A Diagonal Paper-Electrophoretic Technique for Studying Amino Acid Sequences around the Cysteine and Cystine Residues of Proteins

By R. N. PERHAM

Department of Biochemistry, University of Cambridge

(Received 11 May 1967)

1. A diagonal electrophoretic technique for studying the amino acid sequence around cysteine and cystine residues in proteins is described. The residues are first converted into *S*-aminoethylcysteine, and the protein is then treated with *S*-ethyl trifluorothioacetate, which trifluoroacetylates all the protein amino groups. The modified protein is digested enzymically and the resulting peptides are separated by paper electrophoresis. After exposure of the peptides on the paper to ammonia vapour, the electrophoresis is repeated, this time at right angles to the original direction. Peptides from which a trifluoroacetyl group is removed by the ammonia treatment will vacate the 45° diagonal formed by all other unaffected peptides owing to the exposure of an additional amino group and consequent increased electrophoretic mobility towards the cathode. Peptides containing lysine or *S*-aminoethylcysteine are readily purified by this technique. 2. The successful application of the technique to bovine insulin is described. 3. Various methods for distinguishing peptides containing lysine from those containing *S*-aminoethylcysteine in more complicated proteins are suggested and discussed.

Many of the recent advances in our knowledge of the chemistry of proteins, and hence of their biological behaviour, have stemmed from a study of the function of certain amino acid residues in the expression of the biological activity of the proteins. The identification of part, at least, of the active site of a number of enzymes is a particularly well documented example of this approach (Sanger, 1963; Goodwin, Harris & Hartley, 1964). More generally, there is now considerable interest in the existence of amino acid-sequence (or alternatively messenger-RNA base-sequence) homology between proteins of related function, e.g. proteolytic enzymes (Walsh & Neurath, 1964; Hartley, Brown, Kauffman & Smillie, 1965) and the haemoglobins and myoglobin (Braunitzer, 1966). Short of a complete determination of the amino acid sequence of the proteins concerned, each one in itself a major undertaking, this problem is generally attacked by investigating selectively the structure of corresponding areas of the relevant proteins. Thus considerable use of the paper-electrophoretic diagonal technique for determining disulphide bridges described by Brown & Hartley (1963, 1966) was made in studies of the disulphide bridges of human immunoglobulin light chains (Milstein, 1966) and the *C*-terminal sequences of mouse myeloma proteins (Perham, Appella & Potter, 1966). In addition, Tang & Hartley (1967) have introduced a

diagonal electrophoretic method for the selective purification of methionine peptides, with the same purpose in mind.

Perham & Jones (1967) have described a new electrophoretic diagonal technique for the rapid ordering of the lysine-containing tryptic peptides of proteins, which allows the easy preferential isolation of peptides containing lysine, together with the *N*-terminal peptide, from enzymic digests of a protein. This method can also, in favourable cases, enable the lysine (and *N*-terminal) sequences of proteins to be readily compared. The present work was undertaken to determine whether the technique could be extended to the cysteine and cystine residues of proteins through their conversion into *S*-aminoethylcysteine (analogous to lysine) by reduction and aminoethylation (Raftery & Cole, 1963), followed by submission to the paper-electrophoretic diagonal procedure of Perham & Jones (1967). Such an extension of technique would be valuable on three grounds. First, cysteine and cystine residues are often of considerable interest in the catalytic behaviour of enzymes. Secondly, the ability to study these residues selectively in proteins would also facilitate investigations of amino acid-sequence homology. Thirdly, reduction and aminoethylation of cysteine and cystine residues is becoming increasingly common as a means of promoting tryptic cleavage at these

residues in protein chains during structural studies (Raftery & Cole, 1966) and it would be convenient if the 'overlapping' technique of Perham & Jones (1967) could be applied to the tryptic peptides derived from such cleavages.

MATERIALS AND METHODS

Enzymes and reagents. Bovine insulin (lot no. 2345) was purchased from Allen and Hanbury Ltd., London, W. 1, and subtilisin (EC 3.4.4.16) from the British Drug Houses Ltd., Poole, Dorset. Ethyleneimine was a product of Koch-Light Laboratories Ltd., Colnbrook, Bucks. *S*-Ethyl trifluoroethioacetate was purchased from Eastman Kodak, Rochester, N.Y., U.S.A.

Preparation of aminoethylated insulin. A 31 mg. sample of insulin was dissolved in 4 ml. of fresh 8M-urea solution in *m*-tris-HCl buffer, pH 8.9, containing 0.05M-2-mercaptoethanol. After 2 hr. at room temperature (20°), 1 ml. of *m*-tris-HCl buffer, pH 8.9, was added, followed by 0.15 ml. of ethyleneimine. After a further 2 hr. at room temperature the reaction mixture was applied to a column of Sephadex G-25 (35 cm. × 2 cm.) equilibrated with 0.2N-acetic acid. The column was developed at a flow rate of 30 ml./hr. and 3 ml. fractions were collected. The modified insulin, detected in the column effluent by its extinction at 280 mμ, was recovered by freeze-drying. The yield was 28 mg. No cystine could be detected in amino acid analyses of the insulin after reaction.

Reaction with *S*-ethyl trifluoroethioacetate. A 25 mg. sample of aminoethylated insulin was dissolved in 5 ml. of water and the pH adjusted to 9.5 with *N*-NaOH. A fine precipitate appeared at this stage. Sufficient solid urea was added to cause this precipitate to dissolve and 0.5 ml. of *S*-ethyl trifluoroethioacetate was added. The pH was maintained at 9.5 by the addition of *N*-NaOH. After 1½ hr., when the uptake of alkali had practically ceased, the reaction mixture was applied to a column of Sephadex G-25 (32 cm. × 2 cm.) equilibrated with 0.5% NH₄HCO₃, pH 8.0. The column was developed at a flow rate of 30 ml./hr. and 3 ml. fractions were collected. The trifluoroacetyl derivative of aminoethylated insulin was recovered by freeze-drying. The yield was 12 mg. No free amino group could be detected by reaction with fluorodinitrobenzene (Sanger, 1945).

Digestion with subtilisin. A 10 mg. sample of trifluoroacetyl aminoethylated insulin was suspended in 2 ml. of 0.5% NH₄HCO₃, pH 8.0. (After freeze-drying from this solvent trifluoroacetyl aminoethylated insulin was difficult to redissolve. This poor solubility may explain the relatively poor recovery from the Sephadex G-25 column used in the preparation method described above.) Digestion with 0.1 mg. of subtilisin at 37° was continued for 3 hr., during which time the peptides rapidly dissolved, and the buffer was then removed by freeze-drying.

Paper electrophoresis and chromatography. High-voltage paper electrophoresis was carried out by the procedure of Ryle, Sanger, Smith & Kitai (1955), in a modification of the apparatus of Michl (1951). The buffers used were: pH 6.5, pyridine-acetic acid-water (25:1:225, by vol.); pH 3.5, pyridine-acetic acid-water (1:10:189, by vol.); pH 2.0, acetic acid-formic acid-water (4:1:45, by vol.). Electrophoresis was generally performed at a potential gradient of 60 v/cm. for 50–100 min.

Paper chromatography of peptides was done in the descending system (butan-1-ol-acetic acid-water-pyridine 15:3:12:10, by vol.) of Waley & Watson (1953).

Peptides were eluted with 0.05N-NH₃ solution.

Preparation of electrophoretic diagonal peptide 'maps'. A 10 mg. sample of the subtilisin digest of trifluoroacetyl aminoethylated insulin was subjected to electrophoresis at pH 6.5 as a band 12 cm. wide on Whatman 3MM paper. When the paper was dry, a guide strip 1.5 cm. wide was cut from the edge and placed in a closed desiccator over a dish of 5N-NH₃, as described by Perham & Jones (1967). After 3–4 hr. the strip was removed and dried. It was then sewn on to a fresh sheet of paper and resubjected to electrophoresis at pH 6.5, this time at right angles to the original direction. In this way peptides from which a trifluoroacetyl group is removed by NH₃ treatment will acquire an additional positive charge as a result of the exposure of an amino group, and thereby vacate the 45° diagonal formed by peptides not so affected. The increased electrophoretic mobility will occur, of course, towards the negative electrode (cathode).

Peptides on the paper were detected by a ninhydrin-cadmium reagent (Heilmann, Barrolier & Watzke, 1957).

Amino acid analysis. Samples of proteins and peptides were hydrolysed for 24 hr. at 105° with 6N-HCl in sealed evacuated tubes. The HCl was removed *in vacuo* over fresh NaOH pellets and the amino acids were analysed on a Technicon auto-analyser. The values for serine and threonine were corrected, by 6% and 3% respectively, for destruction losses during 24 hr. hydrolysis.

***N*-Terminal groups of peptides.** *N*-Terminal amino acids of peptides were determined by the DNS* chloride method of Gray & Hartley (1963). For routine qualitative identification of *N*-terminal groups the hydrolysis of the DNS-peptide was carried out with 6N-HCl in sealed tubes for 12 hr. at 105°. With peptides containing aminoethylcysteine, however, special precautions were taken to evacuate the tube carefully on an oil pump before it was sealed (see the Results section).

RESULTS

An electrophoretic diagonal peptide 'map' for the subtilisin digest of trifluoroacetyl aminoethylated insulin is shown in Fig. 1. A number of peptides were seen to move off the diagonal formed by peptides unaffected by ammonia treatment. This may be compared with a similar 'map' prepared from the subtilisin digest of the trifluoroacetyl derivative of performic acid-oxidized insulin (Perham & Jones, 1967) shown in Fig. 2. The peptides shown hatched in Figs. 1 and 2 gave a yellow colour with the ninhydrin-cadmium stain, consistent with *N*-terminal glycine or threonine, whereas the other peptides gave the more usual red coloration. It is evident that several additional peptides were affected by ammonia treatment in the case of the aminoethylated insulin compared with performic acid-oxidized insulin.

* Abbreviations: DNS, 1-dimethylaminonaphthalene-5-sulphonyl; AEC and CySO₃H (in amino acid sequences), *S*-aminoethylcysteine and cysteic acid residues.

Fig. 1. Diagonal peptide 'map' of the subtilisin digest of trifluoroacetyl aminoethylated insulin. Electrophoresis was at pH 6.5 at 60 v/cm. for 50 min. in each direction. The major 'off-diagonal' peptides are shown bounded by continuous lines and minor peptides by broken lines. Off-diagonal peptides shown hatched gave a yellow colour with the ninhydrin-cadmium stain. ↑, Origin.

Fig. 2. Diagonal peptide 'map' of the subtilisin digest of trifluoroacetyl performic acid-oxidized insulin. Electrophoresis was at pH 6.5 at 60 v/cm. for 50 min. in each direction. The major 'off-diagonal' peptides are shown bounded by continuous lines and minor peptides by broken lines. Off-diagonal peptides shown hatched gave a yellow colour with the ninhydrin-cadmium stain. ↑, Origin.

The peptides moving off the diagonal shown in Fig. 2 have already been identified (Perham & Jones, 1967) and are listed in Table 1. To identify the peptides moving off the diagonal in the aminoethylated insulin 'map' (Fig. 1) the following procedure (described in detail by Brown & Hartley, 1966) was adopted. The areas of the main body of the ionogram containing peptides affected by ammonia treatment, as located by the diagonal electrophoresis, were excised and exposed to ammonia vapour in the normal way. The paper bands were then sewn on to fresh sheets of paper and

Table 1. Amino acid sequences of 'off-diagonal' peptides derived from performic acid-oxidized insulin (Fig. 2)

Peptide	Amino acid sequence
S1*	Gly-Ile-Val-Glu-Gln-CySO ₃ H-CySO ₃ H-Ala-Ser-Val-CySO ₃ H
S2†	Phe-Val-Asn
S3	Thr-Pro-Lys-Ala
	* <i>N</i> -Terminal sequence of the A-chain.
	† <i>N</i> -Terminal sequence of the B-chain.

subjected to electrophoresis at pH 6.5 in a direction at right angles to the bands. The peptides were located by staining guide strips with the ninhydrin-cadmium reagent. Only those peptides moving out of the 'diagonal' position at pH 6.5 were studied further. These peptides were submitted successively to electrophoresis at pH 3.5, chromatography and electrophoresis at pH 2.0, and were finally eluted from the paper. Suitable samples of each peptide were subjected to end-group and amino acid analysis, the results of which are given in Table 2. On the basis of these analyses and from the known amino acid sequence of insulin (Sanger, 1959), the peptides may be unequivocally identified and are listed in Table 3.

There was some initial difficulty in establishing the *N*-terminal residue of peptides with *N*-terminal aminoethylcysteine. On DNS chloride treatment and acid hydrolysis these peptides gave predominantly one yellow fluorescent derivative, which migrated with DNS-glutamic acid in the electrophoretic separation at pH 4.40 devised by Gray & Hartley (1963). Hydrolysis of the DNS-peptide in carefully evacuated tubes, however, led to the disappearance of this artifact and to the expected appearance of α,ω -bis- and ω -mono-substituted derivatives of aminoethylcysteine, which behaved electrophoretically like the corresponding bis- and ϵ -DNS-lysine. It would seem likely that acid hydrolysis in unevacuated tubes caused oxidative cleavage of the amino acid side chain at the sulphur atom. (In support of this interpretation, it has been found that in some cases aminoethylcysteine does not survive the performic acid oxidation and acid hydrolysis of peptides containing it.) Hydrolysis of DNS derivatives of all peptides containing aminoethylcysteine was subsequently carried out in carefully evacuated tubes and no further difficulty in end-group identification was encountered. (Some of these experiments were done in collaboration with Mr A. Holmgren, to whom I express my thanks.)

The amino acid sequence of the A- and B-chains of insulin and the 'off-diagonal' peptides found in the present study are shown in Fig. 3. It will be

Table 2. *Amino acid compositions and N-terminal analyses of 'off-diagonal' peptides derived from aminoethylated insulin (Fig. 1)*

Peptide ...	S1a1	S1a2	S1b	S2a1	S2a2	S2b1	S2b2	S2b3	S2c	S2d
Amino acid	Amino acid composition (μ moles/mole)									
Asp	1.03	1.97				0.97				
Thr†					0.98*					
Ser†				1.05				0.99*		
Glu		1.10*	1.76							
Pro					1.09					
Gly			1.12*							
Ala				1.03	1.03		0.96			
Val	0.96		0.53†					0.96	1.00*	
Ile			0.50†							
Tyr		0.91								
Phe	0.94*									
AEC		1.00	0.94	0.99*		1.04*	1.05*	1.08	0.98	1.00*
Lys					0.97					

* N-Terminal amino acid.

† Incompletely released on 24 hr. hydrolysis.

‡ Corrected for destruction losses during acid hydrolysis.

Table 3. *Amino acid sequences of 'off-diagonal' peptides derived from aminoethylated insulin (Fig. 1)*

Peptide	Amino acid sequence
S1a1	Phe-Val-Asn*
S1a2	Glu-Asn-Tyr-AEC-Asn
S1b	Gly-Ile-Val-Glu-Gln-AEC†
S2a1	AEC-Ala-Ser
S2a2	Thr-Pro-Lys-Ala
S2b1	AEC-Asn
S2b2	AEC-Ala
S2b3	Ser-Val-AEC
S2c	Val-AEC
S2d	AEC

* N-Terminal sequence of the B-chain.

† N-Terminal sequence of the A-chain.

seen that, with one reservation, it is possible to ascribe off-diagonal peptides to all aminoethylcysteine residues with the exception of residue 7 in the B-chain. No peptide containing this residue was detected. It is entirely possible, of course, that the free aminoethylcysteine (peptide S2d) is derived from this sequence, thereby explaining the apparent absence of a corresponding peptide. The reservation concerns residue 19 in the B-chain. Peptide S2c can theoretically be derived both from this residue and from residue 11 in the A-chain, as indicated. No peptide that could be uniquely assigned to residue 19 of the B-chain was found, and to this extent the dual placing of peptide S2c remains conjecture. However, the enzymic cleavages of peptide bonds necessary to this interpretation are very reasonable in view of the observed specifi-

city of subtilisin elsewhere in the modified insulin molecule.

DISCUSSION

The experiments described show clearly that peptides containing *S*-aminoethylcysteine can be readily isolated from enzymic digests of proteins and peptides by the diagonal electrophoretic procedure originally devised for lysine peptides. However, with a protein or peptide containing many more lysine residues than insulin, confusion may occur between off-diagonal peptides containing lysine and those containing aminoethylcysteine. The diagonal peptide 'map' may, indeed, look very complicated. (The increase in complexity for aminoethylated insulin compared with performic acid-oxidized insulin is noteworthy.) Two ways round this problem immediately present themselves, although it should be emphasized that neither has yet been applied as a routine in practice.

One solution would be to use ^{14}C -labelled ethyleneimine for the aminoethylation reaction. Off-diagonal peptides containing aminoethylcysteine would then be radioactive and could easily be resolved from those containing lysine by radioautography of the peptide 'maps'. This technique would have the additional advantage of making it less easy to overlook an off-diagonal aminoethylcysteine peptide, particularly during studies of amino acid-sequence homology carried out on large proteins.

An alternative (and less expensive) method would be to acetylate all the amino groups of the protein before the disulphide bridges (or free thiol groups) are converted into aminoethylcysteine and

Fig. 3. Amino acid sequence of the aminoethylated A- and B-chains of bovine insulin and the 'off-diagonal' peptides produced by digestion with subtilisin.

trifluoroacetylated. Acetylation is not reversed by exposure to ammonia and hence the off-diagonal peptides should be only those containing aminoethylcysteine, with a correspondingly simple peptide 'map'. It is likely that most totally acetylated proteins will have low solubility in ordinary aqueous solvents and buffers but, in view of the fact that the aminoethylation and trifluoroacetylation reactions are best carried out in strongly denaturing solution, e.g. 8M-urea, this should present little hindrance. If, however, solubility problems do become insuperable, the amino groups of the protein could be succinylated (Habeb, Cassidy & Singer, 1958) instead of acetylated, which should yield a very soluble protein derivative for further experiments with diagonal electrophoresis.

As with previous electrophoretic diagonal methods, the final yield of peptides is not good (generally 10–30% of the theoretical yield) but, again, the advantages of speed, convenience and ease of manipulation remain to offset this disadvantage. It is noteworthy that the subtilisin used in these experiments has frequently caused cleavage of the polypeptide chain at ω -N-trifluoroacetylaminocysteine residues, a point that may be of some further use in structural studies on proteins.

Since the completion of this work, an alternative diagonal electrophoretic method for lysine peptides, based on the reversible reaction of protein amino groups with maleic anhydride, has been suggested (Butler, Harris, Hartley & Leberman, 1967). There appears to be no reason why this technique could not also be used to obtain S-aminoethylcysteine peptides by using the approaches described above.

I am grateful to Professor F. G. Young for his interest, to Dr C. J. R. Thorne for performing the amino acid analyses and to Mrs Louise Oliver for skilled technical assistance.

REFERENCES

- Braunitzer, G. (1966). *J. cell. Physiol.* **67**, (Suppl. 1), 1.
 Brown, J. R. & Hartley, B. S. (1963). *Biochem. J.* **89**, 59 p.
 Brown, J. R. & Hartley, B. S. (1966). *Biochem. J.* **101**, 214.
 Butler, P. J. G., Harris, J. I., Hartley, B. S. & Leberman, R. (1967). *Biochem. J.* **103**, 78 p.
 Goodwin, T. W., Harris, J. I. & Hartley, B. S. (Eds.) (1964). *Structure and Activity of Enzymes*. London: Academic Press (Inc.) Ltd.
 Gray, W. R. & Hartley, B. S. (1963). *Biochem. J.* **89**, 59 p.
 Habeb, A. F. S. A., Cassidy, H. G. & Singer, S. J. (1958). *Biochim. biophys. Acta*, **29**, 587.
 Hartley, B. S., Brown, J. R., Kauffman, D. L. & Smillie, L. B. (1965). *Nature, Lond.*, **207**, 1157.
 Heilmann, J., Barrolier, J. & Watzke, E. (1957). *Hoppe-Seyl. Z.* **309**, 219.
 Michl, H. (1951). *Mh. Chem.* **82**, 489.
 Milstein, C. (1966). *Biochem. J.* **101**, 338.
 Perham, R. N., Appella, E. & Potter, M. (1966). *Science*, **154**, 391.
 Perham, R. N. & Jones, G. M. T. (1967). *Europ. J. Biochem.* **2**, 84.
 Raftery, M. A. & Cole, R. D. (1963). *Biochem. biophys. Res. Commun.* **10**, 467.
 Raftery, M. A. & Cole, R. D. (1966). *J. biol. Chem.* **241**, 3457.
 Ryle, A. P., Sanger, F., Smith, L. F. & Kitai, R. (1955). *Biochem. J.* **60**, 541.
 Sanger, F. (1945). *Biochem. J.* **39**, 507.
 Sanger, F. (1959). *Science*, **129**, 1340.
 Sanger, F. (1963). *Proc. chem. Soc., Lond.*, p. 76.
 Tang, J. & Hartley, B. S. (1967). *Biochem. J.* **102**, 593.
 Waley, S. G. & Watson, J. (1953). *Biochem. J.* **55**, 328.
 Walsh, K. A. & Neurath, H. (1964). *Proc. nat. Acad. Sci., Wash.*, **52**, 884.