Amino Acid Sequences around the Disulphide Bridges and Methionine Residues of Porcine Pepsin

BY J. TANG AND B. S. HARTLEY

Oklahoma Medical Research Foundation and Department of Biochemistry and Molecular Biology, University of Oklahoma School of Medicine, Oklahoma City, Okla. 73104, U.S.A., and Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

(Received 26 January 1970)

1. The amino acid sequences around three disulphide bridges and four methionine residues of porcine pepsin were studied by using diagonal electrophoresis methods. 2. Two of the three disulphide bridges were in small loops of five and six residues. The sequence around one of the two half-cystine residues of the third disulphide bridge had a large number of acidic residues. 3. The sequence of a tetrapeptide containing phosphoserine was also determined. 4. Four unique methionine-containing sequences were constructed. The information is sufficient for the determination of the overlaps in the cyanogen bromide fragments of pepsin. 5. The usefulness of diagonal methods in the study of protein structure, the relative positions of cystinyl and methionyl residues in porcine pepsin and the homology between pepsin and rennin are discussed.

Diagonal paper electrophoresis (Brown & Hartley, 1966) has been used successfully in the elucidation of the disulphide bridges of several proteins and has also proved useful in the selective purification of peptides containing cysteic acid (Brown, Kauffman & Hartley, 1967). This made it possible to compare the sequences around disulphide bridges in homologous proteins (Hartley, Brown, Kauffman & Smillie, 1965) as well as to explore the partial structure of proteins with molecular weights too large for conventional sequence methods (Weeds & Hartley, 1968). Development of diagonal electrophoretic methods for methionine peptides (Tang & Hartley, 1967) and lysine or *N*-terminal peptides (Perham & Jones, 1967; Butler, Harris, Hartley & Leberman, 1969) has extended the range of this strategy.

Porcine pepsin is a protein whose sequence is difficult to study by conventional techniques, since it contains only one lysine and two arginine residues in a total of 321 residues (Rajagopalan, Stein & Moore, 1966), and these are concentrated at the *C*-terminus of the molecule (Perham & Jones, 1967; Stepanov, Ostoslavskaya, Krivtsov, Muratova & Levin, 1967; Dopheide, Moore & Stein, 1967). Tryptic digestion therefore yields an *N*-terminal peptide of 302 residues, and it is obvious that alternative methods of cleaving the chain of pepsin must be used. Enzymes of lower specificity than trypsin naturally tend to give more complex digests, and selective purification techniques are

particularly welcome in such a case. We therefore applied the cystine diagonal technique of Brown & Hartley (1966) to determine the sequences around the disulphide bridges of pepsin, and the methionine diagonal technique of Tang & Hartley (1967) to determine the methionine-containing sequences in this enzyme.

MATERIALS AND METHODS

Chemicals. Crystalline porcine pepsin, pepsinogen, papain, bovine α -chymotrypsin and trypsin were products from Worthington Biochemical Corp., Freehold, N.J., U.S.A. Subtilisin was purchased from Novo Terapeutisk Laboratorium, Copenhagen, Denmark. Other chemicals used were of the highest purity purchased from commercial sources and used without further purification.

Peptic digests of pepsin. Pepsin (120mg) was dissolved in 4ml of water. This solution was made alkaline by the addition of a few drops of aq. NH_3 (sp.gr. 0.88) and left standing at room temperature for 30min to denature the pepsin. Then 2ml of 0.8M-formic acid was added to the enzyme solution, causing the appearance of a large amount of white precipitate, and next 3mg of crystalline pepsin was added. Most of the precipitate disappeared during 12h incubation at 37°C. After 20h the digest was centrifuged and the remaining insoluble material was discarded.

Digestion of pepsin with other proteinases. A procedure that avoids autolysis was devised to denature pepsin before digestion by other proteinases. Pepsin (500mg) was suspended in 5ml of acetone in a 25ml beaker and 3M- NH_3 was added dropwise with constant stirring until most of the pepsin had dissolved to give a 'gel-like' solution,

and thereafter more rapidly until the total volume reached about 20ml. The solution was stirred at room temperature for 30min, dialysed against several changes of water at 2°C overnight and freeze-dried.

A portion of the resulting powder (80mg) was dissolved in 5ml of 0.2M-*N*-ethylmorpholine-acetate buffer, pH8.0, containing 1mg of α -chymotrypsin, and incubated overnight at 37°C.

Digestion with papain was carried out as follows: 80mg of the denatured pepsin was dissolved in 5ml of 0.1M-pyridine-acetate buffer, pH6.5. Mercaptoethanol (0.2ml) and papain (about 1mg of enzyme in aqueous suspension) were added and the solution was incubated at 37°C for 20h.

The conditions used for digestion with subtilisin were identical with those used with papain except that mercaptoethanol was omitted. The substrate/enzyme ratio was about 120:1 by wt.

Purification of peptides. The digests were applied as bands on Whatman 3MM paper at loadings of approx. 1mg of peptides/cm and subjected to high-voltage paper electrophoresis at approx. 60V/cm in pyridine-acetate buffer, pH6.5 or pH3.5, as described by Brown & Hartley (1966). Marker strips (2cm wide) were cut from these sheets and either exposed to the vapour of performic acid to oxidize cysteine residues to cysteic acid, or treated with iodoacetamide solutions at pH3.5 to convert methionine into CAM*-methionine residues. When re-run at right-angles to the original direction of electrophoresis, these strips gave diagonal electrophoretic 'maps' of cysteic acid peptides and of CAM-methionine peptides, which were used as a guide for the selective purification of these peptides from the rest of the bands remaining on the original sheet (Brown & Hartley, 1966; Tang & Hartley, 1967).

Peptides on paper were stained by the cadmium-ninhydrin reagent of Heilmann, Barollier & Watzke (1957), or the Pauly reagent (Dent, 1947) to reveal peptides containing histidine or tyrosine.

Mobilities of peptides (*m*) after high-voltage electrophoresis were determined relative to aspartic acid at pH6.5 ($m = -1.0$) or to serine at pH2 ($m = 1.0$), so as to allow calculation of the number of acidic groups (Offord, 1966).

In the preparative runs, the peptides were applied in a band at the origin. After location of the peptide positions, the bands were cut out from the paper and eluted in a moist chamber. The recovery was found to be about 80% as tested with known dipeptides (Tang & Hartley, 1967).

Sequence methods. Amino acid analyses were made by the method of Spackman (1963) in an automatic amino acid analyser. Peptides were hydrolysed in 5.7M-HCl for 18-20h at 105°C in evacuated tubes. Hydrolysates containing CAM-methionine were first heated at 100°C for 2h at pH7 to convert the sulphonium salt into homoserine (Tang & Hartley, 1967).

N-Terminal analyses were made by the 'dansyl' technique (Gray & Hartley, 1963); DNS-amino acids were identified by paper electrophoresis (Gray 1967) or by t.l.c. on polyamide sheets (obtained from Gallard-Schlesinger Chemical Mfg. Corp., Carle Place, L.I., N.Y.,

U.S.A., a product of Cheng Chin Co., Taipei, Taiwan) with the solvents of Woods & Wang (1967) followed by redevelopment in the second dimension with ethyl acetate-methanol-acetic acid (20:1:1, by vol.) (Crowshaw, Jessup & Ramwell, 1967) followed by 0.05M- Na_3PO_4 -ethanol (3:1, v/v). Sequential analysis was made by the 'dansyl'-Edman technique (Gray, 1967). Mobilities of the free peptides or DNS-peptides at pH6.5 were occasionally determined during the course of the sequential degradation in order to distinguish removal of aspartic acid or asparagine and glutamic acid or glutamine.

C-Terminal residues were identified by digesting the peptide with carboxypeptidase A (Ambler, 1967). Liberated amino acids were identified qualitatively by paper electrophoresis at pH2 or quantitatively by amino acid analysis of the digests. Hydrazinolysis (Fraenkel-Conrat & Tsung, 1967) was occasionally used, and the liberated *C*-terminal amino acid identified in the automatic amino acid analyser.

Enzymic digests of peptides were generally made for 18h at 20°C by using the following enzyme/substrate molar ratios in the following buffers: pepsin (enzyme/substrate ratio 1:20) in 5% (v/v) formic acid; trypsin, chymotrypsin or subtilisin (enzyme/substrate ratio 1:100) in 0.2M-*N*-ethylmorpholine-acetate buffer, pH8.0; papain (enzyme/substrate ratio 1:100) in 0.1M-pyridine-acetate buffer, pH6.5, containing 0.2M-mercaptoethanol.

RESULTS

Disulphide bridges of porcine pepsin

Fig. 1 is a diagram of a pH3.5 cystine diagonal 'map' of a chymotryptic digest of pepsin. There are three major peptide spots that lie off this diagonal (C1, C2 and C3). The corresponding peptides were purified from the remaining digest by the method of Brown & Hartley (1966). Amino acid analysis showed that peptides C1 and C2 contained cysteic acid as expected, but that peptide C3 did not. Later sequence studies showed that the latter had a *C*-terminal methionine sulphone, so it is apparent that the fall in *pK* of the carboxyl group when methionine is oxidized to its sulphone causes the peptide to move slightly off the diagonal. There are also a number of peptides lying off the diagonal in Fig. 1 that were present in a quantity insufficient for further sequence studies. The sequences of peptides C1 and C2 were studied and are described below.

A pair of cystine-bridged peptides should give rise to two spots vertically beneath each other in a diagonal 'map' such as Fig. 1, but no partner appears for either peptide C1 or peptide C2. The partners might lie beneath the 'contaminating' peptides in the relevant diagonal positions, so peptides were eluted from such positions and subjected to amino acid analysis. No cysteic acid was found. Moreover, peptides C1 and C2 were both found to be pure on electrophoresis at pH6.5 and by end-group analysis.

* Abbreviations: CAM-, carboxamidomethyl-; Cya, cysteic acid; Hsr, homoserine; Hsl, homoserine lactone; ACy, amino-ethylcysteine.

Fig. 1. Cystine diagonal 'map' of a chymotryptic digest of denatured pepsin, separated by electrophoresis at pH 3.5. DNS-OH and DNS-Arg show positions of bands of internal fluorescent markers applied along the strip after oxidation and before running in the second (vertical) dimension. An external marker of a blue dye (Xylene-Cyanol FF) was applied at one end of the strip before the second electrophoresis. The major spots lying off the diagonal after dipping in cadmium-ninhydrin reagent are shaded and labelled C1, C2 and C3.

Hence both peptides must contain internal disulphide bridges. This deduction was supported by the amino acid analysis of the peptides (Table 1), which showed that each contained two cysteic acid residues.

Sequence of peptide C1. Amino acid analysis (Table 1) showed that peptide C1 contained 20 residues. The results of 'dansyl'-Edman analysis and carboxypeptidase digestion allow us to write the following sequence:

Cya-Ser-Ser-Leu-Ala-Cya-Ser-Asx-His-
(Asx₄, Ser₂, Thr, Glx, Pro, Phe)Phe

A digest of peptide C1 with chymotrypsin at a substrate/enzyme ratio of 10:1 at 20°C for 18 h gave two bands on paper electrophoresis at pH 2. Peptide C1C1 ($m=0$) was further purified by electrophoresis at pH 3.5, but peptide C1C2 ($m=0.45$) was effectively pure. Amino acid analysis (Table 1) revealed that peptide C1C1 represented the *N*-terminal 12 residues of peptide C1, and peptide C1C2 the remaining residues.

Hydrazinolysis showed that peptide C1C1 had

C-terminal phenylalanine, and digestion with carboxypeptidase A released asparagine, glutamine and phenylalanine, which were recognized as their DNS derivatives. A subtilisin digest of peptide C1C1 gave a Pauly-positive peptide, C1C1S1, which was purified by electrophoresis at pH 3.5 and pH 2. Amino acid analysis (Table 1), 'dansyl'-Edman analysis (Fig. 2) and mobilities of the peptides during the Edman degradation proved the sequence Ser-Asp-His-Asn. Hence peptide C1C1 has the sequence:

Cya-Ser-Ser-Leu-Ala-Cya-
Ser-Asp-His-Asn-Gln-Phe

Peptide C1C2 gave the *N*-terminal sequence Asn-Pro-Asp- by the 'dansyl'-Edman procedure with high-voltage electrophoresis of the DNS-peptides. Carboxypeptidase showed *C*-terminal phenylalanine. A partial acid hydrolysate of the peptide in 6M-hydrochloric acid for 20 min at 100°C was fractionated by electrophoresis at pH 3.5. One fragment, C1C2A1, had the sequence Pro-Asp-Ser (Table 1 and Fig. 2). Another type of partial acid

Table 1. *Peptides from pepsin containing cysteic acid*

Amino acid analyses are uncorrected molar ratios. Electrophoretic mobilities are relative to Asp (= -1.0) at pH 6.5 and to Ser (= 1.0) at pH 2. Colours with cadmium-ninhydrin are shown after 1 h or 18 h incubation at room temperature.

Peptide	Amino acid composition (molar ratios)										Electrophoretic mobility		Cadmium-ninhydrin colour
											pH 6.5	pH 2	
C1	Cya	Asp	Thr	Ser	Glu	Pro	Ala	Leu	Phe	His	-0.62		Yellow → Red
	1.7	4.6	1.0	5.4	1.2	1.1	1.0	1.1	1.9	1.0			
C1C1	1.5	1.8	—	3.1	1.0	—	1.0	1.0	0.8	1.0	-0.61	0.00	Yellow → Red
C1C2	—	2.9	1.0	2.8	—	0.8	—	—	1.0	—		0.45	Yellow → Red
C1C1S1	—	2.2	—	1.0	—	—	—	—	—	0.9	—		Red
C1C2A1	—	1.0	—	0.9	—	0.8	—	—	—	—	—		Yellow → Orange
C1C2A2	—	—	0.9	1.9	—	—	—	—	1.0	—	—		Red
C2	Cya	Ser	Glu	Gly							-0.74		Yellow → Red
	1.6	1.0	1.1	2.2									
S1a	Cya	Ser									-0.81		Yellow → Red
	0.8	1.0											
S2a	Cya	Asp	Ser	Glu	Ile						-0.80		Yellow → Red
	0.7	1.0	1.8	1.0	0.9								
S1b	Cya	Asp	Thr	Ser	Glu						-1.09		Red
	1.0	3.3	1.0	2.6	1.4								
S1bA1	—	1.9	—	1.1	1.0						-1.11		Red
A2	1.0	—	—	1.1	—						-0.78		Orange → Red
A3	—	1.0	—	—	—						-1.22		Orange → Red
A4	—	1.0	—	1.2	—						-0.95		Orange → Red
A5	—	2.0	—	1.4	—						-1.05		Orange → Red
S1bT1	ACy	Asp	Ser	Glu							-1.05		Red
	0.6	1.8	1.0	1.3									
S2b	Cya	Asp	Thr	Ser	Glu						-1.17		Red
	1.0	4.4	1.2	3.6	1.4								
S2bA1	0.9	—	1.0	2.2	—						-0.44		Orange → Red
Sp1:	Ser	Glu	Leu								-0.90		Orange → Red
	0.9	1.9	1.0								-0.85		Red
Sp2:	0.9	1.0	—										

hydrolysis of peptide C1 (0.25 M-acetic acid at 100°C for 18 h) gave, among others, peptide C1C2A2, which was purified by electrophoresis at pH 3.5 and had the sequence Ser-Thr-Ser-Phe. The mobility of peptide C1C2 at pH 6.5 indicates a net content of two acidic groups, so we may write the sequence:

Asn-Pro-Asp-Ser-Asp-Ser-Thr-Ser-Phe

Peptide C1 has an electrophoretic mobility that suggests a net content of 4–5 acidic groups, consistent with the above allocation of amide groups, and must have arisen from a peptide with an internal disulphide bridge with the sequence shown in Fig. 2.

Sequence of peptide C2. Amino acid analysis (Table 1) and sequential degradation (Fig. 2) showed the sequence Cya-Ser-Gly-Gly-Cya-Glx. The electrophoretic mobility at pH 6.5 indicates a net content of two acidic groups. A pH 6.5 diagonal 'map' (Brown & Hartley, 1966) showed that the cystine precursor of peptide C1 was neutral, so that the C-terminal residue must be glutamine. D. Marinkovic & B. S. Hartley (unpublished work), using a similar diagonal electrophoretic method, had derived the same sequence for a peptic fragment of pepsin, namely:

Cya-Ser-Gly-Gly-Cya-Gln

Cystine peptides from a subtilisin digest. The third disulphide bridge of pepsin did not appear as a major spot in the diagonal 'map' shown in Fig. 1, but there were a large number of minor peptides lying off the diagonal. We thought that these might arise from a series of non-specific chymotryptic cleavages in the sequence around this third bridge, and therefore investigated diagonal 'maps' of other enzymic digests of denatured pepsin.

Fig. 3 shows the diagonal electrophoretic 'map' at pH 6.5 of a subtilisin digest of pepsin. A collection of minor peptides in area I must arise from neutral precursors such as the internal bridges represented above, but area II indicates a pattern of acidic cystine peptides that could be the third bridge. The partners of peptides S1b and S2b are, of course, not obvious in the 'map', since they are in the diagonal position. Area II was therefore cut out from the original preparative pH 6.5 electrophoresis, oxidized with performic acid vapour and subjected to electrophoresis at pH 2, where peptides containing cysteic acid and no basic amino acid will be neutral and other contaminants will be cationic. The neutral band from this pH 2 electrophoresis was cut out, stitched to another sheet and re-run at pH 6.5.

Fig. 4(a) shows the peptides in a marker strip cut from the electrophoresis at pH 6.5. Peptides S1b, S2b and Sp1 were well separated and were eluted.

Fig. 3. pH 6.5 cystine diagonal 'map' of a subtilisin digest of denatured pepsin. The shaded spots represent the major peptides containing cysteic acid. Peptides S1a and S2a were recognized after further electrophoresis at pH 2 of the peptides in the diagonal position of area II.

Fig. 4. Purification of cysteic acid peptides and phosphoserine peptides from area II in Fig. 3. (a) pH 6.5 electrophoresis of the neutral band obtained after pH 2 electrophoresis of oxidized area II; (b) pH 3.5 electrophoresis of band Sa+Sp2 from (a).

of peptide. Assuming that aspartic acid release is faster than asparagine hydrolysis (Light, 1967), this result could be obtained only from a C-terminal -Asp-Ser sequence.

To further substantiate the above sequence, the diagonal peptides at the position of S1b were eluted after electrophoresis in the first dimension. The mixture of peptides was subjected to reduction and aminoethylation as described by Cole (1967). After passage of the mixture through a column of Sephadex G-15 (50 cm $\times$ 2 cm) which was eluted with 0.1 M-ammonia, the recovered peptides were free of salts and reagents that were used in the reduction and aminoethylation. The aminoethyl-S1b (AE-S1b) was further purified by electrophoresis at pH 6.5 ($m = -0.65$) and then at pH 2 ($m = 0.92$). Peptide AE-S1b had an amino acid composition of: Asp (2.6), Thr (0.8), Ser (2.7), Glu (1.3), ACy (0.7). The N-terminal was found to be glutamic acid, which agreed with that of S1b. Peptide AE-S1b was subjected to tryptic hydrolysis (substrate/enzyme molar ratio about 10:1) and the resulting peptide mixture was separated by electrophoresis at pH 2. A major peptide band (S1bT1, $m = -1.05$) was observed and further purified by electrophoresis at pH 3.5 from minor contaminants. Peptide S1bT1 gave the N-terminal sequence Glu-Asx- by the 'dansyl'-Edman procedure. The composition

of peptide S1bT1 (Table 1) clearly showed that it was derived from the *N*-terminal part of peptide S1b (Fig. 2). This peptide must be Glu-Asx(Asx, Ser)ACy. Attempts to purify the *C*-terminal portion of AE-S1b after tryptic digest were unsuccessful, since there were a number of slow-moving contaminating bands in the electrophoresis at pH 2 that probably derived from trypsin autodigestion.

From the sequence discussed above, we therefore write peptide S1b, tentatively, as:

Glu-Asx-Asx-Ser-Cya-Thr-Ser-Asp-Ser

Peptide S2b has a higher mobility at pH 6.5 than peptide S1b and must therefore have the sequence:

Glu-Asx-Asx-Ser-Cya-Thr-Ser-Asp-Ser-Asp-Ser

Sequence of phosphoserine peptides. Peptides Sp1 and Sp2 contain no cysteic acid (Table 1), but were shown to contain phosphoserine by partial acid hydrolysis in 6M-hydrochloric acid at 100°C

for 1 h. High-voltage paper electrophoresis at pH 3.5 with authentic phosphoserine standards established its presence in both peptides in good yield. This explains why both peptides are neutral at pH 2.

Amino acid analysis (Table 1), sequence analysis (Fig. 2) and the mobility of the peptides at pH 6.5 (Table 1) establish their structure as:

Sequences around the methionine residues of porcine pepsin

Methionine diagonal electrophoresis of peptic and papain digest of porcine pepsin. The methionine

Fig. 5. Methionine diagonal 'map' of a peptic digest of denatured pepsin separated by electrophoresis at pH 6.5. Spots of markers (DNS-OH, DNS-Arg, Xylene-Cyanol FF and Orange G) were applied to the strip before the final electrophoresis to aid determination of mobilities. Major CAM-methionine peptides lying off the diagonal are shaded. P2b indicates the position of a CAM-methionine peptide found in diagonal 'maps' of commercial pepsin but not of pepsinogen.

Fig. 6. Methionine diagonal 'map' of a papain digest of denatured pepsin separated by electrophoresis at pH 6.5. Details were as described for Fig. 5.

Table 2. *Peptides from pepsin containing CAM-methionine*

Details were as described for Table 1.			Mobility at pH 6.5	Cadmium-ninhydrin colour
Peptide	Amino acid composition (molar ratios)			
P1b:	Thr (0.9), Ser (1.0), Hsr (0.7), Gly (1.0)		0.42	Red
P1c:	Thr (1.0), Ser (1.0), Hsr (0.9), Gly (1.7)		0.38	Yellow
P1d:	Thr (1.0), Ser (1.0), Hsr (0.9), Gly (2.1), Tyr (0.8)		0.34	Red
P1dS1:	Thr (0.9), Gly (1.2), Tyr (0.8)			Red
P1dS2:	Gly (1.0), Tyr (0.8)			Red
P1dS3:	Ser (1.0), Hsr (0.9), Gly (1.0)			Red
P1dS4:	Ser (1.0), Hsr (0.8)			Red
P2a:	Asp (1.0), Thr (0.9), Ser (1.0), Hsr (0.7), Ile (1.0)		0.00	Red
P2b:	Asp (1.0), Ser (1.0), Hsr (0.7), Glu (1.1), Gly (1.1), Tyr (0.8)		−0.05	Yellow → Red
P3a:	Hsr (1.0), Glu (1.4), Gly (1.4)		0.00	Yellow → Red
P4:	Asp (1.1), Thr (1.0), Ser (2.2), Hsr (0.4), Glu (1.0), Gly (1.2), Val (0.8), Leu (0.9)		−0.37	Orange → Red
P5:	Asp (1.9), Ser (1.1), Hsr (0.7), Glu (1.1), Gly (1.0)		−0.39	Orange → Red
P6:	Asp (1.1), Hsr (1.0), Glu (1.0), Gly (1.0), Ser (0.3)		−0.43	Red
P7:	Asp (2.0), Hsr (1.0), Glu, (2.0), Gly (2.0), Ser (1.8), Thr (0.8), Val (0.7)		−0.44	Red
P8:	Asp (1.9), Ser (2.2), Hsr (0.9), Glu (1.9), Gly (1.2), Ala (1.1)		−0.50	Red
N5:	Asp (1.0), Thr (1.0), Ser (1.0), Hsr (1.0), Pro (1.0), Gly (1.0), Val (1.1)		0.00	Yellow
N6:	Asp (1.0), Thr (1.0), Ser (1.0), Hsr (0.7), Ile (1.0)		0.00	Red
N7:	Asp (1.0), Thr (1.0), Ser (1.0), Hsr (0.8), Pro (1.3), Val (0.9)		−0.22	Red
N8 + 8a:	Asp (1.8), Ser (1.7), Glu (1.7), Gly (1.1), Ala (0.8), Val (1.0), Met (0.6), Tyr (1.0)		−0.41	Red
N9:	Asp (2.0), Thr (1.2), Ser (1.2), Glu (1.1), Gly (1.3), Ala (0.9), Hsr+Met (0.6), Tyr (1.0)		−0.45	Red

diagonal electrophoresis 'maps' at pH 6.5 of peptic and of papain digests of porcine pepsin are shown in Figs. 5 and 6 respectively. In both cases several

CAM-methionine peptides emerged from the neutral band and a large number of the precursor methionine peptides were acidic. Most of these

CAM-methionine peptides were purified preparatively by the method of Tang & Hartley (1967) and their amino acid compositions and other properties are presented in Table 2. Further studies on the sequence of these peptides revealed that they can be grouped into four unique sequences. For convenience of comparison, peptides that arise from the same sequence are discussed together hereafter.

Methionine sequence I. (a) In the diagonal 'map' of the peptic digest (Fig. 5) peptide P1a was found to be free CAM-methionine, since its electrophoretic mobility was identical with that of an authentic sample and it gave homoserine only after heating at 105°C for 2h at pH 7 (Tang & Hartley, 1967).

The compositions of bands P1b, P1c and P1d (Table 2) were related to that of peptide P1d, which

Fig. 7. Sequences of methionine peptides. Symbols are as defined for Fig. 2. In the peptide nomenclature P represents a product of pepsin digestion, N of papain digestion, S of subtilisin digestion and H of heating the CAM-methionine peptide at 105°C for 2h at pH 7.

was isolated in best yield. Subtilisin digestion of the latter followed by electrophoresis at pH 2 gave four fragments, P1dS1 to P1dS4. The compositions (Table 2) and sequences (Fig. 7) of all these peptides establish the unique sequence:

Tyr-Gly-Thr-Gly-Ser-Hsr

Peptide C3, which emerged from the cystine diagonal of a chymotryptic digest of pepsin (Fig. 1), had the sequence Gly-Thr-Gly-Ser-Met(SO₂), and must arise from the same parent methionine sequence.

(b) In the diagonal 'map' of the papain digest (Fig. 6) peptides N5 and N7 were found to arise from the same sequence. Their compositions are given in Table 2 and the sequence evidence is summarized in Fig. 7.

Peptide N7 gave no end-group by the 'dansyl' method, but Tang & Hartley (1967) showed that this was a property of *N*-terminal CAM-methionine. After heating at pH 7 at 105°C for 2h free homoserine was released and was recognized by pH 2 electrophoresis. 'Dansylation' of the heated peptide gave DNS-homoserine directly and DNS-homoserine + DNS-aspartic acid after acid hydrolysis. Hence the *N*-terminal sequence is probably CAM-Met-Asx-, and further sequential analysis (Fig. 7) established the sequence CAM-Met-Asx-Val-Pro-Thr-Ser. The parent methionine peptide has a mobility consistent with a single negative charge (Fig. 6), so the peptide must contain aspartic acid and not asparagine. This is supported by the results of incubating peptide N7 in 0.3M-acetic acid at 105°C for 18h, when free homoserine, free aspartic acid and a peptide with *N*-terminal valine (*m* = 0.83 at pH 2) were detected after pH 2 electrophoresis. The mobility of peptide N7 itself is, however, puzzling. Perhaps the anionic character arises from depression of the *pK_a* of the *N*-terminal amino group by the adjacent positive charge of the sulphonium salt. We must nevertheless write the sequence of peptide N7 as:

CAM-Met-Asp-Val-Pro-Thr-Ser

Peptide N5 has a composition that clearly overlaps that of peptide N7 (Table 2). 'Dansyl'-Edman analysis on the peptide itself gave Gly-Ser-, but after heating at 105°C for 2h at pH 7 the end groups were glycine+aspartic acid and serine+valine after one round of Edman degradation. The sequence is therefore Gly-Ser(CAM-Met,Asp,Val,Pro,Thr) and Gly-Ser-Hsr plus Asp-Val(Pro,Thr) after heating.

Peptides P1d, N5 and N7 obviously overlap each other, so we can write the following sequence around methionine I:

Tyr-Gly-Thr-Gly-Ser-Met-Asp-Val-Pro-Thr-Ser

Methionine sequence II. Peptides P3a, P5, P8, N8a and N8 were purified and gave the compositions shown in Table 2. The sequence analysis shown in Fig. 7 clearly indicates that they arise from the same methionine sequence. When peptide N8 was heated at pH 7 at 105°C for 2h a mixture of free isoleucine and a peptide (Ile,Val,Tyr) was liberated, showing that a variety of peptide N8 exists with the *C*-terminal sequence -CAM-Met-Ile (peptide N8a).

The mobilities of these peptides indicate that peptide P3a contains glutamic acid, and peptide P5 has one extra acidic group and must therefore contain both aspartic acid and asparagine. Peptides P8 and N8 have mobilities consistent with three acidic groups in the methionine form and two in the CAM-methionine form. We can therefore write the partial structure around methionine II:

Ala-Ser-Glu-Asx-Ser-Asx-
Gly-Glu-Met-Ile(Val,Tyr)

where one Asx is Asp and another Asn.

Methionine sequence III. Table 2 and Fig. 7 show that peptides P2a and N6 are identical, and are overlapped by peptide N9. The electrophoretic mobilities indicate aspartic acid in peptides P2a and N6 and three acidic residues in peptide N9. Hence we may write the partial sequence around methionine III:

Asp-Ser-Ile-Thr-Met(Asp,Glu,Gly)(Ala,Tyr)

Methionine sequence IV. Peptide P6 was found to have the sequence Glx-Gly-Asx-Hsr after conversion of the CAM-methionine into homoserine by heating. The electrophoretic mobilities show that there are two acidic residues, so the sequence must be Glu-Gly-Asp-CAM-Met. Two small contaminants, peptides 6a and 6b, were found in this fraction on further purification by pH 3.5 electrophoresis. Both had similar compositions; Asx (1.6), Ser (0.9), Hsr (0.7), Glu (0.9), Gly (1.0); and both had the *N*-terminal sequence Asx-Ser-. It is likely that they arise from methionine sequence II and are similar to peptide P5, one having lost an amide group.

Only preliminary results are available for peptide P7. 'Dansyl'-Edman analysis indicated that it had the *N*-terminal sequence Glx-Gly-Asx-Ser-. When heated to cleave the CAM-methionine residue, no fragment was released, suggesting that methionine was *C*-terminal. The composition (Table 2) does not allow the peptide to arise from methionine sequences I, II or III, so we may write the tentative partial structure Glx-Gly-Asx-Ser(Thr,Ser,Val)-Glu-Gly-Asp-CAM-Met, on the assumption that it arises from methionine sequence IV.

Peptide P4 gave no *N*-terminal residue on dansylation, but gave aspartic acid/asparagine after

Fig. 8. Partial amino acid sequence of porcine pepsin (P) and homologues with calf rennin (R). Sequences mentioned in this work are shown in **bold** type. Residues that are identical in porcine pepsin and calf rennin are shown in *italics*. Residues that can be approximately located in the chain of pepsin are numbered. Sets of peptides that can be ascribed to a particular region of the chain are enclosed in square brackets [] and individually separated by/. Pepsin sequences are from the summary of Keil (1970) and rennin sequences from Foltmann & Hartley (1967), Foltmann (1968) and B. Foltmann (personal communication).

one round of Edman degradation. When the peptide was heated at 105°C for 2 h in water, a new *N*-terminal aspartic acid/asparagine residue appeared. Hence this peptide is probably CAM-Met-Asx(Thr, Ser₂, Glu, Gly, Val, Leu) and may be the *C*-terminal continuation of the methionine IV sequence. In this case, the partial sequence around methionine IV would be:

Glx-Gly-Asx-Ser(Thr, Ser, Val)Glu-Gly-Asp-
Met-Asx(Thr, Ser₂, Glu, Gly, Val, Leu)

Methionine peptide impurity?

In methionine diagonal 'maps' of peptic digests of commercial crystalline pepsin, a spot shown as peptide P2b in Fig. 5 was observed. This peptide was isolated and found to have the sequence Asn-Ser-Gly-Glu-Tyr-CAM-Met. Rajagopalan *et al.* (1966) have shown that commercial crystalline pepsin may be impure, but that commercial crystalline pepsinogen is essentially pure and contains only four methionine residues. When diagonal 'maps' of peptic digests of crystalline pepsinogen were prepared, peptide P2b was absent, although the pattern was otherwise similar to that shown in Fig. 5. We therefore believe that peptide P2b is a methionine sequence derived from an impurity in commercial crystalline pepsin.

DISCUSSION

The purpose of this work was to test a new strategy for determining the partial amino acid sequence of a protein. Selective cleavages can be made at methionine residues by cyanogen bromide cleavage (Gross & Witkop, 1962) and the *N*-terminal sequences and compositions of such fragments could be determined relatively easily. Selective purification of methionine peptides by diagonal electrophoresis allows an easy access to peptides that will overlap the cyanogen bromide pieces and thereby arrange them in the correct sequence. Tryptic digests of the cyanogen bromide fragments would then allow one to allocate the tryptic peptides to appropriate regions of the chain. Selective cleavage at cysteine or cystine residues in a tryptic peptide can be achieved by reaction with ethyleneimine (after reduction of cystine) and subsequent tryptic cleavage of the aminoethyl-cysteinyl bands (Raftery & Cole, 1963). Selective purification of cysteine acid peptides by the cystine diagonal technique again provides the required overlap peptides. In this way the positions of disulphide bridges and methionine residues in the polypeptide chain could be established without determining the complete amino acid sequence.

We chose pepsin as our model, partly because of the difficulties it presents for more conventional sequence strategies, and partly because we wished to establish whether it showed sequence homologies with another gastric proteinase, rennin. Most of the work described in this paper was done in 1966-67, and since then the value of this strategy has become apparent. The partial amino acid sequences described above have already established that two of the three disulphide bridges are homologous in pepsin and rennin (Foltmann & Hartley, 1967) and have allowed the correct ordering of fragments obtained from cleavage of pepsin at methionyl or aminoethyl-cysteinyl bonds (Keil, 1970). We had hoped to complete the methionine sequences before submitting this paper, but have been persuaded that the partial sequences will be of value to the many groups investigating the primary structure of pepsin.

Fig. 8 shows the position of the peptides mentioned in this paper in the partial sequence of pepsin, summarized by Keil (1970) from the results of several workers. The methionine in sequence III can be located as position 158, and the cystine bridge I at positions 105-110 from the amino acid compositions of the *N*-terminal cyanogen bromide fragment and a 105-residue tryptic fragment obtained after aminoethylation of the latter (Trufanov, Kostka, Keil & Šorm, 1969) that appears to contain the phosphoserine sequence (Stepanov, 1968; Keil, 1970). Isolation of the *C*-terminal cyanogen bromide fragment (Stepanov *et al.* 1967) shows that the methionine in sequence I lies at position 272. The remaining two cystine bridges and the other two methionine sequences must therefore lie between residues 164 and 266, which also contains the active-centre aspartic acid sequence (Bayliss, Knowles & Wybrandt, 1969), and we expect that the above sequences will be useful in elucidating this third of the molecule.

The homologies between pepsin and rennin are also illustrated in Fig. 8. Cystine bridges I and II are highly homologous, as are the *C*-terminal 27 residues, but cystine bridge III shows only little identity in the two enzymes (Foltmann & Hartley, 1967). The *N*-terminal sequence of rennin (Foltmann, 1968) shows no similarity to that of pepsin, consistent with the fact that rennin contains 270 residues compared with about 320 residues in pepsin, and that both have homologous *C*-terminal sequences.

Part of this work was done during the tenure by J. T. of a John Simon Guggenheim Fellowship at the Medical Research Council Laboratory of Molecular Biology, Cambridge. The continuation of this work at Oklahoma Medical Research Foundation was supported by U.S. Public Health Service Grants AM-01107 and AM-06487 from the National Institute of Health.

REFERENCES

- Ambler, R. P. (1967). In *Methods in Enzymology*, vol. 11, p. 417. Ed. by Hirs, C. H. W. New York: Academic Press Inc.
- Bayliss, R. S., Knowles, J. R. & Wybrandt, G. B. (1969). *Biochem. J.* **113**, 377.
- Brown, J. R. & Hartley, B. S. (1966). *Biochem. J.* **101**, 214.
- Brown, J. R., Kauffman, D. L. & Hartley, B. S. (1967). *Biochem. J.* **105**, 1125.
- Butler, P. J. G., Harris, J. I., Hartley, B. S. & Leberman, R. (1969). *Biochem. J.* **112**, 679.
- Cole, D. (1967). In *Methods in Enzymology*, vol. 11, p. 315. Ed. by Hirs, C. H. W. New York: Academic Press Inc.
- Crowshaw, K., Jessup, S. J. & Ramwell, P. W. (1967). *Biochem. J.* **108**, 79.
- Dent, C. E. (1947). *Biochem. J.* **41**, 240.
- Dopheide, D. A. A., Moore, S. & Stein, W. H. (1967). *J. biol. Chem.* **242**, 1833.
- Foltmann, B. (1968). *Abstr. FEBS 5th Meet., Prague*, p. 727.
- Foltmann, B. & Hartley, B. S. (1967). *Biochem. J.* **104**, 1064.
- Fraenkel-Conrat, H. & Tsung, C. M. (1967). In *Methods in Enzymology*, vol. 11, p. 151. Ed. by Hirs, C. H. W. New York: Academic Press Inc.
- Gray, W. R. (1967). In *Methods in Enzymology*, vol. 11, p. 469. Ed. by Hirs, C. H. W. New York: Academic Press Inc.
- Gray, W. R. & Hartley, B. S. (1963). *Biochem. J.* **89**, 59P.
- Gross, E. & Witkop, B. (1962). *J. biol. Chem.* **237**, 1856.
- Hartley, B. S., Brown, J. R., Kauffman, D. L. & Smillie, L. B. (1965). *Nature, Lond.*, **207**, 1157.
- Heilmann, J., Barollier, J. & Watzke, E. (1957). *Hoppe-Seyler's Z. physiol. Chem.* **309**, 219.
- Keil, B. (1970). *Phil. Trans. R. Soc. B*, **257** (in the Press).
- Light, A. (1967). In *Methods in Enzymology*, vol. 11, p. 417. Ed. by Hirs, C. H. W. New York: Academic Press Inc.
- Offord, R. E. (1966). *Nature, Lond.*, **211**, 591.
- Perham, R. N. & Jones, G. M. T. (1967). *Eur. J. Biochem.* **2**, 84.
- Raftery, M. A. & Cole, R. D. (1963). *Biochem. biophys. Res. Commun.* **10**, 467.
- Rajagopalan, T. G., Stein, W. H. & Moore, S. (1966). *J. biol. Chem.* **241**, 4295.
- Spackman, D. H. (1963). *Fedn Proc. Fedn Am. Socs exp. Biol.* **22**, 244.
- Stepanov, V. M. (1968). *Abstr. FEBS 5th Meet., Prague*, p. 1089.
- Stepanov, V. M., Ostoslavskaya, V. I., Krivtsov, V. F., Muratova, G. L. & Levin, E. D. (1967). *Biochim. biophys. Acta*, **140**, 182.
- Tang, J. & Hartley, B. S. (1967). *Biochem. J.* **102**, 593.
- Trufanov, V. A., Kostka, V., Keil, B. & Šorm, F. (1969). *Eur. J. Biochem.* **7**, 544.
- Weeds, A. G. & Hartley, B. S. (1968). *Biochem. J.* **107**, 531.
- Woods, K. R. & Wang, K. T. (1967). *Biochim. biophys. Acta*, **133**, 369.