

MICRO-SEQUENCE ANALYSIS OF PEPTIDES AND PROTEINS USING 4-*NN*-DIMETHYLAMINOAZOBENZENE 4'-ISOTHIOCYANATE/PHENYLISOTHIOCYANATE DOUBLE COUPLING METHOD

Jui Yoa CHANG, Dieter BRAUER and Brigitte WITTMANN-LIEBOLD

Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, 1 Berlin 33 (Dahlem), Germany

Received 26 June 1978

1. Introduction

The phenylisothiocyanate (Edman) degradation has been a most valuable technique in amino acid sequence determination [1]. One of its inadequacies is, however, the unsatisfactory sensitivity of the identification of PTH-amino acids. This shortcoming has prevented the conventional Edman method from being a practical micro-sequencing technique. The combined use of the sensitive dansyl method [2] with the subtractive Edman degradation has largely overcome this problem and contributed, to a great extent, to the sequence analysis of proteins which are available in only limited quantities. Recently, the development of high sensitivity sequencing methods has centered on the direct identification of PTH-amino acids. These attempts, mostly in conjunction with automated Edman degradations [3,4] include:

- (1) Use of radioactive Edman reagent [5-7];
- (2) Incorporation of radioactivity into proteins by *in vivo* cell culturing in the presence of radioactive amino acids [8];
- (3) Application of high performance liquid chromatography for highly sensitive identification of PTH-amino acids [9,10].

Abbreviations: DABITC, 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate; DABTC, 4-*NN*-dimethylaminoazobenzene 4'-thiocarbonyl; DABTZ, 4-*NN*-dimethylaminoazobenzene 4'-thiazolinone; DABTH, 4-*NN*-dimethylaminoazobenzene 4'-thiohydantoin; PITC, phenylisothiocyanate; PTC, phenylthiocarbonyl; PTH, phenylthiohydantoin; dansyl, 5-dimethylamino-1-naphthalenesulfonyl; TLC, thin-layer chromatography

More recently, we have introduced a colored Edman reagent 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate (DABITC) [11]. The preliminary use of this new reagent in the sequence determination of peptides and proteins [12,13] has proven to be very promising in micro-sequence analysis.

We should like to describe here a novel manual method for micro-sequencing of peptides and proteins. This method employs a double coupling technique, i.e., a first coupling with DABITC and a second coupling with PITC to achieve a quantitative coupling reaction. After the acid cleavage and conversion reaction, only the colored DABTH-amino acids are identified by TLC. This new modification avoids the use of the high coupling temperature (75°C) required when employing only single coupling with DABITC [12].

The new sequencing method is simple, sensitive and inexpensive. One degradation cycle takes ~140 min, needs no special instrumentation or radioactive materials and thus should provide a generally applicable method for micro-sequence analysis of peptides and proteins.

2. Materials and methods

Ribosomal proteins S9, S11, L29 and L30 from *E. coli* were provided by Professor H. G. Wittmann. Their purities were checked by two-dimensional polyacrylamide gel electrophoresis [14]. Tryptic peptides of ribosomal protein L24 were isolated from an ion exchange column or a G-50 Sephadex column (L24-T12, eluted with 50% HCOOH) (B. W.-L., in preparation). Both tryptic peptides of L1 and *Staphylococcus aureus* protease-digested peptide of S3 were isolated

and recovered from cellulose TLC [15]. Melittin, insulin B chain (oxidized) and lysozyme were purchased from Sigma, USA and Boehringer, FRG.

DABITC was prepared by the method in [11,12]. PITC (Merck) was redistilled under nitrogen at reduced pressure (b. p. $\sim 92^{\circ}\text{C}/12$ mm Hg). Pyridine (Merck) was redistilled 3 times over: (i) KOH (10 g/l); (ii) ninhydrin (1 g/l) and (iii) KOH (10 g/l). Trifluoroacetic acid (Fluka) was redistilled over $\text{CaSO}_4 \cdot 0.5 \text{H}_2\text{O}$ (10 g/l). Glacial acetic acid saturated with HCl was prepared by bubbling HCl gas through acetic acid for 2 h. The other solvents used were of commercial analytical grade (Merck) and were used without further purification.

Polyamide sheets were purchased from Schleicher and Schüll, FRG and silica gel plates (G60, without fluorescent indicator, 0.25 mm) were obtained from Merck.

2.1. Sequence determination

Peptides or proteins (2–8 nmol), placed in an acid-washed tube (0.6–0.8 cm i.d. \times 5 cm) fitted with a Quickfit glass stopper, were dissolved in 80 μl aq. 50% (v/v) pyridine and treated with 40 μl DABITC solution (10 nmol/ μl pyridine (2.82 mg/ml), freshly prepared, see section 3). The tube was flushed with N_2 for 10 s, sealed with a glass stopper and placed in a heating block at 52°C for 50 min. After the first coupling, 10 μl PITC was added and the second coupling reaction was allowed to proceed at 52°C for 30 min. After the reaction, the excess reagents and by-products were removed by mixing the reaction mixture with two portions of 0.5 ml heptane/ethyl acetate (2:1, v/v) on a vortex mixer and centrifuging (in some cases, a third washing is also recommended, see section 3). The organic phase was removed with a fine pipette and discarded. The mixture was evaporated in a high vacuum. The dried residue was dissolved in 50 μl anhydrous trifluoroacetic acid, flushed with N_2 , sealed with a glass stopper and heated in a heating block at 52°C for 15 min. The sample was evaporated in a vacuum desiccator and dissolved in 50 μl water. Extraction of the cleaved DABTH-amino acids (and PTZ-amino acids) was performed by mixing with one portion of 200 μl butyl acetate on a vortex mixer and centrifuging. After removal of the butyl acetate extract, the peptide in the water phase was evaporated in a desiccator and subjected to the next

degradation cycle.

The butylacetate extract was evaporated and the residue was redissolved in water (20 μl) and acetic acid saturated with HCl (40 μl). Conversion of the thiazolinones of amino acids into thiohydantoin was carried out in the 52°C heating block for 50 min. The sample was dried and redissolved in a suitable volume of ethanol (5–30 μl) and 1/40–1/5 of the ethanol extract was used for TLC identification.

2.2. Identification of DABTH-amino acids

The identification of DABTH-amino acids on both polyamide sheet and silica gel plate have been reported [12,16,17]. The blue synthetic marker, DABTC-diethylamine (10–20 pmol), was co-chromatogrammed with each unknown sample to facilitate the identifications. For identification of DABTH-amino acids on polyamide sheet (2.5 \times 2.5 cm), acetic acid/water (1:2, v/v) was used as solvent for the first dimensional separation and toluene/*n*-hexane/acetic acid (2:1:1, by vol.) was used as solvent for the second dimensional separation. For identification of DABTH-Leu and DABTH-Ile on silica gel plates, chloroform/ethanol (100:3, v/v) was used as solvent for the one-dimensional separation. After TLC separation, the plates were dried and exposed to HCl vapour. The sensitivity of this TLC technique is ~ 5 –25 pmol for polyamide sheets and 25–50 pmol for silica gel plates.

3. Results

3.1. Performance of sequence analysis

The kinetics of reaction of DABITC with amino acids, peptides and proteins have been studied in detail [18]. Owing to the limited solubility of DABITC in organic solvents, which prevents the arbitrary increase of the molar ratio of DABITC : peptide, the quantitative coupling of N-terminal group with DABITC can only be achieved under rather drastic condition (75°C , 1.5 h). This high temperature promotes the rapid hydrolysis of DABITC and could cause, possibly, some unexpected side reactions.

The use of the DABITC–PITC coupling method enables quantitative coupling to be carried out at moderate temperature (52°C). From the recoveries of the colored DABTH-amino acids, it is estimated that

only 25–50% of NH_2 termini were reacted with DABITC at the first coupling reaction. The addition of excess PITC in the second coupling served to complete the coupling reaction and converted the hydrolyzed DABITC into a blue colored thiourea (U, fig.1) which could be removed together with excess DABITC and PITC by extractions.

The choice of solvent used to extract by-products and excess reagents after the coupling reaction during extended sequence analysis of non-polar peptides is crucial. Experiments in [18] indicate that the solvent heptane/ethyl acetate (2:1, v/v) suggested [19] is the best choice. Two times extraction by this solvent from 67% aqueous pyridine was able to wash out 97–98% of the excess reagents and by-products. Nevertheless, due to the incomplete removal of trace amount of the organic phase (after the second extraction) which might interfere with the later identification of DABTH-amino acids, a third washing is also recommended when the yields of DABTH-amino acids become very low. This extraction procedure has allowed us to establish the sequence of all the non-polar peptides analysed up to the C-terminal ends (except for C-terminal lysine) without significant decline of recoveries (fig.3). The presence of C-terminal lysine, on the other hand, could be deduced from the amino acid analysis or recognized from the red color of the aqueous phase after extraction of DABTZ-amino acids.

Cyclization of DABTC-peptides to form DABTZ-amino acids in anhydrous trifluoroacetic acid appeared to be complete within 2–3 min at 52°C, with the exception of DABTC-proline peptides which take 10–12 min (J. Y. C., unpublished results). It is a common observation that proline gives an abnormally slow cleavage reaction (in anhydrous acid) during the Edman degradation [20,21]. We therefore take 15 min at 52°C as routine conditions for acid cleavage.

The conditions for conversion reactions have also been investigated [18]. In HCl-saturated acetic acid/water (2:1, v/v) at 50°C, the half-life ($t_{1/2}$) for the conversion of DABTZ-amino acids to DABTH-amino acids (via DABTC-amino acids) range from 2–4 min, except for glycine [22], for which $t_{1/2}$ is considerably longer (8–9 min). The half-lives of conversion reactions in 1 N HCl at 50°C are ~3–4 times greater than those in HCl-saturated acetic acid/water. Moreover, the DABTZ-amino acids do not dissolve very well in 1 N

HCl. We use 52°C, 50 min in HCl-saturated acetic acid/water (2:1, v/v) as the optimum conditions for conversion reactions. Under these conditions, only 8–15% of asparagine and glutamine were deaminated. It is believed that another acid solution, such as 50% aq. trifluoroacetic acid, could also be used for conversion reactions.

One should pay careful attention to the following points during the operation:

- (1) DABITC is not very stable in pyridine and is better prepared freshly daily or at every degradation cycle. A large stock solution of DABITC in acetone (1.4 mg/ml) is prepared. Suitable volumes (e.g., 0.4 ml) of this solution are pipetted into clean test tubes, dried under vacuum pump and redissolved in pyridine (0.2 ml) shortly before use.
- (2) Some hydrophobic peptides tend to precipitate as fine particles or thin films during the extractions (either after coupling or after cleavage). These precipitates are suspended at the interface of the aqueous and organic phases after centrifugation and could be inadvertently removed by pipetting.
- (3) Each drying should be very thorough and care must be taken that the samples do not bump.

3.2. Identification of DABTH amino acids

The tactics of identifying each DABTH-amino acid have been discussed [12,16]. Figure 1 gives the two-dimensional separation of DABTH-amino acids on a Schleicher and Schüll polyamide sheet developed by the solvents in section 2.

Several points should be observed during the identifications:

- (1) Leucine and isoleucine can only be distinguished on silica gel plates.
- (2) The discrimination between valine, leucine (or isoleucine), methionine and phenylalanine could be achieved by relating their positions to the marker(e).
- (3) Serine normally appeared as 4 distinct spots; namely DABTH-Ser (S), a dehydro-product (S^Δ), a polymerized-product (S^\square) and a hydrated dehydro-product with an OH group in the α -position (S°) [18].
- (4) Threonine normally appeared as DABTH-Thr (T) and dehydro threonine (T^Δ). The latter overlaps with DABTH-Phe (F). In many cases, a blue colored spot (denoted as T^\times in fig.1, near thiourea U)

Table 1
Proteins and peptides sequenced by the micro-sequencing procedure

Protein/peptide	Quantity ^a (nmole)	Sequence	Reference
L24-T ₁₄ ^b	4.7	ADR	unknown sequence
L24-T ₉ ^b	3.7	NVLSSGK ^d	" "
L24-T ₁₀ ^b	3.4	VIVEGINLVK ^d	" "
L24-T _{11a} ^b	4.2	HQKPVPALNQPGGIVEK ^d	" "
L24-T ₁₃ ^b	5	EAAIQVSNVAIFNAATGK ^d	" "
L1-T ₂₀ ^b	3.2	KSDQNVR	" "
L1-T ₁₇ ^b	3	VVGQLGQVLGPR	" "
L1-T ₁₁ ^b	2.5	GATVLPHTGR	" "
S3-SP ₅ ^c	2.7	LDAKLVADSITSQLE	" "
S9	5.7	AENQYYGTGRR-SSAARVF-	[25]
S11	6.6	NMA ^e KAPIRARKRVRKQV-DGVA-	[23,24]
L29	6.4	MKAKELREKSVEELNTELLNLLREQ-	[26]
L30	8.2	AKTIKITQTRSAIGRLPKH-	[27]
Insulin B chain	3.6	FVNQHL-GSHLVEALYLIV-GERGFFYTP-	[32]
Lysozyme	3	KVFGR-ELAAAMKRHGLDNY-	[28]
Melittin	2.5	GIGAVLKVLTTGLPALIS-I-R-	[29]

^a The quantities were calculated from amino acid analysis

^b Tryptic peptide

^c Peptide from digestion with *Staphylococcus aureus* protease

^d C-terminal lysine was deduced from the amino acid analysis

^e NMA is *N*- α -monomethyl-alanine

identification of the 25th residue required 1/4 DABTH-Gln extract. L29 is the only protein analyzed which remains soluble during the coupling reactions throughout the sequence analysis.

4. For L30 [27], the N-terminal sequence up to

20 residues was analyzed. The first 19 residues were established unambiguously. However, substantial overlapping by alanine and glycine started to appear after the 18th residue and the identification of amino acids after the 19th degradation

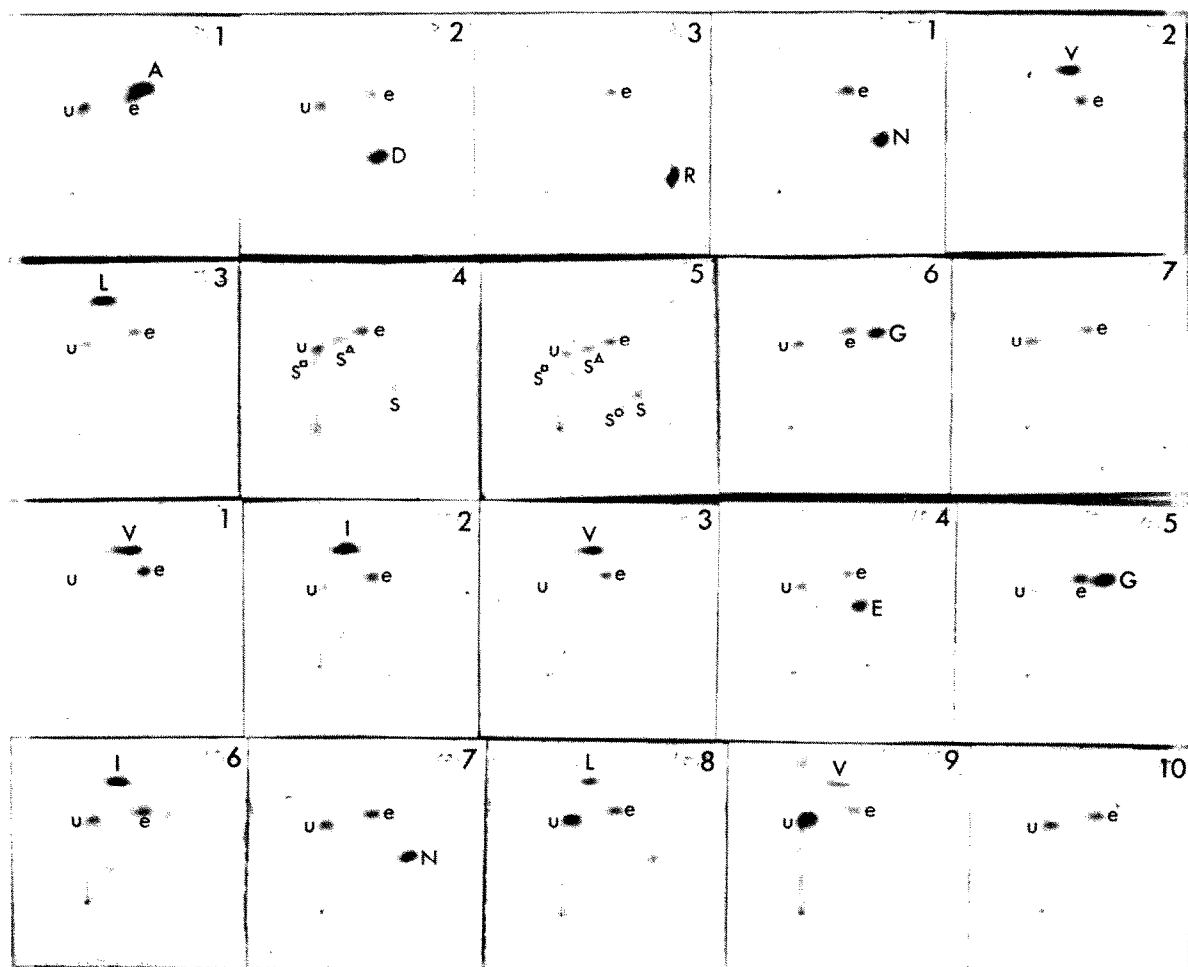


Fig.2. Photograph of the polyamide TLC sheets resulting from the sequence analysis of three tryptic peptides (one tripeptide T_{14} , one heptapeptide T_7 and one decapeptide T_{10} , ~4 nmol each) of ribosomal protein L24. The sequences are listed in table 1. The C-terminal lysine of the heptapeptide and the decapeptide were not detected. Only DABTH-amino acids appeared red. Both by-products (U) and added marker (e) were blue.

was not possible.

5. For native lysozyme [28], the expected residues appeared as the major spot on TLC only up to the 16th residue. After the 16th residue, the sequence could only be deduced by careful examination of the newly appearing spots (like S11, fig.5).
6. For melittin [29], the N-terminal sequence could be established for the first 18 residues. Residue 19 and 21 were not detected. However, residue 20 (Ile) and 22 (Arg) could, again, be distinctly identified when 1/4 of the extracts were used.

Most of the intact proteins (except L29) remained

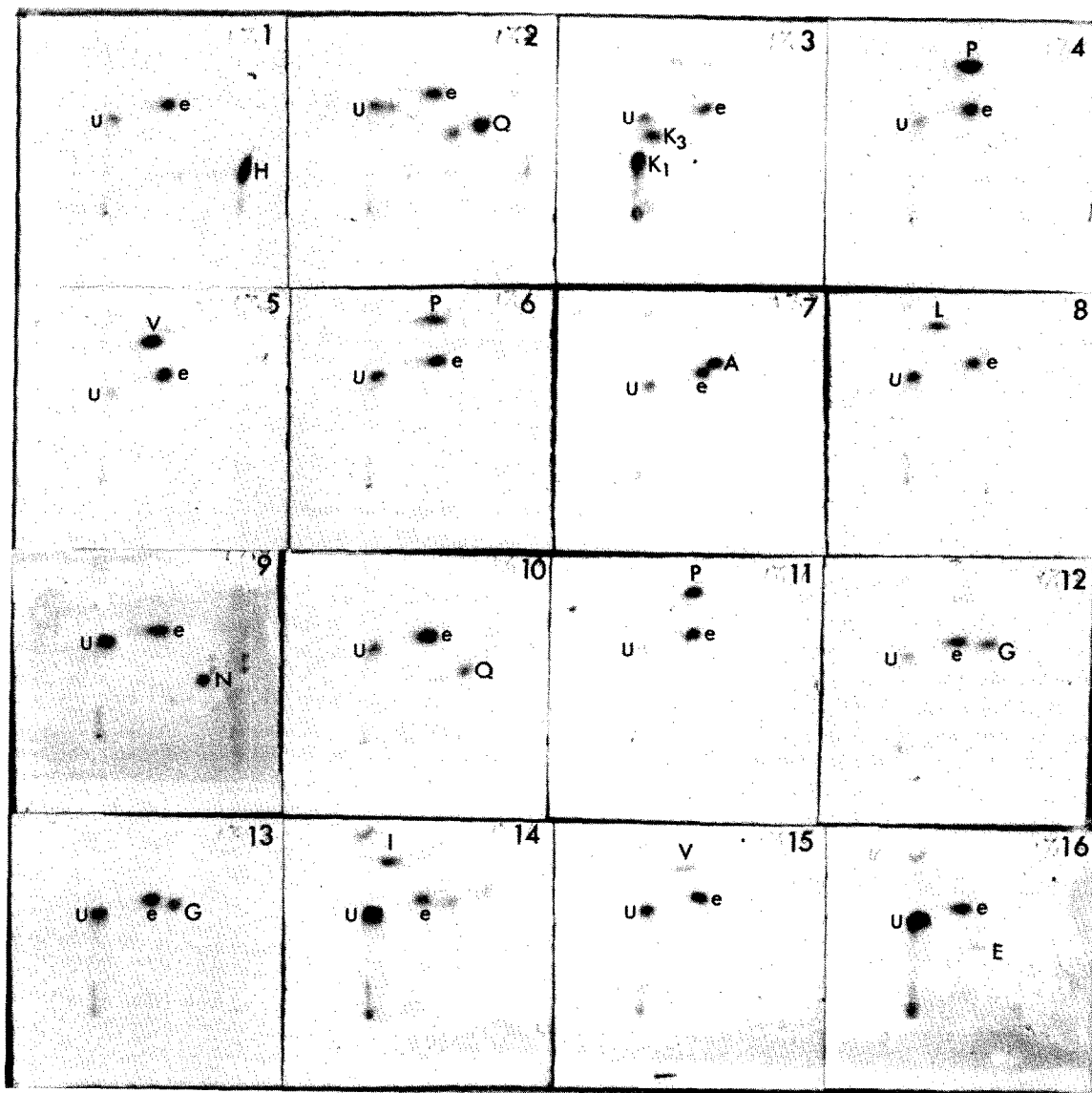


Fig.3. Photograph of the polyamide TLC sheets resulting from the sequence analysis of a seventeen residue peptide T_{11a} derived from the tryptic digest of ribosomal protein L24. 4.2 nmol peptide was analyzed. The portions of the DABTH-amino acid extracts used for TLC identification, corresponding to the residue numbers, are $(\frac{1}{40})$ $(\frac{1}{40})$ $(\frac{1}{30})$ $(\frac{1}{30})$ $(\frac{1}{30})$ $(\frac{1}{30})$ $(\frac{1}{30})$ $(\frac{1}{20})$ $(\frac{1}{20})$ $(\frac{1}{20})$ $(\frac{1}{20})$ $(\frac{1}{20})$ $(\frac{1}{20})$ $(\frac{1}{8})$ $(\frac{1}{8})$. One can notice the sharp decline of the recoveries at the 15th and 16th residues. The C-terminal residue lysine, which was deduced from the amino acid analysis, was not detected.

insoluble and attached to the tube wall during the coupling reactions after the first degradation cycle, presumably due to the incorporation of the hydrophobic DABITC at the lysine side chains which rendered the proteins insoluble in 67% aqueous pyridine.

This new sequencing method has also been successfully applied to the sequence analysis of peptides derived from ribosomal proteins S8 and L13 (G. Allen and L. Mende, personal communication). Sequences of up to 22 residues were obtained.

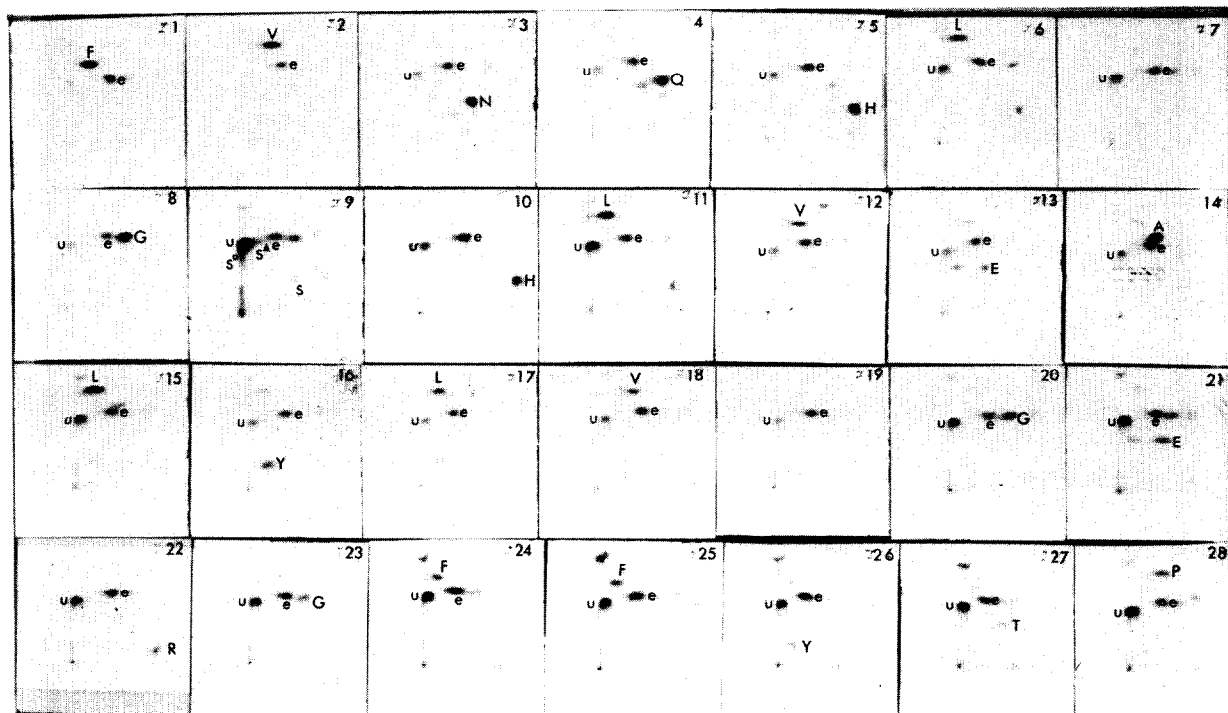


Fig.4. Photograph of the polyamide TLC sheets resulting from the sequence analysis of 3.6 nmol insulin B chain (oxidized). Residues 7 and 19 (both cysteic acid) were not detected. The portions of the DABTH-amino acid extracts applied for TLC identifications were ranged from 1/30 (1st residue) to 1/3 (28th residue). At residue 9 (serine), the discrimination between the by-product, U (blue) and the presumed polymerized product of DABTH-Ser, S[□] (red) could be clearly made from their color difference on the original sheet.

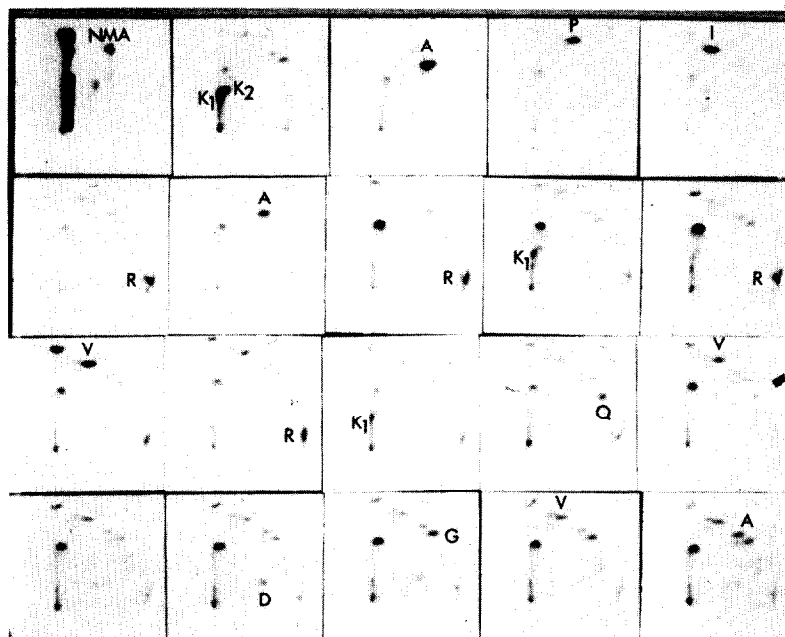


Fig.5. Photograph of the polyamide TLC sheets resulting from the sequence analysis of ribosomal protein S11 (6.6 nmol). The marker (e) was not used in this case. 20 residues were obtained from 19 degradations. This is due to the abnormal behavior of the N-terminal monomethylalanine which cleaved to form thiohydantoin derivatives at the coupling stage of the first degradation/(first sheet). The first degradation, therefore, resulted in the appearance of the second amino acid residue lysine [33]. Residue 16 could not be confirmed. Residues 17–20 were tentatively established as Asp–Gly–Val–Ala by examination of the newly appearing spots.

4. Discussion

With the increasing interest in the study of the structure–function relationships of proteins which can only be obtained in limited quantities, there has been a growing demand for the development of a highly sensitive technique for amino acid sequence determination. To meet this demand, a number of modifications based on the automatic Edman degradation have been perfected which allow sequence determination to be carried out on only 0.1–10 nmol peptides and proteins [6–8,10]. These methods, however, require either expensive instruments or hazardous radioactive materials and, in our opinion, are unlikely to become generally acceptable.

The manual dansyl-Edman method [30,31] originally introduced [2] is, so far, the most widely used method for sequence determination of small amount of peptides. Unfortunately, its sensitivity and efficiency have been hampered by the necessity of removing an aliquot of the peptide at each degradation cycle and the intrinsic disadvantage of total hydrolysis in the dansyl method which destroys asparagine, glutamine and tryptophan. In practice, one needs ≥ 20 nmol peptide to establish the amino acid sequence up to 20 residues. The incomplete hydrolysis at some hydrophobic area which resulted in dansyl-dipeptides is also an undesirable side reaction.

The results reported here show that the sequencing method described possesses several distinct advantages that overcome most of the shortcomings encountered in other reported micro-sequencing methods:

1. An amount of 2–8 nmol peptide or protein is sufficient for sequence determination up to 20–30 residues. This marks a 10–20-fold increase of sensitivity over the conventional manual technique. Even under these conditions, one needs only 1/40–1/5 of the DABTH-amino acid extract for TLC identification. This is important especially when ambiguous results appear and repeated identification becomes necessary.
2. A single polyamide TLC identification is adequate to distinguish all the common amino acids (as their DABTH derivatives) except for leucine and isoleucine which can be distinguished on silica gel

plates. Although TLC is not quantitative, it is simple, efficient and low cost.

3. The color difference between the amino acid derivatives (DABTH, red color) and contaminants (usually DABTC derivative, blue color) greatly facilitates identifications.
4. The direct identification of degraded DABTH-amino acids (including Trp, Asn, Gln, Ser and Thr) enabled the sequence determination to be carried out in a simple and efficient way.

The method described has been applied to the sequence analysis of peptides recovered from gel filtration, ion exchange column chromatography and cellulose TLC. The combined use of the cellulose TLC technique, which has proven to be an elegant technique for separation of peptides in the scale of 2–5 nmol, with the new sequencing method has been found to be a most useful one in the micro-sequencing of peptides in our laboratory.

This new method has also been found to work on intact proteins. Although unambiguous results were normally obtained for the first 19–25 residues, this manual method would provide a simple and inexpensive alternative in addition to the automatic sequencer for N-terminal sequence determination of intact proteins.

Still, there are some aspects of the new method that have not yet been perfected. The recoveries of the acid labile DABTH-Ser and DABTH-Trp were usually too low for their satisfactory identification beyond the 15th residue when starting with < 5 nmol peptide. One way of circumventing this problem would be the use of larger amounts of material, e.g., 15–20 nmol peptides. A new solvent system to discriminate DABTH-Leu and DABTH-Ile on a polyamide sheet is presently under development as identification of DABTH-amino acids on a silica gel plate normally takes 3–5 times more material than on a polyamide sheet.

Acknowledgements

We thank Professor H. G. Wittmann for encouragement during this work and Dr G. Allen for reading this manuscript.

References

- [1] Edman, P. (1956) *Acta Chem. Scand.* 10, 761–765.
- [2] Gray, W. R. and Hartley, B. S. (1963) *Biochem. J.* 89, 379–380.
- [3] Edman, P. and Begg, G. (1967) *Eur. J. Biochem.* 1, 80–91.
- [4] Laursen, R. A. (1971) *Eur. J. Biochem.* 20, 89–102.
- [5] Silver, J. and Hood, L. (1975) *Anal. Biochem.* 67, 392–396.
- [6] Jacobs, J. W. and Niall, H. D. (1975) *J. Biol. Chem.* 250, 3629–3636.
- [7] Bridgen, J. (1976) *Biochemistry* 15, 3600–3604.
- [8] McKean, D. J., Peters, E. H., Waldby, J. I. and Smithies, O. (1974) *Biochemistry* 13, 3048–3051.
- [9] Zimmerman, C. L., Appella, E. and Pisano, J. J. (1977) *Anal. Biochem.* 77, 569–573.
- [10] Margolies, M. N. and Brauer, A. W. (1978) *J. Chromatog.* 148, 429–439.
- [11] Chang, J. Y., Creaser, E. H. and Bentley, K. W. (1976) *Biochem. J.* 153, 607–611.
- [12] Chang, J. Y. (1977) *Biochem. J.* 163, 517–520.
- [13] Chang, J. Y., Creaser, E. H. and Hughes, G. J. (1977) *FEBS Lett.* 78, 147–150.
- [14] Kaltschmidt, E. and Wittmann, H. G. (1970) *Anal. Biochem.* 36, 401–412.
- [15] Heiland, I., Brauer, D. and Wittmann-Liebold, B. (1976) *Hoppe-Seyler's Z. Physiol. Chem.* 357, 1751–1770.
- [16] Chang, J. Y. and Creaser, E. H. (1977) *J. Chromatog.* 132, 303–307.
- [17] Chang, J. Y., Creaser, E. H. and Hughes, G. J. (1977) *J. Chromatog.* 140, 125–128.
- [18] Chang, J. Y. (1977) Ph. D. Thesis, The Australian National University, Canberra.
- [19] Tarr, G. E. (1975) *Anal. Biochem.* 63, 361–370.
- [20] Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H. and Walsh, K. A. (1972) *Biochemistry* 11, 4493–4502.
- [21] Brandt, W. F., Edman, P., Henschen, A. and Von Holt, C. (1976) *Hoppe-Seyler's Z. Physiol. Chem.* 357, 1505–1508.
- [22] Ilse, D. and Edman, P. (1963) *Aust. J. Chem.* 16, 411–416.
- [23] Wittmann-Liebold, B. (1973) *FEBS Lett.* 36, 247–249.
- [24] Chen, R., Brosius, J., Wittmann-Liebold, B. and Schäfer, W. (1977) *J. Mol. Biol.* 111, 173–181.
- [25] Chen, R. and Wittmann-Liebold, B. (1975) *FEBS Lett.* 52, 139–140.
- [26] Bitar, K. G. (1975) *Biochim. Biophys. Acta* 386, 99–106.
- [27] Ritter, E. and Wittmann-Liebold, B. (1975) *FEBS Lett.* 60, 153–155.
- [28] Canfield, R. E. (1963) *J. Biol. Chem.* 238, 2698–2707.
- [29] Habermann, E. and Jentsch, J. (1967) *Hoppe-Seyler's Z. Physiol. Chem.* 348, 37–50.
- [30] Weiner, A. M., Platt, T. and Weber, K. (1972) *J. Biol. Chem.* 247, 3242–3251.
- [31] Gray, W. R. (1972) *Methods Enzymol.* 25, 333–343.
- [32] Sanger, F. and Tuppy, H. (1951) *Biochem. J.* 49, 463–481.
- [33] Chang, J. Y. (1978) *FEBS Lett.* 91, 63–68.