CARBOXYL GROUP MODIFICATION AND AMIDE ASSIGNMENTS IN AUTOMATED SEQUENCING OF PROTEINS

D. Gibson

Département de Biochimie, Université de Sherbrooke, Sherbrooke, Qué. Canada.

P.J. Anderson

Department of Biochemistry, University of Ottawa, Ottawa 2, Canada.

Received August 30, 1972

Summary

A procedure is described for locating the amidated and non-amidated residues in automated sequence analysis of proteins. The method requires no additional identification procedure beyond hydrolysis of the thiazolinone derivatives followed by amino acid analysis. The free carboxyl groups of ribonuclease A were coupled with serine methyl ester or glycine ethyl ester in the presence of a water soluble carbodiimide. The derivatized proteins were then subjected to automated Edman degradation in the protein sequenator. At positions where the residue in the native protein was aspartic or glutamic acid, the coupled amino acid was recovered in addition to aspartic acid or glutamic acid after hydrolysis of the thiazolinone derivatives.

A number of hydrolysis methods have been used to regenerate amino acids from their phenylthiohydantoin (PTH-) (1-3) or 2-anilino-5-thiazolinone (TH-) (4) derivatives. Smithies et. al. (4) showed that characteristic amino acids can be quantitatively recovered from virtually all of the derivatives normally encountered in the protein sequenator (5) if one uses two complementary hydrolysis procedures.

While the hydrolysis approach is attractive in that it permits separation and quantitation of the products by amino acid analysis, there are a number of outstanding difficulties. The major problems with the method are that the thiazolinone derivatives of serine and cysteine cannot be distinguished and the assignment of amides to aspartic and glutamic acid residues is difficult at best. The former problem can be overcome if one employs ^{14}C-carboxymethylated or performic acid oxidized proteins. Theoretically, quantitation of ammonia in the acid hydrolysis should permit distinction between the PTH derivatives of aspartic acid and asparagine or glutamic acid and glutamine. In practive however, precise quantitation of ammonia is difficult since trace amounts of buffer may be carried over from the coupling reaction in the sequenator and give rise to variation in the recovery of ammonia after hydrolysis.

An alternative method for locating the amide groups in automated sequencing of proteins is presented here. This method is based on the use of the car-

boxyl group modification procedure developed by Hoare and Koshland (6, 7). These authors demonstrated that the total number of free carboxyl groups in a protein (and consequently amides by difference) could be accurately determined by measuring the amount of glycine methyl ester or amino methane sulfonate coupled with the protein in the presence of a water-soluble carbodiimide. In the present study, the suitability of such derivatized proteins for use in the protein sequenator has been investigated. It was reasoned that if the side chains of aspartic and glutamic acid were coupled with another amino acid, the presence of the second amino acid upon hydrolysis of the thiazolinone derivative would clearly indicate that the residue had a free carboxylic acid side chain. The absence of the "marker" amino acid would indicate that the side chain was in the amide form. The results presented here with both the serine methyl ester and glycine ethyl ester derivatives of ribonuclease A indicate that the assignment of amides with this method can be made with a high degree of confidence.

Materials and Methods

Ribonuclease A (type XII-A) and glycine ethyl ester-HCl were obtained from Sigma Chemical Company. 1-ethyl-3-dimethylaminopropylcarbodiimide(EDI) lot number 12271-6 was purchased from Pierce Chemical Company. Serine methyl ester-HCl was obtained from Mann Research Labs and taurine, A grade was obtained from Calbiochem. Reagents for the sequenator and solvent 3, (chlorobutane) were obtained from Pierce Chemical Company. Solvents 1 and 2 (benzene and ethyl acetate) were obtained from Burdick and Jackson laboratories, Muskegon Mich. Dithiothreitol, (DTT) (Calbiochem) was added at a concentration of 1 mM to all of the solvents as an antioxidant.

Coupling reaction (6)

Typically 30 mg of ribonuclease A was dissolved in 3.0 ml of freshly prepared 1.0 M glycine ethyl ester-HCl or serine methyl ester-HCl in 6 M guanidine-HCl. The protein solution was then titrated to pH 4.75 if necessary by the addition of 0.1 N NaOH. 60 mg of the carbodiimide (EDI) was dissolved in about 0.2 ml of 6 M guanidine HCl and added to the protein solution. The pH of the reaction was maintained at 4.75 by the occasional addition of 0.1 N HCl, but little change in pH occurred after the first 15 minutes of reaction. The mixture was incubated with stirring for $2\frac{1}{2}$ to 3 hours at room temperature. The solution was then dialized against 3 changes of 1 L of distilled water and lyophilized.

Sequenator Procedure

5 mg of protein was dissolved in 0.2 ml of 0.1 M formic acid containing 1 mM ammonium formate and 1 mg of DTT and was pre-cycled as described by

Smithies et. al. (4). The sequenator employed in this work was an Illitron model 9001 and it was programmed essentially as described by Edman and Begg (5) with the exception that the second chlorobutane extraction was eliminated.

Hydrolysis of thiazolinone derivatives

The hydrolysis of samples from the sequenator was carried out in an evacuated dessicator in an autoclave as described by Smithies et. al. (4). The HCl hydrolysis was done in exactly the same manner as was the HI hydrolysis, except that constant boiling HCl was used in place of 57% HI. Amino acid analysis of all hydrolysed samples was performed on a Technicon TSM-1 amino acid analyser modified for a single column procedure.

Results

Amino acid analysis of the glycine ethyl ester and serine methyl ester derivatives of ribonuclease A showed the incorporation of 10.7 and 7 residues, respectively. The expected number of residues coupled, based on the amino acid sequence of ribonuclease (8) was 11, indicating that coupling had not gone to completion with serine methyl ester. Since the coupling reaction was carried out in the presence of 6M guanidine however, it is unlikely that specific carboxyl groups would remain completely uncoupled. Several attempts were made to couple the amino acid taurine with ribonuclease A at pH 4.75, but the degree of coupling was unsatisfactory. An improved coupling was obtained when the reaction was maintained at pH 7 but the resulting protein product gave low yields in the sequenator suggesting that it might have been partly blocked at the amino terminus.

Figure 1 gives the relevant hydrolysis data for the first 17 steps in the degradation of the serine methyl ester derivative of ribonuclease. The use of serine methyl ester adds absolutely no ambiguity to the results, since the amino acid serine is never recovered as such after hydrolysis of the thiazolinone derivative of serine. The data illustrated is from the HCl hydrolysis, in which the serine methyl ester coupled on the side-chain carboxyl groups is recovered as serine. With the HI hydrolysis, it was found that the coupled serine residues were quantitatively converted to alanine.

Figure 2 shows analogous data for the degradation of the glycine ethyl ester derivative of ribonuclease. The recovery of glycine at positions 2, 9 and 14 clearly indicates that these groups had free carboxyl side chains, in contrast with the groups at positions 11, 24, 27 and 28 which failed to couple. With the glycine ethyl ester derivative the HI hydrolysis procedure has been used and found to give identical results to those shown here. The conclusion based on these results and with the serine methyl ester derivative (fig. 1) are completely in accord with the established sequence of ribonuclease

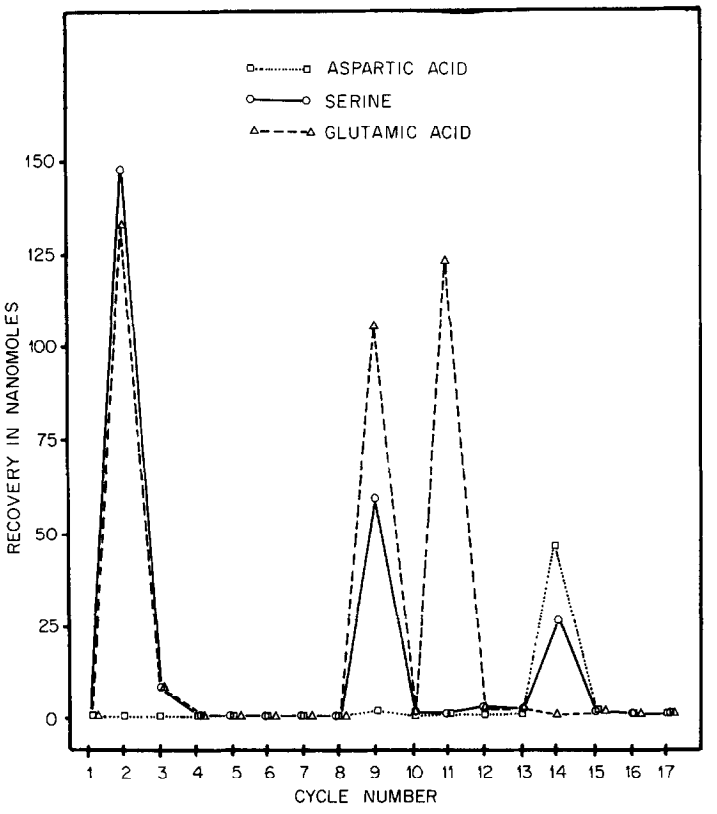

Figure 1. Recovery of the aspartic acid □·····□, serine ○——○ and glutamic acid Δ- - -Δ after HCl hydrolysis of fractions from the first 17 steps in the automated degradation of the serine methyl ester derivative of ribonuclease A. The primary data was corrected for background and an out-of-step factor of 0.01 as described by Smithies et. al. (4). The presence of serine at positions 2, 9 and 14 and its absence at position 11, permits the assignments Glu-2, Glu-9, Gln-11 and Asp-14.

(8), namely that there are amide groups at positions 11, 24, 27 and 28. It was noted that the yields of glycine after hydrolysis at positions 2, 9 and 14 (fig. 2) were higher than the yields of glutamic and aspartic acids. It is likely that this reflects the difference in the rates of hydrolysis of the side chain peptide linkage and the more resistant thiohydantoin ring structure. In the case of the serine methyl ester derivative (fig. 1), the amount of serine recovered at positions 2, 9 and 14 corresponded to 1.1, 0.56 and 0.56 moles per mole of glutamic or aspartic acid. In a subsequent experiment using the same protein preparation, but using ethylacetate (+ 0.1% acetic acid) for improved extraction of the coupled residues, the molar ratios at these three positions were 1.03, 0.69 and 0.81. While slow loss of the coupled seryl

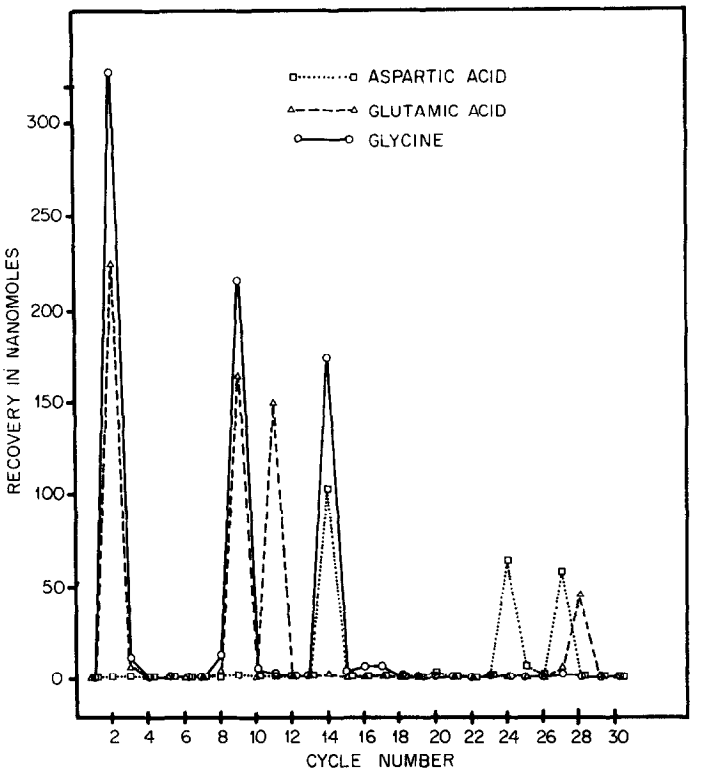

Figure 2. Recovery of aspartic acid □·····□, glutamic acid Δ– – –Δ and glycine O——O after HCl hydrolysis of fractions from the first 30 steps in the automated degradation of the glycine ethyl ester derivative of ribonuclease A. The primary data was corrected for background and an out-of-step factor of 0.01, as described by Smithies et. al. (4). The presence of glycine at positions 2, 9 and 14, but its absence at positions 11, 24, 27 and 28 permits the amide assignments Glu-2, Glu-9, Gln-11, Asp-14, Asn-24, Asn-27 and Gln-28.

groups during the degradation cannot be entirely ruled out, the latter result would indicate that the loss must be less than 1.5% per cycle (19% over 14 cycles).

In figure 3, which presents all of the hydrolysis data from the experiment illustrated in figure 2, it can be seen that no glycine residues were recovered with any amino acid other than aspartic and glutamic acids. Over the first 30 positions of ribonuclease A, all amino acids with the exception of proline, glycine, tryptophan, valine, isoleucine and leucine were encoutered. Of these it is unlikely that any would react or be destroyed under the conditions of the coupling reaction. The alanine residues underlined indicate a low yield of alanine from the hydrolysis of TH-serine (positions 15, 16, 18, 21, 22, 23) or of TH-cysteine (position 26). The low yields of threonine at positions 3 and 17 represent a low recovery of α-aminobutyric acid at these

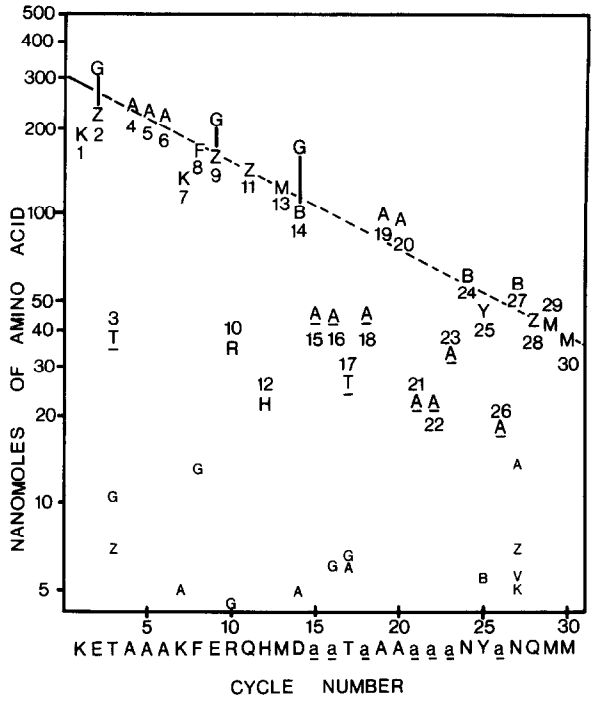

Figure 3. The recovery of amino acids after HCl hydrolysis of fractions from the first 30 steps in the automated degradation of the glycine ethyl ester derivative of ribonuclease A. The primary data was corrected for background and an out-of-step factor of 0.01 (4). The one letter amino acid code used is that of Dayhoff (9) and the positions of the letters indicate the recovery in nanomoles of that amino acid. The code letters used are A = ala, B = asx, C = cys, D = asp, E = glu, F = phe, G = gly, H = his, K = lys, M = met, N = asn, Q = gln, R = arg, T = thr, V = val, Y = tyr, Z = glx. Smaller letters were used in the lower part of the figure since these residues are not considered significant. The amino acid sequence deduced completely from this data is given below the figure. The letter a, used at positions 15, 16, 18, 21, 22, 23 and 26 indicates a low yield of alanine. It could represent either serine or cysteine.

positions. Although these residues are normally completely destroyed on HCl hydrolysis (2), in the present work approximately 0.6 mg of DTT was present in each sample after evaporation of the chlorobutane, so that partial reduction occurs. The low recovery of arginine at position 10 and histidine at position 12 is most certainly due to poor extraction of these residues in the chlorobutane. When solvent 2 (ethylacetate plus 0.1% acetic acid) was used in place of chlorobutane, the yields of both of these residues were quantitative.

Discussion

Carboxyl group modification as described by Hoare and Koshland (6) has been shown here to be a valuable preliminary step in automated sequencing of

proteins. The glycine ethyl ester derivative appears to be completely stable under the conditions employed in the sequenator and the presence of one mole of glycine in addition to aspartic or glutamic acid after hydrolysis clearly identifies those residues which had free carboxylic acid side chains. Glutamine and asparagine residues were completely unreactive.

There is an inherent advantage in using a derivatizing amino acid which is not normally recovered in hydrolysates of thiazolinone or PTH-derivatives of protein-amino acids. One such residue is serine, since serine is never recovered as such on hydrolysis of PTH-serine or TH-serine. The present results show that serine methyl ester can be effectively used in place of glycine ethyl ester to couple protein carboxyl groups. The serine residue coupled with the side chain in this way is recovered after HCl hydrolysis as serine, so that no ambiguity is introduced.

In the studies presented here with ribonuclease, there appeared to be negligible side effects of the prior treatment of the protein with the carbodiimide reagent. It is likely that the method will be of general usefulness but particular proteins may present problems. The method has been tested on two other proteins, one of which had an N-terminal serine, the other an N-terminal glutamic acid or glutamine (Glx.). In the second case preliminary results showed that essentially complete blocking of the protein occurred. This would indicate that treatment of proteins with N-terminal glutamic acid with the carbodiimide reagent may cause cyclization of the glutamic acid to yield an N-terminal pyroglutamic acid thus blocking the Edman degradation. A solution to this problem might be found in altering the conditions of the coupling reaction, perhaps by using a different pH or a higher concentration of coupling amino acid.

Acknowledgements

The authors would like to thank Mrs. Francine Girard for her help with the hydrolyses and amino acid analyses. This work was supported by the Medical Research Council of Canada.

REFERENCES

(1) Africa, B. and F.H. Carpenter, Biochem. Biophys. Res. Commun. 24, 113 (1966).
(2) Van Orden H.O. and F.H. Carpenter, Biochem. Biophys. Res. Commun. 14, 399 (1964).
(3) Inglis, A.S., Nicholls, P.Q. and Roxburgh, C.M., Aust. J. Biol. Sci. 24, 1247 (1971).
(4) Smithies, O., Gibson, D., Fanning, E.M. Goodfliesh, R.M., Gilman, J.G. and Ballantyne, D.L., Biochem. 10, 4912 (1971).
(5) Edman, P. and Begg, G., Eur. J. Biochem. 1, 80 (1967).
(6) Hoare, D.G. and Koshland, D.E., Jr., J. Biol. Chem. 242, 2447 (1967).
(7) Lin, T-Y., and Koshland, D.E., Jr., J. Biol. Chem. 244, 505 (1969).
(8) Smyth, D.G., Stein, W.H. and Moore, S., J. Biol. Chem. 238, 227 (1963).
(9) Dayhoff, M.O., "Atlas of Protein Sequence and Structure" vol. 4, National Biomedical Research Foundation, Silver Spring Md., 1969 p. D-2.