, and 96 h of hydrolysis in 6 N HCl at 110 °C. AAA: compositions obtained by amino acid analysis. The AAA composition of CB5A was obtained by subtracting the compositions of CB5B from that of pool CB5. Total AAA composition was obtained after the values of each fragment were rounded off to nearest integer. SEQ: compositions indicated from sequence. [b] Determined by extrapolating to zero-hydrolysis time. [c] Values were obtained from analysis of 96-h hydrolysates. [d] Determined after 24-h hydrolysis in 4 N methanesulfonic acid at 110 °C. [e] Half-cystine determined as *S*-pyridylethylcysteine. [f] Values are not corrected for homoserine lactone.

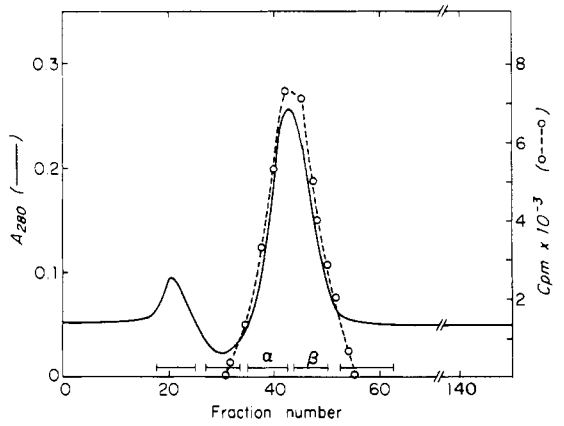

FIGURE 2: Ion-exchange chromatography of pool CB5 (Figure 1A) on a 1 × 30 cm column of SP-Sephadex C-25 equilibrated with 7 M urea and 0.1 M sodium formate (pH 2.9). A linear salt gradient was applied starting with 100 mL of 0.1 M and 100 mL of 0.6 M sodium formate. Approximately 1.1-mL fractions were collected at a rate of 30 mL per h. Radioactivity was estimated in 50-μL aliquots of each fraction. Fractions were pooled as indicated by bars (|—|). Pools of fractions were desalted on Sephadex G-25 equilibrated with 9% formic acid and then submitted for amino acid analysis. Pool α contained only fragment CB5B; pool β contained equimolar amounts of fragments CB5A and CB6B.

Sephadex C-25, equilibrated with 7 M urea, by applying a linear gradient of sodium formate (0.1 to 0.6 M) as shown in Figure 2. Although a single symmetrical peak (fragment CB5B) was observed by absorbance measurements (280 nm) and by analysis for radioactivity, another fragment (CB5A) lacking tryptophan, tyrosine, and *S*-pyridylethylcysteine was detected by determining the compositions of various pooled fractions (Figure 2). The two fragments were not completely resolved, but the leading half of the major peak contained a single fragment (CB5B), as judged by sequenator analysis. The trailing half had an amino acid composition identical with that of the sample applied to the column, and consisted of equal amounts of fragments CB5A and CB5B.

Fragment CB5B lacked homoserine and was therefore assigned to the carboxyl terminus of the enzyme molecule. Its amino-terminal sequence is Gly-Asp-Ser-Gly-Gly-Pro, and approximately 70% of the radioactivity accompanied the seryl residue. The remaining radioactivity was detected at uniformly low levels throughout the first 10 cycles of the analysis.

For sequence analysis, fragment CB5A was purified by exploiting the reactivity of homoserine lactone toward amino groups (Horn & Laursen, 1973). Thus fragment CB5A was selectively attached to an insoluble support containing amino groups, whereas fragment CB5B lacking homoserine remained in solution.

Amino Acid Compositions of the CNBr Fragments. The amino acid compositions of the major CNBr fragments and of the whole enzyme are shown in Table I. Cysteic acid analyses of hydrolyzates of the performic acid oxidized enzyme were consistent with values for half-cystine determined as *S*-pyridylethylcysteine.

The amino acid composition was referred to a molecular weight of 25 200 as estimated by gel electrophoresis of the enzyme in the presence of sodium dodecyl sulfate and dithiothreitol, using the method of Weber & Osborne (1969). The minimum molecular weights of the CNBr fragments, based on their amino acid compositions, are: 8570, 4516, 3533, 3323, 3078, and 1635 for fragments CB1, CB5B, CB6, CB5A, CB7, and CB8, respectively. The combined molecular weight of these fragments is 24 655, and the sum of their amino acid compositions is in agreement with that of GSP.

Alignment of the CNBr Fragments. Three of the six major CNBr fragments of GSP could be placed directly in the se-

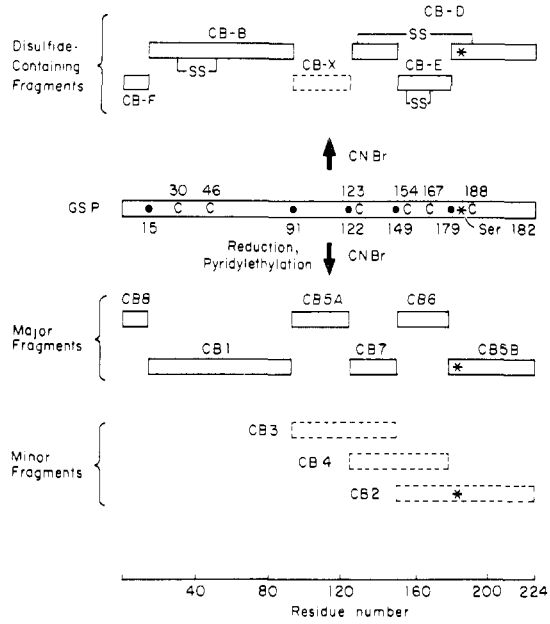

FIGURE 3: Diagrammatic summary depicting the origin of the fragments generated from GSP or pyridylethylated GSP by digestion with CNBr. Locations of the methionyl residues are indicated as black dots (●) and their position numbers are shown below the bar. Locations of half-cystine residues are indicated as C, and their position numbers are shown above the bar. The fragments were purified as shown in Figure 1. Asterisk (*) indicates the location of [^{14}C]DIP in the GSP molecule. No attempt was made to isolate fragment CB-X.

TABLE II: Amino Acid Compositions[a] of Minor CNBr Fragments.

Amino acid	mol of amino acid/mol of fragment					
	CB2	CB6[c] plus CB5B	CB3	CB7[c] plus CB5A	CB4	CB7[c] plus CB6
Asp	5.4	5.6	3.4	3.4	3.5	3.2
Thr	4.4	4.5	4.0	3.9	4.5	4.6
Ser	3.8	3.9	3.7	3.4	2.5	2.0
Glu	3.7	3.1	4.0	4.0	4.6	5.0
Pro	5.7	5.7	5.2	5.2	1.7	2.0
Gly	6.6	7.0	5.3	3.8	5.0	3.9
Ala	6.7	7.7	4.9	3.7	5.1	4.9
Cys/2[b]	ND	2.5	ND	0.8	ND	2.4
Val	6.9	7.4	6.2	6.0	5.2	5.0
Ile	3.4	3.1	2.8	2.4	1.8	1.0
Leu	3.8	3.3	5.6	5.8	4.0	3.0
Tyr	4.4	5.5	2.1	2.0	3.7	4.8
Phe	2.6	3.0	1.3	1.0	2.0	2.0
His	2.0	2.0	1.4	1.0	0	0
Lys	3.3	3.2	3.6	3.8	3.2	3.2
Arg	2.8	3.0	3.0	2.7	4.2	4.8

[a] Duplicate samples were analyzed after 24 h of hydrolysis in 6 N HCl. [b] Determined as *S*-pyridylethylcysteine. Amino-terminal sequence analyses of fragments CB2, CB3, and CB4 indicated the presence of contaminating peptides in low amounts (10–15% total). [c] From Table I. ND, not determined.

quence. Two were aligned by amino-terminal analysis of whole GSP (30 cycles) which placed CB8 at the amino terminus, followed by fragment CB1. The third fragment (CB5B) was placed at the carboxyl terminus by its lack of homoserine.

The other three major fragments (CB5A, CB7, and CB6) were aligned on the basis of analyses of three minor overlapping fragments (CB2, CB3, and CB4) which resulted from incomplete cleavage of methionyl bonds by CNBr (Figure 3). When pools of each minor peak were collected within narrow limits, the amino acid compositions of the overlapping fragments (Table II) are consistent with the following combinations: CB6 plus CB5B equals CB2; CB5A plus CB7 equals CB3; and CB7 plus CB6 equals CB4. The composition of fragment CB2 provides an explanation for its radioactivity (Figure 1A), since this fragment contains the seryl residue of the active site.

Sequence analyses revealed that each of these pools contained a single major sequence, accounting for at least 70% of the total material. The amino-terminal sequence of fragment CB2 is Asp-Glu-Lys-Ala-Cys-Val, as is that of fragment CB6. Thus, considering the amino acid compositions (Table II) fragment CB2 must consist of fragment CB6 followed by fragment CB5B. Similarly, fragment CB3 having an amino-terminal sequence of Leu-Leu-Lys-Leu-Glu-Lys, consists of fragment CB5A followed by fragment CB7. Finally, fragment CB4 with an amino-terminal sequence of Cys-Trp-Ala-Ala-Gly-Trp, contains fragment CB7 followed by fragment CB6. On the basis of these conclusions the fragments are aligned in the order: CB8-CB1-CB5A-CB7-CB6-CB5B. The molecular weights of the minor fragments, based on their amino acid compositions, are consistent with their relative elution position in Figure 1A.

Sequence Analysis of the Major CNBr Fragments. Sequence analyses of the major fragments of intermediate size (CB5A, CB6, CB7, and CB8) were facilitated by their carboxyl-terminal homoseryl residues which could be specifically attached to solid supports. Each subsequent degradation proceeded through the carboxyl-terminal homoserine; the products were analyzed by HPLC and confirmed by gas–liquid chromatography. Typically, the yield of attachment to the insoluble support was of the order of 20% to 30%. Most sequence analyses were performed on approximately 100 nmol of attached fragment. The stepwise degradation yield averaged 90% for these analyses.

Although fragment CB1 also contained a carboxyl-terminal homoseryl residue, this fragment (76 residues) was too large for complete solid-phase sequence analysis. Extended sequence analyses were obtained, however, in the Beckman Sequencer using a peptide program. Optimum degradation was achieved by reducing the time allowed for the acid cleavage step at position 22, which increased the stepwise yield at the glutaminyl residue at position 23 from 50% to approximately 90%. Three extended analyses, using approximately 400 nmol of fragment CB1 in each case provided an unambiguous sequence of 52 residues. The remaining 24 residues in the carboxyl-terminal region of fragment CB1 were then placed by solid-phase sequence analyses of homoseryl peptides obtained from subdigests of fragment CB1 with either trypsin or staphylococcal protease. In each digest, the homoseryl peptide was selectively attached to an insoluble support without further purification. The results of these analyses are summarized in Figure 4. Since sequencer analyses of intact CB1 had already placed all five arginyl residues among the first 48 residues, a fragment representing residues 49–76 was generated by succinylating the lysyl residues of 500 nmol of fragment CB1 and cleaving at arginyl residues with trypsin. The homoserine-containing peptide (designated TS-1 in Figure 4) was selectively attached to resin without further purification. The overall yield of attachment was 15%. Sequence analysis revealed a single peptide with an amino-terminal sequence corresponding to residues 49 to 52 of intact fragment CB1. Peptide TS-1 contained three glutaminyl residues among the first 12, which significantly lowered the overall stepwise yield. For this reason, the sequence analysis was terminated after 16 cycles. The remaining three succinylated lysyl residues were placed (by HPLC) within this

TABLE III: Amino Acid Compositions[a] of CNBr Fragments from GSP with Intact Disulfide Bonds.

Amino acid	mol of amino acid/mol of fragment							
	CB-B[b]	CB1[c]	CB-D[b]	CB5B[c] plus CB7	CB-E[b]	CB6[c]	CB-F[b]	CB8[c]
Asp	5.2	5.2	4.8	5.1	2.0	2.1	0	0
Thr	4.0	3.9	5.2	5.3	2.0	2.0	0	0
Ser	4.0	3.9	3.9	4.1	1.2	1.1	1.6	1.8
Glu	8.9	9.2	2.3	2.1	3.1	3.1	1.1	1.1
Pro	1.6	1.6	5.8	5.9	1.1	1.1	1.7	1.7
Gly	5.6	5.7	8.6	8.4	1.4	1.2	1.7	2.0
Ala	4.4	4.3	6.4	6.9	3.0	3.0	0	0
Cys/2[d]	ND	1.4	ND	1.7	ND	1.6	ND	0
Val	6.1	6.2	6.3	6.6	2.9	3.1	1.1	1.0
Ile	6.7	6.9	4.3	4.2	0	0	1.7	2.0
Leu	6.0	6.0	3.9	3.9	1.0	1.3	0	0
Tyr	1.1	1.0	2.8	2.7	3.4	3.7	1.0	0.8
Phe	2.0	2.0	1.1	1.1	2.0	2.0	0	0
His	4.7	4.8	1.9	2.0	0	0	0.8	1.0
Lys	6.0	5.8	2.0	2.0	1.9	2.2	0	0
Arg	5.7	4.9	3.5	3.7	2.0	2.1	1.2	1.2
Hse[e]	0.5	0.5	0.6	0.4	0.5	0.5	0.5	0.6

[a] Duplicate samples were analyzed after 24 h of hydrolysis in 6 N HCl. [b] Generated from GSP with intact disulfide bonds. [c] Generated from *S*-pyridylethyl-GSP. [d] Determined as *S*-pyridylethylcysteine. [e] Values are not corrected for homoserine lactone. ND, not determined. Table does not include all of the fragments generated in CNBr digests of GSP.

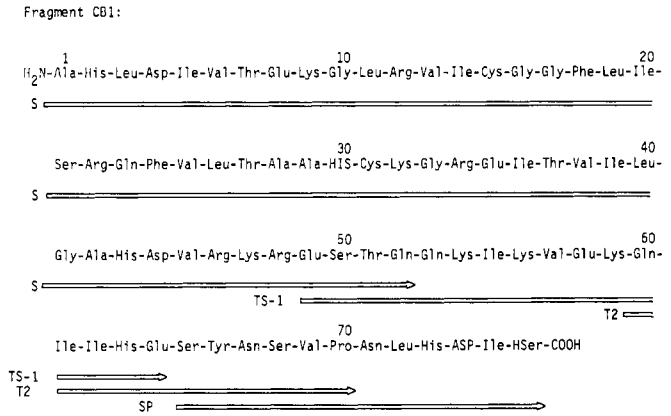

FIGURE 4: Summary of the sequence analysis of fragment CB1. Analyses performed in the Beckman sequencer are indicated by S. The peptides generated from the subdigestion of fragment CB1 and analyzed by solid-phase sequence methods, are: TS-1, generated from succinylated CB1 by tryptic digestion; T2, generated by tryptic digestion of fragment CB1; SP, generated from the digestion of fragment CB1 by the protease from *Staphylococcus aureus*.

region. Fragment CB1 (100 nmol) was then digested by trypsin without prior succinylation, and the peptide containing the homoseryl residue was selectively removed from the digest for solid-phase sequence analysis. A peptide (T2) was attached, but in apparently low yield (15%) as judged by subsequent degradation, probably reflecting both cyclization of the amino-terminal glutamine and difficulty in digesting fragment CB1, which was largely insoluble at pH 8. Nonetheless, analysis did identify 11 residues, 5 of which overlapped the proven sequence of peptide TS-1, and the other six extended the sequence to residue 70. The remaining sequence of fragment CB1 was obtained by sequence analysis of a homoseryl peptide (SP) obtained after digestion of fragment CB1 (100 nmol) by Staphylococcal protease. This peptide was attached in 25% overall yield. Sequence analysis proceeded through the carboxyl-terminal homoseryl residue, thereby completing the analysis of fragment CB1.

Fragment CB5B was not suitable for solid-phase sequence analysis since it lacked homoserine. During preliminary sequence analysis using the sequencer, the stepwise yield averaged only 90% through the first 11 residues. This low efficiency was due to poor release (approximately 50%) of a Ptc-prolyl residue in the sequence Pro-Leu at positions 6 and 7. Poor release of Ptc-proline is not uncommon when that residue is followed by an aliphatic amino acid residue (Hermodson et al., 1972). Cleavage was improved to 85% by increasing the temperature by 1.5 °C during the reactions with the prolyl residue (Brandt et al., 1976). The temperature was then returned to normal for the duration of the sequence analysis. Under these conditions the overall stepwise yield was raised to 95% through the first 11 residues. A single sequence analysis provided data which unambiguously placed all 45 residues of fragment CB5B. This was possible, in part, due to the increased efficiency of the analysis as described above. The stepwise yield was 95% for the first 11 residues, and 92% for the remaining 34 residues. Approximately 210 nmol of Pth-glycine was released at position 1 of fragment CB5B, and 14 nmol of the carboxyl-terminal Pth-asparagine was obtained in the last cycle of the analysis.

The sequence of the carboxyl-terminal region of fragment CB5B (Ala-Val-Ile-Asn-COOH) was supported by digestion of the fragment with carboxypeptidase Y (Figure 5). Pyridylethylated GSP was similarly digested by carboxypeptidase Y. The results indicated the sequence (Ala, Val)-Ile-Asn-COOH at the carboxyl terminus of GSP, thus confirming that fragment CB5B originated from that portion of the protein molecule.

The Location of the Disulfide Bonds in the GSP Molecule. Since GSP has been shown to lack sulfhydryl groups (Katunuma et al., 1975), the six half-cystinyl residues must form three disulfide bonds. In order to determine their locations, [^{14}C]DIP enzyme was digested with CNBr. The resulting fragments were purified by gel filtration on Sephadex G-50 Super Fine, equilibrated with 9% formic acid, as shown in Figure 1B. Fractions were pooled as indicated under each peak, and the amino acid compositions of material in these pools were compared (Table III) with those of the fragments obtained from pyridylethylated GSP (Figure 1A).

TABLE IV: Alignment of the Serine Proteases.[a]

```
                1                   20                  40
GSP             I I G G V E S I P H S R P Y M A H L D I V T E K G L R V I C G G F L I S R Q F V L T A A H C K
E PIG           V V G G T E A Q R N S W P S Q I S L Q Y R S G S S W A H T C G G T L I R Q N W V M T A A H C V
               16      20                  40
CA COW          I V N G E E A V P G S W P W Q V S L Q D K T G F - - - H F C G G S L I N E N W V V T A A H C G
CB COW          I V N G E D A V P G S W P W Q V S L Q D S T G F - - - H F C G G S L I S E D W V V T A A H C G
Tr COW          I V G G Y T C G A N T V P Y Q V S L - - N S G Y - - - H F C G G S L I N S Q W V V S A A H C Y
Tr PIG          I V G G Y T C A A N S I P Y Q V S L - - N S G S - - - H F C G G S L I N S Q W V V S A A H C Y
Tr DOGFISH      I V G G Y E C P K H A A P W T V S L - - N V G Y - - - H F C G G S L I A P G W V V S A A H C Y
Factor X COW    I V G G R D C A E G E C P W Q A L L V - N E E N E - - G F C G G T I L N E F Y V L T A A H C L
TH COW          I V E G Q D A E V G L S P W Q V M L F R K S P Q E - - L L C G A S L I S D R W V L T A A H C L

                                                    60                                                  80
GSP             - - - - - - - - - - G R E I T V I L - - G A H D V R K R E S T Q Q K I K - V E K Q I I H E S Y
E PIG           - - - - - - - - - - D R E L T F R V V V G E H N L N Q N N G T E Q Y V G - V Q K I V V H P Y W
                                  60                                                  80
CA COW          - - - - - - - - - - V T T S D V - V V A G E F D Q G S S S E K I Q K L K - I A K V F K N S K Y
CB COW          - - - - - - - - - - V T T S D V - V V A G E F D Q G L E T E D T Q V L K - I G K V F K N P K F
Tr COW          - - - - - - - - - - K S G I Q V R L - - G Q D N I N V V E G N E Q F I S - A S K S I V H P S Y
Tr PIG          - - - - - - - - - - K S R I Q V R L - - G E H N I D V L E G N E Q F I N - A A K I I T H P N F
Tr DOGFISH      - - - - - - - - - - Q R R I Q V R L - - G E H D I S A N E G D E T Y I D - S S M V I R H P N Y
Factor X COW    - - - - - - - - H Q A K R F T V R V - - G D R N T - - Q E G D E E M A H E V E M T V K H S R F
TH COW          L Y P P W D K N F T V D D L L V R I - - G K H S R T R Y E R K V E K I S M L D K I Y I H P R Y

                                                100                                                 120
GSP             N - S V P N L - - H D I M L L K L E K K V E L T P A V N V V P L P - S - - P S D F I H P G A M
E PIG           N - T D D V A A G Y D I A L L R L A Q S V T L N S Y V Q L G V L P - R - - A G T I L A N N S P
                      100                                                 120
CA COW          N - S L T I N - - N D I T L L K L S T A A S F S Q T V S A V C L P - S - - A S D D F A A G T T
CB COW          S - I L T V R - - N D I T L L K L A T P A Q F S E T V S A V C L P - S - - A D E D F P A G M L
Tr COW          N - S N T L N - - N D I M L I K L K S A A S L N S R V A S I S L P - T S C A S - - - - A G T Q
Tr PIG          N - G N T L D - - N D I M L I K L S S P A T L N S R V A T V S L P - R S C A A - - - - A G T E
Tr DOGFISH      S - G Y D L D - - N D I M L I K L S K P A A L N R N V D L I S L P - T G C A Y - - - - A G E M
Factor X COW    V - K E T Y D - - F D I A V L R L K T P I R F - R N V A P A C L P E K D W A A E T L Q T - K Y
TH COW          N W K E N L D - - R D I A L L K L K R P I E L S D Y I H P V C L P - K Q T A A K L L H A G F K

                                                140                                                 160
GSP             C W A A G W G K - T G V R D P T - S Y T L R E V E L R I M D E K A C V D Y R Y Y E Y K F - - -
E PIG           C Y I T G W G L - T R T N G Q L - A Q T L Q Q A Y L P T V D Y A I C S S S S Y W G S T V K N S
                      140                                                 160
CA COW          C V T T G W G L - T R Y T N A N T P D R L Q Q A S L P L L S N T N C K K - - Y W G T K I K D A
CB COW          C A T T G W G K - T K Y N A L K T P D K L Q Q A T L P I V S N T D C R K - - Y W G S R V T D V
Tr COW          C L I S G W G N - T K S S G T S Y P D V L K C L K A P I L S N S S C K - S A Y P G - Q I T S N
Tr PIG          C L I S G W G N - T K S S G S S Y P S L L Q C L K A P V L S D S S C K - S S Y P G - Q I T G N
Tr DOGFISH      C L I S G W G N - T M D G A - V S G D Q L Q C L D A P V L S D A E C K - G A Y P G - M I T N N
Factor X COW    G I V S G F G - R T H E K G R L - S S T L K M L E V P Y V D R S T C K L S S S F - - T I T P N
TH COW          G R V T G W G N R T T S V A E V Q P S V L Q V V N L P L V E R P V C K A S T R I - - R I T D D

                                                180
GSP             Q V C V G S - - P T T L R - A A F M G D S G G P L L C - - - A G - - - V - - A H G I V S Y G -
E PIG           M V C A G - - - G N G V R - S G C Q G D S G G P L H C - L V N G - - - Q Y A V H G V T S F V S
               180                                                 200
CA COW          M I C A G - - - A S G V - - S S C M G D S G G P L V C - K K N G - - - A W T L V G I V S W G S
CB COW          M I C A G - - - A S G V - - S S C M G D S G G P L V C - Q K N G - - - A W T L A G I V S W G S
Tr COW          M F C A G Y - - L E G G K - D S C Q G D S G G P V V C - - - S G - - - K - - L Q G I V S W G S
Tr PIG          M I C V G F - - L E G G K - D S C Q G D S G G P V V C - - - N G - - - Q - - L Q G I V S W G Y
Tr DOGFISH      M M C B G Y - - M E G G K - D S C Q G D S G G P V V C - - - N G - - - M - - L Q G I V S W G Y
```

TABLE IV: (*Continued*)

Factor X COW	M F C A G Y - P D T G G V - D T C Q G D S G G P - - M F R K D N - A D E W I Q V G I V S W G Y
TH COW	M F C A G Y K P G E G K R G D A C E G D S G G P F V M - - K S P Y N N R W Y Q M G I V S W G E
	200 220
GSP	- - - - H P D A K P P A I F T R V S T Y V P T I N A V I N - - - - - - - - - - - - - - - - - -
E PIG	R L G C N V T R K - P T V F T R V S A Y I S W I N N V I A S N - - - - - - - - - - - - - - - -
	220 240
CA COW	S - T C S - T S T - P G V Y A R V T A L V N W V Q Q T L A A N - - - - - - - - - - - - - - - -
CB COW	S - T C S - T S T - P A V Y A R V T A L M P W V Q E T L A A N - - - - - - - - - - - - - - - -
Tr COW	- - G C A Q K N K - P G V Y T K V C N Y V S W I K Q T I A S N - - - - - - - - - - - - - - - -
Tr PIG	- - G C A Q K N K - P G V Y T K V C N Y V N W I Q Q T I A A N - - - - - - - - - - - - - - - -
Tr DOGFISH	- - G C A E R D H - P G V Y T R V C H Y V S W I H E T I A S V - - - - - - - - - - - - - - - -
Factor X COW	- - G C A R K G K - F G V Y T K V S N F L K W I D K I M K A R A G - - - - - - (21 residues) L
TH COW	- - G C D R N G K - Y G F Y T H V F R K L K W I Q K V I D R L G S - - - - - - - - - - - - - -

[a] The sequences of these serine proteases, other than that of GSP, were compiled by de Haën et al. (1975). E pig, porcine elastase; CA cow, bovine chymotrypsin A; CB cow, bovine chymotrypsin B; Tr cow, bovine trypsin; Tr pig, porcine trypsin; Tr dogfish, dogfish trypsin; Th cow, bovine thrombin. The one letter code for amino acid residues is: A = Ala, B = Asx, C = Cys, D = Asp, E = Glu, F = Phe, G = Gly, H = His, I = Ile, K = Lys, L = Leu, M = Met, N = Asn, P = Pro, Q = Gln, R = Arg, S = Ser, T = Thr, V = Val, W = Trp, Y = Tyr.

The amino acid composition of fragment CB-B is identical with that of fragment CB1. Further, sequence analysis of fragment CB-B placed Ala-His-Leu at the first 3 positions as in fragment CB1. Thus, these two fragments originate from the same region of the GSP molecule, and their two half-cystinyl residues must be joined by a disulfide bond (residues 30 and 46). Similarly, the amino acid composition and the amino-terminal sequence (Asp-Glu-Lys-Ala) of fragment CB-E are identical with those of fragment CB6. Fragments CB6 and CB-E, therefore, represent the same region of the protein molecule, and their two half-cystinyl residues must form a single disulfide bond (residues 154 and 167). The amino acid composition of fragment CB-D is consistent with the conclusion that the third disulfide bond in the enzyme molecule links the half-cystinyl residues 123 and 188 in fragments CB7 and CB5B, respectively. Further, fragments corresponding to CB7 and CB5B (Figure 1A) were replaced by a single fragment, CB-D, of higher molecular weight (Figure 1B) in accord with an interfragment disulfide bond. The locations of these three disulfide bonds are shown in Figure 3.

The complete covalent structure of GSP is shown in Figure 6. The enzyme molecule consists of 224 residues, and the labeled seryl residue is at position 182. The residues assumed to participate in the charge-relay system of the active site are shown in capital letters (HIS-46, ASP-89, SER-182).

The Evolutionary Relationship of GSP to Other Serine Proteases. The amino acid sequence of the GSP molecule was compared with the sequences of the other known serine proteases (Table IV) with the aid of computer analysis (de Haën et al., 1975). This comparison provides an optimum alignment which minimizes the number of necessary gaps (insertions or deletions) along the sequences. The computer analysis is based on several scoring systems which evaluate compatibility of different amino acid residues. Elementary dendrograms were constructed using the appropriate inferred gaps, and the resulting algorithms were solved in order to ascertain the correct topology. The resulting evolutionary relationship of the serine proteases (Figure 7) is derived solely on the basis of gap events in the sequences. This evolutionary tree can be considered only in relative terms, since the constructed dendrograms were trunkless and thus lack a coordinate of time.

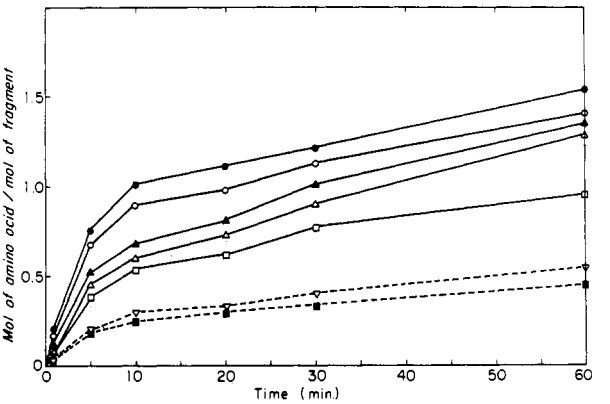

FIGURE 5: Time course of the digestion of fragment CB5B by carboxypeptidase Y. Digestion was carried out in 0.1 M pyridine acetate (pH 6.0) at 37 °C, using an enzyme to substrate ratio of 1/50 (w/w). Norleucine was included in the digest as an internal standard. Asn (●—●), Ile (○—○), Val (▲—▲), Thr (△—△), Ala (□—□), Pro (▽—▽), and Tyr (■- - -■). The high Thr levels compared with Ala reflect the presence of two residues of Thr and only one residue of Ala in the carboxyl-terminal region of the fragment.

Discussion

The covalent structure of "group-specific" protease from rat intestine presented in this paper clearly establishes the homologous relationship of this enzyme to the mammalian serine proteases. When the protein sequences are optimally aligned (Table IV), it is found that the degree of sequence identity with bovine chymotrypsin, trypsin, and porcine elastase is 33, 34, and 34%, respectively, and 30% when the sequences of GSP and bovine thrombin are compared. In addition, the residues forming the "charge-relay" system of the active site in serine proteases (His-57, Asp-102, and Ser-195 in bovine chymotrypsin A) are found in corresponding regions of the GSP sequence. As in all other serine proteases, the sequences around these residues are highly conserved in GSP. Since the enzyme contains an amino-terminal isoleucine and an aspartic acid at position 181 (Asp-194 in chymotrypsin), it is likely that these residues form an ion pair as in chymotrypsin.

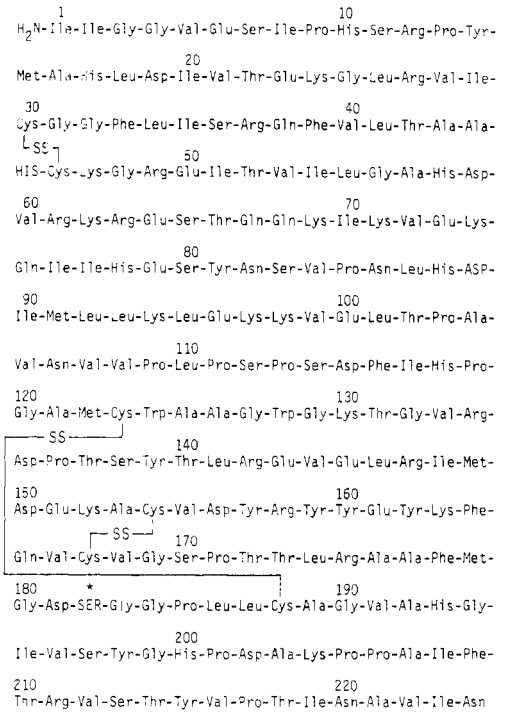

FIGURE 6: The complete covalent structure of GSP. Components of the "charge-relay" system of the active site (HIS-45, ASP-89, and SER-182) are in capital letters. Asterisk (*) indicates the location of the [^{14}]DIP group. The disulfide bonds are indicated as -SS-.

In addition, the sequence Ser-Tyr-Gly-199 in GSP corresponds to the region Ser-Trp-Gly-216 in chymotrypsin. This sequence is thought to form hydrogen bonds with the P_2 and P_3 regions of tripeptide substrates in an antiparallel β-pleated sheet arrangement (Segal et al., 1971).

The three disulfide bonds of GSP occur in analogous positions in chymotrypsin, trypsin, and elastase. They link Cys-42 to Cys-58 (the "His-loop"), Cys-136 to Cys-201, and Cys-168 to Cys-182 (the "Met-loop") in chymotrypsin. These structures are highly conserved in all mammalian serine proteases and are probably important in stabilizing the conformation of these enzymes.

Despite these structural similarities, there are a number of obvious differences. One is the lack of the disulfide bond linking Cys-191 (in the region of the primary binding site) to Cys-220 (near the secondary binding site) in chymotrypsin. This bond is present in all other known serine proteases, including the α-lytic protease from Myxobactor 495 (Olsen et al., 1970), and two proteases from *Streptomyces griseus* (Johnson & Smillie, 1974; Jurásek et al., 1974). Knights & Light (1976) were able to selectively break this disulfide bond in trypsinogen and to show that trypsin generated from the modified zymogen had a larger (1000-fold) k_n, but an unchanged k_{cat} toward small substrates. This disulfide bridge may serve in the other serine proteases to maintain the configuration of the primary binding site and its absence in GSP may indicate a significant change in that site. The half-cystine corresponding to residue 191 in chymotrypsin is replaced by a phenylalanyl residue in GSP (with retention of hydrophobic character), whereas the other half-cystinyl residue (Cys-220 in chymotrypsin) has been deleted along with four other residues immediately following the antiparallel binding site (Ser-Tyr-Gly-199). Hence, the lack of this disulfide bond in GSP could affect both the primary and the extended substrate binding sites of the protease.

A second difference in the GSP structure is the substitution of asparagine-86 for isoleucine at position 99 of chymotrypsin.

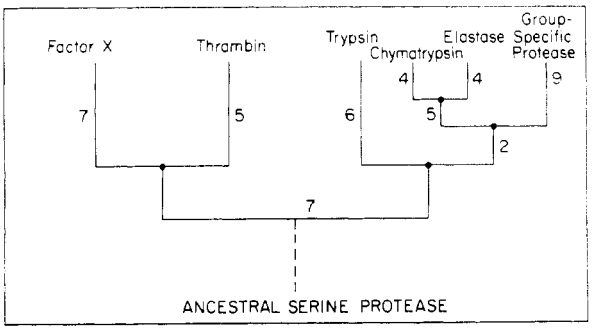

FIGURE 7: The evolutionary relationship of the serine proteases as determined by the number of insertions and deletions (gaps) arising from the optimally aligned sequences of the proteases. Figures indicate the number of sequence gaps which arise when the sequences are compared. The dots (●) indicate points of divergence. This evolutionary tree lacks a coordinate of time.

In other mammalian serine proteases this residue is usually hydrophobic (Table IV). If, in fact, this is the S_2 binding subsite of the serine proteases, as suggested by Segal et al. (1971), one might predict that the most favorable substrates for GSP would contain a hydrophilic residue at the P_2 position capable of interacting with asparagine-86.

Yet another significant difference is an alanyl residue at position 176 in GSP, in place of serine-189 in chymotrypsin, and aspartic acid-177 in trypsin. In the latter two enzymes this residue forms the bottom of the S_1 binding site, and it is likely that this change in GSP indicates a hydrophobic amino acid at the P_1 position in the substrate. All of these structural differences occur in regions of GSP assumed to participate in substrate binding, suggesting a mode of interaction with substrate which is significantly different from those of other mammalian serine proteases.

There is no evidence that GSP is synthesized as a zymogen. On the contrary, studies by Katunuma & co-workers (1973, 1975) and observations in our laboratory have indicated that the protease is fully active immediately following tissue homogenization. It has been suggested by Freer et al. (1970), and considered by de Haën et al. (1975) that the zymogens of the pancreatic serine proteases may be stabilized, in part, by a histidyl residue (His-40 in chymotrypsinogen) which interacts with an aspartyl residue (Asp-194 in chymotrypsinogen). If this is so, a zymogen form of GSP would lack this interaction (as do those of thrombin and factor X) since the histidyl residue has been replaced by valine-28.

It is clearly impossible to estimate the relative position of GSP in the evolution of mammalian serine proteases on the basis of sequence identities. Since insertions and deletions (gaps) of residues in the sequences of homologous proteins occur less frequently than simple amino acid substitution (Dayhoff, 1972, Table 5-2), the evolutionary relationship of homologous proteins may be more reliably estimated from the number and positions of inferred gaps which arise when the sequences are optimally aligned (de Haën et al., 1975). By this criterion, the divergence of GSP may have occurred sometime after that of trypsin, but before that of elastase and chymotrypsin (Figure 7). Comparison of the disulfide bonds among the serine proteases, however, suggests that the progenitors of GSP and elastase may have evolved together for a long period of time.

Since it is required that two bases of the specified codon undergo simultaneous change in order to directly replace an aspartyl with a seryl residue, the presence of an alanyl residue in the primary binding pocket of GSP poses several interesting questions. Did the substitution of aspartic acid by serine occur

in one event, requiring the simultaneous mutation of two bases, or in two separate events? If the latter is the case, did the transition from aspartic acid to serine occur with alanine as the intermediate residue? This sequential change would be consistent with the hypothesis that the ancestor molecule of GSP represents an evolutionary intermediate between those of trypsin and chymotrypsin. It may also suggest that the progenitor of GSP represents one of the earlier attempts to establish a functional serine protease lacking trypsin-like specificity.

Acknowledgments

The authors are indebted to Dr. David C. Teller for determining the evolutionary relationship of GSP to the other serine proteases. We are also grateful to Dr. Kenneth A. Walsh for his interest and helpful discussions throughout the course of this work and to Roger D. Wade, Lowell H. Ericsson, Richard R. Granberg, and Santosh Kumar for technical assistance.

References

Brandt, W. F., Edman, P., Henschen, A., & Von Holt, C. (1976) *Z. Physiol. Chem. 357,* 1505–1508.
Bridgen, P. J., Cross, G. A. M., & Bridgen, J. (1975) *Nature (London) 263,* 613–614.
Coffey, J. W., & de Duve, C. (1968) *J. Biol. Chem. 243,* 3255–3263.
Crewther, W. G., & Inglis, A. S. (1975) *Anal. Biochem. 68,* 572–585.
Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, Vol. 5, National Biomedical Research Foundation, Washington, D.C.
de Haën, C., Neurath, H., & Teller, D. C. (1975) *J. Mol. Biol. 92,* 225–259.
Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1,* 80–91.
Freer, S. T., Kraut, J. Robertus, J. D., Wright, H. T., & Xuong, Ng. H. (1970) *Biochemistry 9,* 1997–2009.
Goldberg, A. L., & St. John, A. C. (1975) *Annu. Rev. Biochem. 45,* 747–803.
Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., & Walsh, K. A. (1972) *Biochemistry 11,* 4493–4502.
Hermodson, M. A., Ericsson, L. H., Neurath, H., & Walsh, K. A. (1973) *Biochemistry 12,* 3146–3153.
Hirs, C. H. W. (1967) *Methods Enzymol. 11,* 197–199.
Horn, M. J. & Laursen, R. A. (1973) *FEBS Lett. 36,* 285–288.
Johnson, P., & Smillie, L. B. (1974) *FEBS Lett. 47,* 1–6.
Jurásek, L., Carpenter, M. R., Smillie, L. B., Gertler, A., Levy, S., & Ericsson, L. H. (1974) *Biochem. Biophys. Res. Commun. 61,* 1095–1100.
Katunuma, N. (1973) *Curr. Top. Cell. Regul. 7,* 175–203.
Katunuma, N., Kominami, E., Kobayashi, K., Banno, Y., Suzuki, K., Chichibu, K., Hamaguchi, Y., & Katsunuma, T. (1975) *Eur. J. Biochem. 52,* 35–50.
Knights, R. J., & Light, A. (1976) *J. Biol. Chem. 251,* 222–228.
Laursen, R. A. (1975) in *Solid-Phase Methods in Protein Sequence Analysis,* Proceedings of the First International Conference, Pierce Chemical Co., Rockford, Ill.
Lui, T.-Y., & Chang, Y. H. (1971) *J. Biol. Chem. 246,* 2842–2848.
Olsen, M. O. J., Nagabhushan, N., Dzwiniel, M., Smillie, L. B., & Whitaker, D. R. (1970) *Nature (London) 228,* 438–442.
Robinson, N. C. (1971) Ph.D. Thesis, University of Washington.
Schimke, R. T. (1970) *Mamm. Protein Metab. 4,* 177–228.
Schimke, R. T. (1973) *Adv. Enzymol. 37,* 135–187.
Segal, D. M., Powers, J. C., Cohen, G. H., Davies, D. R., & Wilcox, P. E. (1971) *Biochemistry 10,* 3728–3738.
Weber, K., & Osborne, M. (1969) *J. Biol. Chem. 244,* 4406–4412.

Appendix: Crystallographic Data for a Group-Specific Protease from Rat Intestine

To allow detailed comparisons of the structure of the group-specific protease (Woodbury et al., 1978) with other serine proteases, x-ray crystallographic studies have been started.

Group-specific protease was purified from rat small intestine by a procedure which included an affinity adsorption chromatography step (to be described in detail elsewhere). The enzyme isolated in this manner is identical in chemical, physical, and immunologic properties to that purified by the method of Katunuma et al. (1975).

The best crystals were obtained using a batch method. An ammonium sulfate suspension of the enzyme was centrifuged and dissolved by adding distilled water to give a solution which was 20% saturated in ammonium sulfate and contained 10 mg of protein per mL. Aliquots (0.03 mL) of the enzyme solution were placed in small vials and saturated ammonium sulfate was gradually added until the solution was very slightly turbid (the final concentration of ammonium sulfate was about 46% of saturation). The vials were then left for 2 to 4 weeks at 4 °C. The best crystals were obtained between pH 7 and 8. Below neutral pH, intergrown crystals were usually obtained. The crystals were equilibrated before use with 0.02 M phosphate buffer pH 6.8 containing 55% saturated ammonium sulfate.

Preliminary x-ray diffraction experiments revealed that the *hk*0 zone has sixfold symmetry while the *hkl* zone has threefold symmetry. The *h*0*l* and *hhl* zones have only centers of symmetry, and the only 00*l* reflections observed have $l = 3n$. The space group is, therefore, $P3_1$ or its enantiomorph. The cell dimensions are $a = b = 78.2$ Å and $c = 96.8$ Å. Crystals as large as 0.8 mm in each direction are readily obtained, and intensities can be recorded to 2.5 Å using conventional Buerger precession photographs.

The density of the crystals is 1.201 g cm^{-3} and that of the mother liquor 1.157 g cm^{-3} which, assuming 0.3 g of bound water per g of protein, corresponds to 46 900 daltons of protein per asymmetric unit (Matthews, 1974). Since the molecular weight calculated from the amino acid sequence (Woodbury et al., 1978) is 24 665, there must be two molecules per asymmetric unit. The solvent content parameter V_M is 3.5 Å^3/dalton, which is quite high for a protein of molecular weight in the range 20 000 to 40 000. Although such loose packing is unusual for a protein of this size, it is not outside the range which has been observed (Matthews, 1976).

A search for heavy atom derivatives is currently underway.

References

Katunuma, N., Kominani, E., Kobayashi, K., Banno, Y., Suzuki, K., Chichibu, K., Hamaguchi, Y., & Katsunuma, T. (1975) *Eur. J. Biochem. 52,* 35–50.
Matthews, B. W. (1974) *J. Mol. Biol. 82,* 513–526.
Matthews, B. W. (1976) *Annu. Rev. Phys. Chem. 27,* 493–523.
Woodbury, R. G., Katunuma, N., Kobayashi, K., Titani, K., & Neurath, H. (1978) *Biochemistry 17* (main text of this paper).