THE AMINO ACID SEQUENCE AND PREDICTED STRUCTURE OF *STREPTOMYCES GRISEUS* PROTEASE A

P. JOHNSON and L. B. SMILLIE
Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2E1, Canada

Received 27 May 1974

1. Introduction

Streptomyces griseus Protease A (SGPA) is a protease of wide specificity [1] isolated from Pronase [1,2]. Together with *Streptomyces griseus* Trypsin (SGT) [3], α-lytic protease of *Myxobacter 495* [4] and *Streptomyces griseus* Protease B [2,5], SGPA is a microbial Asp–Ser–Gly protease which has features in common with the mammalian pancreatic Asp–Ser–Gly protease family (trypsin, elastase and the chymotrypsins). The present paper describes the complete sequence of this enzyme and the alignment of this sequence with the other Asp–Ser–Gly protease sequences. Comparisons made from this alignment indicate that SGPA, like *Myxobacter* α-lytic protease, probably retains the same internal hydrophobic cores which are present in the mammalian proteases, but has different structural features on the surface of the molecule.

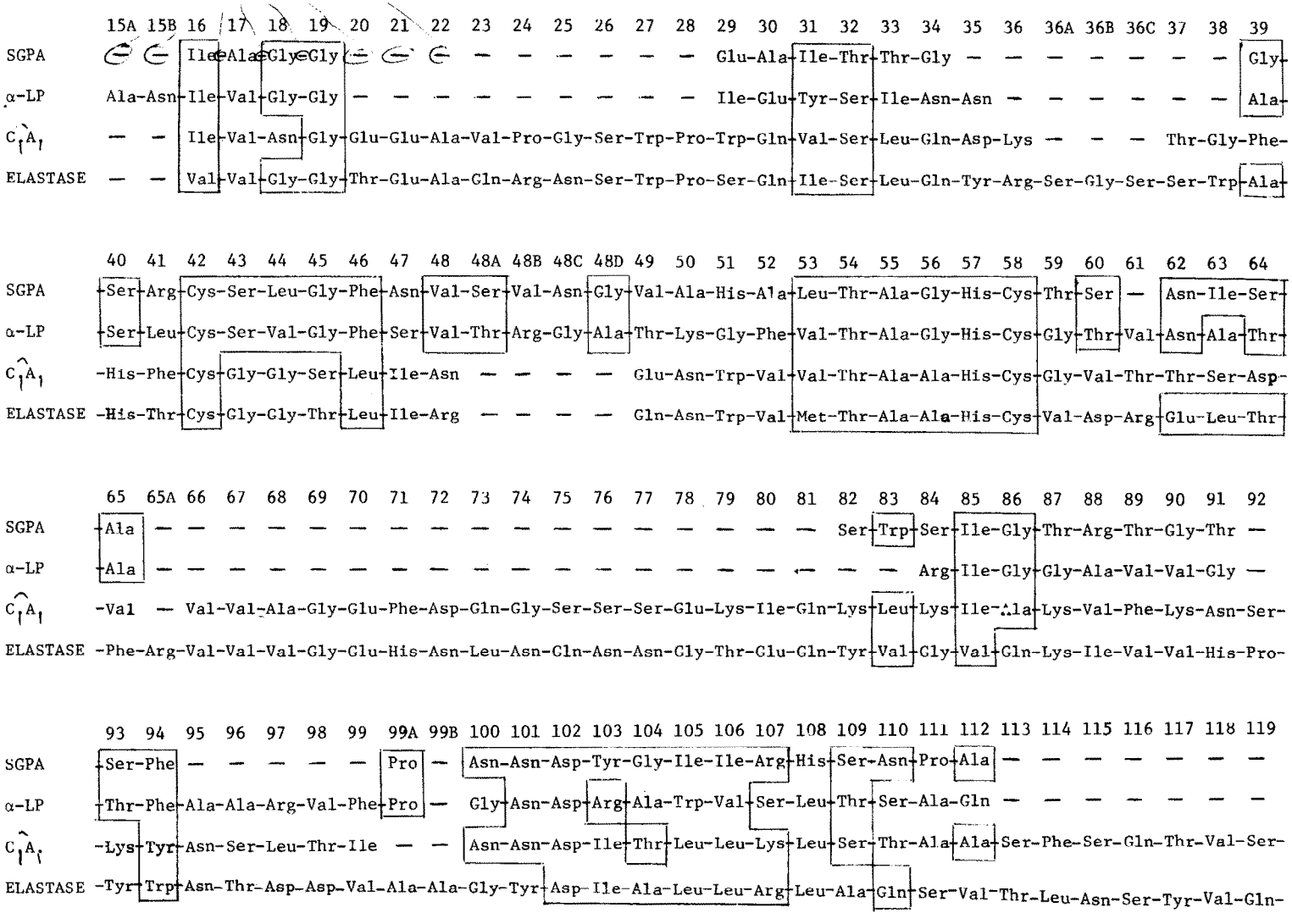
2. Experimental

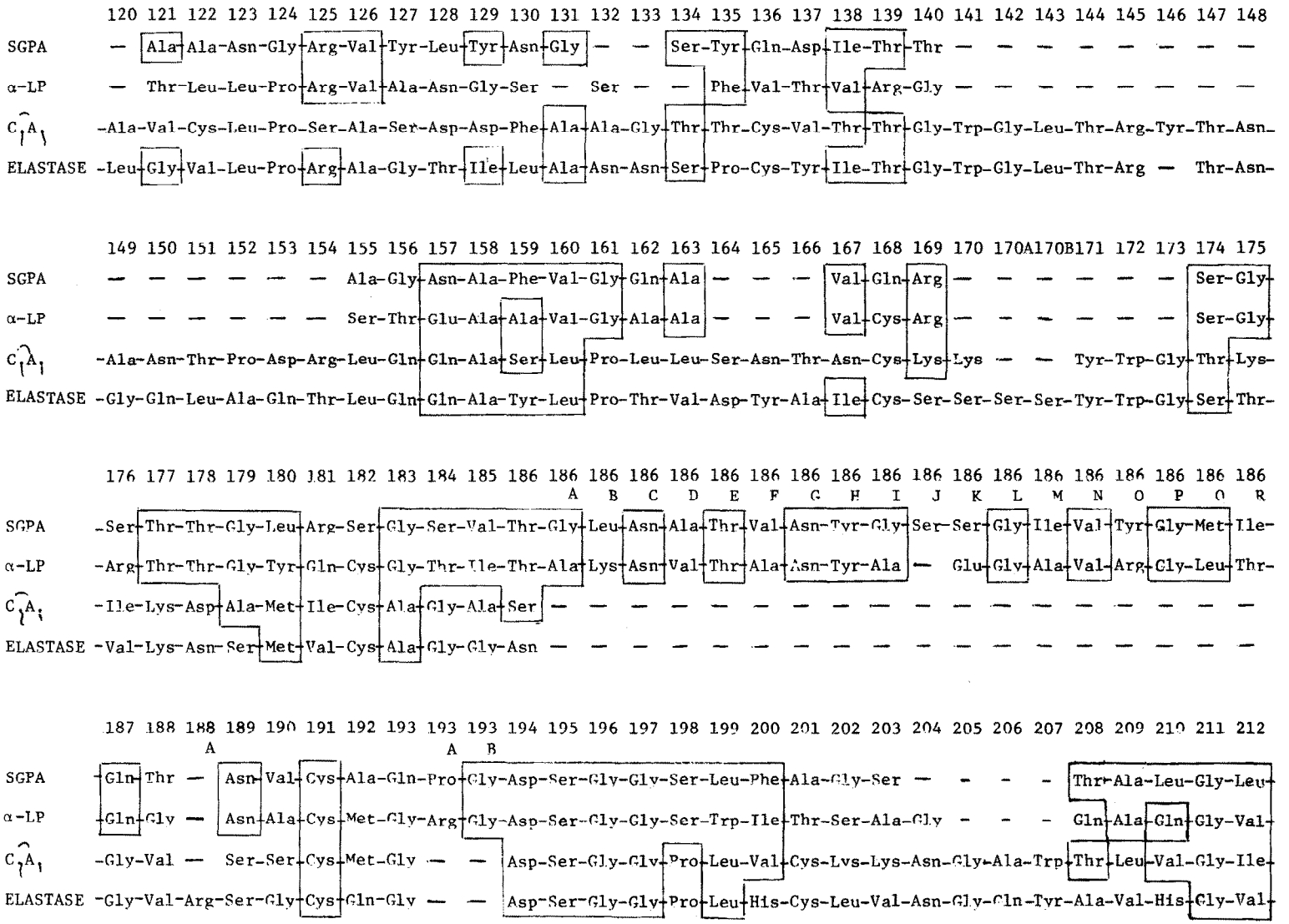
SGPA was purified from crude Pronase (Grade B, Calbiochem, lot 801930) by chromatography on CM-Sephadex and Bio-Rex 70 as described previously [1,2]. Peptides derived from SGPA by proteolytic or chemical cleavage were purified as necessary by molecular sieving, ion-exchange chromatography on Dowex 1 or Dowex 50 resin, high-voltage electrophoresis on paper, or by descending chromatography on paper. As previously described [1,6,7], the purified peptides were subjected to sequence analysis by the dansyl–Edman method, and amide assignments were determined by the method of Offord [8]. In automated sequence analysis using the Beckman Model 890 B instrument, residues were identified as PTH-derivatives [9] and as free amino acids after HI treatment of the thiazolinone derivatives [10]. The sequence of SGPA was established by automated sequencing and peptic digestion of the native enzyme, and from a CNBr cleavage of the inactivated enzyme, with subsequent redigestions of the two CNBr fragments by α-lytic protease, trypsin, chymotrypsin and thermolysin.

3. Results

SGPA is a single polypeptide chain of 182 amino acid residues with two disulphide bridges linking residues 42 to 58 and 191 to 220. Its sequence is shown in fig. 1.where it has been aligned with *Myxobacter* α-lytic protease, bovine chymotrypsinogen and porcine elastase. This alignment is based on the known tertiary structures of bovine trypsin [13], procine elastase [14] and bovine α-chymotrypsin [15] and the predicted structure of *Myxobacter* α-lytic protease [4,16]. Inspection of the alignment reveals a great similarity of sequence between SGPA and *Myxobacter* α-lytic protease, particularly in sequence deletions and insertions which previously were unique to the latter enzyme.

Table 1 shows comparisons of residue similarities and identities between SGPA and the other proteases and also distinguishes between conservative buried residues and the more variable surface regions defined from X-ray studies on the mammalian enzymes [12]. These results emphasize the close homology between SGPA and *Myxobacter* α-lytic protease in that residue similarity and identity between the two enzymes is


15A 15B 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 36A 36B 36C 37 38 39
SGPA — — Ile-Ala-Gly-Gly — — — — — — — — — — Glu-Ala-Ile-Thr-Thr-Gly — — — — — — — Gly-
α-LP Ala-Asn-Ile-Val-Gly-Gly — — — — — — — — — Ile-Glu-Tyr-Ser-Ile-Asn-Asn — — — — — — Ala-
C1A1 — — Ile-Val-Asn-Gly-Glu-Glu-Ala-Val-Pro-Gly-Ser-Trp-Pro-Trp-Gln-Val-Ser-Leu-Gln-Asp-Lys — — — Thr-Gly-Phe-
ELASTASE — — Val-Val-Gly-Gly-Thr-Glu-Ala-Gln-Arg-Asn-Ser-Trp-Pro-Ser-Gln-Ile-Ser-Leu-Gln-Tyr-Arg-Ser-Gly-Ser-Ser-Trp-Ala-
40 41 42 43 44 45 46 47 48 48A 48B 48C 48D 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64
SGPA -Ser-Arg-Cys-Ser-Leu-Gly-Phe-Asn-Val-Ser-Val-Asn-Gly-Val-Ala-His-Ala-Leu-Thr-Ala-Gly-His-Cys-Thr-Ser — Asn-Ile-Ser-
α-LP -Ser-Leu-Cys-Ser-Val-Gly-Phe-Ser-Val-Thr-Arg-Gly-Ala-Thr-Lys-Gly-Phe-Val-Thr-Ala-Gly-His-Cys-Gly-Thr-Val-Asn-Ala-Thr-
C1A1 -His-Phe-Cys-Gly-Gly-Ser-Leu-Ile-Asn — — — — Glu-Asn-Trp-Val-Val-Thr-Ala-Ala-His-Cys-Gly-Val-Thr-Thr-Ser-Asp-
ELASTASE -His-Thr-Cys-Gly-Gly-Thr-Leu-Ile-Arg — — — — Gln-Asn-Trp-Val-Met-Thr-Ala-Ala-His-Cys-Val-Asp-Arg-Glu-Leu-Thr-
65 65A 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92
SGPA -Ala — — — — — — — — — — — — — — — — — Ser-Trp-Ser-Ile-Gly-Thr-Arg-Thr-Gly-Thr —
α-LP -Ala — — — — — — — — — — — — — — — — — Arg-Ile-Gly-Gly-Ala-Val-Val-Gly —
C1A1 -Val — Val-Val-Ala-Gly-Glu-Phe-Asp-Gln-Gly-Ser-Ser-Ser-Glu-Lys-Ile-Gln-Lys-Leu-Lys-Ile-Ala-Lys-Val-Phe-Lys-Asn-Ser-
ELASTASE -Phe-Arg-Val-Val-Val-Gly-Glu-His-Asn-Leu-Asn-Gln-Asn-Asn-Gly-Thr-Glu-Gln-Tyr-Val-Gly-Val-Gln-Lys-Ile-Val-Val-His-Pro-
93 94 95 96 97 98 99 99A 99B 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119
SGPA -Ser-Phe — — — — — Pro — Asn-Asn-Asp-Tyr-Gly-Ile-Ile-Arg-His-Ser-Asn-Pro-Ala — — — — — — —
α-LP -Thr-Phe-Ala-Ala-Arg-Val-Phe-Pro — Gly-Asn-Asp-Arg-Ala-Trp-Val-Ser-Leu-Thr-Ser-Ala-Gln — — — — — — —
C1A1 -Lys-Tyr-Asn-Ser-Leu-Thr-Ile — — Asn-Asn-Asp-Ile-Thr-Leu-Leu-Lys-Leu-Ser-Thr-Ala-Ala-Ser-Phe-Ser-Gln-Thr-Val-Ser-
ELASTASE -Tyr-Trp-Asn-Thr-Asp-Asp-Val-Ala-Ala-Gly-Tyr-Asp-Ile-Ala-Leu-Leu-Arg-Leu-Ala-Gln-Ser-Val-Thr-Leu-Asn-Ser-Tyr-Val-Gln-


120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148
SGPA — Ala Ala-Asn-Gly Arg-Val Tyr-Leu Tyr Asn Gly — — Ser-Tyr Gln-Asp Ile-Thr Thr — — — — — — — —
α-LP — Thr-Leu-Leu-Pro Arg-Val Ala-Asn-Gly-Ser — Ser — — Phe Val-Thr Val Arg-Gly — — — — — — — —
C A
-Ala-Val-Cys-Leu-Pro-Ser-Ala-Ser-Asp-Asp-Phe Ala Ala-Gly Thr Thr-Cys-Val-Thr Thr Gly-Trp-Gly-Leu-Thr-Arg-Tyr-Thr-Asn-
ELASTASE -Leu Gly Val-Leu-Pro Arg Ala-Gly-Thr Ile Leu Ala Asn-Asn Ser Pro-Cys-Tyr Ile-Thr Gly-Trp-Gly-Leu-Thr-Arg — Thr-Asn-
149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 170A 170B 171 172 173 174 175
SGPA — — — — — — Ala-Gly Asn-Ala-Phe-Val-Gly Gln Ala — — — Val Gln Arg — — — — — — Ser-Gly
α-LP — — — — — — Ser-Thr Glu-Ala Ala Val-Gly Ala Ala — — — Val Cys Arg — — — — — — Ser-Gly
C A
-Ala-Asn-Thr-Pro-Asp-Arg-Leu-Gln Gln-Ala Ser Leu Pro-Leu-Leu-Ser-Asn-Thr-Asn-Cys Lys Lys — — Tyr-Trp-Gly Thr Lys-
ELASTASE -Gly-Gln-Leu-Ala-Gln-Thr-Leu-Gln Gln-Ala-Tyr-Leu Pro-Thr-Val-Asp-Tyr-Ala Ile Cys-Ser-Ser-Ser-Ser-Tyr-Trp-Gly Ser Thr-
176 177 178 179 180 181 182 183 184 185 186 186A 186B 186C 186D 186E 186F 186G 186H 186I 186J 186K 186L 186M 186N 186O 186P 186Q 186R
SGPA -Ser Thr-Thr-Gly-Leu Arg-Ser Gly-Ser-Val-Thr-Gly Leu Asn Ala Thr Val Asn-Tyr-Gly Ser-Ser Gly Ile Val Tyr Gly-Met Ile-
α-LP -Arg Thr-Thr-Gly-Tyr Gln-Cys Gly-Thr-Ile-Thr-Ala Lys Asn Val Thr Ala Asn-Tyr-Ala — Glu Gly Ala Val Arg Gly-Leu Thr-
C A
-Ile-Lys-Asp Ala-Met Ile-Cys Ala Gly-Ala Ser — — — — — — — — — — — — — — — —
ELASTASE -Val-Lys-Asn Ser Met Val-Cys Ala Gly-Gly-Asn — — — — — — — — — — — — — — — —
187 188 188A 189 190 191 192 193 193A 193B 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212
SGPA Gln Thr — Asn Val Cys Ala-Gln-Pro Gly-Asp-Ser-Gly-Gly-Ser-Leu-Phe Ala-Gly-Ser — — — — Thr Ala-Leu-Gly-Leu
α-LP Gln Gly — Asn Ala Cys Met-Gly-Arg Gly-Asp-Ser-Gly-Gly-Ser-Trp-Ile Thr-Ser-Ala-Gly — — — Gln Ala Gln Gly-Val
C A
-Gly-Val — Ser-Ser Cys Met-Gly — — Asp-Ser-Gly-Gly Pro Leu-Val Cys-Lys-Lys-Asn-Gly-Ala-Trp Thr Leu Val-Gly-Ile
ELASTASE -Gly-Val-Arg-Ser-Gly Cys Gln-Gly — — Asp-Ser-Gly-Gly Pro Leu His-Cys-Leu-Val-Asn-Gly-Gln-Tyr-Ala-Val-His Gly-Val

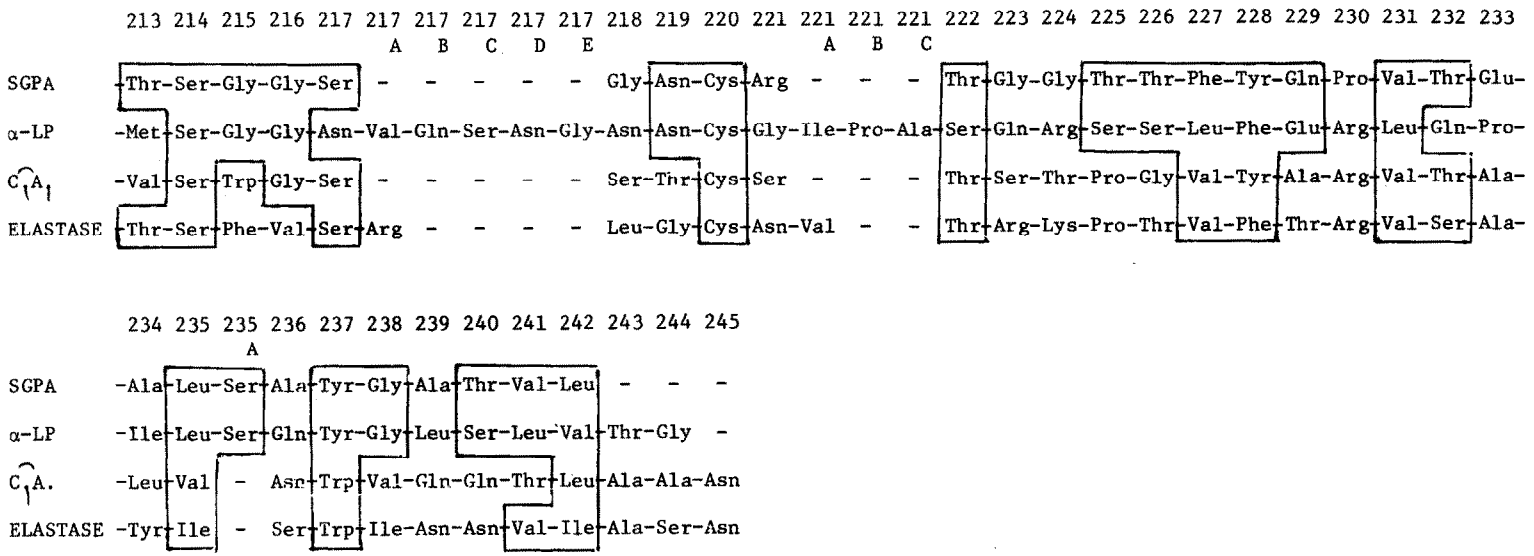


Fig. 1. The amino acid sequence of SGPA aligned with other Asp–Ser–Gly proteases. The numbering system is that of bovine chymotrypsinogen A, and chemically similar residues [4] are enclosed in solid lines. The proteases compared are *Myxobacter* α-lytic protease (α-LP) [4], bovine chymotrypsinogen A (CA) [11], and porcine elastase [12].

Table 1
Residue Homologies Between SGPA and Other Asp–Ser–Gly Proteases

	All residues		Internal residues*		External residues*	
	(%) identity	(%) similarity†	(%) identity	(%) similarity†	(%) identity	(%) similarity†
Myxobacter α-lytic protease	31	49	34	60	29	44
Porcine elastase	11	23	20	41	7	16
Bovine chymotrypsinogen A	12	22	23	43	7	13

* Defined from X-ray studies on elastase [14] and α-chymotrypsin [15].
† Defined by Olson et al. [4].

maintained even in the presumed surface sequences, where there is much less similarity between SGPA and the mammalian enzymes.

As the prediction of tertiary structure based on the amino acid sequences of a protein and the known tertiary structure of a homologous protein is valid [17], we predict that the SGPA molecule, like the *Myobacter* α-lytic protease molecule, will closely resemble the structures of the mammalian enzymes in the sequences around the two disulphide bridges (Cys 42 to Cys 58, and Cys 191 to Cys 220 in fig. 1) and in the region of the active site triad Asp 102, His 57, and Ser 195. The only significant sequence alteration in the two microbial enzymes in these regions is a dipeptide insertion between residues 193 and 194 which can probably be accommodated into the structure without distortion of the active centre [16].

The microbial enzymes however are predicted to differ from the mammalian enzymes as they possess alterations in three surface regions. One of the major surface alterations is near the 'autolysis loop' of chymotrypsin. In SGPA and *Myxobacter* α-lytic protease, the loop sequence (residues 141 to 154 in fig. 1) is deleted, and long insertions are present between residues 186 and 187. These are located in the same surface area of the molecule as the deleted autolysis loop present in α-chymotrypsin. Also missing from this surface area of the microbial enzymes are residues 20 to 28. On the basis of the mammalian protease structures, it has been postulated [16] that in *Myxobacter* α-lytic protease, the large insertion in this region functionally compensates for the missing autolysis loop which is thought to be involved in the burial of the NH_2-terminal residue into the interior of the molecule by means of a salt-linkage between the terminal α-NH_2 group and the side-chain COOH group of Asp 194.

The second region of alteration is near the surface 'uranyl loop' of chymotrypsin. In SGPA, there are three neighboring surface deletions; residues 66 to 81 of the uranyl loop, residues 35 to 38 and residues 113 to 120. A tetrapeptide insertion also exists in this region between residues 48 and 49 which can probably be accommodated as a small surface loop.

The third region of alteration is at the methionyl loop where both SGPA and *Myxobacter* α-lytic protease have residues deleted. In SGPA, residue 168 is Gln rather than Cys as in the other proteases, and the disulphide mate Cys 182 in the other proteases is Ser in SGPA. Thus, SGPA is dissimilar to the other Asp–Ser–Gly proteases in that it possesses only two disulphide bridges, both of which are intimately involved in the structure of the catalytic site. Another difference between SGPA and α-lytic protease in this region is the deletion in SGPA between residues 94 and 100. This probably occurs in a small surface loop, and although it may be sterically permissible, the correctness of the residue alignments is uncertain as both residues 94 and 99 (one of which is missing from SGPA) have been implicated in the burial of Asp 102 in other Asp–Ser–Gly proteases [13,18]. Furthermore, residue 99 has been identified as part of the substrate binding site in chymotrypsin [19,20].

These and other structural predictions for SGPA

will only be resolved when the detailed tertiary structure of the enzyme is solved, a project which is currently under investigation in this department by Dr. M. N. G. James.

Acknowledgements

The expert technical assistance of Mr M. Nattriss, Mr M. R. carpenter and Miss A. Thomsom is gratefully acknowledged. This work was supported by the Medical Research Council of Canada.

References

[1] Johnson, P. and Smillie, L. B. (1972) Can. J. Biochem. 49, 548–562.
[2] Jurášek, L., Johnson, P., Olafson, R. W. and Smillie, L. B. (1971) Can. J. Biochem. 49, 1195–1201.
[3] Jurášek, L., Fackre, D. and Smillie, L. B. (1969) Biochem. Biophys. Res. Comm. 37, 99–105.
[4] Olson, M. O. J., Nagabhushan, N., Dzwiniel, M. and Smillie, L. B. (1970) Nature 228, 438–442.
[5] Wahlby, S. and Engström, L. (1968) Biochim. Biophys. Acta 151, 402–408.
[6] Johnson, P. and Smillie, L. B. (1971) Can. J. Biochem. 49, 1083–1097.
[7] Johnson, P. and Smillie, L. B. (1972) Can. J. Biochem. 50, 589–599.
[8] Offord, R. E. (1966) Nature 211, 591–593.
[9] Pisano, J. J. and Bronzert, T. J. (1969) J. Biol. Chem. 244, 5597–5607.
[10] Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G. and Ballantyne, D. L. (1971) Biochemistry 10, 4912–4921.
[11] Hartley, B. S. (1964) Nature 201, 1284–1287.
[12] Shotton, D. M. and Hartley, B. S. (1970) Nature 225, 802–806.
[13] Stroud, R. M., Kay, L. M. and Dickerson, R. E. (1971) Cold Spring Harb. Symp. Quant. Biol. 36, 125–140.
[14] Shotton, D. M. and Watson, H. C. (1970) Nature 225, 811–816.
[15] Mathews, B. W., Sigler, P. B., Henderson, R. and Blow, D. M. (1967) Nature 214, 652–656.
[16] McLachlan, A. D. and Shotton, D. M. (1971) Nature New Biol. 229, 202–205.
[17] Hartley, B. S. (1970) Phil. Trans. Roy. Soc. Ser. B 257, 77–86.
[18] Blow, D. M., Birktoft, J. J. and Hartley, B. S. (1969) Natur 221, 337–340.
[19] Segal, D. M., Cohen, G. H., Davies, D. R., Powers, J. C. and Wilcox, P. E. (1971) Cold Spring Harb. Symp. Quant. Biol. 36, 85–90.
[20] Segal, D. M,, Powers, J. C., Cohen, G. H., Davies, D. R. and Wilcox, P. E. (1971) Biochemistry 10, 3728–3738.