Journal of Chromatography, 129 (1976) 211–237

CHROM. 9271

IMPROVED TECHNIQUE FOR THE ANALYSIS OF AMINO ACIDS AND RELATED COMPOUNDS ON THIN LAYERS OF CELLULOSE

XI. DIRECT, VISUAL MICRO-METHOD FOR INVESTIGATING THE PRIMARY STRUCTURAL SEQUENCE OF COLLAGEN AND OTHER PROTEINS

J. G. HEATHCOTE and SAMIRAH J. AL-ALAWI

Department of Chemistry and Applied Chemistry, The University of Salford, Salford M5 4WT (Great Britain) and The Skin Hospital, University of Manchester, School of Medicine, Quay Street, Manchester 3 (Great Britain)

(First received November 27th, 1975; revised manuscript received April 5th, 1976)

SUMMARY

The present paper describes a new variation of the "finger printing" technique for examining the primary structure of proteins. It is based on the production of a reproducible, partial acid hydrolysate consisting of amino acids and small peptides which are separated into compact spots by bi-dimensional thin-layer chromatography. The spots are characterized by their R_F values and colour reactions with a variety of reagents. The method was initially developed for collagen but has been applied to other proteins, with interesting results. It has developed logically from earlier, systematic studies on amino acids and peptides, and holds promise for the comparative investigation of normal and diseased human collagen.

INTRODUCTION

Earlier papers in this series[1–3] described methods for the systematic characterization of amino acids by thin-layer chromatography (TLC) and selective staining. The studies were later extended to peptides in order to provide an independent method for the detection of unknown peptides which might interfere with the amino acid analysis of biological fluids by ion-exchange chromatography[4–3]. It was shown that many small peptides could be characterized, in part at least, by these methods, and the suggestion arose that they might form the basis of a direct, visual micro-method for examining the primary structure of a protein. Such a method should be of particular value in the study of abnormal proteins which differ from their normal counterparts only by the substitution of one amino acid in the sequence. Provided that a satisfactory method of partial hydrolysis could be found, TLC separation of the fragments together with staining techniques should provide a refined variation of the finger printing technique. Most finger printing methods have followed the early

technique used by Ingram[9] in which enzymic hydrolysis and a combination of paper chromatography and electrophoresis were used for the separation of the fragments. Acid hydrolysis has usually been avoided in the past and there have even been reports of sequence alteration caused by this method.

We were particularly interested in developing a method for the comparative analysis of collagen in order to study the genetic bone disease, *Osteogenesis imperfecta*, and, therefore, required a method which could be applied to small samples (2–20 mg) of the protein. Because the collagen molecule has some 3,000 amino acid residues in its three chains and because most of the structure which has been elucidated is highly repetitive, it is a difficult protein sequence in which to identify the point of a single amino acid change.

It was essential, if bidimensional TLC was to be used for the separation of fragments, that the method of hydrolysis should yield peptides of a suitably small molecular size to give a reproducible pattern of discrete, small spots. In nearly all structural studies of collagen, cyanogen bromide is the method of choice for splitting the collagen chains into fragments. We abandoned this method when it became clear that the fragments which were obtained were far too large for TLC separation and decided on the use of partial acid hydrolysis[10]. The micro-technique finally developed gave a pattern of 61 small, disparate spots from a hydrolysate of bovine achilles tendon collagen, many of which were readily identifiable on a 20 cm × 20 cm cellulose thin layer. The results proved to be surprisingly consistent and the method was accordingly applied to other proteins of known structure. The results for two of these, glucagon and insulin, are described below, and the value of the method may be judged from the fact that six tryptophanyl peptides, as well as free tryptophan, were detected in a hydrolysate of glucagon which contains only one tryptophan residue in the molecule.

EXPERIMENTAL

Apparatus

The thin-layer equipment, glass tanks and 1-μl "Microcaps" capillary pipettes used throughout this work were supplied by Shandon (London, Great Britain).

Materials and methods

The adsorbent used was cellulose powder MN 300 (without binder), supplied by Macherey, Nagel and Co. (Agents: Camlab, Cambridge, Great Britain). Before use this powder was purified as described previously in Part I[1].

The solvents used throughout this work were of Analar grade, with the exception of 2-methyl-2-butanol (G.P.R.) and butanone and propanone (M.F.R.), and were supplied by Hopkin and Williams (Chadwell Heath, Essex, Great Britain).

Control amino acids and synthetic peptides, chiefly di- and tripeptides, were obtained from Sigma (Kingston-upon-Thames, Surrey, Great Britain). Solutions (0.025 *M*) were made in aqueous 2-propanol (10%, v/v) and kept refrigerated before use. Native insoluble bovine achilles tendon collagen (prepared as described by Einbinder and Schubert[11]), glucagon, and crystalline insulin were also obtained from Sigma.

Standard hydrochloric acid (1 *M*) for partial hydrolysis of proteins was made

from constant-boiling HCl, as described by Foulk and Hollingsworth[12]. The weight of constant-boiling HCl at 747.5 mm Hg pressure required to make 1 l of accurate standard HCl (1 *M*) was estimated to be 179.9302 g.

The following chromatographic solvent systems were used: Solvent A, which we found best for the separation of the fragments of hydrolysates in the first dimension, was that described by Haworth and Heathcote[1]; it consisted of propan-2-ol–butanone–1 *M* HCl (60:15:25). For the separation in the second dimension it was necessary to develop a new solvent system (Solvent B), which consisted of 2-methyl-2-butanol–butanone–propanone–methanol–(0.88) ammonia (25:20:35:5:20).

A list of reagents for the detection and identification of amino acids and peptides is given in Table I.

Partial acid hydrolysis. Samples of bovine achilles tendon collagen (1–25 mg) were subjected to partial acid hydrolysis[10] (1 *M* HCl, 2 ml) under nitrogen at 100° over a period of 3 h; the hydrolysates were freeze-dried and adjusted to a concentration of 10 μg/μl with distilled water.

Thin-layer chromatography. The cellulose layers were prepared as follows: The purified cellulose powder (15 g) was spread as a slurry over five plates (20 cm × 20

TABLE I

CHROMOGENIC REAGENTS FOR AMINO ACIDS AND PEPTIDES

No.	*Reagent*	*Reference*	*Used for the detection of*	*Note*
1	Cd-ninhydrin	1, 3	General amino acids and peptides and N-Gly peptides	10% of glacial acetic acid was used in the preparation of the reagent
2	Platinic iodide	13	Sulphur-containing amino acids and peptides	
3	Zimmerman's *o*-phthalaldehyde	3	Gly, His, Hyl, Arg, Ala, Glu, Asp, and N-His peptides	
4	Pentacyanoaquoferriate	3, 14	Arg and Arg-containing peptides	
5	Pauly's	13	His, Tyr, and their peptides	
6	Folin–Ciocalteu's*	15	Tyr and Tyr-containing peptides	Before spraying with reagent, the plate was sprayed with glacial acetic acid in acetone (50%, v/v) and heated at 60° for 10 min.
7	Ehrlich's	3	Hyp, Trp, and Trp-containing peptides	
8	Periodate/acetyl acetone + ammonium acetate	3	Ser, Hyl, and N-Ser peptides	
9	Cd-isatin	1, 3	Pro, Hyp, and N-Pro or N-Hyp peptides	As for Reagent No. 1
10	Isatin	13	Pro, Hyp, and N-Pro or N-Hyp peptides	As for Reagent No. 1

* Obtained commercially from BDH (Poole, Dorset, Great Britain).

TABLE II
WEIGHT OF HYDROLYSED PROTEIN PER TLC PLATE

Reagent	*Sample weight (μg)*		
	Collagen	*Glucagon*	*Insulin*
No. 1	140	75	180
No. 2	210	75	180
No. 3	140	75	180
No. 4	210	100	180
No. 5	150	100	180
No. 6	—	100	300
No. 7	140	75	—
No. 8	350	75	300
No. 9	140	75	180
No. 10	140	75	180

cm) at an initial thickness of 400 μm. The coated plates were allowed to dry horizontally overnight before use[1].

Samples of collagen, and other protein hydrolysates were applied to individual thin layers of cellulose, as described previously[2]. The weights of the proteins that were found to give the best chromatograms are given in Table II.

The chromatograms were developed bidimensionally as described previously[1]. Solvent A (100 ml) was used for development in the first dimension, and solvent B (170 ml) for the second development at right angles to the first.

For identification of amino acids and peptides the developed chromatograms were sprayed in duplicate with the general (Cd-ninhydrin) and more specific reagents. The conditions of spraying, the assessment of the final colour and the determination of R_F values were as described previously[6]. New reactions were observed with some of the more specific staining reagents, as described below:

Zimmerman's *o*-phthalaldehyde (reagent No. 3) has been used for the identification of N-terminal glycyl peptides as grey spots on silica gel TLC[10]. It was found in the present work that free amino acids could be detected with this stain also. Thus, aspartic acid, glutamic acid, and alanine gave grey colours, while glycine gave a green-grey colour immediately after staining and heating. Arginine gave a light purple colour, and hydroxylysine gave a light pink colour, 18 h after staining. All the above amino acids were detected under visible light; UV light also detected the amino acids, but the colours were different. Therefore, *o*-phthalaldehyde could not be considered specific for N-terminal glycyl peptides. Heathcote *et al.*[4] observed that N-terminal glycyl peptides give a permanent yellow colour with Cd-ninhydrin, which seems to be specific. The free amino acids —proline and hydroxyproline— also give a yellow colour with Cd-ninhydrin, but this turns to orange after some 18 h; the spots corresponding to N-terminal glycyl peptides stay as bright yellow colours.

It was also found in the present work that *o*-phthalaldehyde gave a positive reaction with some N-terminal histidyl peptides. For example, the synthetic dipeptide, His-Leu, gave a purple fluorescence under UV light, a finding made by other workers also[16].

N-Terminal seryl peptides, *e.g.*, Ser-Gly, were found to give yellow fluorescence under UV light with reagent No. 8. This reaction was obtained also with other amino

acids and peptides, but the yellow fluorescence was lost one day after spraying, in contrast with the colour of the N-terminal seryl peptides and the colour of specific amino acids.

RESULTS AND DISCUSSION

It was found in preliminary trials that the number of spots produced by hydrolysis of collagen for 1 h with 1 *M* HCl was quite small. Hydrolysis for 2 h produced a well-defined pattern of spots which did not change in appearance or number when the hydrolysis time was extended to 3 h. The concentrations of the spots had increased, however, and were judged to be optimal for the staining reactions. Accordingly, a time of 3 h was standardized for the hydrolysis during the present study. Subsequent experiments with other collagens and other proteins have confirmed the desirability of the chosen hydrolysis conditions.

Although partial acid hydrolysis was used in a number of early finger printing studies[17–19], certain conditions were found to produce anomalies and the preference has been, in more recent studies, for such methods as tryptic, and other enzymic, digestion. Thus, an interconversion was found to take place in some peptides containing aspartic acid, especially when linked to lysine[17,18]. This conversion appeared to be less marked when the acid concentration was reduced from 12 *M* and also when the temperature was elevated and the time shortened to produce comparable hydrolysis. In the present experiments the acid concentration was considerably lower than this figure (12 *M*), and the consistency observed in the pattern of spots from the collagen hydrolysate when the period of hydrolysis was increased from 2–3 h indicated that little interconversion could be taking place.

A different type of interconversion was reported by Ryle and Sanger[19] when insulin was hydrolysed with cold concentrated HCl. More peptides were found than could be accounted for by the unique structure of the protein, and they later confirmed that interchange was taking place at the disulphide bonds. Again, in the present study of insulin, the conditions of hydrolysis are such that this reaction would be at a minimum according to the data put forward by these workers[19].

When due allowance is made for the above factors, acid hydrolysis is by no means such a random occurrence as was generally thought, and the patterns obtained in the present study have been remarkably consistent. This overall observation is not without support from other recent work on the specificity of HCl. Thus, Schultz *et al.*[20], in a study of the partial hydrolysis of proteins by HCl, showed that cleavage took place preferentially at those peptide bonds in the sequence which are particularly sensitive to acid. This is different from the attack of many proteases which have more specificity for bonds next to specific amino acids. Aspartic acid was found by these workers to be an exceptional amino acid in that its release was found to be related to its abundance in the protein.

Good separation of the peptide spots was achieved with the protein hydrolysates examined, even including that of insulin, if some allowance is made for the fact that the A and B chains were not separated before hydrolysis and that unoxidised cysteine-peptides tend to streak a little anyway. The peptides produced by the acid hydrolysis in the present study appeared to be small, mostly di-, tri- and tetrapeptides. We did not set out to identify the sequences of all the peptide spots,

but rather to see how much structural information could be obtained directly by visual examination of the stained and/or UV-irradiated plates.

The systematic application of the staining reagents represents a novel approach to the partial identification of many of the peptide fragments. It facilitates the identification of the sequence in the vicinity of a specifically colour-labelled residue when only small amounts of material are present. This is similar in many ways to the isotopic labelling of a specific residue employed by Naughton *et al.*[17] in their study of the active centres of enzymes. Some of the more interesting findings from the application of the method to three proteins —collagen, glucagon and insulin— will now be discussed in relation to their structures. The sequence structures of glucagon and insulin have been known for a long time, and much of the structure of collagen has now been established, including the whole of its α_1 chain (at least with the aid of two different species of animal).

Bovine achilles tendon collagen

Some 61 separate spots in all were detected on thin-layer chromatograms of the collagen hydrolysate, and Fig. 1 shows 57 which were detected by Cd-ninhydrin. All the amino acids present in collagen, except lysine and histidine, were identified

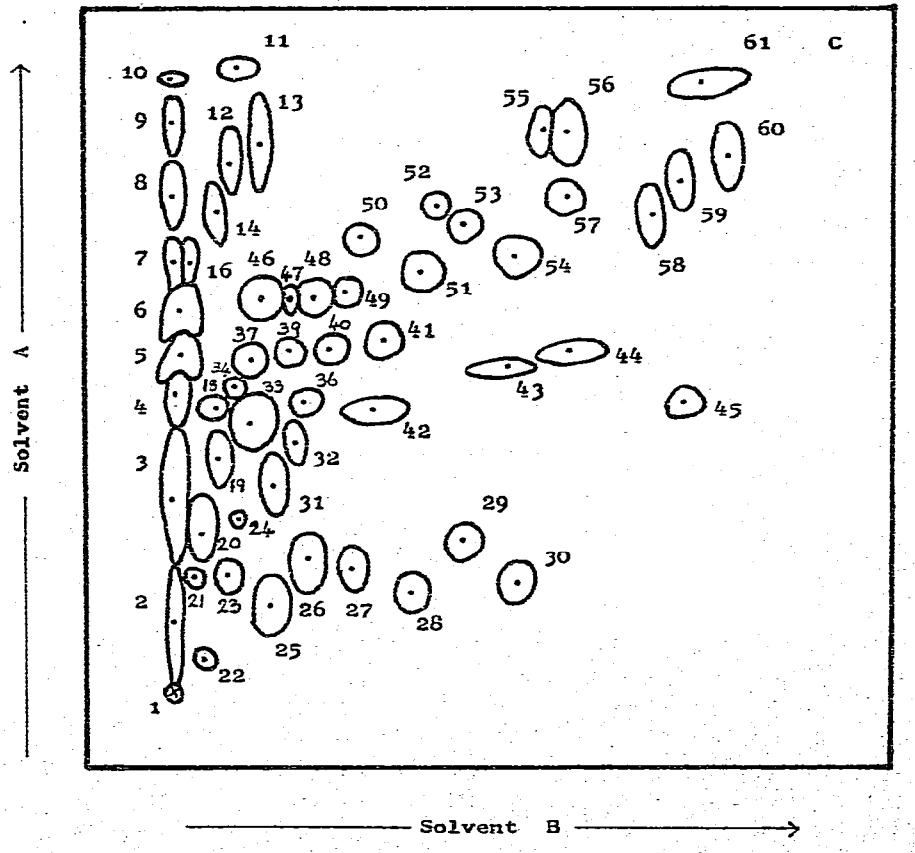

Fig. 1. Chromatogram of bovine collagen hydrolysate stained with Cd-ninhydrin (Reagent No. 1).

TABLE III

BOVINE ACHILLES TENDON COLLAGEN HYDROLYSATE AMINO ACIDS AND PEPTIDES IDENTIFIED

Spot	*Identity*	*Data for conclusion*			
		R_F × 100 values		*Reaction*	
		Solvent A	*Solvent B*	*Reagent No.*	*Colour*
C_1	Residue containing His and Met	0	0	5	Red
				2	White
C_2	His-containing peptide	10	0	5	Red
C_3	N-His peptide containing Arg	28	0	5	Red
				3	Purple (UV)
				4	Red
C_4		44	0		
C_5	Asp	50	1	3	Grey
C_6	Glu	56	1	3	Grey
C_7		63	0		
C_8		73	0		
C_9		84	0		
C_{10}		90	0		
C_{11}		91	9		
C_{12}		77	9		
C_{13}		80	12		
C_{14}		70	6		
C_{15}	N-Pro, or N-Hyp, peptide	64	5	9 or 10	Blue
C_{16}		63	3		
C_{17}	N-Pro, or N-Hyp, peptide	59	5	9 or 10	Blue
C_{18}		41	6		
C_{19}	N-Gly peptide	34	6	1	Yellow
C_{20}	Arg	23	4	3	Purple
				4	Red
C_{21}	Arg-containing peptide	17	3	4	Red
C_{22}	His-containing peptide	4	5	5	Red
C_{23}	N-Gly peptide containing Arg	17	8	1	Yellow
				4	Red
C_{24}		25	9		
C_{25}	Hyl + His-containing peptide	13	14	3	Pink
				5	Red
				8	Blue (UV)
C_{26}	His-containing peptide	19	19	5	Red
C_{27}		18	25		
C_{28}		14	33		
C_{29}		22	40		
C_{30}		16	48		
C_{31}	Arg-containing peptide	30	15	4	Red
C_{32}	N-Gly peptide (Gly-Gly)	36	17	1	Yellow
C_{33}	Gly	39	12	1	Orange
				3	Green-grey
				9 + 7 or 10 + 7	Purple-orange
C_{34}		45	9		

(Continued on p. 218)

TABLE III *(continued)*

Spot	*Identity*	*Data for conclusion*			
		$R_F \times 100$ *values*		*Reaction*	
		Solvent A	*Solvent B*	*Reagent No.*	*Colour*
C_{35}	Arg-containing peptide	45	14	4	Red
C_{36}		42	19		
C_{37}	Hyp	49	12	1	Orange
				9 or 10	Blue
				9 + 7 or 10 + 7	Purple
C_{38}	N-Pro, or N-Hyp, peptide	49	14	9 or 10	Blue
C_{39}	N-Gly peptide (Gly-Ala)	50	16	1	Yellow
C_{40}		50	22		
C_{41}	N-Gly peptide	52	29	1	Yellow
C_{42}	Ser	41	28	8	Yellow under UV
C_{43}	N-Gly peptide	48	47	1	Yellow
C_{44}	Thr	50	55		
C_{45}		42	71		
C_{46}	Ala	58	13	3	Grey
C_{47}	N-Gly peptide	58	17	1	Yellow
C_{48}	Pro	58	20	9 or 10	Dark blue
				1	Orange
				9 + 7 or 10 + 7	Dark blue
C_{49}		58	24		
C_{50}	Tyr	67	26		
C_{51}		61	35		
C_{52}	Val	71	37		
C_{53}	Met	68	41	2	White
C_{54}	N-Pro, or N-Hyp, peptide containing Met	64	48	9 or 10	Blue
				2	White
C_{55}	Ile	82	52		
C_{56}	Leu	82	55		
C_{57}	Phe	73	55	10	Blue
C_{58}		70	67		
C_{59}	N-Gly peptide	76	71	1	Yellow
C_{60}		79	78		
C_{61}		89	74		

directly. These two amino acids, which happened to coincide with peptides, were identified independently by ion-exchange chromatography[4]. The R_F values of the spots and their identity or partial identity, where directly ascertainable, are given in Table III, and the patterns obtained with the various staining reagents are given in Figs. 1–5.

In view of the large molecular size of the collagen molecule and the fact that the chains were not separated before hydrolysis, it might have been expected that a much larger number of peptides would have been produced. The explanation must reside in the fairly regular structure where glycine occupies every third position of

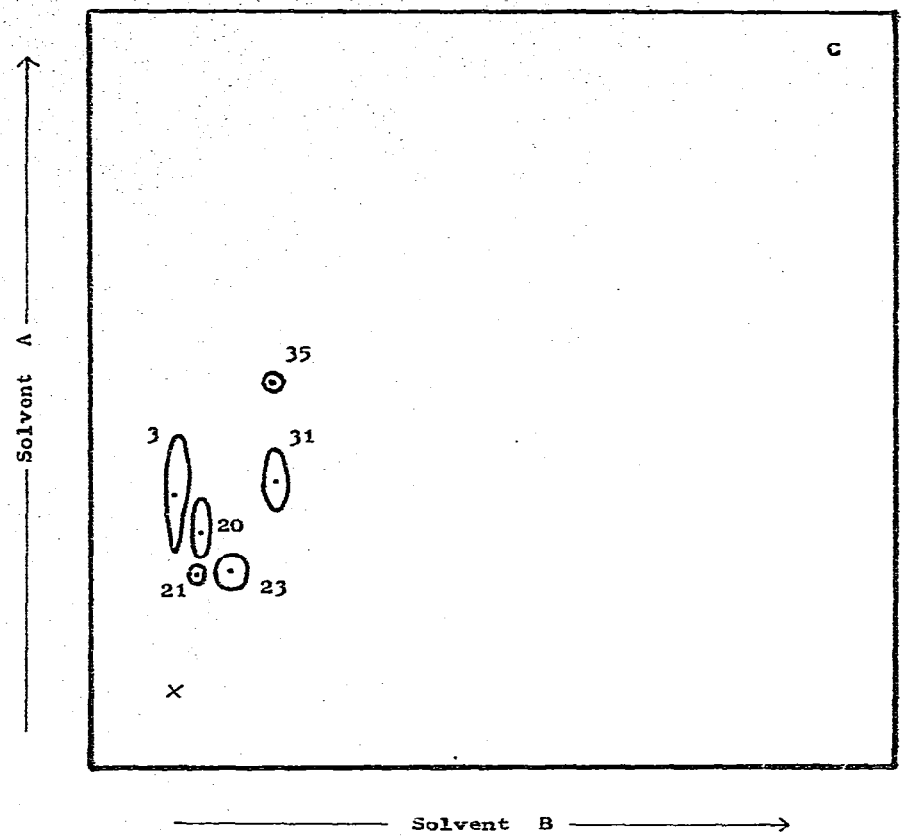

Fig. 2. Chromatogram of bovine collagen hydrolysate stained with pentacyanoaquoferriate (Reagent No. 4).

the large helical regions of the molecule[21], and the fact that the acid is fairly specific in its positions of attack.

Glycine is the most abundant amino acid in collagen and accounts for almost one third of the total amino acid residues in the molecule[21]. Eight of the spots (C_{19}, C_{23}, C_{32}, C_{39}, C_{41}, C_{43}, C_{47}, and C_{59}) were identified directly by Cd-ninhydrin as N-terminal glycyl peptides (Table III, Fig. 1). Two of the spots (C_{32} and C_{39}) were further characterized as Gly-Gly and Gly-Ala, respectively, from their R_{AH} values on ion-exchange chromatography[6]. There are about 54 possible dipeptides (Gly-Ala) in the α_1-chain of rat and calf skin collagen[22–33], and if all these were actually produced by the hydrolysis they could, theoretically, account for 108 amino acid residues (approximately 10%) of the α_1-chain while still producing only one spot, C_{39}, on the chromatogram. In a partial hydrolysis study of steer hide collagen, using concentrated HCl at 37° and ion-exchange separation, Kroner *et al.*[34,35] identified ten N-terminal glycyl peptides of which Gly-Gly and Gly-Ala were two. They also identified the tripeptide Gly-Pro-Hyp, and the dipeptide Gly-Pro. The former represents about 10% of the glycine residues in collagen while the latter is the most repetitive dipeptide in the molecule[22]. None of the N-terminal glycyl peptides identified by Kroner *et al.*[34,35]

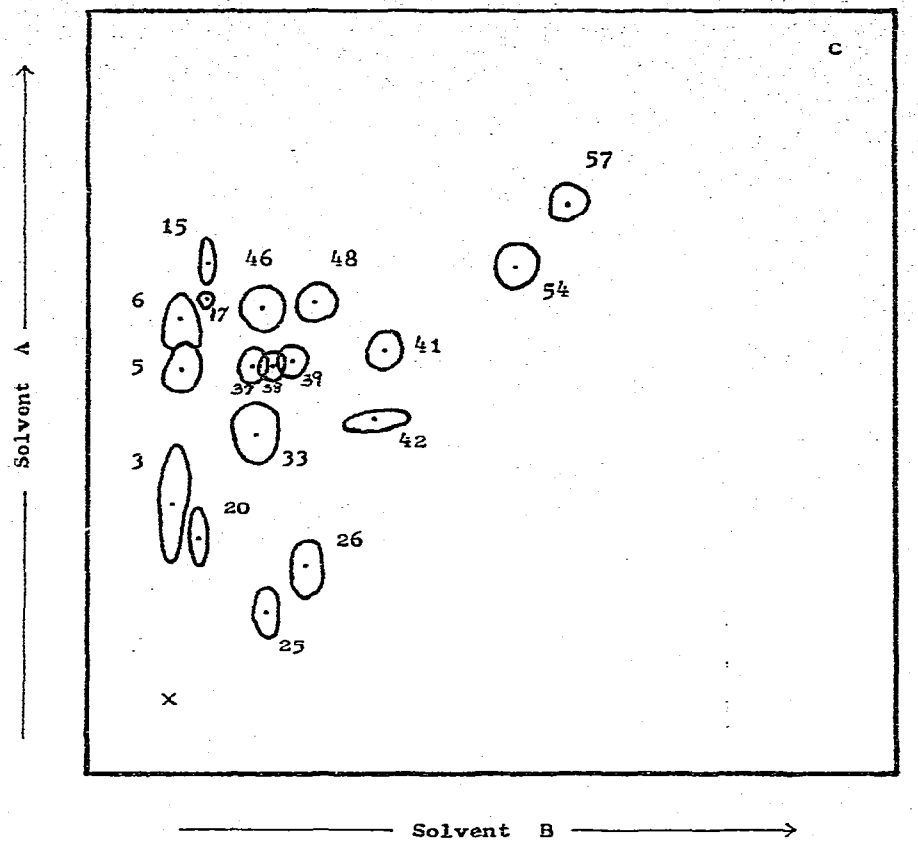

Fig. 3. Chromatogram of bovine collagen hydrolysate stained with isatin (Reagent No. 10).

contained arginine, whereas one, C_{23}, in our study contained this amino acid (Fig. 2).

The imino acids proline and hydroxyproline, taken together, account for two of every nine residues in collagen[21]. Four prolyl, or hydroxyprolyl, N-terminal peptides were specifically identified, three of which (C_{15}, C_{17}, and C_{38}) had not been detected by Cd-ninhydrin. There are about 111 residues of hydroxyproline in the α_1-chain of rat and calf skin collagen[22–33]. All of them exist as Hyp-Gly, except one, which is Hyp-Hyp. Kroner *et al.*[34,35] identified one peptide as Hyp-Gly. This was the only N-terminal hydroxyproline peptide isolated by these workers and no N-terminal prolyl peptide was identified. It is possible that this dipeptide is one of these three spots (C_{15}, C_{17}, or C_{38}) and could account for up to 22% of the α_1-chain. The fourth peptide, C_{54}, contained methionine also (Figs. 3 and 4) and is discussed under methionyl peptides.

Histidine is present in all vertebrate collagen to a small extent[21] and plays an important role in the intermolecular crosslinking[36]. There are about six residues of histidine per 1,000 residues of bovine achilles tendon collagen[37].

Six histidine-containing peptides were identified in the present study (see Fig. 5) compared with none in the earlier study[34,35]. One of the spots, C_3, was found to have histidine as the N-terminal amino acid and also to contain arginine. There are

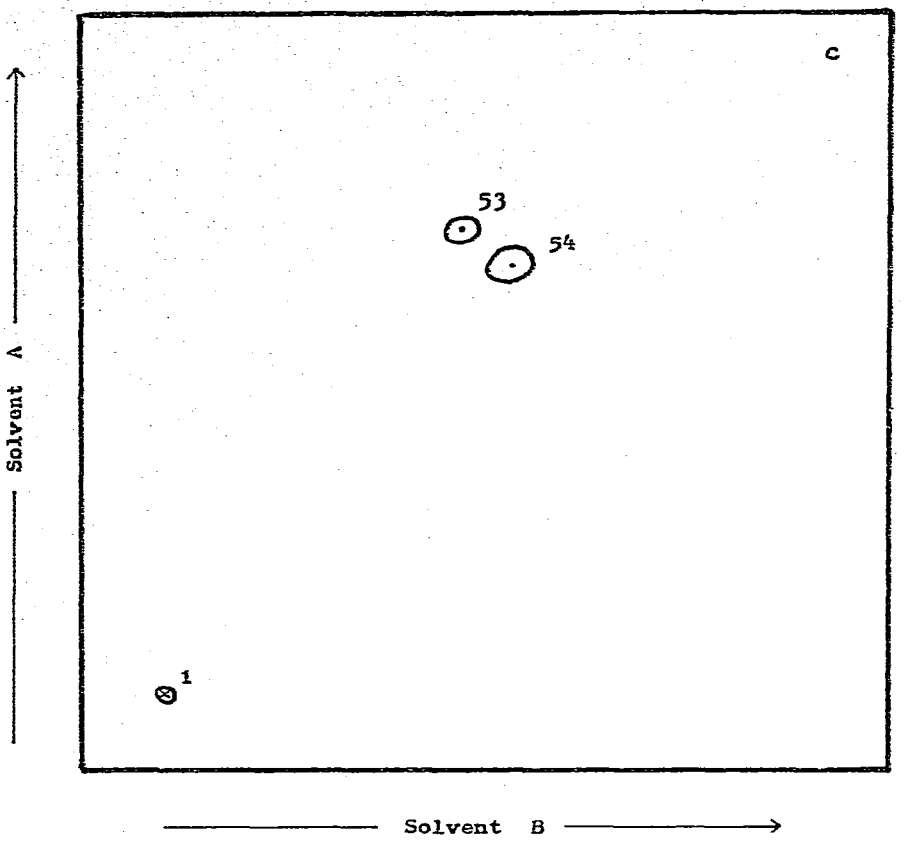

Fig. 4. Chromatogram of bovine collagen hydrolysate treated with platinic iodide (Reagent No. 2).

only two histidine residues in the helical part of the α_1-chain of rat and calf skin collagen[22–33], and both exist as His-Arg; hence, it is possible that spot C_3 could account for this sequence.

There are some 53 to 54 arginine residues per 1,000 residues of bovine achilles tendon collagen[37]. Five arginine-containing peptides were found in the tendon collagen hydrolysate (Fig. 2), one of which (C_{23}) had glycine as the N-terminal amino acid. Eight arginine-containing peptides had been identified previously in the hydrolysate of hide collagen[34,35], but none had histidine or glycine as the N-terminal residue.

There are about 41 Arg-Gly peptides in the α_1-chain of rat and calf skin collagen[22–33] and, if all of these were released as the dipeptide, they would, while forming one spot only on TLC, account for over 80 amino acid residues (8% approximately) of the α_1-chain. We did not attempt to characterize the arginyl peptides completely, but one of the peptides in the study by Kroner *et al.*[34,35] was characterized as Arg-Gly.

The major vertebrate collagens contain five to twelve methionine residues per α chain. As well as the free amino acid, C_{53}, bovine collagen hydrolysate gave two spots which were identified as methionine-containing peptides. One of these, C_{54}, has been shown above to have an N-terminal prolyl (or hydroxyprolyl) residue. Proline

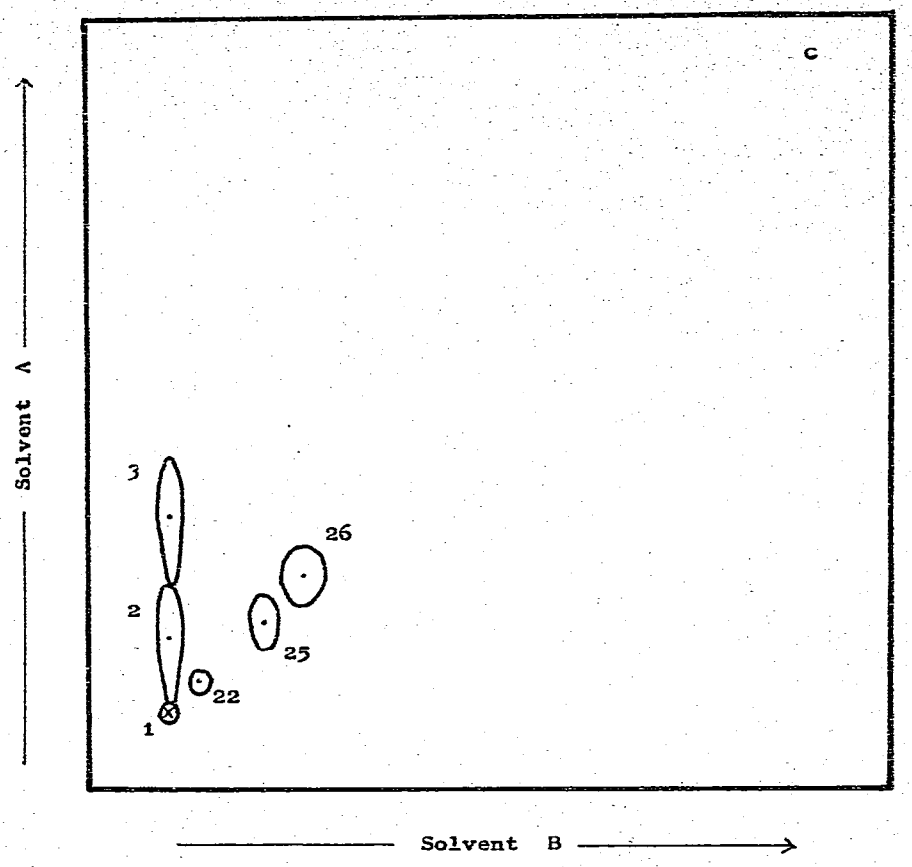

Fig. 5. Chromatogram of bovine collagen hydrolysate stained with Pauly's reagent (No. 5).

is the residue which is usually on the N-terminal side of methionine in the α_1-chain[22–33] and the α_2-chain[38–41] of rat and calf skin collagens. Therefore, the N-terminal residue in spot C_{54} is more likely to be proline than hydroxyproline. Hence, this peptide could correspond to the sequence Pro-Met, present in α_1-CB1[21], α_1-CB2[24] of rat skin collagen (Type I) and α_1-CB7[31] of calf skin collagen; and also of α_2-CB1[38] of calf skin collagen. Thus, Pro-Met could be released from all above positions while still giving only one methionyl spot on TLC.

The other methionyl peptide, C_1, contained histidine somewhere in its sequence. The only region in the α_1-chain of rat skin collagen which contains both methionine and histidine residues in close proximity is that in the helical position where methionine is at position 102 and histidine at position 105[25,27]. The sequence around this region is:

–Arg–	Gly–	(Leu–	Hyp–	Gly–	Thr–	Ala–	Gly–	Leu–	Hyp–	Gly–	Met–	Hyl–	Gly–	His)–	Arg–
(91)	(92)	(93)	(94)	(95)	(96)	(97)	(98)	(99)	(100)	(101)	(102)	(103)	(104)	(105)	(106)

Since there was no reaction for arginine or N-terminal glycine, C_1 could contain any sequence from positions 93–105, the only restriction being that neither glycine nor hydroxyproline could be present at the N-terminal residue. Kroner *et al.*[34,35] found

TABLE IV
GLUCAGON HYDROLYSATE AMINO ACIDS AND PEPTIDES IDENTIFIED

Spot	*Identity*	*Data for conclusion*			
		$R_F \times 100$ values		*Reaction*	
		Solvent A	*Solvent B*	*Reagent No.*	*Colour*
G_1		28	0		
G_2		51	0		
G_3	Tyr-containing peptide	66	0	5	Brown-orange
				6	Blue
G_4	Trp-containing peptide	84	0	7	Purple
G_5	Trp-containing peptide	93	5	7	Purple
G_6		78	16		
G_7		69	14		
G_8		75	10		
G_9	Tyr-containing peptide	68	7	5	Brown-orange
				6	Blue
G_{10}	Met-containing peptide	68	10	2	White
G_{11}		60	3		
G_{12}	Glu	56	1	3	Grey
G_{13}	Asp	49	1	3	Grey
G_{14}	N-Gly peptide	49	8	1	Yellow
G_{15}	Arg-containing peptide	41	3	4	Red
G_{16}	Arg-containing peptide	39	7	4	Red
G_{17}	N-His peptide	31	3	5	Red
				3	Purple (UV)
G_{18}	Arg	23	5	4	Red
G_{19}	Arg-containing peptide	18	1	4	Red
G_{20}	N-Ser peptide containing Arg	11	3	4	Red
				8	Yellow (UV)
G_{21}	Arg-containing peptide	22	10	4	Red
G_{22}	His	16	18	3	Green (visible light) Purple (UV)
				5	Red
G_{23}	Lys	20	17		
G_{24}		21	25		
G_{25}		18	38		
G_{26}	Gly	39	12	1	Orange
				3	Green-grey
				9 + 7 or 10 + 7	Purple-orange
G_{27}		37	15		
G_{28}		32	12		
G_{29}	Met-containing peptide	48	13	2	White
G_{30}	Ala	56	14	3	Grey
G_{31}	Ser	41	27	8	Yellow (UV)
G_{32}	Thr	48	54		
G_{33}		50	72		
G_{34}		36	55		
G_{35}	Tyr	67	26	10	Blue
				5	Brown-orange
				6	Blue

(Continued on p. 224)

TABLE IV *(continued)*

Spot	*Identity*	*Data for conclusion*			
		R_F × 100 values		*Reaction*	
		Solvent A	*Solvent B*	*Reagent No.*	*Colour*
G_{36}		57	39		
G_{37}	Met	69	41	10	Blue
				2	White
G_{38}		85	37		
G_{39}	Trp-containing peptide	92	38	7	Purple
G_{40}	Leu	81	55		
G_{41}	Phe	73	55	10	Blue
G_{42}	Trp	63	48	7	Purple
G_{43}		67	66		
G_{44}		69	76		
G_{45}		69	86		
G_{46}		85	81		
G_{47}		85	88		
G_{48}	Trp-containing peptide	91	89	7	Purple
G_{49}	Trp- and Met-containing peptide	91	80	2	White
				7	Purple
G_{50}	Trp- and Met-containing peptide	92	67	2	White
				7	Purple

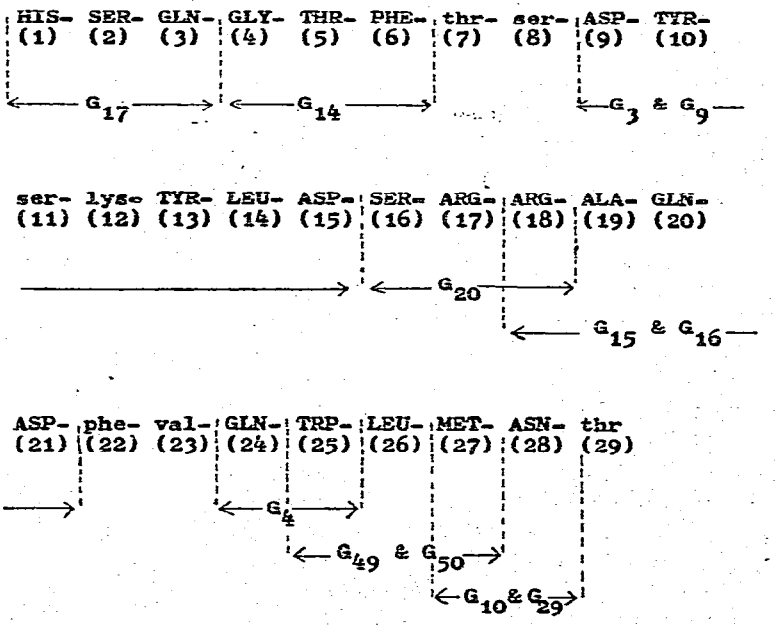

Fig. 6. The amino acid sequence of glucagon.

only one methionine-containing peptide in steer hide collagen, *viz.* Glu-Met. This sequence occurs in α_1-CB0[21] and α_1-CB5[27].

All the above information was obtained from about 2 mg of collagen. Clearly,

much more information could be obtained by eluting and analysing individual peptides further, as might be required in a study of abnormal collagen. The next step was to see whether the technique which had been successfully developed for collagen could be adapted to other proteins of different, but known, structure.

Glucagon

The main function of this pancreatic hormone is the promotion of hyperglycaemia[42]. In this respect, glucagon may be considered as the opposite, in physiological terms, of the other protein hormone, insulin. It is a straight-chain polypeptide of 29 amino acid residues (mol. wt. 4,000 approximately) and its structure does not involve any covalent crosslinks. In all, 50 peptides or amino acids were separately detected on TLC by the method outlined above, and a list of spots is given in Table IV, together with their characterization data.

The known structure of glucagon is given in Fig. 6. The pattern of spots detected by staining with Cd-ninhydrin is shown in Fig. 7. All the amino acids present in glucagon, except valine, glutamine and asparagine, were identified as free amino acids on TLC. Valyl peptides have been reported previously to be fairly stable to acid hydrolysis[43], and asparagine and glutamine would presumably be converted to aspartic and glutamic acids.

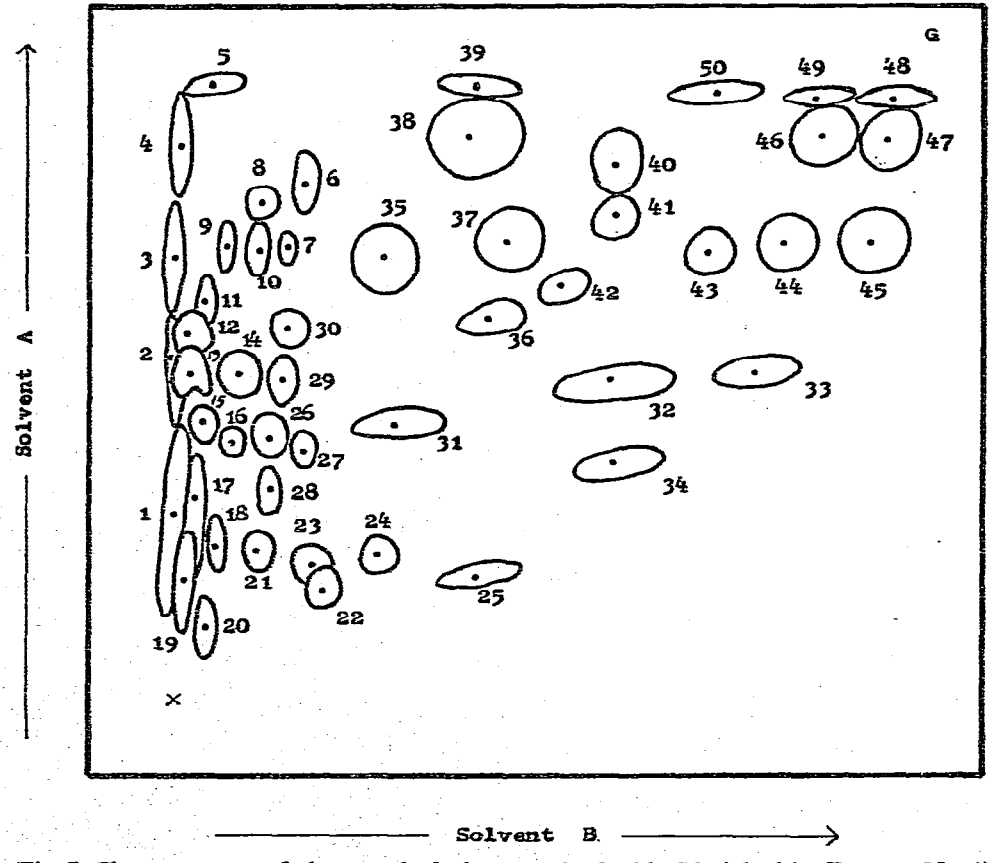

Fig. 7. Chromatogram of glucagon hydrolysate stained with Cd-ninhydrin (Reagent No. 1).

Only one histidyl peptide, G_{17}, was detected on the chromatograms, and this was shown by reagent No. 3 to be an N-terminal histidyl peptide. As there is only one histidine residue in the chain, this must also occupy the N-terminal position in the glucagon molecule. The peptide had different R_F values from the synthetic dipeptide His-Ser, and, because of its acidic nature, it probably contained glutamic acid (from glutamine at position 3, Fig. 6). Hence, –His–Ser–Glu– could account for part, or all, of the peptide in spot G_{17}.
(1) (2) (3)

One N-terminal glycyl peptide, G_{14}, was found, and since there is only one glycine residue which is at position 4, this peptide must commence at this position. The R_F values of this spot were different from those of the synthetic dipeptide Gly-Thr, and so phenylalanine at position 6 must be present in the sequence. Hence, the partial sequence –Gly–Thr–Phe– could account for this peptide.
(4) (5) (6)

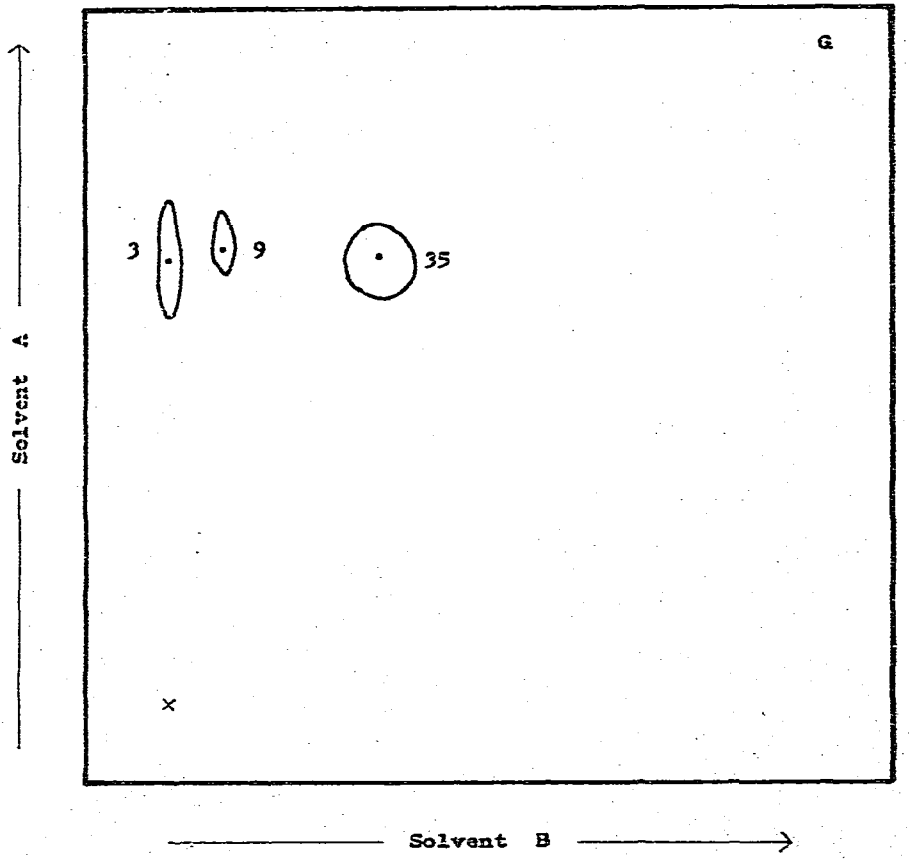

Fig. 8. Chromatogram of glucagon hydrolysate stained with Folin–Ciocalteu's reagent (No. 6).

Two tyrosine-containing peptides, G_3 and G_9, were detected (Fig. 8). They could be of one, or both, tyrosine residues at positions 10 and 13 in the chain.

Five arginyl peptides (G_{15}, G_{16}, G_{19}, G_{20} and G_{21}) were detected, of which three were capable of giving direct information (Fig. 9). All of these must contain one, or both, of the only two arginine residues which are adjacent at positions 17 and

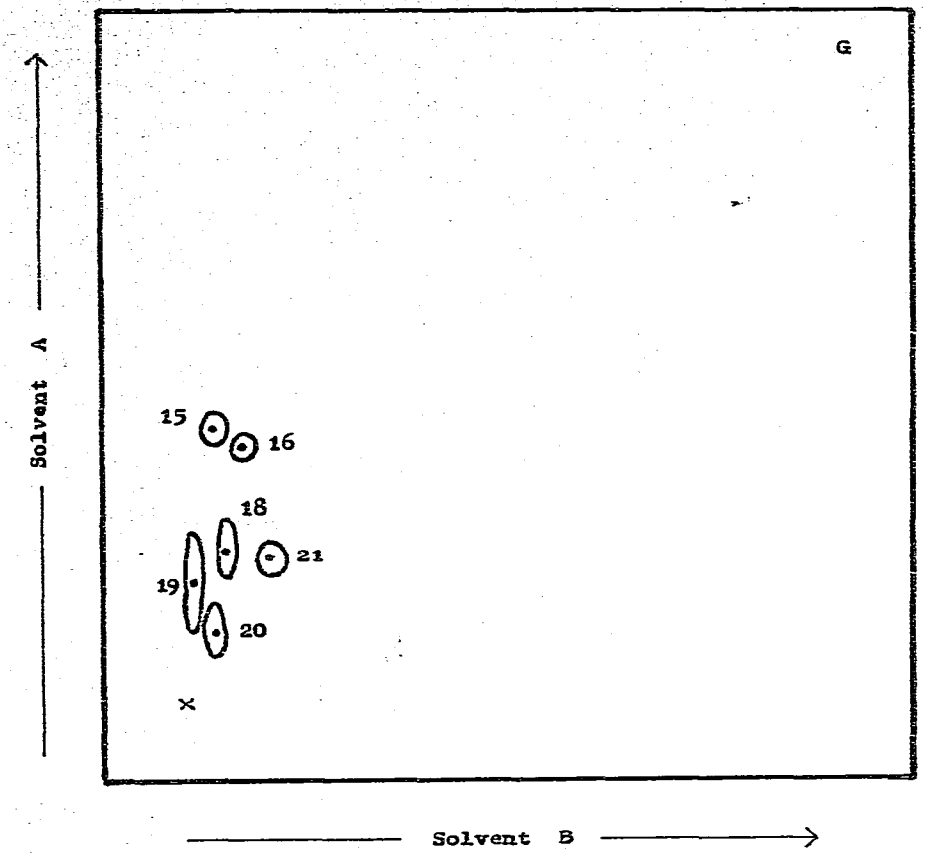

Fig. 9. Chromatogram of glucagon hydrolysate stained with pentacyanoaquoferriate (Reagent No. 4).

18. Spot G_{20} also had serine as the N-terminal residue and had a marked basic nature which could have been caused by the presence of two arginyl residues. The partial sequence –Ser–Arg–Arg– could be accounted for by this peptide. Spots G_{15} and G_{16} had high R_F values in the first solvent system A, which could be accounted for by the presence of at least two acidic amino acids when only one arginyl residue is present. Glutamic acid from glutamine (at position 20) and aspartic acid (at position 21) are the most likely residues to be present in G_{15} and G_{16}, which could tentatively account for the partial sequence –Arg–Ala–Glu–Asp– in glucagon (Fig. 6).

One of the most interesting findings was that, despite the fact that tryptophan is fairly readily destroyed by acid, six well-defined tryptophan-containing peptides (as well as free tryptophan, G_{42}) were identified when a chromatogram was treated with Ehrlich's reagent (Fig. 10). This observation is, however, in agreement with a similar, recent finding by Perry *et al.*[44] of little destruction following mild acid hydrolysis. As there is only one tryptophan residue in the glucagon chain (at position 25), this amino acid acts as a marked residue in the same way as a radioactive label would, and analysis of the simple peptides would give definite characterization of partial se-

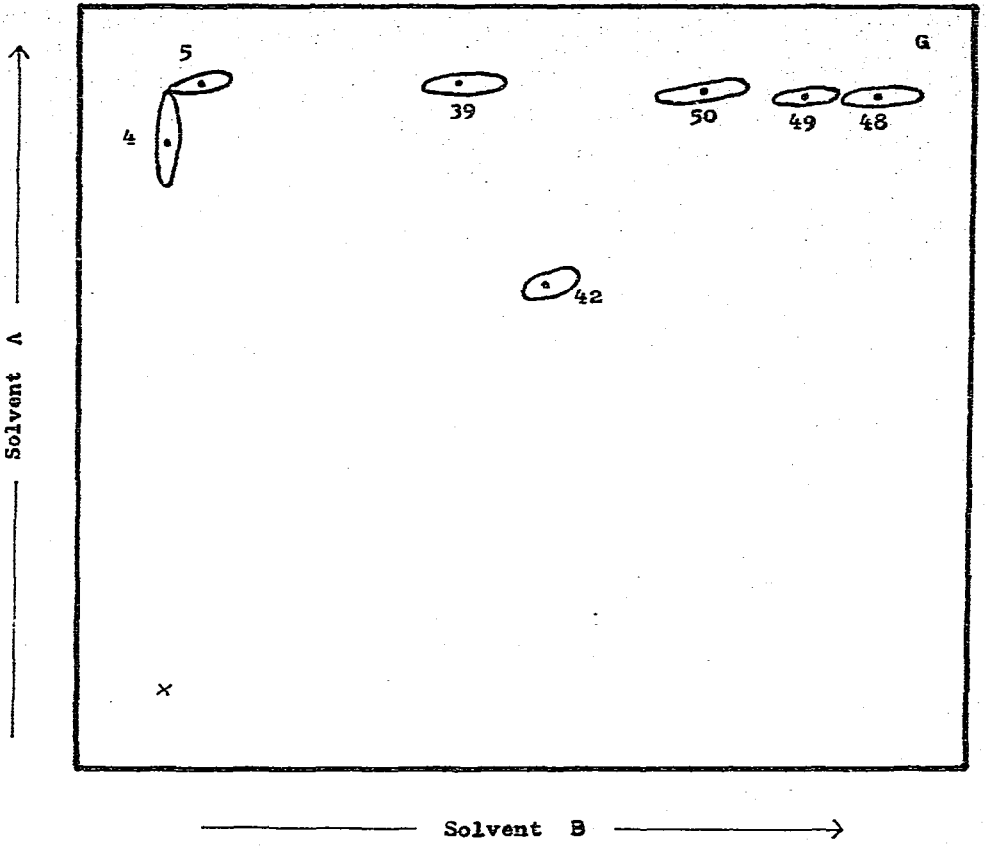

Fig. 10. Chromatogram of glucagon hydrolysate stained with Ehrlich's reagent (No. 7).

quence around the tryptophan. The chromatographic movement of G_4 indicated that it was a highly acidic peptide, and it is most likely that glutamic acid, coming from glutamine, accounts for the acidity. The sequence –Glu–Trp– could be part, or all, of peptide G_4, and this could account for the residues 24 and 25 in the glucagon molecule (Fig. 6). Two of the tryptophanyl peptides, G_{49} and G_{50}, also contained methionine, and these must both contain the residues –Trp–Leu–Met–.
(25) (26) (27)

Besides G_{49}, G_{50} and free methionine, two other methionine-containing peptides, G_{10} and G_{29}, were present (Fig. 11). There is only one methionine residue in the molecule, and specific staining therefore labels its position (27) in the same way as the position of tryptophan (25) was labelled. Since peptides G_{10} and G_{29} did not give any specific reaction for other amino acids and also since they were acidic in nature, their sequences probably lie within residues 26 to 29 of the glucagon chain, and they would both contain the sequence –Met–Asn–.
(27) (28)

The above information on glucagon was obtained directly from the examination of a few peptides on the chromatograms, and it will be clear that analysis of the 36 separated peptides should yield virtually the whole sequence structure. Certainly,

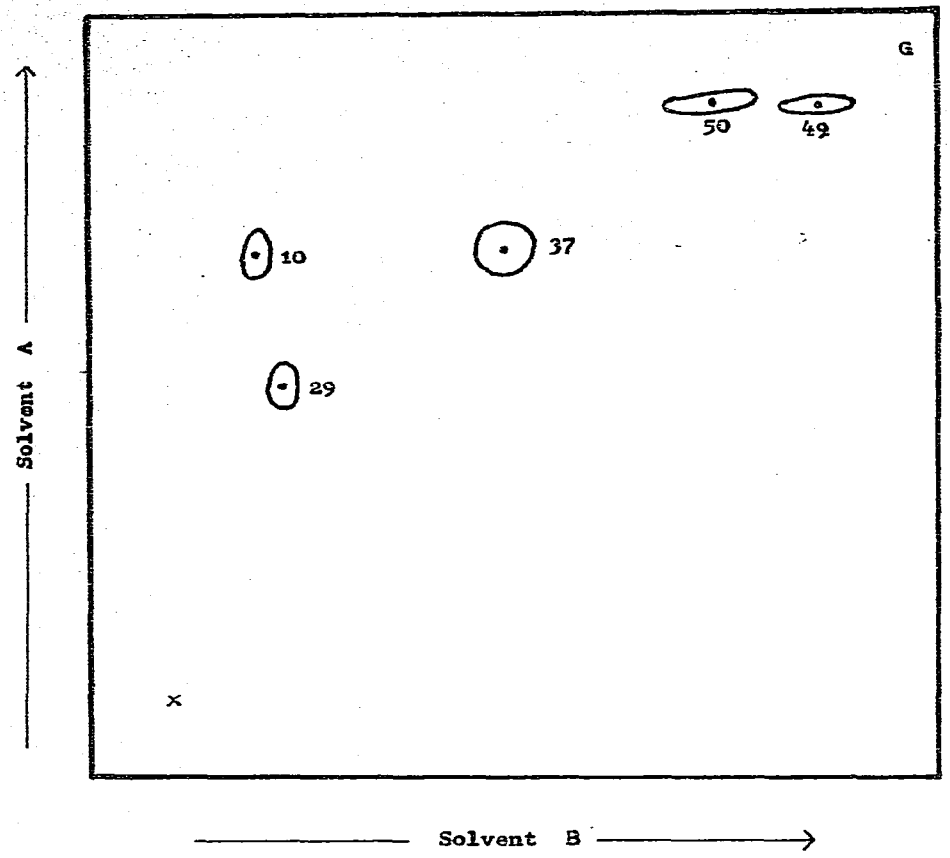

Fig. 11. Chromatogram of glucagon hydrolysate treated with platinic iodide (Reagent No. 2).

differences between two types of glucagon molecule could be readily apprehended without detailed analysis.

Insulin

Insulin was considered to be a good protein on which to test the method since the cysteine bridges present a peculiar difficulty in structural analysis (Fig. 12). Thus, in order to determine its structure, Ryle and Sanger[19] found it necessary to oxidize the A and B chains with performic acid and analyse them separately. The present study was carried out on the protein directly without separation of the chains and without preliminary oxidation of the –SH groups. It was therefore carried out under the most difficult conditions in order to see whether useful information could still be obtained.

Thirty-three individual fragments were detected on the chromatograms of the insulin hydrolysate, including all sixteen amino acids (see Table V and Fig. 13). Seven cysteine-containing peptides were specifically identified in all, which was not large in view of the number (six) of cysteine residues present in the molecule. This, and the consistency of the hydrolysis pattern, tended to confirm that re-arrangement of the cysteinyl peptides, as observed by Ryle and Sanger[19], was not a serious problem under

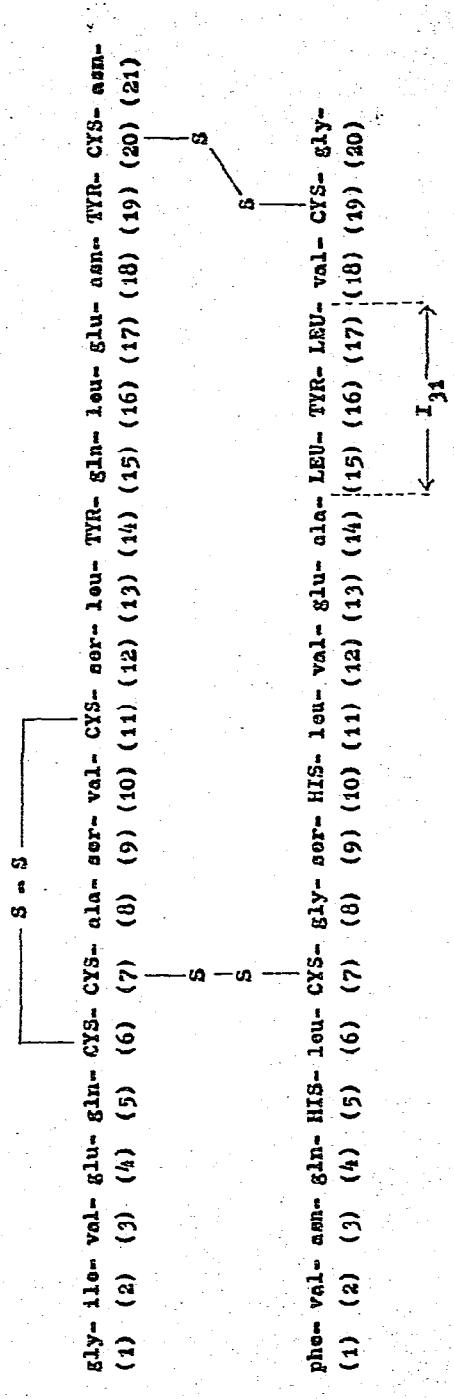

Fig. 12. The amino acid sequence of bovine insulin.

TABLE V

INSULIN HYDROLYSATE AMINO ACIDS AND PEPTIDES IDENTIFIED

Spot	*Identity*	*Data for conclusion*			
		$R_F \times 100$ values		*Reaction*	
		Solvent A	*Solvent B*	*Reagent No.*	*Colour*
I_1	Residue containing Tyr, Arg, and Cys	0	0	5	Brown-orange
				6	Blue
				4	Red
				2	White
I_2	Cys- and His-containing peptide	0	5	5	Red
				2	White
I_3	Cys- and His-containing peptide	14	0	5	Red
				2	White
I_4	Arg- and His-containing peptide	28	0	5	Red
				4	Red
I_5	Cys- and His-containing peptide	37	0	5	Red
				2	White
I_6	Cys- and Tyr-containing peptide	52	0	5	Brown-orange
				6	Blue
				2	White
I_7	Cys- and His-containing peptide	73	0	5	Red
				2	White
I_8	Cys- and Tyr-containing peptide	88	0	5	Brown-orange
				6	Blue
				2	White
I_9		83	9		
I_{10}		69	5		
I_{11}	N-Pro peptide	60	3	9 or 10	Blue
I_{12}	Glu	55	1	3	Grey
I_{13}	Asp	47	1	3	Grey
I_{14}	Arg	23	4	3	Purple
				4	Red
I_{15}	His	15	18	3	Green (visible light) Purple (UV)
				5	Red
I_{16}	Lys	18	17		
I_{17}		27	32		
I_{18}	Gly	38	10	3	Grey-green
I_{19}	Ser	40	25	8	Yellow (UV)
I_{20}		46	14		
I_{21}	Ala	56	12	3	Grey
I_{22}	Pro	57	18	9 or 10	Dark blue
				9 + 7 or 10 + 7	Blue
I_{23}	Tyr	68	24	10	Blue
				5	Brown-orange
I_{24}	Val	72	35		
I_{25}		83	25		
I_{26}	Thr	48	53		
I_{27}	Phe	73	53	10	Blue

(Continued on p. 232)

TABLE V *(continued)*

Spot	*Identity*	*Data for conclusion*			
		R_F × 100 values		*Reaction*	
		Solvent A	*Solvent B*	*Reagent No.*	*Colour*
I_{28}	Leu	82	53		
I_{29}	Ile	82	49		
I_{30}	Tyr-containing peptide	86	73	5	Brown-orange
				6	Blue
I_{31}	Tyr-containing peptide	87	85	5	Brown-orange
				6	Blue
I_{32}	Cys	13	0	2	White
I_{33}		63	92		

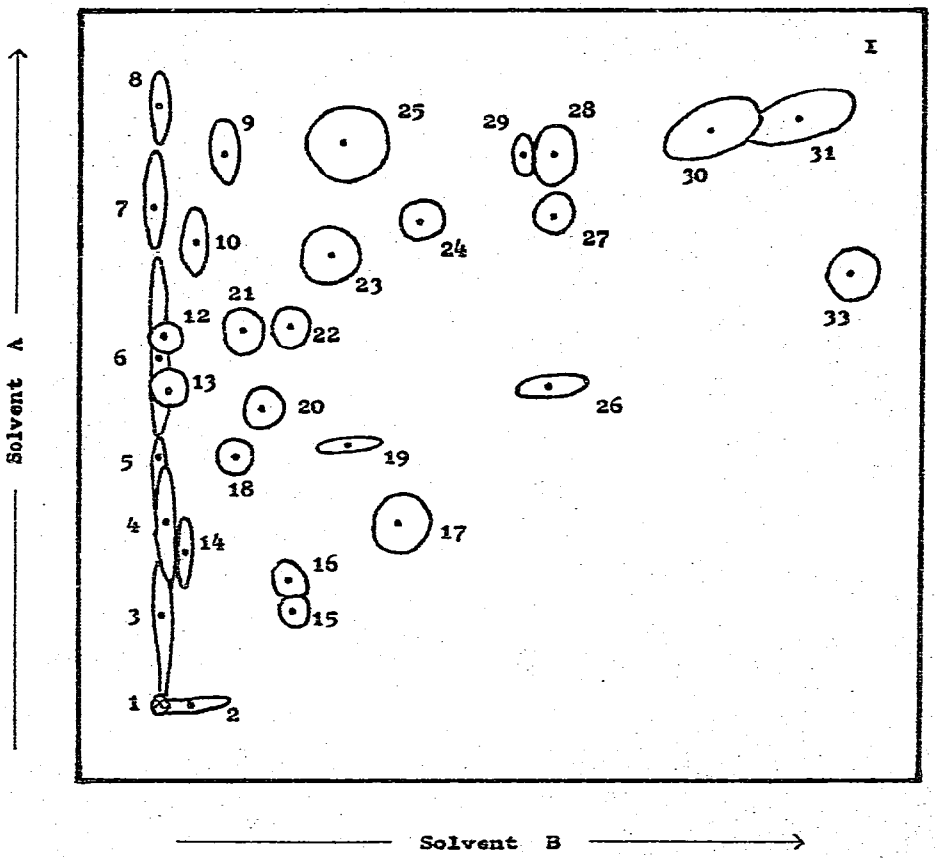

Fig. 13. Chromatogram of insulin hydrolysate stained with Cd-ninhydrin (Reagent No. 1).

the present conditions of hydrolysis.

For the separation of the free amino acid, cysteine, from the sulphur-containing peptide I_3, it was necessary to use a different solvent system (C) in the sec-

ond dimension (Fig. 14). This system was Solvent No. 2 described in ref. No. 2. It was also used on the chromatograms shown in Figs. 15 and 17.

It is clear that, even without separation of the A and B chains, full analysis of the seventeen distinct peptides would go a long way towards defining the full sequence of insulin. However, some deductions could be made directly from visual examination of the plates, and these will now be related to the known structure of insulin (Fig. 12).

Spot I_4 contained both arginine and histidine. Since there is only one arginine residue in the insulin molecule, this must be at position 22 in the B chain. Since no specific reactions were given by this peptide for cysteine or tyrosine, it could not consist of the whole sequence of positions from 10–22 in the B chain. Spot I_4, therefore, consisted of two peptides in coincidence which were identified as follows:

I_4 arginine-containing peptide —Since glycine was not detected and the presence of an acidic amino acid was indicated by the chromatographic movement (Fig. 13), the sequence pointed to that of –Glu–Arg– in the B chain.

(21) (22)

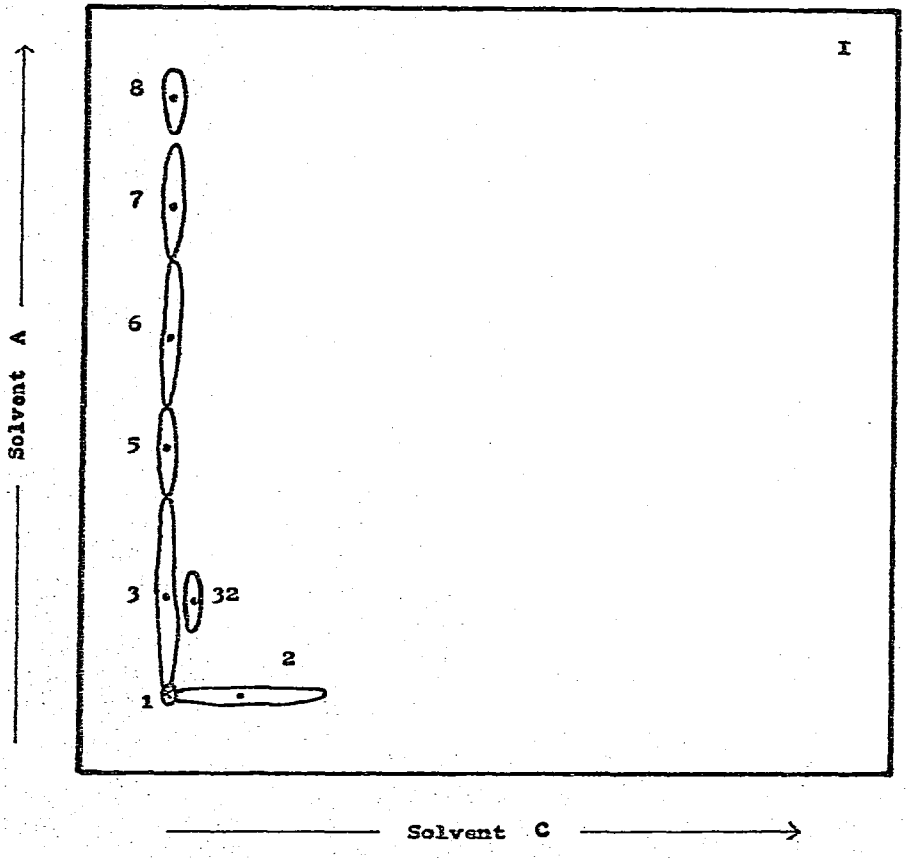

Fig. 14. Chromatogram of insulin hydrolysate treated with platinic iodide (Reagent No. 2).

I_4 histidine-containing peptide —The histidine residue in this peptide could be at either position 5 or position 10 in the B chain.

The four spots I_2, I_3, I_5 and I_7 were identified as both cysteine- and histidine-containing peptides (Figs. 14 and 15). It would seem that both histidine residues at positions 5 and 10 and cysteine at position 7 in the B chain are involved; possible involvement of cysteine residues at positions 6, 7, and 11 in the A chain is also likely.

Spot I_{11} was identified as an N-terminal prolyl peptide (Fig. 16). This spot was not detected by Cd-ninhydrin. There is only one proline residue which is at position 28. Hence, this peptide must contain at least the sequence –Pro–Lys– in the B chain of insulin. (28) (29)

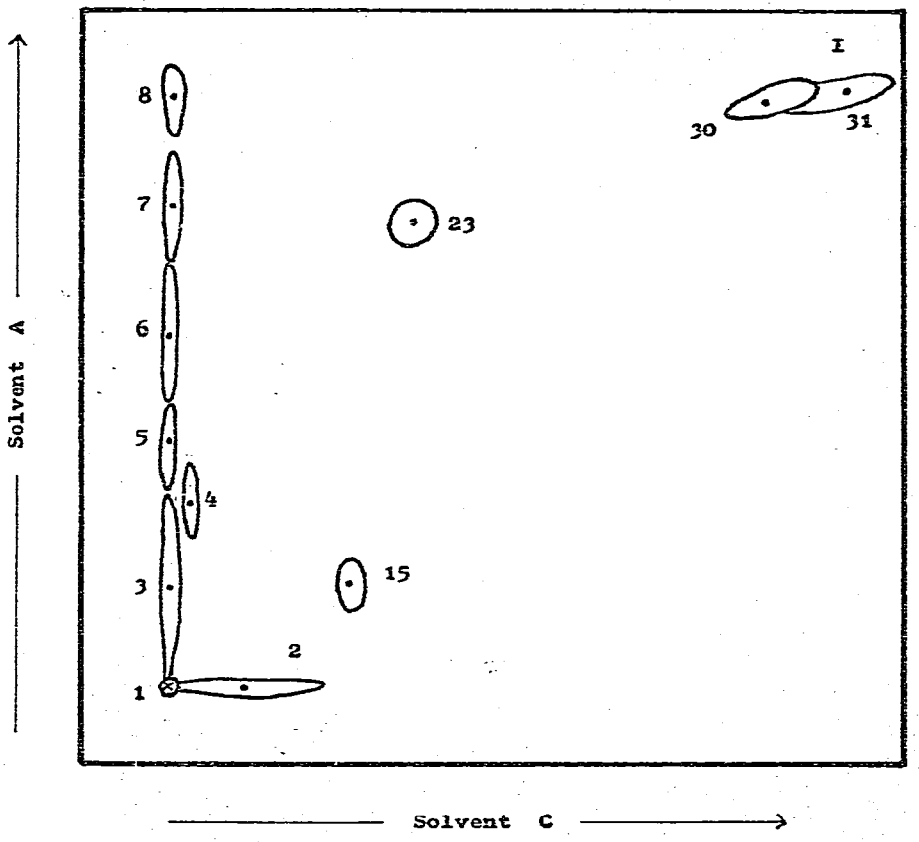

Fig. 15. Chromatogram of insulin hydrolysate stained with Pauly's reagent (No. 5).

Four tyrosine-containing peptides were detected in addition to free tyrosine (I_{23}). The chromatogram (Fig. 17) of peptides I_{30} and I_{31} points to the presence of a very mobile amino acid in the sequence of each peptide, which is probably leucine[4,6]. Spot I_{31} travelled further than I_{30} in both solvent systems, which could be explained by the presence of a further leucyl group. Therefore, –Leu–Tyr–Leu– could be part, or all, of peptide I_{31}. The latter sequence accounts for the partial sequence –Leu–Tyr–Leu– in the B chain of insulin. Peptides I_6 and I_8, in addition to containing (15) (16) (17)

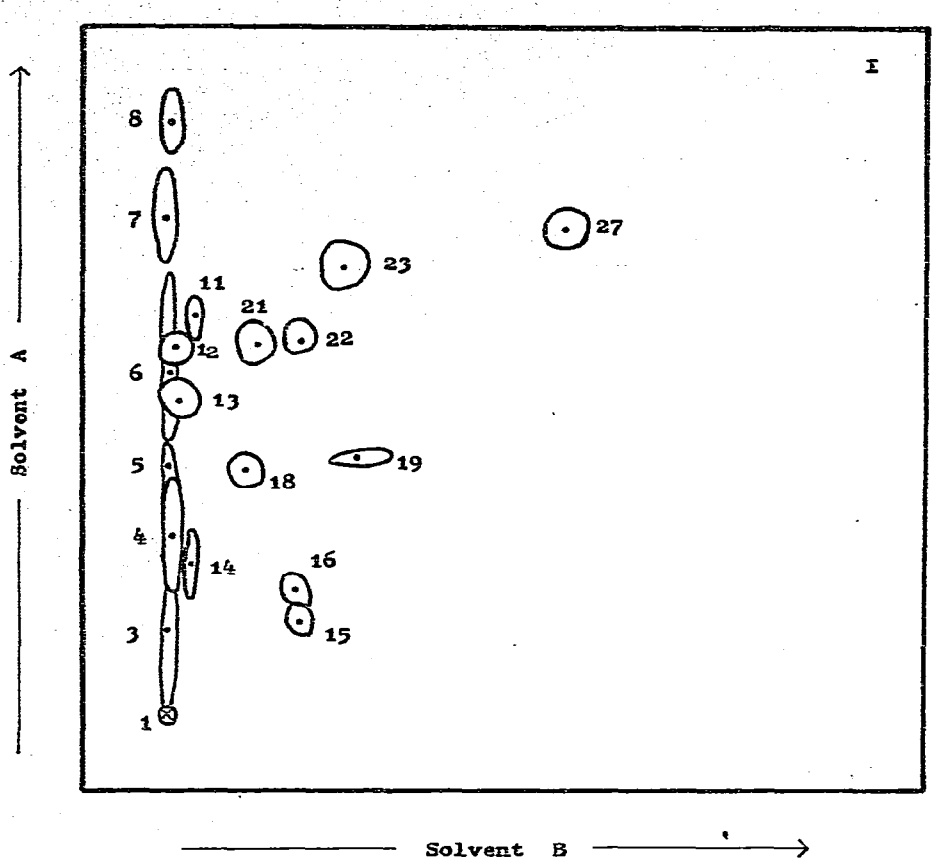

Fig. 16. Chromatogram of insulin hydrolysate stained with isatin (Reagent No. 10).

tyrosine, also contained cysteine, which could be either at position 20 in the A chain or at position 19 in the B chain, or at both. Spots I_6 and I_8 had an acidic nature. Therefore, the acidic amino acids in the sequence near to cysteine (positions stated above) and tyrosine at positions 14 or 19 in the A chain, or position 16 in the B chain, could be present in the peptides I_6 and I_8. Tyrosine, at position 26 in the B chain, could not be involved in spots I_6 and I_8 because arginine was not present in these peptides. Separation of the A and B chains of insulin would clearly produce much more information, but the study was not pursued.

CONCLUSION

It will be seen that a considerable amount of useful information about the amino acid sequence of proteins can be gained from a rapid visual examination of their partial hydrolysates by TLC. While it is not suggested that this micro-technique can replace modern automated methods of total sequence determination, nevertheless it seems to be well suited to the comparative analysis of abnormal proteins which

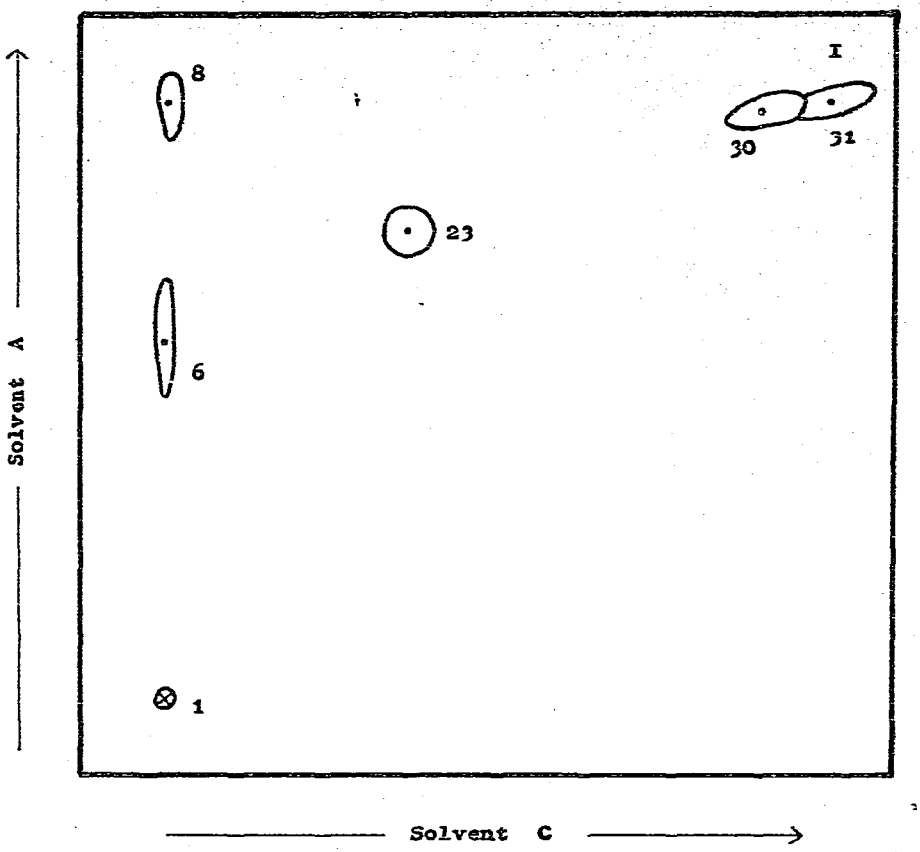

Fig. 17. Chromatogram of insulin hydrolysate stained with Folin–Ciocalteu's reagent (No. 6).

differ only slightly in structure from their normal analogues. In view of the results which have been obtained with collagen, the technique is now being applied to a study of the dermis collagen in *Osteogenesis imperfecta*[45].

ACKNOWLEDGEMENTS

We wish to thank Dr. F. S. Steven for helpful discussions and advice. We are indebted to the Medical Research Council for the support of a pilot study project.

REFERENCES

1 C. Haworth and J. G. Heathcote, *J. Chromatogr.*, 41 (1969) 380.
2 J. G. Heathcote and C. Haworth, *J. Chromatogr.*, 43 (1969) 84.
3 J. G. Heathcote, R. J. Washington, C. Haworth and S. Bell, *J. Chromatogr.*, 51 (1970) 267.
4 J. G. Heathcote, R. J. Washington, B. J. Keogh and R. W. Glanville, *J. Chromatogr.*, 65 (1972) 397.
5 J. G. Heathcote, B. J. Keogh and R. J. Washington, *J. Chromatogr.*, 78 (1973) 181.
6 J. G. Heathcote, B. J. Keogh and R. J. Washington, *J. Chromatogr.*, 79 (1973) 187.
7 J. G. Heathcote, R. J. Washington and B. J. Keogh, *J. Chromatogr.*, 92 (1974) 355.
8 J. G. Heathcote, R. J. Washington and B. J. Keogh, *J. Chromatogr.*, 104 (1975) 141.

9 V. M. Ingram, *Nature (London)*, 180 (1957) 327.
10 O. de Jongh, A. J. Perry and J. J. Hefferren, *J. Dent. Res.*, 50 (1971) 1526.
11 J. Einbinder and M. Schubert, *J. Biol. Chem.*, 188 (1951) 335.
12 C. W. Foulk and M. Hollingsworth, *J. Amer. Chem. Soc.*, 45 (1923) 1220.
13 R. M. C. Dawson, D. Elliott, W. H. Elliott and K. M. Jones, *Data for Biochemical Research*, Oxford University Press, London, 2nd ed., 1969.
14 I. Smith, *Chromatographic and Electrophoretic Techniques*, Vol. I, *Chromatography*, Heinemann, London, 2nd ed., 1960.
15 D. Folin and V. Ciocalteau, *J. Biol. Chem.*, 73 (1927) 627.
16 L. Edvinsson, R. Håkanson, A. L. Rönnberg and F. Sundler, *J. Chromatogr.*, 67 (1972) 81.
17 M. A. Naughton, F. Sanger, B. S. Hartley and D. C. Shaw, *Biochem. J.*, 77 (1960) 151.
18 D. L. Swallow and E. P. Abraham, *Biochem. J.*, 70 (1958) 364.
19 A. P. Ryle and F. Sanger, *Biochem. J.*, 60 (1955) 535.
20 J. Schultz, H. Allison and M. Grice, *Biochemistry*, 1 (1962) 694.
21 J. E. Eastoe, in G. N. Ramachandran (Editor), *Treatise on Collagen*, Vol. I, Academic Press, New York, 1967.
22 W. Traub and K. A. Piez, *Advan. Protein Chem.*, 25 (1971) 243.
23 J. Rauterberg, R. Timpl and H. Furthmayr, *Eur. J. Biochem.*, 27 (1972) 231.
24 P. Bornstein, *Biochemistry*, 6 (1967) 3082.
25 W. T. Butler and S. L. Ponds, *Biochemistry*, 10 (1971) 2076.
26 J. Rauterberg, P. P. Fietzek, F. Rexrodt, U. Becker, M. Stark and R. Kühn, *FEBS Lett.*, 21 (1972) 75.
27 W. T. Butler, *Biochemistry*, 9 (1970) 44.
28 G. Bailian, E. M. Click and P. Bornstein, *Biochemistry*, 10 (1971) 4470.
29 G. Bailian, E. M. Click, M. A. Hermodson and P. Bornstein, *Biochemistry*, 11 (1972) 3798.
30 W. T. Butler, S. P. Underwood and J. E. Finch, Jr., *Biochemistry*, 13 (1974) 2946.
31 P. P. Fietzek, F. W. Rexrodt, K. E. Hopper and K. Kühn, *Eur. J. Biochem.*, 38 (1973) 396.
32 P. Wendt, K. V. D. Mark, F. Rexrodt and K. Kühn, *Eur. J. Biochem.*, 30 (1972) 169.
33 P. P. Fietzek, F. W. Rexrodt, P. Wendt, M. Stark and K. Kühn, *Eur. J. Biochem.*, 30 (1972) 163.
34 T. D. Kroner, W. Tabroff and J. J. MacGarr, *J. Amer. Chem. Soc.*, 75 (1953) 4084.
35 T. D. Kroner, W. Tabroff and J. J. MacGarr, *J. Amer. Chem. Soc.*, 77 (1955) 3356.
36 M. L. Tanzer, T. Housley, L. Berube, R. Fairweather, C. Franzblau and P. M. Galiop, *J. Biol. Chem.*, 248 (1973) 393.
37 R. M. Osborne, R. W. Longton and B. L. Lamberts, *Anal. Biochem.*, 44 (1971) 317.
38 P. P. Fietzek, D. Breitreutz and K. Kühn, *Biochim. Biophys. Acta*, 365 (1974) 305.
39 P. P. Fietzek, H. Furthmayr and K. Kühn, *Eur. J. Biochem.*, 47 (1974) 257.
40 P. P. Fietzek and K. Kühn, *FEBS Lett.*, 36 (1973) 289.
41 P. P. Fietzek and K. Kühn, *Hoppe-Seyler's Z. Physiol. Chem.*, 355 (1974) 647.
42 C. P. Kimball and J. R. Murlin, *J. Biol. Chem.*, 58 (1923–24) 337.
43 R. L. M. Synge, *Biochem. J.*, 39 (1945) 351.
44 A. J. Perry, A. Biel, O. de Jongh and J. J. Hefferren, *Arch. Oral Biol.*, 14 (1969) 1193.
45 J. G. Heathcote, C. E. Dent, S. J. Al-Alawi and J. G. Heathcote, *Lancet*, i (1976) 756.