Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Homologies in Serine Proteinases [and Discussion]

B. S. Hartley and H. Neurath

Phil. Trans. R. Soc. Lond. B 1970 **257**, 77-87
doi: 10.1098/rstb.1970.0010

References Article cited in:
http://rstb.royalsocietypublishing.org/content/257/813/77.citation#related-urls

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. Lond. B* go to: **http://rstb.royalsocietypublishing.org/subscriptions**

This journal is © 1970 The Royal Society

Downloaded from rstb.royalsocietypublishing.org

Phil. Trans. Roy. Soc. Lond. B **257**, 77–87 (1970) [77]
Printed in Great Britain

Homologies in serine proteinases

By B. S. Hartley
M.R.C. Laboratory of Molecular Biology, Cambridge

[Plate 54]

The pancreatic proteinases

Bovine pancreatic juice contains approximately equal amounts of four inactive precursors of endopeptidases (zymogens): chymotrypsinogen A, chymotrypsinogen B, trypsinogen (Keller, Cohen & Neurath 1958) and a component of procarboxypeptidase which resembles a chymotrypsinogen (Brown, Greenshields, Yamasaki & Neurath 1963). In porcine pancreas another

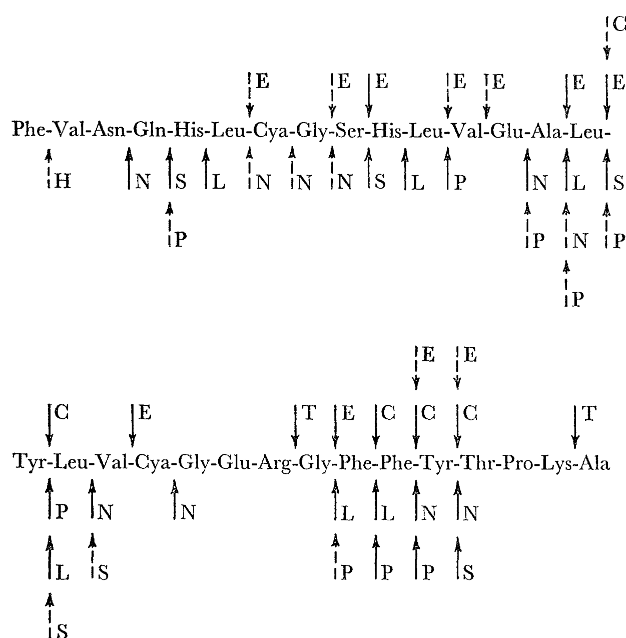


Figure 1. Action of some proteinases on the B chain of oxidized insulin (from Johansen, Ottesen, Svendsen & Wybrandt (1968). Major points of cleavage are shown by full arrows and secondary points by broken arrows. C, Bovine α-chymotrypsin; E, porcine elastase; L, thermolysin; N, papain; P, porcine pepsin; S, subtilisin (Novo); T, bovine trypsin.

endopeptidase, elastase, is found which is uniquely effective against elastin, the elastic protein of ligaments. Chymotrypsin A and chymotrypsin B are almost identical in enzyme activity (Enenkel & Smillie 1963), but the chymotrypsins, trypsin and elastase have widely different substrate specificities, as seen, for example, in their action on the B chain of oxidized insulin (Naughton & Sanger 1961; figure 1). Nevertheless, it has long been suspected that they have a common catalytic mechanism, since in each enzyme a unique serine residue, residue 195,† reacts with organophosphorus compounds with complete inhibition of the enzyme activity. This property was proposed as the definition of a class of 'serine proteinases' (Hartley 1960).

† Throughout this article residues are numbered according to the sequence of bovine chymotrypsinogen A. 'Insertions' in the sequence of other enzymes are numbered 187A, 187B, etc.

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences.* www.jstor.org

Downloaded from rstb.royalsocietypublishing.org

The same serine residue is acylated as a step in the hydrolysis of some ester substrates (Hartley & Kilby 1954; Oosterbaan & van Adrichem 1958) and a unique histidine residue has long been suspected as a catalyst for both acylation and deacylation steps (Gutfreund & Sturtevant 1956). Reaction with reagents which are analogues of substrates showed that this residue was His-57 in both chymotrypsin (Ong, Shaw & Shoellmann 1964; Smillie & Hartley 1964; Pošpíšilova, Meloun & Šorm 1964) and trypsin (Shaw, Mares-Guia & Cohen 1965).

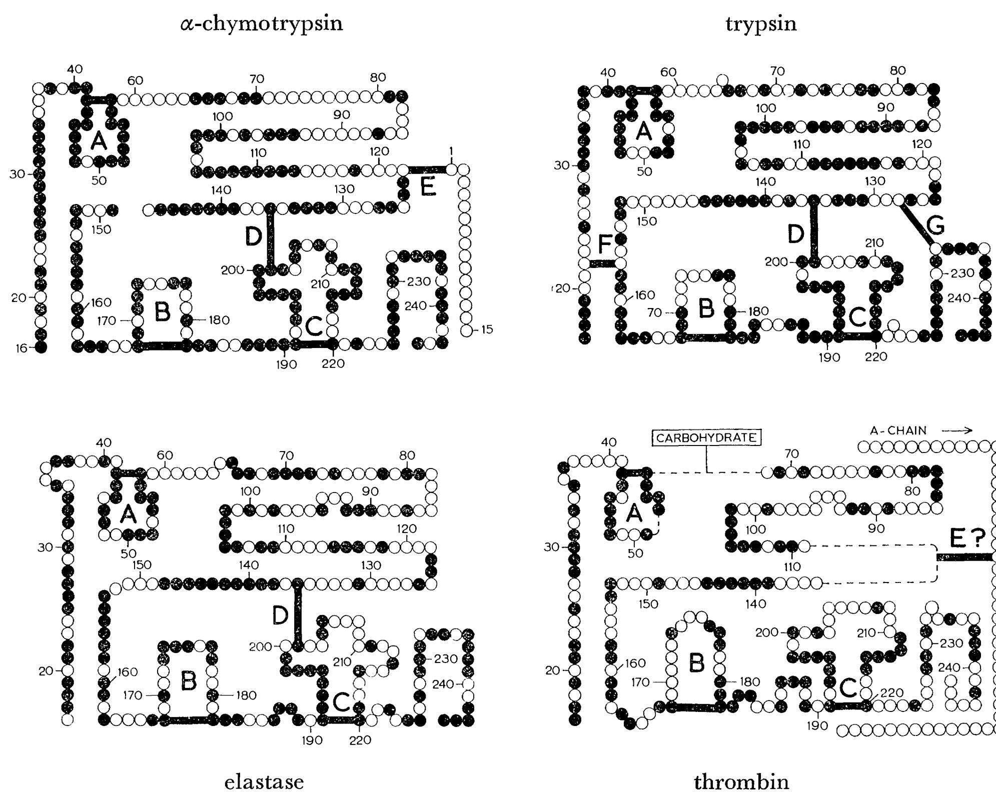


FIGURE 2. Disulphide bridges in mammalian serine proteinases. Black circles show residues which are identical in any pair of these four proteins.

The amino acid sequences around the active centre serine were found to be the same in chymotrypsin (Turba & Gundlach 1955; Schaffer, Simel, Harshman, Engle & Drisko 1957), trypsin (Dixon, Kauffman & Neurath 1958) and elastase (Hartley, Naughton & Sanger 1959), suggesting that despite the differences in specificity the enzymes may have descended from a common ancestor. Comparison of the complete primary structures of bovine chymotrypsinogen A (Hartley 1964) and trypsinogen (Walsh, Kauffman, Kumar & Neurath 1964) strengthened this hypothesis since 40 % of the residues were found to be identical.

Moreover, the disulphide bridges in the pancreatic proteinases were strikingly similar (Hartley, Brown, Kauffman & Smillie 1965). Four bridges were common to chymotrypsinogen and trypsinogen (figure 2):

A. The 'histidine loop' (Cys-42 to Cys-58).
B. The 'methionine loop' (Cys-168 to Cys-182).
C. The 'serine loop' (Cys-191 to Cys-220).
D. The 'B-C chain' bridge (Cys-136 to Cys-201).

Downloaded from rstb.royalsocietypublishing.org

Chymotrypsinogen had a unique bridge joining the A chain 'activation peptide' to the B chain:

E. The 'A-B' chain bridge.

And trypsinogen had two unique bridges:

F. The 'N-terminal' bridge (Cys-22 to Cys-157).

G. The 'C-terminal' bridge (Cys-128 to Cys-232).

Chymotrypsinogen B had the same five bridges (A, B, C, D and E) as chymotrypsinogen A and of the 71 residues in tryptic peptides including these bridges, 63 (89 %) were identical in both proteins.

Porcine elastase had only four bridges which were the same four (A, B, C and D) common to chymotrypsin and trypsin, and 52 % of the residues around these bridges were identical.

The primary structure of all these enzymes, now almost complete, is shown in figure 3. Four corrections have been made to the sequence of bovine chymotrypsinogen A (Hartley 1964) at positions 18, 19, 215 (Hartley & Kauffman 1966) and 102 (Blow, Birktoft & Hartley 1969). The complete sequence of bovine chymotrypsinogen B has recently been published by Smillie, Furka, Nagabhushan, Stevenson & Parks (1968). The bovine trypsinogen sequence (Walsh *et al.* 1964; Mikeš, Tomášek, Holeyšovský & Šorm 1966) has been corrected at position 189 (see below). The sequence of porcine elastase (Shotton & Hartley 1970) is now complete, but that

FIGURE 3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	**16**
CA	Cys	Gly	Val	Pro	Ala	Ile	Gln	Pro	Val	Leu	Ser	Gly	Leu	Ser	ARG	**Ile**
CB	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	Ala	,,	,,
T										Val	Asp	Asp	Asp	Asp	LYS	**Ile**
E																VAL
Th																**Ile**

	17	18	**19**	20	21	**22**	23	24	25	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**
CA	**Val**	Asn	**Gly**	GLU	**Glu**	**Ala**	Val	Pro	**Gly**	**Ser**	**Trp**	**Pro**	**Trp**	**Gln**	**Val**	**Ser**	**Leu**
CB	,,	,,	,,	,,	**Asp**	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,
T	**Val**	**Gly**	**Gly**	Tyr	Thr	Cys	Gly	Ala	**Asn**	THR	Val	**Pro**	TYR	**Gln**	**Val**	**Ser**	**Leu**
E	**Val**	**Gly**	**Gly**	Thr	**Glu**	**Ala**	GLN	Arg	**Asn**	**Ser**	**Trp**	**Pro**	Ser	**Gln**	ILE	**Ser**	**Leu**
Th	**Val**	Glu	**Gly**	GLN	**Asp**	**Ala**	GLU	Val	**Gly**	Leu	Ser	**Pro**	**Trp**	**Gln**	**Val**	Met	**Leu**

	34	35	36	36A	36B	36C	37	38	39	**40**	41	**42**	**43**	**44**	**45**	**46**	**47**
CA	**Gln**	Asp	LYS	——	——	——	THR	**Gly**	PHE	**His**	**Phe**	**Cys**	**Gly**	**Gly**	**Ser**	**Leu**	**Ile**
CB	,,	,,	Ser	——	——	——	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,
T	ASN	——	——	——	——	——	**Ser**	**Gly**	TYR	**His**	**Phe**	**Cys**	**Gly**	**Gly**	**Ser**	**Leu**	**Ile**
E	**Gln**	TYR	**Arg**	Ser	Gly	**Ser**	**Ser**	Trp	Ala	**His**	Thr	**Cys**	**Gly**	**Gly**	THR	**Leu**	**Ile**
Th	——	PHE	**Arg**	Lys	——	**Ser**	Pro	Glx	Glx	Leu	Leu	**Cys**	**Gly**	Ala	**Ser**	**Leu**	**Ile**

	48	49	50	**51**	**52**	**53**	**54**	**55**	56	57	**58**	59	**60**	61	62	63	64	65
CA	**Asn**	GLU	**Asn**	**Trp**	**Val**	**Val**	**Thr**	**Ala**	**Ala**	**His**	**Cys**	Gly	Val	THR	Thr	Ser	Asp	**Val**
CB	**Ser**	,,	ASP	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,
T	**Asn**	Ser	GLN	**Trp**	**Val**	**Val**	SER	**Ala**	**Ala**	**His**	**Cys**	Tyr	Lys	SER	Gly	ILE	Gln	**Val**
E	Arg	GLN	**Asn**	**Trp**	**Val**	Met	**Thr**	**Ala**	**Ala**	**His**	**Cys**	Val	Asp	Arg	Glu	LEU	Thr	Phe
Th	**Ser**	ASP	Arg	**Trp**			**Thr**	**Ala**	**Ala**	**His**	**Cys**							

	65A	**66**	**67**	**68**	69	70	71	72	73	74	75	76	77	78	79	80	81
CA	——	**Val**	**Val**	Ala	**Gly**	**Glu**	Phe	ASP	Gln	Gly	Ser	Ser	Ser	Glu	Lys	Ile	**Gln**
CB	——	,,	,,	,,	,,	,,	,,	,,	,,	,,	Leu	Glu	Thr	,,	ASP	Thr	,,
T	**Arg**	Leu	——	——	**Gly**	GLN	Asp	**Asn**	ILE	**Asn**	Val	Val	**Glu**	**Gly**	ASN	GLN	**Gln**
E	**Arg**	**Val**	**Val**	VAL	**Gly**	**Glu**	**His**	**Asn**	LEU	**Asn**	Gln	Asn	Asn	**Gly**	Thr	**Glu**	**Gln**
Th				ILE	**Gly**	Lys,	**His**	Ser	Arg	Thr	Arg	Tyr	**Glu**	Arg	Val	**Glu**	**Gln**

Downloaded from rstb.royalsocietypublishing.org

	82	**83**	84	**85**	86	87	88	89	90	91	92	93	94	95	96	97	98
CA	**Lys**	LEU	Lys	ILE	Ala	**Lys**	VAL	Phe	Lys	Asn	Ser	Lys	**Tyr**	**Asn**	**Ser**	Leu	**Thr**
CB	Val	,,	,,	,,	Gly	,,	,,	,,	,,	,,	**Pro**	,,	PHE	Ser	ILE	,,	,,
T	PHE	**Ile**	**Ser**	Ala	Ser	**Lys**	Ser	**Ile**	**Val**	**His**	**Pro**	Ser	**Tyr**	**Asn**	**Ser**	ASN	**Thr**
E	TYR	VAL	Gly	VAL	Gln	**Lys**	ILE	VAL	**Val**	**His**	**Pro**	Tyr	TRP	**Asn**	THR	ASP	Asp
Th	**Lys**	**Ile**	**Ser**	Met	Leu	Asp	LYS	**Ile**	Tyr	(**His**,	**Pro**,	Ile,	Arg)	Tyr	Asn	Trp	Lys

	99	99A	99B	100	101	**102**	**103**	**104**	**105**	**106**	107	**108**	109	110	111	**112**	113
CA	ILE	——	——	**Asn**	**Asn**	**Asp**	**Ile**	Thr	**Leu**	**Leu**	**Lys**	**Leu**	**Ser**	**Thr**	**Ala**	**Ala**	**Ser**
CB	**Val**	——	——	Arg	,,	,,	,,	,,	,,	,,	,,	,,	**Ala**	,,	Pro	,,	Gln
T	LEU	——	——	**Asn**	**Asn**	**Asp**	**Ile**	Met	**Leu**	ILE	**Lys**	**Leu**	Lys	SER	**Ala**	**Ala**	**Ser**
E	**Val**	Ala	Ala	Gly	Tyr	**Asp**	**Ile**	**Ala**	**Leu**	**Leu**	ARG	**Leu**	**Ala**	Gln	Ser	Val	THR
Th	Glu	Asn	Leu	ASP	Arg	**Asp**	**Ile**	**Ala**	**Leu**	**Leu**	**Lys**,	Ala	**Ser**	**Thr**	Arg		

	114	115	116	117	**118**	119	120	**121**	122	123	**124**	125	126	127	128	129	130	131
CA	Phe	Ser	Gln	Thr	**Val**	Ser	Ala	VAL	Cys	**Leu**	**Pro**	SER	**Ala**	**Ser**	Asp	Asp	Phe	**Ala**
CB	,,	,,	Glu	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	Asp	Glu	,,	,,	Pro
T	**Leu**	**Asn**	**Ser**	Arg	**Val**	Ala	Ser	ILE	Ser	**Leu**	**Pro**	THR	——	**Ser**	Cys	Ala	Ser	——
E	**Leu**	**Asn**	**Ser**	Tyr	**Val**	Gln	Leu	Gly	Val	**Leu**	**Pro**	Arg	**Ala**	Gly	Thr	Ile	Leu	**Ala**
Th																		

	132	133	134	135	**136**	**137**	**138**	**139**	**140**	**141**	**142**	143	144	145	146	147	148
CA	**Ala**	**Gly**	**Thr**	Thr	**Cys**	Val	Thr	**Thr**	**Gly**	**Trp**	**Gly**	**Leu**	**Thr**	**Arg**	Tyr	**Thr**	**Asn**
CB	,,	,,	Met	Leu	,,	Ala	,,	,,	,,	,,	,,	Lys	,,	**Lys**	,,	Asn	Ala
T	**Ala**	**Gly**	**Thr**	Gln	**Cys**	Leu	**Ile**	SER	**Gly**	**Trp**	**Gly**	**Asn**	**Thr**	**Lys**	Ser	SER	Gly
E	Asn	Asn	SER	Pro	**Cys**	Tyr	**Ile**	**Thr**	**Gly**	**Trp**	**Gly**	**Leu**	**Thr**	**Arg**	——	**Thr**	**Asn**
Th				Lys	Gly	Arg	VAL	**Thr**	**Gly**	**Trp**	**Gly**	**Asn**	——	**Arg**,	Leu	Leu	His

	149	150	151	152	153	154	**155**	156	157	**158**	159	**160**	161	**162**	162A	163	164
CA	**Ala**	ASN	Thr	**Pro**	**Asp**	Arg	**Leu**	**Gln**	**Gln**	**Ala**	Ser	**Leu**	**Pro**	LEU	——	**Leu**	**Ser**
CB	Leu	Lys	,,	,,	,,	Lys	,,	,,	,,	,,	Thr	,,	,,	**Ile**	——	**Val**	,,
T	Thr	Ser	TYR	**Pro**	**Asp**	Val	**Leu**	Lys	Cys	**Leu**	**Lys**	Ala	**Pro**	**Ile**	——	**Leu**	**Ser**
E	Gly	GLN	Leu	Ala	**Gln**	**Thr**	**Leu**	**Gln**	**Gln**	**Ala**	Tyr	**Leu**	**Pro**	Thr	——	**Val**	Asp
Th	**Ala**	Gly	PHE	LYS,	**Gln**	(**Thr**,	Ala,	Ala)	Lys,	**Leu**	**Lys**	Arg	**Pro**	**Ile**	Glu	**Leu**	**Ser**

	165	166	167	**168**	169	170	170A	170B	170C	171	172	173	174	175	**176**	177
CA	**Asn**	THR	Asn	**Cys**	**Lys**	Lys	——	——	——	**Tyr**	**Trp**	**Gly**	THR	LYS	**Ile**	**Lys**
CB	,,	,,	Asp	,,	ARG	,,	——	——	——	,,	,,	,,	**Ser**	**Arg**	**Val**	**Thr**
T	**Asn**	SER	Ser	**Cys**	**Lys**	**Ser**	——	——	——	Ala	TYR	Pro	Gly	Gln	**Ile**	**Thr**
E	Tyr	Ala	**Ile**	**Cys**	Ser	**Ser**	Ser	Ser	——	**Tyr**	**Trp**	**Gly**	**Ser**	Thr	**Val**	**Lys**
Th	ASX-	(His,	**Ile**,	**Cys**,	Pro,	Val,	Leu,	Pro,	Asx,	**Tyr**)-	Lys,	Arg	Ile	**Arg**	**Ile**	**Thr**

	178	179	**180**	**181**	**182**	**183**	**184**	184A	185	186	187	187A	187B	188	188A	188B
CA	ASP	Ala	**Met**	ILE	**Cys**	**Ala**	**Gly**	——	Ala	Ser	**Gly**	——	——	**Val**	——	——
CB	,,	Val	,,	,,	,,	,,	,,	——	,,	,,	,,	——	——	,,	——	——
T	Ser	ASN	**Met**	**Phe**	**Cys**	**Ala**	**Gly**	**Tyr**	Leu	Glu	**Gly**	——	**Gly**	**Lys**	——	——
E	ASN	Ser	**Met**	VAL	**Cys**	**Ala**	**Gly**	——	Gly	Asn	**Gly**	——	——	**Val**	**Arg**	——
Th	ASX	ASX	**Met**	**Phe**	**Cys**	**Ala**	**Gly**	**Tyr**	Lys	Pro	**Gly**	Glu	**Gly**	**Lys**	**Arg**	Gly

	189	*190*	*191*	*192*	**193**	**194**	195	**196**	**197**	**198**	**199**	**200**	**201**	202	203	204
CA	**Ser**	**Ser**	**Cys**	Met	**Gly**	**Asp**	**Ser**	**Gly**	**Gly**	**Pro**	**Leu**	**Val**	**Cys**	**Lys**	Lys	**Asn**
CB	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	,,	Gln	,,	,,
T	**Asp**	**Ser**	**Cys**	**Gln**	**Gly**	**Asp**	**Ser**	**Gly**	**Gly**	**Pro**	Val	**Val**	**Cys**	Ser	Gly	Lys
E	**Ser**	Gly	**Cys**	**Gln**	**Gly**	**Asp**	**Ser**	**Gly**	**Gly**	**Pro**	**Leu**	His	**Cys**	Leu	Val	**Asn**
Th	**Asp**	Ala	**Cys**	GLU	**Gly**	**Asp**	**Ser**	**Gly**	**Gly**	**Pro**	Phe	**Val**	Met	**Lys**	Ser	Pro

	204A	204B	205	206	**207**	208	**209**	**210**	**211**	**212**	**213**	**214**	**215**	*216*	217	217A
CA	——	——	**Gly**	Ala	**Trp**	Thr	**Leu**	Val	**Gly**	**Ile**	**Val**	**Ser**	**Trp**	**Gly**	**Ser**	——
CB	——	——	,,	,,	,,	,,	,,	Ala	,,	,,	,,	,,	,,	,,	,,	——
T	——	——	——	——	——	——	**Leu**	Gln	**Gly**	**Ile**	**Val**	**Ser**	**Trp**	**Gly**	**Ser**	——
E	——	——	**Gly**	Gln	TYR	Ala	Val	His	**Gly**	VAL	Thr	**Ser**	PHE	Val	**Ser**	Arg
Th	Tyr	Asn	Asn	Arg	**Trp**	Tyr	Gln	Met	**Gly**	**Ile**	**Val**	**Ser**	**Trp**	**Gly**	Glu	——

Downloaded from rstb.royalsocietypublishing.org

	218	219	*220*	221	221A	222	223	*224*	*225*	*226*	**227**	***228***	**229**	230	**231**	232
CA	Ser	Thr	**Cys**	Ser	——	**Thr**	Ser	Thr	**Pro**	**Gly**	**Val**	**Tyr**	Ala	**Arg**	**Val**	THR
CB	,,	,,	,,	,,	——	,,	,,	,,	,,	Ala	,,	,,	,,	,,	,,	,,
T	——	**Gly**	**Cys**	Ala	Gln	Lys	Asn	**Lys**	**Pro**	**Gly**	**Val**	**Tyr**	**Thr**	LYS	**Val**	Cys
E	Leu	**Gly**	**Cys**	ASN	Val	**Thr**	Arg	**Lys**	**Pro**	Thr	**Val**	PHE	**Thr**	**Arg**	**Val**	SER
Th	——	**Gly**	**Cys**	ASP	Arg	Asn	Gly	**Lys**	Tyr	**Gly**	Phe	**Tyr**	**Thr**	His	**Val**	Trp

	233	**234**	**235**	236	**237**	**238**	239	240	241	242	243	244	245	245A	245B
CA	**Ala**	**Leu**	**Val**	**Asn**	**Trp**	VAL	**Gln**	**Gln**	**Thr**	LEU	**Ala**	Ala	**Asn**	——	——
CB	,,	,,	Met	Pro	,,	,,	,,	GLU	,,	,,	,,	,,	,,	——	——
T	Asn	**Tyr**	**Val**	**Ser**	**Trp**	**Ile**	Lys	**Gln**	**Thr**	**Ile**	**Ala**	**Ser**	**Asn**	——	——
E	**Ala**	**Tyr**	ILE	**Ser**	**Trp**	**Ile**	ASN	ASN	**Val**	**Ile**	**Ala**	**Ser**	**Asn**	——	——
Th	(Arg, Lys, LEU)			Lys	**Trp**	**Ile**	**Gln**	Lys	**Val**	**Ile**	Asp	Arg	Leu	Gly	Ser

FIGURE 3. Amino acid sequences of mammalian serine proteinases: CA, bovine chymotrypsinogen A; CB, bovine chymotrypsinogen B; T, bovine trypsinogen; E, porcine elastase and Th, bovine thrombin. Residues which are 'chemically similar' (Lys=Arg, Asp=Glu, Asp=Asn, Glu=Gln, Ser=Thr, Val=Ile, Ile=Leu, Tyr=Phe=Trp) in any pair of trypsin, elastase, thrombin and chymotrypsinogen A *or* chymotrypsinogen-B are shown in capitals. Identical residues are in bold type. In chymotrypsinogen B only residues which differ from those in chymotrypsinogen A are indicated. Numbering is that of bovine chymotrypsinogen A with 'insertions' indicated 36A, 36B, etc. The numbers in bold represent side chains which are 'internal' in the tertiary structure of α-chymotrypsin. The numbers in italics indicate side-chains which line the 'tosyl-hole' of α-chymotrypsin.

of thrombin (S. M. Magnusson & B. S. Hartley, unpublished evidence) must be considered tentative.

The homology between these enzymes is summarized in table 1 and leads to the following conclusions:

(*a*) Chymotrypsinogen A and B appear to be 'isoenzymes'—the products of gene doubling—with an overall identity (78 %) much higher than that for the α- and β-chains of human haemoglobin (44 %).

(*b*) The enzymes with different specificity are remarkably similar in sequence (40–46 % identity), suggesting relatively recent descent from a common ancestor; contrast, for example, sperm whale myoglobin and human haemoglobin α-chain (18 % identity).

(*c*) The homology, particularly for example in the loops bridged by cystine, leads one to expect considerable similarities in tertiary structure.

TABLE 1. HOMOLOGIES IN SERINE PROTEINASES

	percentage identity (or chemical similarity†) to				
	CB	T	E	Th	any other
bovine chymotrypsin A‡ (CA)	78 (86)	46 (57)	42 (55)	37 (43)	63 (72) §
bovine chymotrypsin B‡ (CB)	—	41 (53)	40 (51)	36 (39)	57 (66) ‖
bovine trypsin (T)	—	—	40 (55)	40 (47)	62 (74)
porcine elastase (E)	—	—	—	31 (41)	55 (69)
bovine thrombin (Th)	—	—	—	—	51 (56)

† Chemical similarities are scored on the basis of Lys=Arg; Asp=Glu; Asp=Asn; Glu=Gln; Ser=Thr; Val=Ile; Ile=Leu; Tyr=Phe=Trp.

‡ Residues 16–245.

§ Excluding chymotrypsinogen B.

‖ Excluding chymotrypsinogen A.

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Thrombin

The homologies between the pancreatic enzymes might, however, represent descent from a relatively recent *pancreatic* ancestor rather than the essential skeleton of a *proteinase*. Liver (which has a common embryological origin with pancreas) secretes into plasma a zymogen, prothrombin, whose activation to the serine proteinase, thrombin, is the control step of blood clotting. The activity and specificity of thrombin are high towards its natural substrate, fibrinogen, but much lower towards most peptides or synthetic substrates such as benzoyl arginine amide (Sherry & Troll 1954). Some properties of prothrombin and thrombin are summarized in table 2. The molecular weight of prothrombin is much greater than those of the pancreatic

Table 2. Some properties of bovine prothrombin and thrombin

See review by Harmison & Mammen (1967).

	prothrombin	thrombin
molecular weight	68 500	33 700
N-terminal residue (mol/mol)	Ala (1.0)	Ile (0.9), Thr (0.5)
C-terminal residue	none found	Arg (0.8), Ser (0.1)†
carbohydrate	10 to 12 %	5 to 6 %‡
mol/mol		
sialic acid	9.5	1.7
galactose	6.0	1.0
mannose	5.3	1.5
glucosamine	6.8	2.5

† Magnusson & Steele (1965).
‡ S. M. Magnusson (private communication).

zymogens, and is essentially unchanged after performic oxidation followed by maleylation (S. M. Magnusson & B. S. Hartley, unpublished experiments), so the zymogen must be a single peptide chain as indicated by the end groups. Both prothrombin and thrombin are glycoproteins. The activation process is extremely complex (review by Harmison & Mammen 1967) but it is possible that the activation of chymotrypsinogen, so convincingly explained by Sigler, Blow, Matthews & Henderson (1968) may be an adequate model for its essential features. For these reasons Dr S. M. Magnusson and I have begun to determine the primary structure of bovine thrombin.

We find that the enzyme consists of two chains: an A chain of 49 residues linked by a single disulphide bridge to a B chain of about 250 residues. The sequence of the A chain (figure 4) (S. M. Magnusson, E. Merler, J. Wootton & B. S. Hartley, unpublished evidence) exhibits no detectable homology with any part of any of the pancreatic proteinases, and may be an 'activation peptide' attached to the main chain by a disulphide bridge, like the A chain of chymotrypsin. We have isolated and sequenced tryptic peptides from the B chain accounting for 205 of the 250 residues, and have been able to extend and 'overlap' some of these with fragments from chymotryptic, peptic or cyanogen bromide cleavage. The latter allows us to guess the arrangement of the chain, as shown in figures 2 and 3.

Of the four disulphide bridges common to the pancreatic enzymes, the histidine loop (A), the methionine loop (B) and the serine loop (C) are probably identical in thrombin, but bridge D (residues 136 to 201) is absent, since residue 201 is methionine in thrombin. Another bridge joins Cys-A22 of the A chain to some position in the region 112–134 of the B chain. All the

Downloaded from rstb.royalsocietypublishing.org

carbohydrate is found in a single peptide derived from the region 59 to 67 close to the 'histidine loop'. Those sequences in the B chain which we can so far locate by homology are shown in figure 3. The N-terminal homology and that around Asp-194 suggest that thrombin probably has the internal ion pair involved in the activation of chymotrypsinogen (Sigler *et al.* 1968). Though we have not yet definitely located His-57, the homology around Asp-102 and Ser-195 indicates that thrombin has all the components of the 'charge relay system' (Blow, Birktoft & Hartley 1969). The C-terminal sequence suggests that the B chain of thrombin may be the C-terminal portion of prothrombin, since the C-terminal serine is an unlikely point of cleavage in the activation process. The overall homology between the residues so far tentatively located and comparable regions of chymotrypsin, trypsin and elastase is summarized in table 1. Thrombin is clearly almost as similar to the pancreatic proteinases as they are to each other, and resembles trypsin rather more closely than chymotrypsin or elastase.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Thr -	Ser -	Glu -	Asn -	His -	Phe -	Glu -	Pro -	Phe -	Phe -	Asx -	Glx -	Lys -	Thr -	Phe -	Gly -	Ala -
18	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**
Gly -	Glu -	Ala -	Asp -	Cys -	Gly -	Leu -	Arg -	Pro -	Leu -	Phe -	Glu -	Lys -	Lys -	Glx -	Val -	Glx -
35	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**		
Asx -	Glx -	Thr -	Gln -	Lys -	Glu -	Leu -	Phe -	Glu -	Ser -	Tyr -	Ile -	Glu -	Gly -	Arg		

FIGURE 4. Amino acid sequence of the A chain of bovine thrombin.

HOMOLOGIES OF TERTIARY STRUCTURE

The sequences shown in figure 3 appear in a new light in view of the tertiary structure of bovine α-chymotrypsin, discussed in the preceding paper. This figure indicates side chains in α-chymotrypsin which may be considered 'internal', judged by the rough criterion that they appear inaccessible to solvent. The homology is clearly greatest in these areas which contain, in each enzyme, predominantly hydrophobic (Pro, Cys, Val, Met, Ile, Leu, Tyr, Phe or Trp) or 'ambiphilic' side chains (Ser, Thr, Gly, Ala or His). The hydrophilic side chains (Asn, Asp,

TABLE 3. DISTRIBUTION OF HOMOLOGIES IN RESIDUES 16 TO 245

	chymotrypsin A			trypsin			elastase		
	int.	ext.	total	int.	ext.	total	int.	ext.	total
identical†	70	70	140	62	72	134	60	66	126
similar‡	5	17	22	7	21	28	12	19	31
different	8	60	68	11	50	61	11	72	83
total	83	147	230	80	143	223	83	157	240
hydrophobic §	54	29	83	52	33	85	49	43	92
'ambiphilic' ‖	26	67	93	23	64	87	30	61	91
hydrophilic ¶	3	51	54	5	46	51	4	53	57
total	83	147	230	80	143	223	83	157	240

† In any pair of these three enzymes.
‡ Lys=Arg, Asp=Glu, Asp=Asn, Glu=Gln, Ser=Thr, Val=Ile, Ile=Leu, Tyr=Phe=Trp.
§ Pro, Cys, Val, Met, Ile, Leu, Tyr, Phe, Trp.
‖ Ser, Thr, Gly, Ala, His.
¶ Asn, Asp, Gln, Glu, Lys, Arg.

Downloaded from rstb.royalsocietypublishing.org

Gln, Glu, Lys or Arg) are found almost exclusively in positions which are 'external' in α-chymotrypsin. Exceptions to this latter rule include Asp-102 and Asp-194 which are the 'buried' acidic groups referred to by Birktoft *et al.* (this volume, p. 67), and Gln-30 which is internally hydrogen-bonded to Thr-139 and to the main chain. The other exceptions, Lys-60 and Gln-210 in trypsin and Asp-60 in elastase are residues which lie very close to the surface in chymotrypsin. Table 3 summarizes the distribution of 'internal' and 'external' residues defined on this basis. It is apparent that the whole internal arrangement of polypeptide chains could be almost identical.

Models of trypsin and elastase

Dr J. Birktoft and Mr R. J. Wolverson have therefore constructed, in our laboratory, models of trypsin and elastase to fit the electron density map of α-chymotrypsin. The arrangement of polypeptide chain and of homologous side chains was made identical to that in chymotrypsin except where 'deletions' or 'insertions' occur in the sequences. It was obvious from these models that the differences in amino acid sequences can be readily accommodated without any serious steric problem. Nowhere in the trypsin or elastase models did a hydrophilic group replace an internal hydrophobic group of chymotrypsin. Indeed, the only problem which arose in the model building was traced to an error in the elastase sequence! As an example of the changes which occur, an internal cluster composed of Trp-29, Ser-45, Val-53, Val-200, Leu-209, Val-210 and Ile-212 in chymotrypsin becomes Ser-29, Thr-45, Met-53, His-200, Val-209, His-210 and Val-212 in elastase, without any necessary disturbance of the polypeptide chain. One wonders by what evolutionary steps this internal compensation has been achieved.

The models are also revealing in sections of sequence where 'deletions' or 'insertions' are presumed to occur. In all cases these occur at external loops of the peptide chain where it is easy to add or remove residues without disturbing the general course of the chain. An example, shown in figure 5, plate 54, represents the deletion in the trypsin model of residues 205–208 which form the external portion of the 'serine loop' in chymotrypsin illustrated by Birktoft *et al.* (figure 3, facing p. 78).

The models allow, and indeed positively seem to encourage, the interactions which are important for the activity of chymotrypsin, e.g. the 'charge–relay system' of Asp-102, His-57 and Ser-195 and the internal ion pair from Asp-194 to the N-terminal amino group. It is therefore almost certain that the catalytic activity and mechanism of action of all these enzymes are identical. In support of this latter point, a new method for determining the pK_a and reactivity of individual amino groups in native proteins (Kaplan, Stevenson & Hartley 1969) has shown that the pK_a of the N-terminal amino group in elastase is 9.7 and that the reactivity of the uncharged nucleophile towards acetic anhydride is about 1 % of that of the same group in the free N-terminal peptide, Val-Val-Gly-Gly-Thr-Glu. For comparison, the reactivities and pK_a of Lys-87 and Lys-228 are normal (both pK_a 10.3). Hence the N-terminus is 'buried' in both the protonated and unprotonated form.

The 'tosyl hole'

Perhaps the most dramatic features to emerge from the models concerns the structures corresponding to the 'tosyl-hole' of chymotrypsin, which appears to be the binding site for substrates of chymotrypsin (see preceding paper). In the trypsin model, the side chains which line this hole are identical except for Ser-189 of chymotrypsin. In the sequences for trypsin of both

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

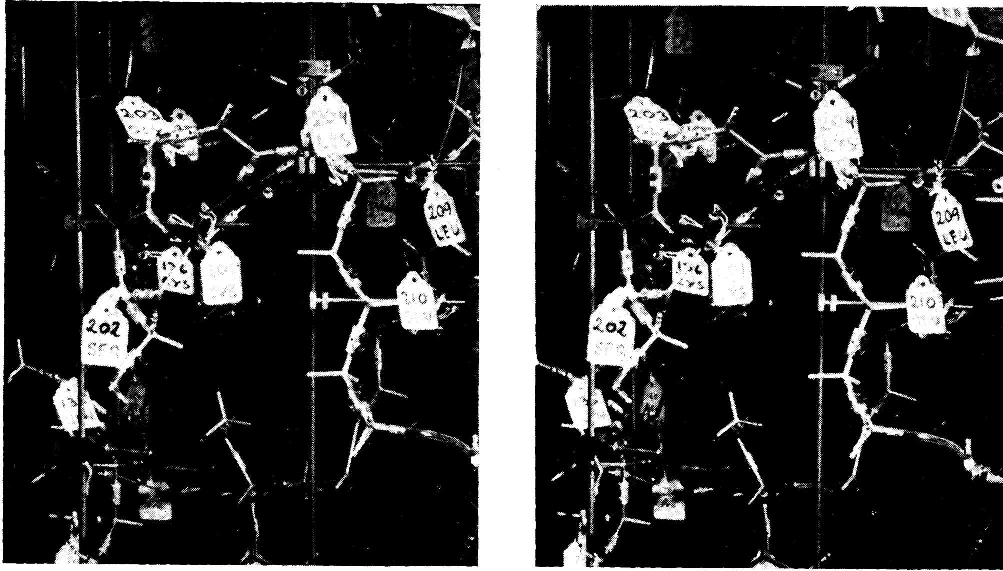


FIGURE 5. Structure of the 'serine loop deletion' in the hypothetical model of trypsin.

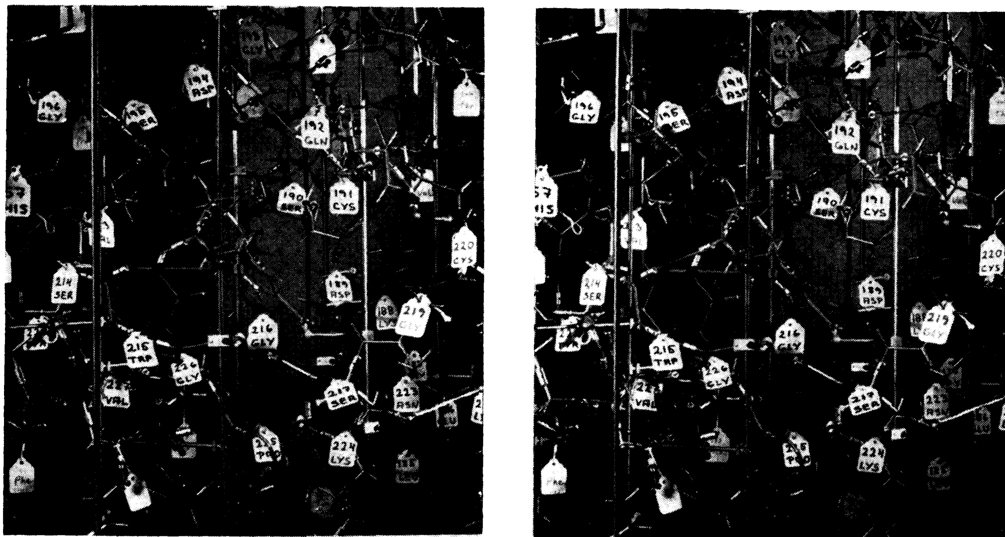


FIGURE 6. The 'tosyl-hole' of the trypsin model containing the substrate *N*-formyl-L-argininamide. Note the interaction between the guanidine group and Asp-189 and the carbonyl group of the substrate and Ser-195.

FIGURE 7. The 'tosyl-hole' of the hypothetical model of elastase. Note that Val-216 (replacing Gly-216 in chymotrypsin) blocks the entrance to this hole.

Downloaded from rstb.royalsocietypublishing.org

Walsh *et al.* (1964) and Mikeš *et al.* (1966) this residue is asparagine. This offered no obvious explanation of the specificity of trypsin for arginine and lysine substrates. We therefore isolated, by mild techniques, the peptide Asx-Ser-Cya-Glx-Gly-Asx-Ser-Gly-Gly-Pro-Val-Val-Cya-Ser-Gly-Lys (residues 189–204) in high yield from a tryptic digest of performic oxidized trypsin. The amino acid analysis and electrophoretic mobility of this peptide indicated a net charge of −3 to −4 at pH 6.5 (Offord 1966). The N-terminal sequence was Asx-Ser-Cya- by the 'dansyl-Edman' technique (Gray & Hartley 1963) and after removal of Asx- by Edman degradation the mobility decreased to indicate a net charge of −2 to −3 at pH 6.5. Hence residue 189 must be aspartic acid. Figure 6, plate 54, demonstrates how a substrate such as formyl-L-arginyl amide fits perfectly into the tosyl hole in the trypsin model to allow a salt-bridge between the guanidino-group and Asp-189, orienting the substrate to allow attack of Ser-195 on the carbonyl group.

In the elastase model, the tosyl hole did not appear since the side-chain of Val-216 (replacing Gly-216 in chymotrypsin or trypsin) blocked the entrance. Moreover, Thr-226 (replacing Gly-226) appeared to fill up the bottom of this hole. These features are illustrated in figure 7, plate 54. We therefore predicted that the tosyl group in tosyl-elastase must lie outside this hole. Our models were built before, and without knowledge of the crystallographic evidence of Shotton & Watson (this volume, p. 111). A close comparison between our hypothetical model and the electron density map of these authors has revealed that the model is adequate in all but minor particulars. The tosyl hole, for example, in the Shotton–Watson model is almost identical to that in ours (see figure 7 of Shotton & Watson, facing p. 116). The absence of this binding site probably accounts for the low residue specificity of elastase, and it is probable that other structures may be involved in the binding of the enzyme to elastin. The good agreement between the hypothetical model and the true structure of elastase gives us confidence that the trypsin model may be adequate, and encourages one to think of building models of other serine proteinases on the same basis. We have not yet attempted to do this for bovine thrombin since the sequence work is not complete, but we believe that such a model will prove particularly illuminating in respect of the A–B chain disulphide bridge, the position of the glycopeptide and the possible binding sites for fibrinopeptides.

Serine proteinases in micro-organisms

At the risk of destroying a tidy story, it is necessary to mention the serine proteinases found in micro-organisms. It was found that the reactive serine in proteinases from *Proteus vulgaris*, *Aspergillus oryzae* and several strains of *Bacillus subtilis* (subtilisin) occurred in the sequence Thr-Ser-Met-Ala (Shaw 1962). Moreover, subtilisin contained no disulphide bridges (Fuke, Matsubara & Okunuki 1959). This naturally led to the conclusion that another class of serine proteinase exists in micro-organisms, which may have evolved from a different ancestor to that common to the mammalian enzymes. The complete sequence of two types of subtilisin (Smith *et al.* 1966) showed that though 67 % of the two sequences were identical, they showed no detectable homology with any of the mammalian serine proteinases.

We were thus left with the hypothesis of two separate lines of evolution for serine proteinases, from a 'Gly-Asp-Ser-Gly' ancestor in mammals, which might be related to a 'Gly-Glu-Ser-Ala' ancestor of mammalian esterases (reviewed by Oosterbaan & Jansz 1965) and a 'Thr-Ser-Met-Ala' ancestor common to a wide range of micro-organisms.

Downloaded from rstb.royalsocietypublishing.org

Even this modified hypothesis has recently taken a severe knock. A serine proteinase has been discovered in *B. sorangium* which appears to be homologous with the mammalian enzymes, both in its active serine sequence (Whitaker & Roy 1967) and a disulphide-bridged sequence homologous with His-57 (Smillie & Whitaker 1967). This bacterium seems to have a 'mammalian' gene! To finish on a note of high and not very serious speculation, I would like to suggest that an ancestor of *B. sorangium* might have acquired such a gene from an ancestor of a cow, perhaps by the accidental introduction into a mammalian cell of a lysogenic phage directed towards *B. sorangium*. Such a phage could then have lysogenized with the mammalian DNA, and thereby picked up a serine proteinase gene. Returning to the soil by the usual route, the phage would inject the mammalian gene into *B. sorangium*. In other words, the bacterium might have been infected by a cow!

References (Hartley)

Blow, D. M., Birktoft, J. J. & Hartley, B. S. 1969 *Nature, Lond.* **221**, 337.

Brown, J. R., Greenshields, R. N., Yamasaki, M. & Neurath, H. 1963 *Biochemistry* **2**, 867.

Dixon, G. H., Kauffman, D. L. & Neurath, H. 1958 *J. Am. chem. Soc.* **80**, 1260.

Enenkel, A. G. & Smillie, L. B. 1963 *Biochemistry* **2**, 1449.

Fuke, I., Matsubara, H. & Okunuki, K. 1959 *J. Biochem.* **46**, 1513.

Gray, W. R. & Hartley, B. S. 1963 *Biochem. J.* **89**, 379.

Gutfreund, H. & Sturtevant, J. M. 1956 *Biochem. J.* **63**, 656.

Harmison, C. R. & Mammen, E. F. 1967 In *Blood clotting enzymology* (ed. W. H. Seegers), p. 23. London: Academic Press.

Hartley, B. S. 1960 *Ann. rev. Biochem.* **29**, 45.

Hartley, B. S. 1964 *Nature, Lond.* **201**, 1284.

Hartley, B. S., Brown, J. R., Kauffman, D. L. & Smillies, L. B. 1965 *Nature, Lond.* **207**, 1157.

Hartley, B. S. & Kauffman, D. L. 1966 *Biochem. J.* **101**, 229.

Hartley, B. S. & Kilby, B. R. 1954 *Biochem. J.* **56**, 288.

Hartley, B. S., Naughton, M. A. & Sanger, F. 1959 *Biochim. biophys. Acta* **34**, 243.

Johansen, J. T., Ottesen, M., Svendsen, I. & Wybrandt, G. 1968 *C.r. Trav. Lab. Carlsberg* **36**, 365.

Kaplan, H., Hartley, B. S. & Stevenson, K. J. 1969 *Fedn Proc. Fedn Am. Socs exp. Biol.* **28**, 533.

Keller, P. J., Cohen, E. & Neurath, H. 1958 *J. biol. Chem.* **233**, 344.

Magnusson, S. M. & Steele, B. 1965 *Ark. Kemi* **24**, 359.

Mikeš, O., Tomášek, V., Holeyšovský, V. & Šorm, F. 1966 *Biochim. biophys. Acta* **117**, 281.

Naughton, M. A. & Sanger, F. 1961 *Biochem. J.* **78**, 156.

Offord, R. E. 1966 *Nature, Lond.* **211**, 591.

Ong, E. B., Shaw, E. & Shoellmann, G. 1964 *J. Am. chem. Soc.* **86**, 1271.

Oosterbaan, R. A. & van Adrichem, M. E. 1958 *Biochim. biophys. Acta* **27**, 423.

Oosterbaan, R. A. & Jansz, H. S. 1965 In *Comprehensive biochemistry*, vol. 16, p. 1 (ed. M. Florkin and E. H. Stotz), p. 1. Amsterdam: Elsevier.

Pospíšilová, D., Meloun, B. & Šorm, F. 1964 *Abstr. 1st Meeting Fed. European Biochem. Soc.* A 31. London: Academic Press.

Schaffer, N. K., Simel, L., Harshman, S., Engle, R. R. & Drisko, R. W. 1957 *J. biol. Chem.* **225**, 197.

Shaw, D. C. 1962 Ph.D. thesis, University of Cambridge.

Shaw, E., Mares-Guia, M. & Cohen, W. 1965 *Biochemistry* **4**, 2219.

Sherry, S. & Troll, W. 1954 *J. biol. Chem.* **208**, 95.

Shotton, D. M. & Hartley, B. S. 1970 To be published.

Sigler, P. B., Blow, D. M., Matthews, B. W. & Henderson, R. 1968 *J. molec. Biol.* **35**, 143.

Smillie, L. B. & Hartley, B. S. 1964 *Abstr. 1st Meeting Fed. European Biochem. Soc.* A 30. London: Academic Press.

Smillie, L. B. & Whitaker, D. R. 1967 *J. Am. chem. Soc.* **89**, 3350.

Smillies, L. B., Furka, A., Nagabhushan, N., Stevenson, K. J. & Parkes, C. O. 1968 *Nature, Lond.* **218**, 343.

Smith, E. L., Markland, F. S., Kasper, C. B., De Lange, R. J., Landon, M. & Evans, W. H. 1966 *J. biol. Chem.* **241**, 5974.

Turba, F. & Gundlach, G. 1955 *Biochem. Z.* **327**, 186.

Walsh, K. A., Kauffman, D. L., Kumar, K. S. V. S. & Neurath, H. 1964 *Proc. natn. Acad. Sci. U.S.A.* **51**, 301.

Whitaker, D. R. & Roy, C. 1967 *Can. J. Biochem.* **45**, 917.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

Discussion

Professor H. Neurath (*University of Washington, U.S.A.*). If I may expand on the observations presented here by Dr Hartley, I should like to add fraction II of the bovine procarboxypeptidase A complex to the intra-special serine proteinases cited by Dr Hartley. The active form of this zymogen has similar substrate specificity to chymotrypsin. Partial peptide sequences around the reactive serine and histidine are homologous to bovine trypsin and chymotrypsin and the same type of histidine peptides have been observed in the enzymically inactive fraction III of bovine procarboxypeptidase A (W. D. Behnke, unpublished).

The earliest indications of the existence of at least two subclasses of serine proteinases were indeed the two different types of DIP-peptides characteristic of the mammalian and bacterial types. Today, we have more complete structural data to substantiate the hypothesis that these two subclasses have fundamentally distinct structures, representing, as it seems, the products of independent evolution. The three dimensional models of α-chymotrypsin discussed this morning by Dr Blow, and of subtilisin BPN′ to be discussed by Dr Kraut, leave no doubt that in these two enzymes essentially the same active site has been formed by two different structural approaches. One wonders how many additional 'prototypes' of serine proteinases exist in nature? There are several other enzymes that meet the criteria of 'serine proteinases' that bear no obvious phylogenetic relation to bovine trypsin or to bacterial subtilisin. For example, the enzyme 'cocoonase' which aids the escape of the moth from the cocoon is a trypsin-like enzyme (Kafatos, Tartakoff & Law 1967). In our own laboratory, we have isolated a trypsin-like enzyme from the starfish *Evasterias trochelii* (W. P. Winter, unpublished) and a trypsinogen-like zymogen, anionic at neutral pH, from an elasmobranch, the Pacific Spiny Dogfish (R. Haynes & W. P. Winter, unpublished). An anionic trypsinogen also occurs in pancreatic extracts of the African lungfish *Protopterus* (G. R. Reeck & W. P. Winter, unpublished). The α-lytic proteinase from *Sorangium* (Smillie & Whitaker 1967) shows traits of homology to bovine chymotrypsin. It will indeed be important to determine whether the linear and three dimensional structures of these various serine proteinases are similar or whether, in addition to the mammalian and bacterial types, additional subclasses of serine proteinases exist, each perhaps characterizing an independent evolutionary event.

Several years ago, Hartley (1960) proposed a classification of proteolytic enzymes based on four different mechanisms of action known at that time: the serine proteinases, the metallo exopeptidases, the sulphydryl proteases and the acid proteinases. These represent different functional solutions to the same chemical problem of enzymic peptide bond cleavage, each requiring a unique active site. I am not aware of any additional class of proteinases that has since been found. It is quite possible, however, as in the case of the serine proteinases, that the same active site may have been arrived at in more than one way, and hence it will be most important to compare, for instance, the three-dimensional structure presented by Dr Drenth and co-workers for papain with that of other sulphydryl enzymes, such as ficin or the streptococcal proteinase of Ferdinand, Stein & Moore (1965). While at one time it may have appeared that 'if you have seen one, you have seen them all', I believe that it is becoming increasingly evident that the structural differences within a given class of enzymes will be as revealing as are the similarities which define the class.

References (Neurath)

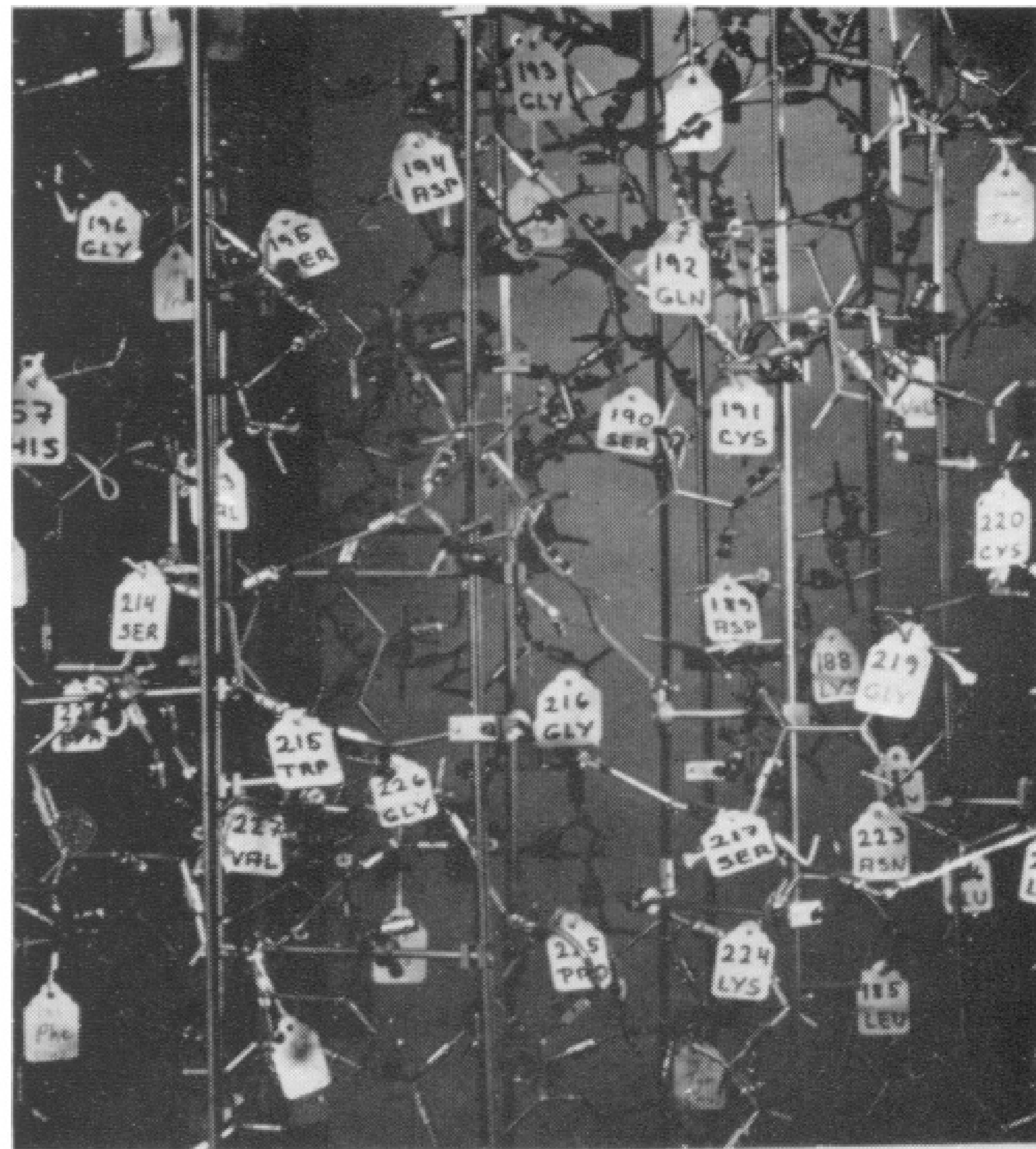
Ferdinand, W., Stein, W. H. & Moore, S. 1965 *J. biol. Chem.* **240**, 1150.
Hartley, B. S. 1960 *Ann. Rev. Biochem.* **29**, 45.
Kafatos, F. C., Tartakoff, A. M. & Law, J. H. 1967 *J. biol. Chem.* **242**, 1477, 1488.
Smillie, L. B. & Whitaker, D. R. 1967 *J. Am. chem. Soc.* **89**, 3350.

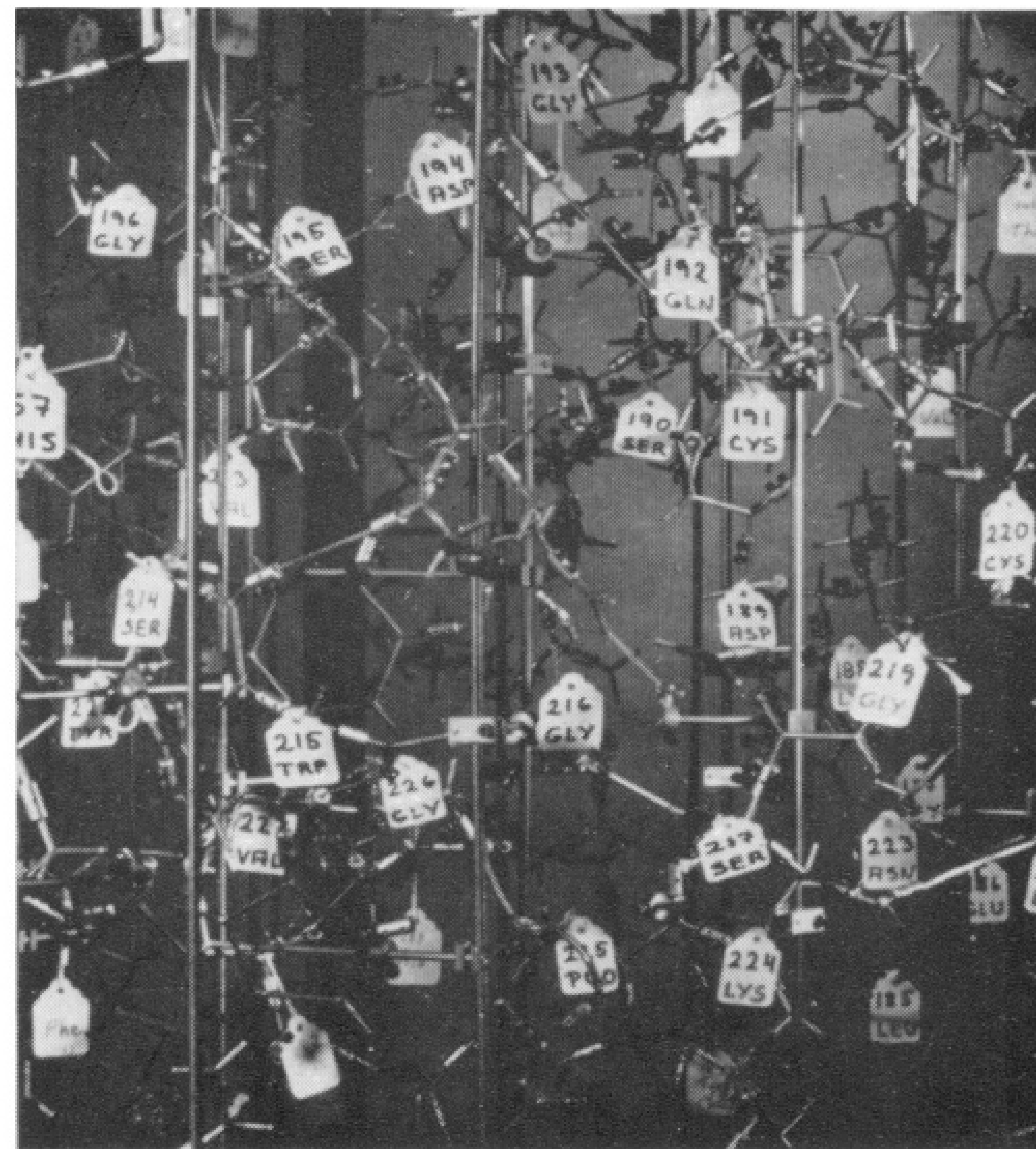
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B
BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

FIGURE 5. Structure of the 'serine loop deletion' in the hypothetical model of trypsin.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B
BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B





IGURE 6. The 'tosyl-hole' of the trypsin model containing the substrate *N*-formyl-L-argininamide. Note the interaction between the guanidine group and Asp-189 and the carbonyl group of the substrate and Ser-195.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

FIGURE 7. The 'tosyl-hole' of the hypothetical model of elastase. Note that Val-216 (replacing Gly-216 in chymotrypsin) blocks the entrance to this hole.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY