Tsao, T. C. (1953), *Biochim. Biophys. Acta 11*, 368.

Weber, G. (1952), *Biochem. J. 51*, 155.

Weber, K., and Osborn, M. (1969), *J. Biol. Chem. 244*, 4406.

Weltman, J., Frackelton, A., Jr., Szaro, R. P., and Dowben, R. (1972), *Biophys. Soc. Abstr. 12*, 277a.

Weltman, J. K., Szaro, R. P., Frackelton, R., Jr., Bunting, J. R., Cathou, G. E., and Dowben, R. M. (1973), *J. Biol. Chem. 248*, 3173.

Woods, E. F. (1966), *J. Mol. Biol. 16*, 581.

Wu, C., and Stryer, L. (1972), *Proc. Nat. Acad. Sci. U. S. 69*, 1104.

Amino Acid Sequence Determination with Radioactive Proteins†

David J. McKean,* Erwin H. Peters, James I. Waldby, and O. Smithies

ABSTRACT: A procedure is described for quantitative analysis of amino acids based only on their radioisotope content. The effluent from a conventional amino acid analyzer is collected so that each amino acid is placed in a discrete fraction for subsequent scintillation counting. This procedure is applied to the analysis of the hydrolytic products of thiazolinones formed by the automatic degradation of proteins in the Edman–Begg sequenator. Sequence data derived solely from the radioisotope content of the protein under investigation are then obtained. We tested the procedure for 30 degradations with a uniformly ^{14}C-labeled peptide of 47 residues in length obtained from Brome Mosaic virus coat protein. The radioisotope sequence agreed completely with that derived from conventional chemical analysis. The radioisotope sequencing procedure can also be used with ^{3}H labels. The procedure is insensitive to the presence of nonradioactive proteins introduced accidentally or deliberately during the purification of the radioactive protein.

Determining the primary structure of proteins using the automated Edman–Begg sequenator is currently restricted to proteins available in milligram quantities. These sequence determinations are limited by the sensitivity of the methods used to identify and quantitate the resulting phenylthiazolinones or the amino acids obtained from them by hydrolysis.

In contrast, a method for sequencing radiolabeled proteins would not effectively be limited by the availability of any fixed amount of protein but rather by its specific activity. With such a method, amino acid sequence information could be obtained on proteins available in only trace quantities provided that sufficient radiolabel could be introduced. An important additional advantage inherent in procedures using radiolabeled protein is that the protein under study need be purified only to radiochemical homogeneity; the presence of adventitious nonradioactive contaminants is unimportant.

This paper presents a method for identification and quantitation of radioactive amino acids and illustrates how these methods can be automated and applied to sequencing radiolabeled proteins.

Materials and Methods

Proteins. The [^{14}C]Brome Mosaic virus peptide ([^{14}C]BMV-D)[1] was prepared from a cyanogen bromide digest of [^{14}C]Brome Mosaic plant virus coat protein (BMV) which was uniformly labeled by growing a virus-infected plant in a $^{14}CO_2$ atmosphere (Shih and Kaesberg, 1973). This 47 residue peptide was a generous gift of Mr. Joe Moosic (Biophysics Laboratory, University of Wisconsin). The human λ Bence-Jones protein (Eddy) which has a free α-amino group was prepared by diethylaminoethyl-cellulose chromatography of the patient's urinary proteins. The preparation of [^{3}H]*S*-carboxymethyl T_1 ribonuclease has been previously described (McKean and Smith, 1974).

Sequence Procedures. Uniformly radiolabeled [^{14}C]BMV-D (0.076 μCi in 32 nmol of protein, corresponding to 150,000 cpm ^{14}C) was combined with 1 mg of unlabeled carrier protein (λ chain), 0.53 μCi of [^{3}H]T_1 ribonuclease, 1 mg of Braunitzer's III reagent (Pierce Chemical Co., Rockford, Ill.), and 1 mg of dithioerythritol (Pierce Chemical Co.). The [^{3}H]T_1 ribonuclease, with radiolabeled cysteines at positions 2, 6, 10, and 103, was added to the sequenator sample to permit monitoring of the sequenator chemistry efficiency; its radiolabel was not involved in determining the sequence of [^{14}C]BMV-D. The combined protein sample was sequenced with an Edman–Begg sequenator (Illitron Division, Illinois Tool Works, Chicago, Ill.) using procedures and chemicals previously described (Edman and Begg, 1967; Smithies *et al.*, 1971).

A portion (5%) of the thiazolinone sample from each degradative step was counted for ^{3}H and ^{14}C to determine the degradative efficiency of the [^{3}H]T_1 ribonuclease and the total amount of ^{14}C in each sample, respectively. A known amount of [^{14}C]-β-alanine (7×10^{-3} μCi) was then added to each of the thiazolinone samples as in internal standard to control for sample loss and dilution. The thiazolinone samples were dried and hydrolyzed with HI (Smithies *et al.*, 1971). Unlabeled amino acid standards (Pierce Chemical Co., Rockford, Ill.), 20 nmol each, supplemented with 20 nmol each of α-aminobutyric acid, β-alanine, *allo*-isoleucine, and ornithine, were added to all hydrolyzed samples. These amino acids were added to act as carriers for the radioactive amino acids and to produce enough ninhydrin color to be detected by the analytical system of the amino acid analyzer. The hydrolysates were then subjected to amino acid analysis on a Durrum Model D500 amino acid ana-

† Paper No. 1736 from the Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706. *Received March 20, 1974.* Supported by the National Institutes of Health Grants GM 20069 and GM 00398.

[1] Abbreviations used are: BMV-D, Brome Mosaic virus peptide; BMV, Brome Mosaic plant virus coat protein.

lyzer (Durrum Instrument Corp., Palo Alto, Calif.) which had been modified for effluent collection.

When sequenator samples are hydrolyzed with HI and quantitated on the amino acid analyzer, positive identification of the thiazolinones of Thr (as α-aminobutyric acid), Pro, Gly, Val, Ile (as allo-ile and Ile), Leu, Tyr, Phe, His, Lys, and Arg can be made. Several residues are indistinguishable from each other: Asn is converted to Asp; Gln is converted to Glu; Ser and Cys are converted to Ala; and Trp is converted to Gly and Ala. Met is completely destroyed. NaOH–dithionite hydrolysis can be used to distinguish Ala from Ser and Cys and to quantitate Met and Trp (Smithies *et al.*, 1971). BMV-D contains no Cys or Trp so only Asn/Asp, Gln/Glu, and Ala/Ser are indistinguishable in the hydrolysis data presented here.

Amino Acid Analyzer Modifications. A Durrum D500 amino acid analyzer, controlled by a PDP-8M computer, was used for the radiolabel amino acid analysis. It was modified to control a fraction collector, causing each amino acid to be collected in a single scintillation vial. The modification for effluent collection was achieved by replacing the output tube of the 440-nm photometer with a 15-cm 316 stainless steel tube (Durrum Instrument Corp.) of 0.007 in. i.d. This tube terminates in port 1 of a 316 stainless steel 3-port mixing block (Durrum Instrument Corp.). The mixing block, which is used to control the final exit of the effluent, is mounted on the outside of the amino acid analyzer near the back pressure regulator. Port 2 of the mixing block is connected to the back pressure regulator by 0.020 in. i.d. 316 stainless steel tubing (Durrum Instrument Corp.). Port 3 of the mixing block is open during the radioisotope quantitation procedures and is connected to the fraction collector by a 30-cm piece of 0.007 in. i.d. stainless steel tubing and a 64-cm piece of 30 gauge Teflon tubing. Port 3 is plugged when the amino acid analyzer is not being used for effluent collection, so the effluent passes out of the system in the usual manner through port 2 to the pressure regulator. The total dead volume of tubing between the flow cell and the fraction collector is about 0.121 ml, which was calculated from the 29-sec delay time required to collect an amino acid into one scintillation vial at a machine flow rate of 0.265 ml/min.

A fraction collector (Model B100, Gilson Medical Electronics, Middleton, Wis.) was modified to change fractions when it received a 110-V ac pulse of approximately 100 msec duration generated by a small interface circuit (O. Hiller Co., Madison, Wis.). The interface circuit detects the "begin integration" and "end integration" signal generated by the Durrum's PDP-8 computer when peak integrations begin and end. A variable time delay with a range of 0–64 sec is incorporated into the interface to permit compensation for the timelag introduced by the additional tubing between the flow cell and the fraction collector. Thus integration signals trigger the delay circuit and after the appropriate delay the fraction collector is advanced one vial. When the fraction collector system is used, dead spaces between the flow cell and the fraction collector must be kept small. This requires use of the analyzer without the usual backpressure system. In order to operate the analyzer without this backpressure the reaction coil temperature is lowered to approximately 97°. The resulting reduction in sensitivity is of no consequence since the ninhydrin color is used only to operate the fraction collector and not to quantitate the amino acids. The amino acid analyzer is currently being operated with a 75-min period between samples using standard Durrum buffers with a pump stroke of 0.115 in.

We have modified the computer printout of the amino acid analyzer to record each time that the integration begins or ends, in addition to the standard listing of time, area, and quantity for each amino acid peak which is eluted from the column. With this additional information we can readily associate each amino acid peak on the chromatogram with its corresponding scintillation vial in the fraction collector.

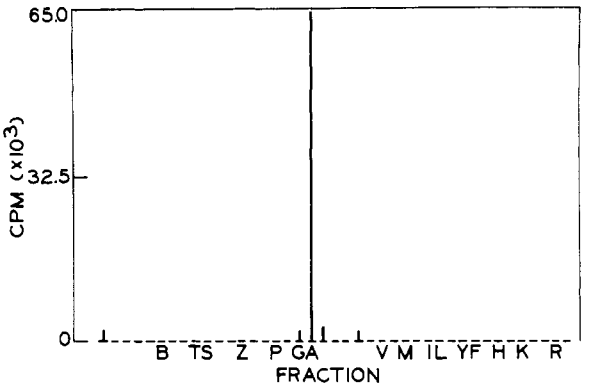

FIGURE 1: A bar graph of the scintillation counter data from an amino acid analysis of a [^{14}C]alanine standard (Amersham/Searle, Arlington Heights, Ill.) using the Durrum D500 amino acid analyzer modified for effluent collection. Each horizontal dash represents one scintillation vial; vertical lines represent counts above background. The graph shows that essentially all of the [^{14}C]alanine goes into one single scintillation vial. Amino acids are labeled with one letter code: B, Asx; T, Thr; S, Ser; Z, Glx; P, Pro; G, Gly; A, Ala; α,α-aminobutyric acid; V, Val; M, Met; I, Ile; L. Leu; Y, Tyr: F, Phe; β, βAla; H, His; O, ornithine; K, Lys; R, Arg.

Scintillation Counting. All scintillation counting was done by adding 10 ml of Bray's solution (Bray, 1960) to each 25-ml plastic scintillation sample vial. Vials were counted in a Packard liquid Tri-Carb scintillation counter (Packard Instrument Corp., Downers Grove, Ill.) using the automatic external standard to correct for ninhydrin and amino acid analyzer buffer quenching.

Data Processing. A Wang Model 720C programable calculator (Wang Laboratories, Natick, Mass.) was used to correct the scintillation counter data for quenching, sample losses (relative to the β-alanine internal standard), and specific activity of each amino acid. Data for each analysis are presented in the form of a bar graph by the Wang (see Figures 1, 2B, and 3). The corrected scintillation data are subsequently handled as conventional amino acid analyzer sequenator data (Smithies *et al.*, 1971) (see Figure 4).

Specific Activity Determination of BMV-D. A 0.34-μCi sample of [^{14}C]BMV coat protein, from which the [^{14}C]BMV-D peptide had been isolated, was hydrolyzed *in vacuo* with 50 μl of constant-boiling HCl at 110° for 24 hr. Amino acid analysis was carried out on the Durrum D500 amino acid analyzer modified for effluent collection. Specific activities of the amino acids were calculated using the amino acid composition data of Stubbs and Kaesberg (1964).

Results

By utilizing the modifications to the Durrum amino acid analyzer presented here, a complete amino acid analysis can be carried out routinely with each amino acid collected into one separate vial. Figure 1 shows an example of a chromatogram with [^{14}C]alanine as the only radioactive amino acid. The glycine-alanine region of the chromatogram is crowded; yet virtually all the label is recovered in a single vial. Effluent, if any, between peaks is collected into one or more vials depending on the time between peaks and the stability of the base line. Thus a complete analysis is collected into 40–50 scintillation vials. An example of a complete radiolabel amino acid analysis of a ^{14}C algal hydrolysate is presented in Figure 2. In some parts of

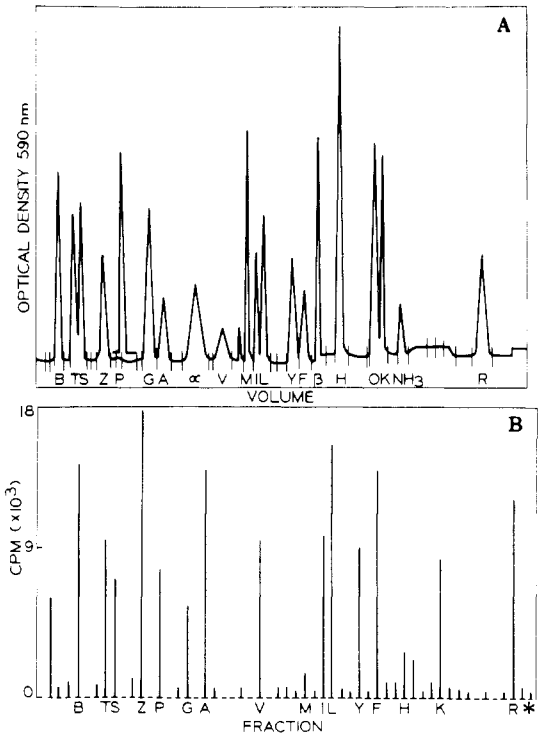

FIGURE 2: (A) A conventional ninhydrin produced chromatogram of an amino acid standard analysis on the Durrum D500 amino acid analyzer. The amino acid standard used here is described in the text. Peaks are labeled with the amino acid one letter code. (B) A bar graph of the scintillation counter data from an amino acid analysis of a ^{14}C-uniformly labeled algal hydrolysate (Amersham/Searle) using the Durrum D500 amino acid analyzer modified for effluent collection under conditions described in the text. The ninhydrin chromatogram controls the fraction collection system so that each amino acid is collected into a separate vial. Amino acids are labeled with the one letter code in both Figure 2A and B (α = α-aminobutyric acid; β = β-alanine). The vial containing NaOH column flush is denoted by *.

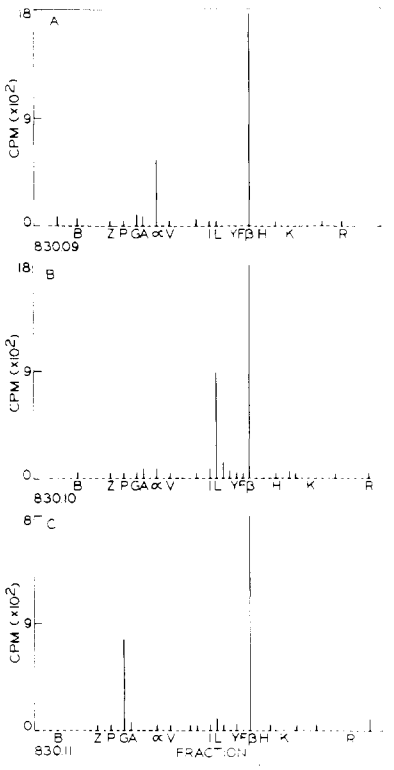

FIGURE 3: Bar graphs of the scintillation counter data (corrected for specific activity, quenching, and losses) from radioisotope amino acid analyses for three representative degradative cycles: 9, 10, and 11. The large peak occurring in all three graphs is the internal standard [^{14}C]β-alanine. Peaks are labeled with the amino acid one letter code.

the chromatogram no fractions occur between consecutive amino acids, *e.g.* Gly and Ala. Each amino acid nonetheless is virtually unmixed with its neighbor.

The amount of aqueous buffer in the scintillation vials must be minimized in order to avoid considerable quenching of the radioactive samples. To do this: (1) the amount of aqueous amino acid analyzer buffer is minimized by the small flow-through volume of the amino acid analyzer;[2] (2) the amount of ninhydrin color is kept relatively constant from analysis to analysis by the addition of a small constant amount of amino acid standard to each analyzer sample; (3) only a small amount of carrier protein is used in the sequenator sample, so that its amino acids in the degradative products add relatively little ninhydrin color to the constant amino acid background. The scintillation quenching which does occur is corrected in subsequent calculations by using the automatic external standard data from the scintillation counter.

The specific activity of the [^{14}C]BMV-D peptide used in our test experiment was very low (0.009 μCi in 32 nmol of protein). The degradative products from this amount of protein in the sequenator sample could be analyzed by conventional amino acid analysis. However, the sequenator degradative cycles from [^{14}C]BMV-D were identified and quantitated exclusively using the ^{14}C radiolabel. Our experience with other radiolabeled proteins of much higher specific activity (with protein quantities too small for conventional amino acid analysis) have shown results which are similar to the results presented here.

Each of the amino acid analyses of the 30 sequenator degradations showed one major radiolabeled amino acid at each cycle, as illustrated in Figure 3. There is also present a small amount of the amino acid from the previous cycle due to the incomplete coupling-cleavage which normally occurs in each cycle of the Edman degradation.

Figure 4A shows the sequence data of [^{14}C]BMV-D for 30 degradations as determined by using only the radiolabel techniques to identify and quantitate the sequenator degradative products. In Figure 4B the sequence data are shown from an experiment in which 50 nmol of nonradioactive BMV-D peptide was quantitated by conventional amino acid analysis. The peptides were not sequenced further than step 30 because experience with the same peptide in previous sequenator runs using comparable sequenator conditions and conventional amino acid analysis had shown that the sequence could not be identified much after the lysine at position 29. (The peptide is rapidly washed out of the sequenator reaction cup after the naphthalenedisulfonic acid derivative of lysine at position 29 is removed by the Edman degradation.) Figure 4 shows that the sequence obtained with the radiolabel technique is identical with that obtained with conventional amino acid analysis.

Discussion

Methods for determining the primary structure of proteins which are available in only trace amounts will be useful for a variety of proteins (such as membrane proteins) which heretofore could not be sequenced. The needed technology can be approached either by increasing the sensitivity of the analytical systems used for identifying the nonradioactive thiazolinones

[2] The flow-through volume of the Durrum analyzer produces not more than 0.7 ml of aqueous volume in each amino acid containing vial. Most other amino acid analyzers have flow-through volumes 5 to 10 times greater than this. The resulting large volumes of aqueous solutions in each amino acid containing vial may result in severe quenching of the radiolabel.

or their derivatives (Zimmerman *et al.*, 1973; Udenfriend *et al.*, 1972) or by developing systems to separate and quantitate the degradative products from radioactive proteins.

We decided to develop radioactive sequencing methods for several reasons. Many proteins which are available in only trace amounts can be intrinsically radiolabeled during their synthesis *in vivo* or *in vitro*. Some of these proteins can be radiolabeled so as to take advantage of a biological purification; a particular protein may be synthesized at a specific time even though it is difficult to separate from many chronologically older or younger proteins. The radiolabel itself is also advantageous to monitor proteins during purification; other physiochemical properties such as optical density are not practical monitors when only small amounts of protein are available. Homologous or heterologous unlabeled protein (such as carrier protein or antibody), which may be added adventitiously or deliberately during purification to minimize purification losses, does not interfere with later sequence determination.

Hembree *et al.*(1973) have described a method which appears to be suitable for extrinsically radiolabeling trace amounts of protein to a very high specific activity by a tritium exchange reaction without changing the protein's primary structure. This method should make our radioactive sequencing procedure useful even for proteins that cannot be intrinsically labeled during their synthesis.

When the thiazolinones are either acid or base hydrolyzed, each amino acid derivative undergoes a certain amount of degradation (Smithies *et al.*, 1971). With conventional ninhydrin-monitored amino acid analysis, this degradation results in decreased peak areas for the particular residue involved. The degradation products themselves, however, are rarely ninhydrin positive and are usually not seen in the amino acid analyzer chromatogram. When radioactive amino acids are used, the degradative products should be eluted somewhere during the sample chromatography unless they are volatilized or are insoluble after hydrolysis. Although our experience is still limited, chromatograms of hydrolysates from sequenator runs of ^{14}C-labeled proteins do not usually contain spurious radioactive peaks. Both the initial flow through volume (unretarded material) and the NaOH column wash (strongly bound products) do, however, often contain radioactivity. These presumably represent [^{14}C]amino acid degradation products.

Amino acid analyses of HCl hydrolysates from ^{14}C proteins sometimes produce a peak of radioactivity at the beginning of the third analyzer buffer. This peak is also routinely seen with commercially available radiolabeled algal hydrolysates. It occurs in the area where histidine normally elutes and interferes somewhat with histidine quantitation. We did not see this peak in any of the 30 analyses from the sequenator run of [^{14}C]BMV-D.

Our experience to date has been mostly limited to ^{14}C-radiolabeled proteins, but preliminary experiments suggest that the method is also applicable to ^{3}H-labeled proteins. Our method for sequencing radiolabeled proteins is consequently a practical way to determine the primary structure of proteins which are available in only trace amounts.

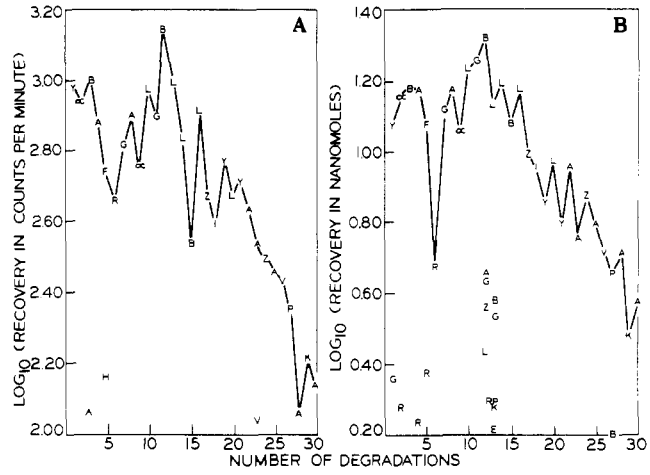

FIGURE 4: Corrected semilog plot of the amino acid recoveries after HI hydrolysis of the thiazolinones obtained in two separate sequenator runs of two different plant virus (BMV-D) preparations. The [^{14}C]BMV-D data (A), presented in counts/min, were obtained solely from radioisotope measurements. The unlabeled BMV-D data (B), presented in nanomoles, were obtained from regular chemical (ninhydrin color) measurements. The lines connect the amino acids which were recovered in the highest yield at each degradation. From the lines one can deduce the sequence of the BMV-D peptide. The sequence of BMV-D in (A), which was obtained by using only radiolabeling sequence techniques, is identical with that in (B), which was obtained by conventional amino acid analysis. The amino acids are presented by the one letter code. The sequence of BMV-D determined in both experiments was: Tyr-Thr-Asx-Ala or Ser-Phe-Arg-Gly-Ala or Ser-Thr-Leu-Gly-Asx-Leu-Leu-Asx-Leu-Glx-Ile-Tyr-Leu-Tyr-Ala or Ser-Ala or Ser-Glx-Ala or Ser-Val-Pro-Ala or Ser-Lys-Ala or Ser (see text for explanation of Ala or Ser).

Acknowledgments

We thank Charles Forge, Otto Hiller, and Joseph Swoboda for their assistance in modifying the Durrum analyzer, and the Durrum Instrument Corp. for supplying parts and advice.

References

Bray, G. A. (1960), *Anal. Biochem. 1,* 279.

Edman, P., and Begg, G. (1967), *Eur. J. Biochem. 1,* 80.

Hembree, W. C., Ehrenkaufer, R. E., Lieberman, S., and Wolf, A. P. (1973), *J. Biol. Chem. 248,* 5532.

McKean, D. J., and Smith, G. P. (1974), *Biochem. J. 139,* 779.

Shih, D., and Kaesberg, P. (1973), *Proc. Nat. Acad. Sci. U. S. 70,* 1799.

Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., and Ballantyne, D. L. (1971), *Biochemistry 10,* 4912.

Stubbs, J. D., and Kaesberg, P. (1964), *J. Mol. Biol. 8,* 314.

Udenfriend, S., Stein, S., Bohlen, P., Dairman, W., Leimgruber, W., and Weigele, M. (1972), *Science 178,* 871.

Zimmerman, C. L., Pisano, J. J., and Appella, E. (1973), *Biochem. Biophys. Res. Commun. 55,* 1220.