ANALYTICAL BIOCHEMISTRY 65, 346–354 (1975)

# A General Method for the Determination of the Carboxyl-Terminal Sequence of Proteins

RONALD G. DUGGLEBY[1] AND HARVEY KAPLAN

*Department of Biochemistry, University of Ottawa, Ottawa, Ontario K1N 6N5*

Received October 10, 1974; accepted November 15, 1974

A method is described for the selective isolation of peptides derived from the carboxyl-terminus of proteins. The method is of general applicability since it is independent of the primary structure of the protein in this region. The principle behind the isolation procedure is as follows: when an enzymic digestion is performed on a protein in which all free carboxyl groups have been modified by coupling with ethanolamine, the carboxyl-terminal peptide will be unique in that it does not have a carboxyl group. This difference is utilized to selectively purify carboxyl-terminal peptides. These peptides have the same electrophoretic mobility at pH 4.4 and at pH 2.1, whereas all other peptides will migrate more rapidly toward the cathode at the lower pH. Using a diagonal electrophoretic procedure based on this property, carboxyl-terminal peptides may be isolated, and their sequence determined by standard procedures.

The sequence of the carboxyl-terminal region of proteins, unlike the amino-terminal, is not readily determined. Several procedures, such as hydrazinolysis (1), reduction (2), treatment with carboxypeptidases (3), and sequential degradation (4) may often give limited information, but these methods are frequently difficult experimentally, and dependent on a favourable sequence of amino acids in the carboxyl-terminal region for success. Currently, the method of choice is to isolate the carboxyl-terminal peptide or fragment, which is identified by its amino acid composition (e.g. by the absence of homoserine in a peptide derived by cyanogen bromide cleavage of the protein). Procedures (5–9) have been suggested for the specific isolation of the carboxyl-terminal peptide, but these have not been widely applied, presumably because an unfavorable sequence of amino acids renders these methods ineffective. Clearly, a simple method for determining the carboxyl-terminal sequence of any protein, irrespective of its primary structure, would be extremely useful. In this communication, we report such a method which is based on the selective isolation of the carboxyl-terminal peptide.

[1] Present address: Department of Biochemistry, John Curtin School of Medical Research, Australian National University, Canberra, A.C.T., Australia.

## EXPERIMENTAL

### *Strategy*

The general principle behind the isolation procedure is similar to that described by Furka *et al.* (6). When a protein is digested with a proteolytic enzyme, each bond cleaved generates a new amino and a new carboxyl group. If the digestion is performed on a protein in which all carboxyl groups have been modified, then the carboxyl-terminal peptide will be unique in that it will be the only peptide which does not have a carboxyl group. This difference may be exploited as a means to selectively purify carboxyl-terminal peptides. Since the $pK_a$ of the $\alpha$-carboxyl group of a peptide is approximately 3.5 (10), only the carboxyl-terminal peptide should have unchanged electrophoretic mobility at pH 4.4 and pH 2.1; all other peptides will migrate more rapidly toward the cathode at the lower pH.

The procedure we have adopted consists of the following steps: (1) all carboxyl groups of the protein are coupled to ethanolamine using a carbodiimide; (2) the modified protein is digested with a proteolytic enzyme; (3) the digest, together with internal markers of taurine and an amino acid amide are subjected to electrophoresis at pH 4.4; (4) a guide strip is removed, turned through 90°, subjected to electrophoresis at pH 2.1, and stained with ninhydrin; (5) any peptides lying on a line connecting the two internal markers are located, cut from the pH 4.4 electrophoretogram, and subjected to electrophoresis at pH 2.1; and (6) bands corresponding to the previously located peptides are cut out and further purified. The presence of ethanolamine coupled to the $\alpha$-carboxyl group of the terminal residue permits positive identification of the carboxyl-terminal peptide.

### *Materials*

Porcine pepsin (two times recrystallized) and bovine $\alpha$-chymotrypsin (three times recrystallized) were purchased from Worthington, chicken egg white lysozyme (three times recryatallized) was obtained from Sigma, thermolysin (three times recrystallized) was obtained from Calbiochem, and porcine elastase was obtained from Whatman Biochemicals. 1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride was purchased from Pierce, alanine amide hydrochloride and $\alpha$-aminobutyric amide hydrochloride were purchased from the Fox Chemical Co., and other amino acid amides were obtained from Sigma. Other chemicals were high purity preparations from commercial sources.

The solvent systems employed were: pH 2.1 buffer, formic acid:acetic acid:water (1:4:45, by volume); pH 4.4 buffer, pyridine:acetic acid:water (6:10:1200, by volume); pH 6.5 buffer, acetic acid:pyri-

dine:water (3:100:900, by volume); and BAWP, *n*-butanol:acetic acid:water:pyridine (15:3:12:10, by volume).

*Methods*

Lysozyme was reduced and carboxymethylated by the method of Sela, White, and Anfinsen (11). Hemoglobin was prepared from human erythrocytes by hemolysis, followed by ammonium sulfate fractionation. Heme was removed by treatment with acid–acetone.

The procedure for isolating carboxyl-terminal peptides from globin is typical: 1.6 $\mu$moles of each subunit (50 mg protein) was dissolved in 3 ml of water containing a few drops of HCl, and 2.6 g urea added. After stirring for 30 min, 0.3 g KCl and 0.2 ml of ethanolamine were added. The pH was immediately adjusted to pH 4.75 with HCl, then 125 mg of 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide added, maintaining pH 4.75. The solution was stirred at room temperature for 4–6 hr, then the reaction terminated by the addition of 1 ml of 4 M sodium acetate buffer, pH 4.75. The solution was dialyzed against several changes of water. Thermolysin digestion was performed in 1% (w/v) ammonium bicarbonate at 37°C for 8 hr, using an enzyme:protein ratio of 1:35 by weight. Digestion was terminated by lyophilization.

The peptide mixture was dissolved in pH 4.4 buffer and applied to a 33-cm band on Whatman 3 MM chromatography paper, together with markers of taurine and alanine amide (50 nmoles/cm). The peptide mixture was subjected to electrophoresis at pH 4.4 for 60 min at 60 V/cm. After drying, a guide strip was removed, stitched on to a full sheet of Whatman 3MM, and subjected to electrophoresis at pH 2.1 for 40 min at 60 V/cm. The electrophoretogram was stained with cadmium–ninhydrin to locate the on-diagonal peptides. The region of the pH 4.4 electrophoretogram corresponding to these on-diagonal peptides was cut out, and subjected to electrophoresis at pH 2.1. The slowest moving peptides, corresponding to the on-diagonal peptides, were cut out and purified by electrophoresis at pH 6.5, and finally by electrophoresis at pH 2.1.

Peptides were hydrolyzed in 6 N HCl at 110°C for 24 hr in sealed, evacuated tubes, and amino acid analyses performed on a Technicon TSM-1 autoanalyzer. Sequence determination was performed by a subtractive Edman procedure. Pepsin digestion was performed in 5% (v/v) formic acid at 37°C for 20 hrs, using an enzyme:protein ratio of 1:50, by weight. Elastase or $\alpha$-chymotrypsin digestion was performed in 1% (w/v) ammonium bicarbonate at 37°C for 24 hr, using an enzyme:protein ratio of 1:25, by weight.

## RESULTS AND DISCUSSION

The success of the specific isolation procedure described herein depends on establishing conditions for two-dimensional electrophoresis

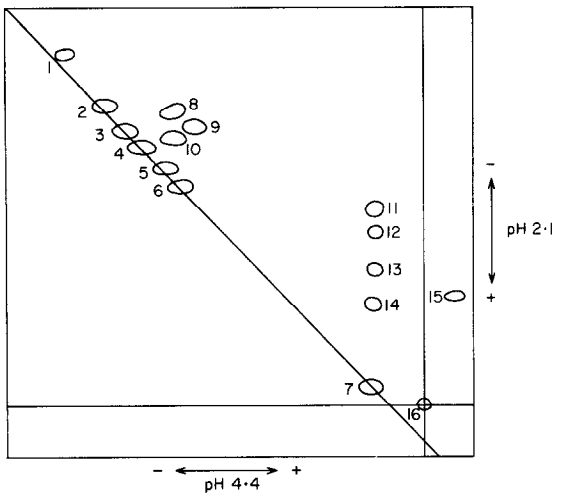


FIG. 1. Two-dimensional electrophoretogram obtained with a series of standard compounds. The indicated areas correspond to the positions of: (1) glycinamide; (2) alanine amide; (3) α-aminobutyric amide; (4) valine amide; (5) leucine amide; (6) phenylalanine amide; (7) taurine; (8) lysine; (9) arginine; (10) histidine; (11) glycine; (12) alanine; (13) leucine; (14) tyrosine; (15) glutamic acid; (16) origin.

such that peptides with a blocked carboxyl-terminus have the same electrophoretic mobility at two pH values, whereas peptides with a free carboxyl group will exhibit altered mobility. This is achieved by working over a pH range in which the carboxyl group, but no other functional group found in proteins, ionizes. If this criterion is met, it might be expected that carboxyl-terminal peptides would be located on a line at 45° passing through the origin. However, due to differences in electrophoresis conditions, both the absolute mobility and the extent of electroosmosis will vary, and the diagonal line will be displaced from its expected position. For this reason, some means of defining the diagonal line is necessary in order to ensure a reliable identification of carboxyl-terminal peptides. We have chosen to define the diagonal line by the use of ninhydrin-positive marker compounds such as taurine and amono acid amides. Figure 1 shows a pH 4.4/pH 2.1 two-dimensional electrophoretogram obtained with taurine, a series of amino acid amines, and a selection of amino acids. With the exception of glycinamide, all of the amino acid amides lie on a straight line. Glycinamide has a slightly higher relative mobility at pH 2.1 than at pH 4.4, but is nevertheless close to the diagonal line. Taurine, which is neutral at both pH values, defines the lower end of the diagonal line and is used to correct for differing amounts of electroosmosis. In practice, taurine and any amino acid amide can be used as markers, but glycinamide or alanine amide are usually the most suitable as they have high mobilities and are unlikely to overlap, and hence mask, carboxyl-terminal peptides. Experience has

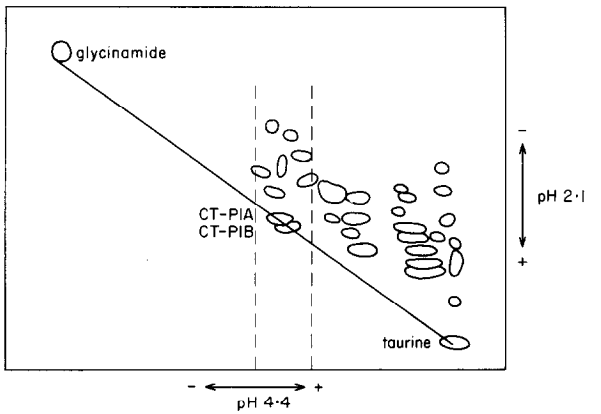


FIG. 2. Two-dimentional electrophoretogram obtained from a pepsin digest of α-chymotrypsin. The region indicated by broken lines was cut from the pH 4.4 electrophoretogram, and subjected to electrophoresis at pH 2.1. The slowest moving peptides were isolated.

shown that without the use of diagonal markers it is not always obvious which peptides lie on the diagonal line. If a particular digestion procedure releases the amino acids lysine, arginine, or more particularly histidine, these can easily be mistaken for carboxyl-terminal peptides (Fig. 1). Diagonal markers minimise the risk of such an error.

We have applied this method to three test proteins. α-Chymotrypsin has three polypeptide chains, and thus has three carboxyl-termini. The pH 4.4/pH 2.1 electrophoretogram obtained from a peptic digest of 1 μmole of α-chymotrypsin is shown in Fig. 2. A single ninhydrin-positive area is located on the line connecting glycinamide and taurine. This region, which clearly contains two peptides, was purified as described in Methods, except that chromatography in BAWP replaced the pH 6.5 electrophoresis step The amino acid compositions of the two peptides isolated, CT-P1A and CT-P1B, is shown in Table 1. CT-P1A is clearly derived from the carboxyl-terminus of the C chain of α-chymotrypsin, and CT-P1B is derived from the carboxyl-terminus of the A chain. No peptide from the carboxyl-terminus of the B chain was found.

The second test protein, human hemoglobulin, was chosen for two reasons. First, the α chain ends in the amino acids tyrosine and arginine, the two amino acids which terminate the B chain of α-chymotrypsin. This provides an opportunity to determine whether the failure to isolate the carboxyl-terminal peptide from the B chain is due to some peculiar property of one of these amino acids. The second reason in that there is a histidine residue at the end of the β chain. This is particularly important as the $pK_a$ of the imidazole size-chain of histidine is closest to that of the carboxyl group. Other methods (6,7) for isolating carboxyl-terminal peptides are ineffective if there is a histidine residue close to the

TABLE 1

AMINO ACID ANALYSES OF CARBOXYL-TERMINAL PEPTIDES FROM α-CHYMOTRYPSIN, HUMAN HEMOGLOBIN, AND CARBOXYMETHYL-LYSOZYME

| Amino acid | Peptide[a] | | | | | | |
|---|---|---|---|---|---|---|---|
| | CT-P1A | CT-P1B | HB-T1 | HB-T2 | L-C1A | L-C1B | L-E1 |
| Cmc | | | | | 0.68(1) | | 0.71(1) |
| Asp | 1.05(1) | | | | 0.05(0) | | |
| Ser | | 0.92(1) | 0.05(0) | | | 0.08(0) | |
| Gly | | 1.07(1) | 0.17(0) | 0.09(0) | 1.01(1) | | |
| Ala | 1.95(2) | | | | | | |
| Leu | | 1.01(1) | 0.10(0) | | 0.94(1) | 0.91(1) | 1.00(1) |
| Tyr | | | 0.91(1) | 0.85(1) | | | |
| His | | | 1.05(1) | | | | |
| Eth[b] | 1.04(1) | 1.09(1) | 1.03(1) | 1.09(1) | 2.23(2) | 1.16(1) | 2.24(2) |
| Arg | | | 0.05(0) | 1.05(1) | 1.05(1) | | 1.05(1) |
| Yield | | | | | | | |
| nmoles | 98 | 119 | 59 | 57 | 37 | 28 | 72 |
| % | 9.8 | 11.9 | 3.7 | 3.6 | 2.5 | 1.9 | 4.8 |

[a] Results are reported as the number of residues found on analysis. All amino acids found at a level of 0.05 residues or greater are reported. Expected values are shown in parentheses.

[b] eth, ethanolamine.

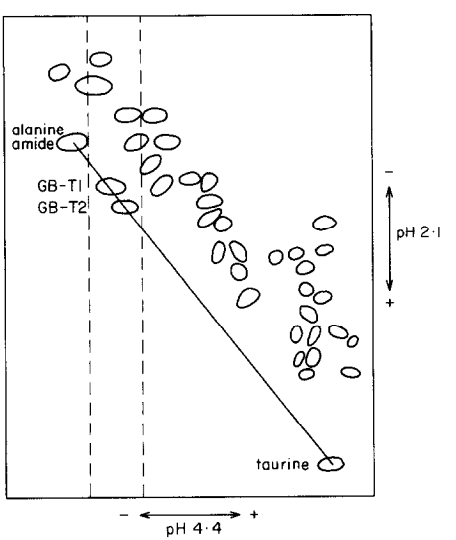


FIG. 3. Two-dimentional electrophoretogram obtained from a thermolysin digest of human hemoglobin. The region indicated by broken lines was cut from pH 4.4 electrophoretogram, and subjected to electrophoresis at pH 2.1. The slowest moving peptides were isolated.

TABLE 2
AMINO ACID SEQUENCE OF THE PEPTIDES ISOLATED FROM α-CHYMOTRYPSIN, HUMAN HEMOGLOBIN, AND CARBOXYMETHYL-LYSOZYME

| Peptide | Amino-terminal | Sequence |
|---|---|---|
| CT-P1A | Ala | Ala-ala-asn-eth[a] |
| CT-P1B | Ser | Ser-gly-leu-eth |
| HB-T1 | Tyr | Tyr-his-eth |
| HB-T2 | Tyr | Tyr-arg-eth |
| L-C1A | Gly | Gly-cmc(eth)-arg-leu-eth |
| L-C1B | Leu | Leu-eth |
| L-E1 | Cmc(eth)[b] | Cmc(eth)-arg-leu-eth |

[a] eth, ethanolamine.

[b] cmc(eth), *S*-[2-acetamido-*N*-(2′-hydroxyethyl)]cysteine (i.e., the compound formed when the side-chain carboxyl of carboxymethyl-cysteine is coupled to ethanolamine).

carboxyl-terminus. The pH 4.4/pH 2.1 electrophoretogram obtained from human hemoglobin is shown in Fig. 3. Two peptides, GB-T1 and GB-T2, were isolated as described in Methods. Amino acid analysis (Table 1) showed that these are derived from the β and α chains, respectively.

Carboxymethyl-lysozyme was chosen as the third test protein as there is an acidic residue close to the carboxyl-terminus (residue 127 is carboxymethylcysteine). Digestion of 1.5 μmoles of protein with α-chymotrypsin yielded peptides L-C1A and L-C1B, and digestion of 1.5 μmoles of protein with elastase gave peptide L-E1. Amino acid analysis (Table 1) showed that each was derived from the carboxyl-terminus of carboxymethyl-lysozyme.

In all cases investigated, peptides located on the diagonal line have proved to be carboxyl-terminal peptides, as judged from their amino acid compositions. Their identity as carboxyl-terminal peptides is confirmed by the presence of ethanolamine, and by Edman degradation (Table 2). Only in the case of the B chain of α-chymotrypsin have we failed to isolate a carboxyl-terminal peptide, but this is undoubtedly due to the particular enzyme used to digest the protein. The peptides isolated contain amino acids representing all the major groups of side-chain type, and thus the method may be applied with confidence to unknown proteins. We are currently using the method to study sequence homologies between bacterial ribosomal proteins (R. G. Duggleby, H. Kaplan, and L. P. Visentin, manuscript in preparation). In one case, three peptides of varying length were isolated, which permitted a partial sequence determination on the basis of amino acid composition alone. In a similar way, the carboxyl-terminal sequence of lysosyme could be deduced to be Gly-(Cys,Arg)-Leu from the data in Table 1.

Ethanolamine was chosen as the coupling nucleophile for several reasons. It is very soluble in water, and thus can be used at a high concentration to ensure good coupling yields. The modified protein, and peptides derived from it, have desirable solubility properties rendering digestion and subsequent manipulations fairly easy. Finally, it is not found in normal proteins, and is readily separated from amino acids on the amino acid analyzer which makes quantitation simple. A disadvantage is the possibility that proteolytic enzymes could remove it from the carboxyl-terminus of the protein, which may account for our failure to find a peptide corresponding to the B chain of $\alpha$-chymotrypsin. Perhaps a secondary amine might be used as the coupling nucleophile where this factor is a problem, but in view of the large number of proteolytic enzymes available, a different digestion procedure will usually produce satisfactory results.

In order for a method to be of general applicability to unknown systems, the success of its application should not be dependent on the system meeting any particular conditions. The method for the isolation of carboxyl-terminal peptides described in this communication meets this criterion of generality since, after derivatization of all the free carboxyl groups with ethanolamine, all peptides derived from the carboxyl-terminus of the protein will lie on the diagonal line, irrespective of their amino acid sequence. Other advantageous features of the method described herein are: (1) unlike some other methods for isolating carboxyl-terminal peptides (7–9), the particular enzyme used for digestion is of no consequence, except insofar as it generates peptides of a useful length; (2) the sequence of the peptide may be determined using standard methodology and, moreover, the presence of amides is determined directly; (3) the presence of ethanolamine at the carboxyl-terminus of the peptides isolated provides a built-in check that the peptide is derived from the carboxyl-terminus of the protein; and (4) while high coupling yields are desirable in order to obtain maximum yields of the peptides, this is not an essential feature of the method, since any fraction of the carboxyl-terminal peptide which has not been coupled with ethanolamine will simply move off the diagonal. A final point worth noting is that a number of proteins have been found to be amidated at the carboxyl-terminus (12). Since such proteins naturally contain a blocking group of just the type that we introduce artificially, we see no reason why the method would not work equally well with such proteins.

## ACKNOWLEDGMENTS

This work was supported by the Medical Research Council of Canada. RGD is an MRC Postdoctorate Fellow.

## REFERENCES

1. Akabori, S., Ohno, K., and Narita, K. (1952) *Bull. Chem. Soc. Jap.* **25,** 214–218.
2. Fromageot, C., Jutisz, M., Meyer, D., and Penasse, L. (1950) *Biochim. Biophys. Acta* **6,** 283–289.
3. Ambler, R. P. (1972) *Methods Enzymol.* **25,** 143–154.
4. Stark, G. R. (1968) *Biochemistry* **7,** 1796–1807.
5. Carlton, B. C., and Yanofsky, C. (1963) *J. Biol. Chem.* **238,** 636–639.
6. Furka, A., Sebestyén, F., and Karácsonyi, T. (1970) *Fed. Eur. Biochem. Soc.* **6,** 34–36.
7. Zabin, I., and Fowler, A. V. (1972) *J. Biol. Chem.* **247,** 5432–5435.
8. Hargrave, P. A., and Wold, F. (1973) *Int. J. Peptide Protein Res.* **5,** 85–89.
9. Naughton, M. A., and Hagopian, H. (1962) *Anal. Biochem.* **3,** 276–284.
10. Tanford, C. (1962) *Advan. Protein Chem.* **17,** 69–165.
11. Sela, M., White, F. H., and Anfinsen, C. B. (1959) *Biochim. Biophys. Acta* **31,** 417–426.
12. Dayhoff, M. O. (1972) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington, DC.