A Novel Manual Method for Protein-Sequence Analysis

By JUI YOA CHANG and ERNEST H. CREASER

*Protein Biochemistry Unit, Research School of Biological Sciences,
The Australian National University, P.O. Box 475, Canberra City, A.C.T. 2601, Australia*

(Received 15 December 1975)

A novel manual method for protein-sequence analysis is described. Three peptides, the hexapeptide (Leu-Trp-Met-Arg-Phe-Ala), insulin A chain and glucagon were used to test this technique. Peptides (1 or 2 nmol) were hydrolysed with acid and their qualitative amino acid compositions were confirmed by reacting with 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl chloride and 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate. Sequence determination of 20–200 nmol of peptide was then performed by the combined use of phenyl isothiocyanate and 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate, a new procedure that is analogous to the dansyl-Edman method with the replacement of dansyl chloride by 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate as the *N*-terminal residue determination reagent. On t.l.c. this new *N*-terminal reagent gave brightly coloured 4-*NN*-dimethylaminoazobenzene-4-thiohydantoin derivatives of amino acids and showed the following advantages: (1) the detection sensitivity is in the pmol range; (2) u.v. observation is not required; (3) there is no destruction of acid-labile amino acids; (4) two-dimensional t.l.c. separation is adequate to identify 24 amino acids, except leucine and isoleucine (this pair of amino acids can be resolved by using 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl chloride); (5) the determination of a new *N*-terminal residue (from coupling to t.l.c. identification) takes only 3 h; (6) the colour difference between isothiocyanate, thiocarbonyl and thiohydantoin derivatives facilitates the identifications.

Composition analysis of peptides by the amino acid automatic analyser and their sequence determination by Edman degradation, either by automatic Edman degradation (Edman & Begg, 1967; Laursen, 1971), subtractive Edman degradation (Konigsberg, 1972), manual Edman degradation with direct identification of phenylthiohydantoin derivatives (Schröder, 1972), or manual Edman degradation plus the dansyl method (Gray, 1972; Hartley, 1970), have been the most important and common techniques used in investigating the primary structure of proteins. However, the high running expenses of the automatic sequencer, the insensitive and time-consuming nature of the subtractive method, the immaculate laboratory technique needed in the direct Edman method and the extra work of positioning amide groups and the destruction of tryptophan in the dansyl-Edman method, justify the search for alternative methods that might overcome these problems.

In the present paper, we describe a novel method of protein-sequence analysis, which includes: (a) qualitative amino acid-composition analysis by analysing the reaction products of proteins or peptide hydrolysates with 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl chloride and 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate respectively; (b) the amino acid-sequence determination of peptides or proteins by the

combined use of phenyl isothiocyanate and 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate, i.e. phenyl isothiocyanate is used only as a degradation reagent and 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate as the *N*-terminal-residue determination reagent. Both 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl chloride (I) and 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate (II) are azo dyes with the functional group of sulphonyl chloride or isothiocyanate (structures as indicated). Their synthesis and effectiveness for amino acid labelling and *N*-terminal residue determination have been reported (Chang & Creaser, 1976; Chang *et al.*, 1976; Lin & Chang, 1975).

This new method has been proved to possess several advantages that overcome most of the shortcomings encountered in the conventional methods.

Materials and Methods

The hexapeptide (Leu-Trp-Met-Arg-Phe-Ala) and bovine insulin A chain were from Schwarz/Mann, Orangeburg, NY, U.S.A. Glucagon was from Sigma, St. Louis, MO, U.S.A. All solvents used in this study were commercial analytical grade and were redistilled before use. Phenyl isothiocyanate was from Fluka, Buchs, Switzerland or Pierce Co., Rockford, IL, U.S.A. Specially purified pyridine (Pierce), anhydrous

trifluoroacetic acid and phenyl isothiocyanate are preferable in the extended sequence determination. 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate was prepared by the method of Chang *et al.* (1976). 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl chloride was synthesized as described by Lin & Chang (1975). Chen-Chin polyamide sheets were from Pierce.

Phenyl isothiocyanate solution (10%, v/v; in pyridine) was flushed with N₂ every time before being re-stoppered and it was kept refrigerated when not in use. Triethylamine/acetic acid buffer was prepared by mixing 50 ml of water, 50 ml of acetone, 0.2 ml of triethylamine and 5 ml of 0.2 M-acetic acid; the observed pH was 9.65 (pH-meter calibrated with sodium borate buffer, pH 9.2).

Qualitative amino acid-composition analysis

Two portions of 1 or 2 nmol of peptides sealed in separate evacuated test tubes (0.6 cm × 8 cm) containing 20 μl of 6 M-HCl were hydrolysed at 105–110°C for 22 h. HCl was then removed in a vacuum desiccator over P₂O₅ and NaOH pellets. Triethylamine/acetic acid buffer (20 μl, described above) was then added to the samples and re-dried to ensure the complete removal of HCl. The dried samples were then treated with 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate or 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl chloride.

One sample was dissolved in 20 μl of 0.2 M-NaHCO₃ and treated with excess of 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl chloride dissolved in the same volume of acetone. The mixture was sealed with a rubber stopper and left at 70°C for 15–20 minutes; then it was ready for t.l.c. identification.

The second sample was dissolved in 40 μl of triethylamine/acetic acid buffer (pH 9.65) and treated with excess of 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate dissolved in one-half the volume of acetone. The mixture was left at 50°C for 75 min; dried in a vacuum desiccator and then dissolved in 20 μl of water and 40 μl of acetic acid saturated with HCl. The mixture was allowed to react at 50°C for 45 min in a sealed tube, dried again and dissolved in 50 μl of ethanol. About 1 μl samples were removed for t.l.c. identification. The molar ratio of reagent to sample should be about 3–4:1. Greater excess of reagent is likely to interfere with the t.l.c. identification; 50 nmol of reagent is a satisfactory upper limit.

Sequence determination of peptides and proteins

The procedures of phenyl isothiocyanate degradation were based on Gray's (1972) method, summarized as follows. Insulin A chain or glucagon (200 nmol), placed in a 6 cm × 1.2 cm tube fitted with a Quick-Fit glass stopper, was dissolved in 200 μl of aq. 50% pyridine, 5 μl was removed for the *N*-terminal residue determination, and the rest was treated with 100 μl of 10% phenyl isothiocyanate solution (in pyridine). The tube was flushed with N₂ for a few seconds, sealed tightly with the glass stopper and placed in an oven at 50°C for 30 min, then the mixture was dried in a vacuum desiccator over P₂O₅ and NaOH pellets. To ensure the complete evaporation of phenyl isothiocyanate, the evacuated desiccator should remain in an oven at 70°C for 15–20 min. The dried sample was treated with 100 μl of anhydrous trifluoroacetic acid, flushed with N₂, sealed and placed in an oven at 50°C for 10 min. The sample was dried in a vacuum desiccator again and dissolved in 150 μl of water. Extraction of non-volatile by-products was performed by mixing with 800 μl of ethyl acetate three times on a vortex mixer and centrifuging in a bench-top centrifuge for a few minutes to separate the water and ethyl acetate phase. After the third portion of ethyl acetate extract was removed, the sample was dried in a desiccator and subjected to the next degradation cycle.

In order to compensate for the gradual decrease of the amount of peptide, due to the removal of small portions at every degradation step, the reaction volume was reduced to three-quarters of the original volume after seven to eight cycles (i.e. 150 μl of aq. 50% pyridine and 75 μl of phenyl isothiocyanate solution were added). However, the portions removed for *N*-terminal residue determination were still kept at 5–7 μl. After 14–15 cycles, the amount was further reduced to one-half of the original volume and the volume of the samples removed was raised to 10 μl.

For the hexapeptide, the whole scale of the reaction was reduced to one-quarter of that used with insulin A chain and glucagon, i.e. starting with 50 nmol of peptide, but the portions removed were kept at 5–7 μl (5 nmol).

For *N*-terminal residue determination, the sample removed (containing approx. 5 nmol of peptide) was placed in a test tube (0.6 cm × 6 cm) and dried in a vacuum desiccator. It was then dissolved in 40 μl of triethylamine/acetic acid buffer (pH 9.65) and mixed with 20 μl of 4-*NN*-dimethylaminoazobenzene 4'-is-

thiocyanate solution (2 nmol/ μ l of acetone). The mixture was flushed with N_2 , sealed with a rubber stopper, heated in an oven at 50°C for 75 min, and then dried in the vacuum desiccator. Water (40 μ l) was added. The extraction of excess of 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate was performed by mixing with one portion of 250 μ l of benzene. The purpose of this extraction is to eliminate the possible overloading of the t.l.c. sheet by 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate when the *N*-terminal residue is found to be very weak and it is necessary to apply more of the sample to the sheet. (For the small peptides, when there is danger of extracting 4-*NN*-dimethylaminoazobenzene-4'-thiocarbonyl peptide derivatives into the benzene, this extraction should be avoided.) After removal of benzene by a fine tip pipette, the sample was dried and dissolved in 20 μ l of water, and 40 μ l of acetic acid saturated with HCl. The tube was flushed with N_2 , sealed with a rubber stopper and left in an oven at 50°C for 45 min. After the reaction period the sample was dried and redissolved in 25 μ l of ethanol. About 0.25–2 μ l samples were applied to a polyamide sheet for t.l.c. identification.

The total time for one degradation cycle is 2.5 h; for one *N*-terminal residue determination it is 3.5 h.

Thin-layer chromatography on polyamide sheets

Polyamide sheets (2.5 cm $\times$ 2.5 cm) were used in both qualitative amino acid-composition analysis and *N*-terminal residue determination. The samples were carefully applied on the origin (about 6 mm from the edges of two adjacent sides) the diameter of the spot being confined to 1.0 mm by using a hair drier. Two-dimensional development in a covered jar (3 cm $\times$ 8 cm $\times$ 8 cm) was by ascending solvent flow. No phase equilibration was necessary. For separating the 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin derivatives of amino acids, solvent 1 [water/acetic acid (2:1, v/v)] was used for the first-dimension development and solvent 2 [toluene/*n*-hexane/acetic acid (2:1:1, by vol.)] was used for the second-dimension development. For separating 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl amino acid derivatives, solvent 3 [water/2-chloroethanol/formic acid (200:120:7, by vol.)] was used for the first-dimension development and solvent 4 [benzene/acetic acid (6:1, v/v)] was used for the second-dimension separation. After the second-dimension separation, the sheet was dried by warm air and exposed to HCl vapour, when all the spots turned from yellow to red, blue or purple.

Results

Qualitative amino acid-composition analysis

The two-dimensional separations of 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl and 4-*NN*-

dimethylaminoazobenzene-4'-thiohydantoin derivatives of amino acids have been reported (Chang & Creaser, 1976; Chang *et al.*, 1976). As indicated in Fig. 1 all the amino acids present, except leucine and isoleucine, could be identified immediately from their 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin derivatives (H_1 , H_2 , H_3); this ambiguity could be resolved by using the 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl derivatives (S_2). On the other hand, all the overlaps in the 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl amino acid derivatives, such as cysteic acid, threonine, methionine sulphone and hydroxyproline, and arginine, α -histidine, α -lysine and ϵ -lysine, could be totally separated as 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin derivatives. We confirmed the existence of arginine directly (Fig. 1, S_1), because there was no appearance of bis-histidine or bis-lysine hence there was no possibility of the presence of mono-histidine or mono-lysine derivatives. This was reconfirmed with derivative H_1 . In derivatives S_2 and H_2 all the amino acids in the insulin A chains appeared as both 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl and 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin derivatives. The only unexpected result was the appearance of the spot at the threonine position; the same results were obtained in repeated experiments for unknown reasons. However, the absence of threonine in insulin A chain was proved with derivative H_2 . The 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl cysteic acid derivative, which could not be unequivocally identified from derivative S_2 as it runs very close to 4-*NN*-dimethylaminoazobenzene-4'-sulphonic acid, was confirmed in derivative H_2 .

In derivatives S_3 and H_3 , the amino acid composition of glucagon was also satisfactorily identified. From both 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl and 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin derivatives of amino acids, we deduced the presence of arginine, histidine, and lysine as well as the presence of leucine, but no isoleucine.

The recovery of tryptophan after 22 h hydrolysis was found to be high in both the hexapeptide and glucagon.

Amino acid-sequence analysis

In our experiments, by this new method, we have been able to sequence the hexapeptide and the first 17 residues of insulin A chain and glucagon without encountering major problems. Fig. 2 shows the results of the sequence determination of the hexapeptide. The blue by-products (B, C, D and E) on the leucine- and tryptophan-containing sheets were too faint to be seen. It was found from these six sheets that there were no incomplete degradations during the phenyl isothiocyanate procedure, and in every sheet we only found one amino acid, i.e. except the

Fig. 1. Schematic representations of qualitative amino acid-composition analysis by treating the acid hydrolysates of hexapeptide (S_1), insulin A chain (S_2), and glucagon (S_3) with 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl chloride and hexapeptide (H_1), insulin A chain (H_2), and glucagon (H_3) with 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate

The solvents 1, 2, 3 and 4 employed were as described in the text. The by-products (dotted circles indicated in hydrolysates H_1 , H_2 and H_3) were all blue or bluish-purple in colour, except by-product (F), which was red. The six major by-products, which always appeared, were denoted by A, B, C, D, E and F. The hydrolysed 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl chloride (converted into 4-*NN*-dimethylaminoazobenzene-4'-sulphonic acid) in hydrolysates S_1 , S_2 and S_3 , and the unchanged 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate are indicated by the hatched areas.

Fig. 2. Results of the sequence determination of hexapeptide by the method described

The expected 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin of amino acids were obtained. Besides the by-products (A, B, C, D, E and F) no other contaminations appeared. No benzene extraction (see the text) was performed in analysing the hexapeptide. Solvents 1 and 2 used for developing the sheets were as described in the text.

red by-product (F), there was only one red spot, which appeared on each sheet. No seventh residue of the hexapeptide was found. One should note the positions of phenylalanine and methionine relative to the blue by-product (E). This major by-product was located at a most critical position and offers a quick discrimination between almost all the 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin of the amino acids by judging their relative position to by-product (E). Blue by-product (C) is also a very valuable marker, which plays a role in distinguishing phenylalanine, methionine, valine, leucine and isoleucine. However, this by-product always gives a weaker spot than by-product (E). In the *N*-terminal residue determinations of the hexapeptide degradation, we did not perform the benzene extraction after the coupling step. This extraction procedure was only applied on the *N*-terminal residue determinations of insulin A chain and glucagon, since there was no problem of the extraction losses of 4-*NN*-dimethylaminoazobenzene-4'-thiocarbonyl peptide derivatives. After this extraction, most of the unchanged 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate and part of blue and red by-products (A, B, C, D, E and F) were removed. However, by-product (E) still appeared in an amount just enough to be used as a marker. By-products (C, D and F) virtually disappeared (see Fig. 3).

The first 17 residues of insulin A chain and glucagon were investigated. Of the 34 *N*-terminal residues being

analysed, 13 residues were found to be the sole *N*-terminal residue on the sheet; the other residues were found to be contaminated either by preceding or succeeding amino acids. Lysine (glucagon residue 12) was undetectable in the repeated experiment for reasons as yet unknown, as reported by Schroeder (1972). However, in this step, minor amounts of tyrosine, serine and aspartic acid appeared instead of the expected 4-*NN*-dimethylaminoazobenzene-4'-(ϵ -phenylthiocarbonyl)thiohydantoin of L-lysine, which was supposed to be a red spot located near the 4-*NN*-dimethylaminoazobenzene-4'-(4-*NN*-dimethylaminoazobenzene-4'-thiocarbonyl)thiohydantoin of L-lysine (bis-Lys, Fig. 1, H₃). Results are shown in Table 1, and Fig. 3 gives the schematic representations of the indicated residues observed on the polyamide sheets. We found that although some expected *N*-terminals were contaminated by their neighbouring amino acids, they were always present as the major spots (Table 1). Even when some contaminants appeared as strong as the expected amino acids, the new residues can still be judged from their orders of appearance. For example, the detection of the ninth amino acid of insulin A chain results in serine (major) and alanine (minor). The detection of the tenth amino acid of insulin A chain was found to be valine and serine (major), and alanine (trace); it could thus be deduced that the tenth amino acid of insulin A chain is valine. Some amino acids appeared as much

stronger spots than others, such as glycine, phenylalanine, aspartate, tyrosine and leucine, probably owing to their higher recoveries. Glycine is a frequently detectable contaminant, despite the fact that it does not occur as the neighbouring amino acid of expected *N*-terminal residues (see Fig. 3, insulin residue 5 and insulin residue 17).

Leucine and isoleucine could be distinguished as their 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin derivatives by their slightly different R_F value in a one-dimensional t.l.c. separated by solvent 1 (Chang, *et al.*, 1976) or by the previous qualitative amino acid composition analysis.

Discussion

Thin-layer chromatography

In our experiments, polyamide sheets have proved to be an excellent medium for identifying 4-*NN*-dimethylaminoazobenzene-4'-sulphonyl and 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin derivatives of amino acids as these compounds give less

diffuse spots on polyamide as compared with silica. The whiter background of polyamide is also an advantage. The diameter of the separated 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin of the amino acids were always found to be around 1.5 mm when the diameter of the origin spot was limited to 1.0 mm. The limitation of the origin spot is a key factor in the procedure.

It is recommended to use a 1 μ l micropipette with a fine, smooth tip to apply the samples on the origin of a 2.5 cm $\times$ 2.5 cm sheet and confine the diameter to 1 mm. The nature of the solvent used to extract 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin of amino acids for t.l.c. is important. Most non-polar solvents such as acetone or benzene are more likely to produce a diffused origin spot on the polyamide sheet. We found that ethanol is a very suitable solvent due to its fast evaporation, and it produces compact spots.

The amount of sample applied to the sheet should not be too great. For a 2.5 cm $\times$ 2.5 cm sheet 20–50 pmol is enough for both 4-*NN*-dimethylamino-

Table 1. The t.l.c. results of the first 17 amino acid residues of insulin A chain and glucagon determined according to the method described

Among the 34 residues 13 appeared on the polyamide sheet as the sole *N*-terminal residue without any contamination except those major by-products indicated in Figs. 1, 2, or 3.

Insulin A chain			Glucagon		
	Expected	Found	Expected	Found	
1	Gly	Gly (only)	His	His (only)	
2	Ile	Ile (only)	Ser	Ser (only)	
3	Val	Val (only)	Gln	Gln and Gly (major)	
4	Glu	Glu (only)	Gly	Gly (only)	
5	Gln	Gln (major), Glu (minor)	Thr	Thr and Gly (major), Phe (trace)	
6	Cya	Cya (major), Glu and Gln (minor)	Phe	Phe (only)	
7	Cya	Cya (only)	Thr	Thr (major), Phe (minor)	
8	Ala	Ala (only)	Ser	Ser (major), Asp and Gly (minor)	
9	Ser	Ser (major), Ala (minor)	Asp	Asp (only)	
10	Val	Val and Ser (major), Ala (trace)	Tyr	Tyr (major), Asp (minor)	
11	Cya	Cya, Val and Ser (major)	Ser	Ser and Tyr (major), Asp (trace)	
12	Ser	Ser (only)	Lys		
13	Leu	Leu (major), Ser (minor)	Tyr	Tyr (only)	
14	Tyr	Tyr (major), Leu (trace)	Leu	Leu and Tyr (major)	
15	Gln	Gln and Tyr (major), Leu (minor)	Asp	Asp (major), Tyr and Leu (minor)	
16	Leu	Leu, Tyr and Gln (major), Gly (trace)	Ser	Ser and Asp (major), Gly, Leu (trace)	
17	Glu	Glu, Leu and Gly (major), Tyr, Ser (minor), Gln (trace)	Arg	Arg (major), Ser and Asp (trace)	

Fig. 3. Schematic chromatograms of 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin of amino acids on polyamide sheets obtained at the indicated steps of sequencing insulin A chain and glucagon

The blackened areas were the expected 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin of the amino acids. The size of all the areas approximately represents their amounts. The by-products (A, B, C, D and E) (blue or bluish-purple), and (F) (red) were also obtained. The unchanged 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate is indicated by the hatched areas.

azobenzene-4'-sulphonyl and 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin derivatives of amino acids. Also, running the first dimension in solvent 1, or solvent 3 is preferable, otherwise tailing will always be observed.

The production of some blue (or bluish-purple) by-products is a very interesting aspect. As shown in Figs. 1, 2 and 3 by-products (A, B, C, D and E) (blue or bluish-purple), and (F) (red) were the spots that always appear. In our previous report (Chang *et al.*, 1976) we adopted the buffer suggested by Sjöquist (1960) and found it generated very heavy by-products (B, C, D, E and F), which always interfere with the t.l.c. separation and especially the methionine identification. In this work, we used the modified triethylamine/acetic acid buffer, which generates by-products (A, B, C, D, E and F) only in amounts small enough to act as markers without interfering with the identification of methionine (see Fig. 2). Under these circumstances, it was found that, in most cases, the application of a standard on the opposite side of polyamide sheet is unnecessary, owing to the easily recognised position of every 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin of an amino acid resulting from the satisfactory separation and the use of the blue spots as the reference points.

The colour change between the 4-*NN*-dimethylaminoazobenzene-4'-thiocarbamoyl amino acid derivative (blue) and 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin amino acid derivative (red) is believed to result from the interaction of thiocarbamoyl and thiohydantoin groups in the resonance structures of 4-*NN*-dimethylaminoazobenzene ring after being exposed to the acid. This character provided quick discrimination against some contaminants caused by amines (blue). This has been demonstrated by reacting various aromatic and aliphatic amines with 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate according to the *N*-terminal residue determination procedure.

In the dansyl method, the major products of lysine and histidine as *N*-terminal residues are α -Dns-Lys or α -Dns-(ϵ -Ptc)-Lys and α -Dns-His. The mono-substituted lysine or histidine derivatives are not always separated from Dns-Arg on the two-dimensional separation, and this normally needs some extra solvents to resolve it (Hartley, 1970). However, this problem was absent from the 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate method. The separations between aspartic acid and glutamic acid, and between serine and threonine were also clearer than those of the dansyl method when using only two solvents.

Qualitative amino acid-composition analysis

We suggest a rapid manual method for the analysis of small peptides, i.e. the combined use of 4-*NN*-

dimethylaminoazobenzene-4'-sulphonyl chloride and 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate, which can offer fairly accurate qualitative amino acid composition with sub-nmol quantities of peptides. This manual method enables one to analyse five to ten peptides at the same time, so it would be very suitable for enzymically digested fragments. In this method, the amount of starting peptide was 1–2 nmol, but the actual amounts applied on the t.l.c. sheets were only 20–30 pmol. If care is taken, one can start with less than 0.1 nmol of peptide.

Amino acid-sequence determination

The automatic sequencer and the dansyl-Edman method are probably the most popular techniques among the conventional sequencing methods. Some researchers prefer the automatic sequencer, partly because of unavoidable drawbacks in the dansyl-Edman method, especially the extra work of assigning amide positions and the non-detection of tryptophan. However, the high cost of the commercial sequencer has restricted it from being widely used, as with the solid-phase sequencer (Laursen, 1971) even though this is considerably cheaper. The identification of the released *N*-terminal residue (after heating in the acid) either by gas chromatography (Pisano *et al.*, 1972) or t.l.c. (Summer *et al.*, 1973; Kuble, 1974) actually makes the sequencer work only as a semi-automatic instrument. This new manual method described here is simple, cheap, sensitive and accurate. In addition to the advantages mentioned above, it is also able to overcome a drawback of the dansyl-Edman method, i.e. the destruction of tryptophan, glutamine and asparagine, because there is no acid hydrolysis step.

Several points should be observed during the operation.

(1) The pH of the coupling mixture (with 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate) is critically important. Some *N*-terminal residues can be missed if the pH is not checked and adjusted before adding the isothiocyanate solution. The samples removed (for *N*-terminal residue determination), after being dried in a vacuum desiccator, are redissolved in triethylamine/acetic acid buffer, and about 0.2 μ l is removed for checking the pH on pH paper. The paper normally shows a pH around 8–9, despite its pH being 9.65 in the pH meter, because of its volatile nature. If the apparent pH, as indicated by the paper, is below 7.5, redrying in the vacuum desiccator should be performed.

(2) Every drying should be very thorough.

(3) Care must be taken not to remove any precipitates or particles at the water/ethyl acetate or water/benzene interfaces during the extraction.

(4) The cleaved 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin of amino acids should be applied to t.l.c. sheets as soon as possible after redissolving in

ethanol, since in a few cases, poor results were obtained when kept overnight in solution.

The use of 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate as the sole reagent in the sequence determination was studied. Owing to the limited solubility of 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate in acetone, which prevents the arbitrary increase of the molar ratio of 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate:peptides, it has been found that there exist some difficulties in achieving complete degradations of *N*-terminal amino acids to date. One possible remedy is to prolong the coupling or cleavage time, but this would make it a time-consuming method and would allow one to determine only one residue along the peptide in 1 day. Also the non-volatile nature of 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate and the accumulated by-products were troublesome problems.

We have also explored unifying the buffer used in both degradation and *N*-terminal residue determination steps. However, all the trials suggest that: (1) Aq. pyridine buffer is more suitable for phenyl isothiocyanate degradation, because of the higher recoveries found in the extended degradations; (2) triethylamine/acetic acid buffer is more suitable for *N*-terminal identifications, because the pyridine/water buffer produces some valueless by-products instead of the blue markers produced when employing triethylamine/acetic acid buffer.

The starting amount of the hexapeptide was 50nmol, and of insulin A chain and glucagon 200nmol. We estimate that only 25–30nmol of hexapeptide was totally removed for six *N*-terminal residue determinations and that only 100–130nmol of insulin A chain and glucagon was removed (see Materials and Method section) for the first 17 amino acid determinations. Further, although 5–7nmol (based on the assumption that all the recoveries from phenyl isothiocyanate degradation are 100%) of shortened peptide was removed for the next *N*-terminal residue determination, only approx. 0.03–0.05nmol was finally applied on the sheet for t.l.c. identification.

That means, for every *N*-terminal residue determination, we have the possibility of 25–100 repeat determinations.

It is feasible to reduce the whole scale down to one-tenth of the experiments described above, i.e. to determine 17 residues by using only approx. 12nmol of peptide. Even under these circumstances, there is still enough material for 2–10 replicates.

This new *N*-terminal reagent has been successfully exploited on tryptic digests of horse heart cytochrome *c* and of lipoproteins (with approx. mol.wt. 10000).

It should be noted that residue 17 of insulin A chain (glutamic acid) and glucagon (arginine) still comes out very distinctly, although there was substantial contamination by their preceding amino acids. It is not possible to completely predict the amount of starting material needed, or the number of cycles one can go through, as this will depend both on the nature of the starting material and the skill of the operator. The purity of chemicals, cleanness of glassware, patience, care as well as experience are all important factors. According to the method described, we have been able to achieve 12 residue determinations per day by starting with four peptides at the same time.

References

- Chang, J. Y. & Creaser, E. H. (1976) *J. Chromatogr.* **116**, 215–217
- Chang, J. Y., Creaser, E. H. & Bentley, K. W. (1976) *Biochem. J.* **153**, 20–21
- Edman, P. & Begg, E. (1967) *Eur. J. Biochem.* **1**, 80–91
- Gray, W. R. (1972) *Methods Enzymol.* **25**, 333–344
- Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822
- Konigsberg, W. (1972) *Methods Enzymol.* **25**, 326–332
- Kuble, K. D. (1974) *Anal. Biochem.* **59**, 564–573
- Laursen, R. A. (1971) *Eur. J. Biochem.* **20**, 89–102
- Lin, J. K. & Chang, J. Y. (1975) *Anal. Chem.* **47**, 1634–1638
- Pisano, J. J., Bronzert, T. J. & Brewer, H. B., Jr. (1972) *Anal. Biochem.* **45**, 43–59
- Schroeder, W. A. (1972) *Methods Enzymol.* **25**, 299–313
- Sjöquist, J. (1960) *Biochim. Biophys. Acta* **41**, 20–30
- Summer, M. R., Smythers, G. W. & Oroszlan, S. (1973) *Anal. Biochem.* **53**, 624–628