Stárka, L., Šulcová, J., Dahm, K., Döllefeld, E., and Breuer, H. (1966), *Biochim. Biophys. Acta 115*, 228.
Šulcová, J., and Stárka, L. (1968), *Steroids 12*, 113.
Tamaoki, B., Inano, H., and Nakano, H. (1969), *in* The Gonads, McKerns, K. W., Ed., New York, N. Y., Appleton-Century-Crofts, p 547.
Tamaoki, B., and Shikita, M. (1966), *in* Steroid Dynamics, Pincus, G., Nakao, T., and Tait, J. F., Ed., New York, N. Y., Academic, p 493.
Zaretskii, V. I., Wulfson, N. S., Zaikin, V. G., Kogan, L. M., Voishvillo, N. E., and Torgov, I. V. (1966), *Tetrahedron 22*, 1399.

Sequence Analysis of Complex Protein Mixtures by Isotope Dilution and Mass Spectrometry*

Thomas Fairwell, William T. Barnes,† Frank F. Richards,‡,§ and Robert E. Lovins§

ABSTRACT: A volatile Edman reagent, methyl isothiocyanate, is used to degrade sequentially polypeptide mixtures from the N-terminal end. Standard mixtures of ^{15}N-enriched amino acids or their methylthiohydantoin derivatives are added to the reaction. Increase in the $^{14}N:^{15}N$ isotope ratio is observed when ^{14}N amino acid methylthiohydantoins derived from the N-termini of the polypeptide chains dilute the standard mixture. The extracted methylthiohydantoins are transferred to the sample probe of a single-focusing mass spectrometer; the temperature of the probe is raised gradually and the methylthiohydantoin amino acids are volatilized sequentially into the ion beam, identified, and their $^{14}N:^{15}N$ isotope ratio established. By these means an artificial mixture of polypeptide chains has been shown to have four components; the quantity of each component present has been measured, the amino acid sequence of the N-terminal ten residues has been established for both of the two major sequences present, and the amino acids present at each locus have been determined for the first ten residues of the two minor sequences which were present in equimolecular proportions. This information is obtained simultaneously from a single operation which is less time consuming and laborious than the analysis of the first ten amino acid residues from a single polypeptide chain by conventional means. The rates of loss of reactive N-terminal groups have been studied when A and B chains of insulin were sequenced simultaneously. The loss rates were 2.5 and 0.9% per residue, respectively. Recovery rates of ^{15}N-labeled methylthiohydantoin amino acids using ^{15}N amino acids or their methylthiohydantoin derivatives established that either may be used for isotope dilution and that losses of the thiazolinone intermediate to compounds other than the methylthiohydantoin derivatives are not quantitatively important. Enolase was shown to have identical sequences for the first eight amino acids from the amino-terminal end in each of the two subunits by this method.

Many proteins occur in a microheterogeneous state. Such proteins form families in which the individual members are distinguished from each other by single or relatively limited amino acid residue sequence changes. These changes may reflect genetic polymorphism in protein structure (Dixon, 1966), the action of enzymes on the protein, or may be concerned directly with the function of the protein. Sometimes a single amino acid replacement may cause a major change in the physical characteristics of the protein, facilitating isolation of the variant (Zuckerkandl, 1968). It has been established that antibodies are collections of heterogeneous proteins, differing in amino acid residue sequence in limited regions of the molecule (Edelman and Gall, 1969). It has not so far proved possible to separate these closely related proteins into single species by conventional means although less heterogeneous populations have been produced (Miller *et al.*, 1967; Richards *et al.*, 1969; Brenneman and Singer, 1969) or isolated from more heterogeneous populations (Eisen and Siskind, 1964).

A method for the quantitative sequential degradation of a mixture of proteins (or peptides) from the N-terminal residue is described in this paper. It is an adaption of the Edman procedure using volatile reagents, coupled with isotope dilution and the use of the single-focusing mass spectrometer employed in its triple capacity as a differential vacuum distillation apparatus, an isotope ratio assay instrument (Biemann, 1962) and employed for the determination of residue structure by means of the molecular ion and fragmentation peak pattern of the mass spectrum. This quantitative sequential degradation method is designed to study microheterogeneity both in antibodies (Richards *et al.*,

* From the Department of Biochemistry, University of Georgia, Athens, Georgia 30601 (T. F. and R. E. L.), and the Departments of Internal Medicine and Microbiology, Yale University School of Medicine, New Haven, Connecticut 06510 (W. T. B. and F. F. R.). *Received January 7, 1970.* Supported by Grants RO1-AI-08614 from the Public Health Service, 9070-41-45538 from the National Science Foundation, and by the American Heart Association.

† Present address: Department of Biochemistry, Tulane University New Orleans, Louisiana.

‡ Established Investigator, American Heart Association.

§ To whom enquiries should be addressed.

1967, 1969) and in other proteins where variants, having subtle differences in amino acid sequence are difficult to separate (Petra and Neurath, 1969).

Experimental Section

Conditions for Sequential Degradation. Reagents. Methyl isothiocyanate (Eastman) was twice redistilled (mp 35–36°). Tetrahydrofuran (Eastman) was redistilled twice from $LiAlH_4$ suspension (bp 66°). Trifluoroacetic acid (Eastman) was twice redistilled. Mass spectroscopy indicated absence of impurities and gave the correct M^+ ion. All reagents were stored under anhydrous conditions. Bovine insulin and insulin A and B chains were obtained from Mann; 3× recrystallized hen's egg lysozyme from Pierce Chemicals, and Type XII A bovine pancreatic ribonuclease from Sigma. The purity of these proteins was checked by amino acid analysis.

Conditions. Formation of the Peptide–Thiocarbamate Adduct. The reduced and carboxymethylated mixture is (100 nM–10 μM) dissolved in 0.5 ml of 80% aqueous pyridine and adjusted to pH 8.5 (if necessary) with pyridine. For the first residue a tenfold molar excess of MeSNC[1] is used (based on the total reactive side-chain groups on the protein, as determined by amino acid analysis and on the total reactive groups of the standard ^{15}N mixture). A fivefold molar excess of MeSNC was used in subsequent steps. Reduction and alkylation of the proteins is performed as described by Hirs (1967). The reaction is carried out under nitrogen in the dark with rapid magnetic stirring for 2 hr at 60° for soluble proteins. Proteins in suspension were treated for 3 hr at 60°. The solvents and excess methyl isothiocyanate are removed at 55° for 25 min at 100-μ pressure.

Formation of Methylthiohydantoin Amino Acid Derivatives from the N Terminus of Peptides. The dried reduced and carboxymethylated proteins are treated with 0.5 ml of trifluoroacetic acid at 55° in the dark under nitrogen with stirring 60 min and then cooled to 20°. Trifluoroacetic acid is removed at 100-μ pressure with stirring.

The dried precipitate is extracted three times with 0.5 ml of H_4-furan or with a 1:1 mixture of H_4-furan and trifluoroacetic acid when arginine, lysine, asparatic acid, or asparagine are present. The extract is concentrated to 30 μl. A 10-μl aliquot was transferred to a capillary tube which was dried in a desiccator and transferred to the mass spectrometer.

Addition of ^{15}N Compounds. A mixture of ^{15}N amino acids is added to the protein mixture at the beginning of each cycle of degradation. The amino acids are 80–95% heavy atom excess in ^{15}N. The exact ratio of ^{14}N:^{15}N is determined in advance for the mixture by mass spectrometry. Greatest accuracy is obtained when the molar ratio of total ^{15}N to ^{14}N is approximately 1:1. However, reasonably good quantitation may be achieved with ratios of ^{15}N:^{14}N as low as 1:50. Accuracy of ratio measurement is also a function of the size of the M^+ ion peak.

The addition at each step of approximately equimolar quantities of ^{15}N amino acid mixtures is expensive. For this reason the mixture to be sequenced is divided into two portions. A qualitative analysis is performed on the first aliquot, followed by a quantitative analysis on the second portion. Only those amino acids occurring in the qualitative analysis are quantitated by addition of the corresponding ^{15}N amino acids.

Synthesis of ^{14}N and ^{15}N Amino Acid Methylthiohydantoin Derivatives. ^{14}N Methylthiohydantoin amino acid derivatives were synthesized according to Stepanov *et al.* (1965). The ^{15}N methylthiohydantoin amino acids were synthesized in 100–200-mg quantities by a micromodification of the above method. In a typical synthesis 1 μmole of L-phenylalanine (165.1 mg) was reacted with 3 μmoles (219.3 mg) of MeSCN in 5.0 ml of 60% aqueous pyridine under N_2 with stirring at 60° for 45 min (adding approximately 4.0 ml of 1 N Na_2CO_3 drop by drop to maintain pH 9.0). The reaction mixture at 20° was extracted three times with 5 ml of benzene, then adjusted to pH 2.0 with 6 N HCl, and refluxed under nitrogen for 30 min. The ^{15}N methylthiohydantoin amino acid derivatives crystallized directly from the mix and could be separated from NaCl crystals by washing in cold water; yield was 60–70%. The derivatives give sharp melting points identical with those of the ^{14}N derivatives (except serine) and the correct infrared and mass spectra.

Reactivity of Methyl Isothiocyanate. [^{14}C]Ala-[^{3}H]Leu-Phe-Gly was synthesized by Christine Attardo using the solid-state technique (Merrifield, 1964). It was modified using a continuous-mixing process (Richards *et al.*, 1967). Amino acid analysis of the product gave the ratios Ala 1.03, Leu 1.00, Phe 1.00, Gly 0.99. Sequential degradation gave the two amino-terminal residues in the correct order; specific radioactivities [^{3}H]Leu = 3.1×10^5 cpm/mmole, [^{14}C]Ala = 6.4×10^4 cpm/mmole.

Quantities (1.0 mg) of [^{14}C]Ala-[^{3}H]Leu-Phe-Gly which had not been cleaved from the supporting resin were reacted with MeSNC and phenyl isothiocyanate and cyclized under the conditions given earlier. After extraction of the thiohydantoin derivatives the resin–peptide was cleaved with 0.2 ml of 20% ethanolic KOH for 4 hr at 40° (Richards *et al.*, 1967); ethanol was removed *in vacuo*, the mixture was neutralized, and the supernatant was counted. After correction for losses (based on ^{3}H counts added) the residual ^{14}C counts were (a) phenyl isothiocyanate, 0.09%, (b) methyl isothiocyanate, 0.07% of initial radioactivity. This experiment indicates that the overall efficiency of the individual reactions leading to the release of amino acid methylthiohydantoin derivatives are approximately the same when methyl isothiocyanate or phenyl isothiocyanate are employed.

Isotope Ratio Assay. Instrumentation. A mass spectrometer giving a resolution of 1:500 or better is required for this method. A CEC Model 21-104, and a CEC Model 21-490 single-focusing mass spectrometer, equipped with a solid sample inlet, variable-temperature probe, variable-ionizing voltage, and an electrical detection system were employed.

Operating Conditions. Source temperature was 250°. Probe temperature was programed to rise at approximately 20°/min between 25 and 250°. During the temperature rise, tht total ion current was monitored and spectra taken at 30° (or 1.5-min) intervals whenever a rise in total ion current was indicated. At each step spectra at 80 eV and 10 eV were recorded.

Sample Size. For routine use, the CEC 21-490 will measure in the sample size range 1×10^{-9}–5×10^{-8} mole as a lower

[1] Abbreviations used are: MeSNC, methyl isothiocyanate; H_4-furan, tetrahydrofuran; MTH, methylthiohydantoin.

TABLE I: Mass Spectroscopic Identification of Amino Acid Methylthiohydantoin Derivatives.[a]

Amino Acid Methylthiohydantoin Derivative	m/e	Relative Abundance 80 eV	Relative Abundance 10 eV	Temp Range (°C)
Valine	172*	65	100	70–80
	130	100	50	
Leucine	186*	70	100	80–90
	143	30	30	
	130	100	72	
Isoleucine	186*	47	44	80–90
	130	100	100	
Glycine	130*	100	100	90–100
	102	32	46	
Threonine	174*	4	2	90–100
	156	76	20	
	130	100	100	
Serine	160*	0?	0?	90–100
	142	100	100	
Alanine	144*	100	100	90–110
Proline	170*	100	100	100–110
Methionine	204*	100	100	110–125
	156	16	16	
	143	43	78	
	130	48	91	
S-Methylcysteine	190	50	50	120–130
	142	100	100	
Phenylalanine	220*	60	30	125–140
	129	6	6	
	91	100	100	
Glutamine	201*	75	100	125–150
	184	100	88	
	156	20	8	
	142	80	44	
	130	17	2	
Glutamic acid	202*	90	100	140–160
	184	65	60	
	156	20	8	
	142	100	30	
	130	5	2	
Tyrosine	236*	20	40	150–170
	152		3	
	130	30	20	
	107	100	100	
Aspartic acid	188*	100	100	160–180
	170	3	3	
	142	80	28	
Asparagine	187*	100	100	170–190
	170	48	30	
	142	80	60	
Tryptophan	259*	20	6	170–190
	130	100	100	
Lysine	274*	3	4	190–210
	243	80	100	
	212	10	60	
	201	35	34	
	184	40	37	
	143	100	25	
	130	30	6	
Arginine	229*	10	9	190–210
	212	70	13	
	170	94	100	
	142	100	17	
	130	88	71	
Histidine	210*	100	100	220–230
	130	25	18	

[a] *m/e* Ratios and relative abundances at 80 and 10 eV are given for the M^+ ions (*) and for those fragmentation ion peaks useful for identification. The temperature ranges given are those for the maximal ion current for each compound.

limit. Optimal sample size is between 5×10^{-8} and 5×10^{-7} mole.

Identification of MTH-amino Acid Spectra. Table I gives the mass over charge ratio (*m/e*), the relative abundance at 80 and 10 eV, the approximate temperature range of appearance of the M^+ peaks, and those fragmentation ion peaks important for the identification of the ^{14}N methylthiohydantoin amino acid derivatives. Isotope ratios were generally measured at the M^+ ion peak (^{15}N derivatives have an M^+ peak at $m/e + 1$ of the ^{14}N derivatives, except for glutamine and asparagine where the corresponding peak is at $m/e + 2$). Mixtures containing both asparagine and aspartic acid give overlapping M^+ ions for their ^{15}N methylthiohydantoin derivatives. However, the ratio of M^+ to their common fragments will identify the derivatives (See Table I). Once identified, asparagine and aspartic acid can be quantitated by adding each ^{15}N amino in turn. Mixtures containing glutamine and glutamine acid may be treated in a similar manner or be quantitated using deuterium-labeled ^{15}N glutamine. Corrections were made for the occurrence of peaks due to the natural abundance of isotopes. Serine has not given an M^+ ion peak in our hands. It is identified from its fragmentation peak at *m/e* 142. We have not been able to synthesize a satisfactory specimen of serine methylthiohy-

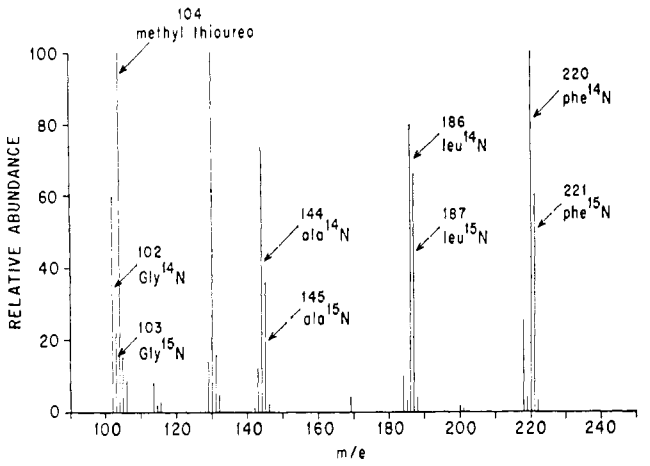

FIGURE 1: Mass spectrum (between *m/e* 100 and 250 at 170° and 80 eV) obtained at locus 8 during simultaneous quantitative sequencing of ribonuclease, lysozyme, and the A and B chains of insulin. The M^+ ion peaks of ^{14}N and ^{15}N glycine, alanine, leucine, and phenylalanine methylthiohydantoin are labeled. The M^+ peak at *m/e* 186 by itself could be either leucine methylthiohydantoin or isoleucine methylthiohydantoin. The peak at *m/e* 143 is too small to be the 80-eV fragmentation peak of isoleucine methylthiohydantoin hence leucine methylthiohydantoin must have been present.

polypeptide mixture
5×10^{-9}–5×10^{-8} M
+
5×10^{-9}–5×10^{-8} M
^{15}N amino acid standard mixture
+

10 M excess methyl isothiocyanate (1st residue) (5 M excess thereafter)	120 min
↓ 2 hr, 60°	
solvent and excess methyl isothiocyanate reagent removed *in vacuo* at 55°	25 min
trifluoroacetic acid added at 20° in dark under N_2	60 min
trifluoroacetic acid removed *in vacuo*	10 min
extracted three times with H_4-furan; aliquot dried in capillary	25 min
mass spectrometer run	12 min
total time	4 hr, 12 min

FIGURE 2: Flow sheet for the sequential degradation procedure.

dantoin free from the dehydro derivative and the assignation of O for the abundance of the M^+ peak for this compound is in doubt.

Criteria for Identification of Methylthiohydantoin Amino Acids. Background. All the reagents used for the sequential degradation provide background peaks in the mass spectrum. This is not a problem in practice, as they occur chiefly on the 80-eV spectrum and not on the 10-eV spectrum, and may be easily subtracted. It has been observed that most volatile impurities present in protein and peptides preparations give extraneous peaks during the identification of the N-terminal residue and do not recur in subsequent residues. A blank run, omitting protein may be useful when the preparation is heavily contaminated.

Identification. The identity of a peak is established: (a) by the mass of the M^+ peak; (b) by its relative abundance ratio compared with one or two of the principal fragmentation ion peaks both at 80 and 10 eV; (c) by the temperature at which the methylthiohydantoin amino acid is volatile (Richards *et al.*, 1968).

While a alone gives rise to ambiguities, *i.e.*, in the M^+ peak of leucine and isoleucine at *m/e* 186; the abundance ratios of the *m/e* 143 peaks are quite different for these amino acids.

Serine methylthiohydantoin shows no M^+ ion and the most prominent fragment ion is at *m/e* 142. Aspartic acid, asparagine, tyrosine, glutamine, glutamic acid, arginine, and *S*-methylcysteine all give fragmentation ions of this mass. M + 1 ions formed by ion–molecule collision or due to natural abundance of ^{13}C may be subtracted by determining their contribution when a standard ^{14}N methylthiohydantoin of comparable concentration is run. However, serine may be unequivocally identified by the volatility of the methylthiohydantoin derivative at 90–100°. Even in the simultaneous presence of other amino acids giving *m/e* 142 peaks, the presence of serine can be deduced and quantitated by the increase in *m/e* 142 to M^+ ratio over that set out in Table I for the other amino acid. The slow raising of the probe temperature assures an uncluttered 10-eV mass spectrum in which the M^+ ion and the principal fragmentation peaks are easy to identify (Richards *et al.*, 1969). The 80-eV spectrum shows a small peak at every mass number and is used to fix the exact mass of the peaks detected in the 10-eV spectrum. Figure 1 shows a typical 80-eV mass spectrum at residue 2 of the ribonuclease, lysozyme, and insulin mixture. A complete quantitative assay can be carried out in 10–12 min. Figure 2 shows a flow sheet of a typical analysis.

Sequence Analyses. Small peptides; the A and B chains of insulin, ribonuclease, lysozyme, an immunoglobulin light chain (PAS human γG myeloma), synthetic amino acid copolymers, and mixtures of these components have been analyzed in the development of this system. The system has also been applied to solve a biological problem: the subunit structure of enolase.

Table II gives the analysis for the 10 N-terminal residues of a mixture containing four peptide chains sequenced simultaneously.

In section A, the quantities of amino acids obtained at each locus are listed in decreasing order of abundance. The maximum number of different amino acids found at any one locus is four (residues 3, 4, 5, 8, and 10). It is concluded that there are probably at least four peptide chains present. In residues 3, 4, 5, 8, and 10 the molar quantity of each amino acid must indicate the molar quantity of each peptide chain. The most abundant amino acids at these residues decrease in value from 4.7 μM at residue 3, to 3.4 μM at residue 10, due to sequential handling losses. We can therefore extrapolate values for the most abundant amino acids at residues 1, 2, 6, 7, and 9. In each case there is only one amino acid residue present of the necessary abundance at each locus. An unambiguous major sequence may therefore be constructed and this is given as sequence 1. It represents, in molar concentration, 47% of the total protein present.

When this sequence is subtracted from the amino acids initially present, the remainder is set out in section B and the same analysis is repeated. A second sequence is obtained

TABLE II: Simultaneous Analysis of a Mixture Containing Four Reduced and Carboxymethylated Polypeptide Chains.

	Amino Acid Residue										
	1	2	3	4	5	6	7	8	9	10	Conclusions
Section A[a] (μmoles)	Lys, 7.0 Gly, 1.3 Phe, 1.2	Glu, 4.7 Val, 3.2 Ile, 1.0	Thr, 4.7 Phe, 2.2 Val, 1.0 Asn, 1.0	Ala, 4.7 Gly, 2.2 Gln, 0.9 Glu, 0.8	Ala, 4.6 Arg, 2.1 His, 1.0 Gln, 0.8	Ala, 4.3 Leu, 1.0 CM-Cys, N/M	Lys, 3.8 Glu, 1.6 CM-Cys, N/M	Phe, 4.6 Leu, 1.7 Ala, 0.8 Gly, 0.8	Glu, 3.3 Ala, 2.3 Ser, 1.5	Arg, 3.4 Ala, 1.5 Val, 0.8 His, N/M	4 Polypeptide chains present
Sequence 1 (μmoles)	Lys, 4.7	Glu, 4.7	Thr, 4.7	Ala, 4.7	Ala, 4.6	Ala, 4.3	Lys, 3.8	Phe, 4.6	Glu, 3.3	Arg, 3.4	Major sequence (47%)
Section B[b] (μmoles)	Lys, 2.3 Gly, 1.3 Phe, 1.2	Val, 3.2 Ile, 1.0	Phe, 2.2 Val, 1.0 Asn, 1.0	Gly, 2.2 Gln, 0.9 Glu, 0.8	Arg, 2.1 His, 1.0 Gln, 0.8	CM-Cys, N/M Ile, 1.0	Glu, 1.6 CM-Cys, N/M	Leu, 1.7 Ala, 0.8 Gly, 0.8	Ala, 2.3 Ser, 1.5	Ala, 1.5 Val, 0.8 His, N/M	
Sequence 2 (μmoles)	Lys, 2.3	Val, 2.2	Phe, 2.2	Gly, 2.2	Arg, 2.1	CM-Cys, N/M	Glu, 1.6	Leu, 1.7	Ala, 2.3	Ala, 1.5	Second sequence (25%)
Section C[c] (μmoles)	Gly, 1.3 Phe, 1.2	Val, 1.1 Ile, 1.0	Val, 1.0 Asn, 1.0	Gln, 0.9 Glu, 0.8	His, 1.0 Gln, 0.8	Ile, 1.0 CM-Cys, N/M	CM-Cys, N/M	Ala, 0.8 Gly, 0.8	Ser, 1.5	Val, 0.8 His, N/M	
Unresolved sequences 3 and 4 (μmoles)	Gly, 1.3 Phe, 1.2	Val, 1.1 Ile, 1.0	Val, 1.0 Asn, 1.0	Gln, 0.9 Glu, 0.8	His, 1.0 Gln, 0.8	Ile, 1.0 Cys, N/M	CM-Cys, N/M CM-Cys, N/M	Ala, 0.8 Gly, 0.8	Ser, 0.8 Ser, 0.8	Val, 0.8 His, N/M	Unresolved sequences 14% each sequence
Section D[d]											
Ribonuclease (5 μmoles)	Lys	Glu	Thr	Ala	Ala	Ala	Lys	Phe	Glu	Arg	
Lysozyme (2.5 μmoles)	Lys	Val	Phe	Gly	Arg	Cys	Glu	Leu	Ala	Ala	
Insulin A (1.4 μmoles)	Gly	Ile	Val	Glu	Gln	Cys	Cys	Ala	Ser	Val	
Insulin B (1.4 μmoles)	Phe	Val	Asn	Gln	His	Leu	Cys	Gly	Ser	His	

[a] The number of micromoles of recovered amino acid methylthiohydantoin derivative at each locus are given. The inferred major sequence (47% of total moles of protein present) is given as sequence 1. N/M = residue identified but not quantitated. [b] Micromoles of amino acid methylthiohydantoin at each locus after subtraction of sequence 1. The second (25%) sequence inferred is given. For ambiguity at residue 9, please see text. [c] The residues at each locus of the two minor sequences (14%) present in equimolar proportions. [d] Composition of test mixture. The composition of each protein added was checked by amino acid analysis.

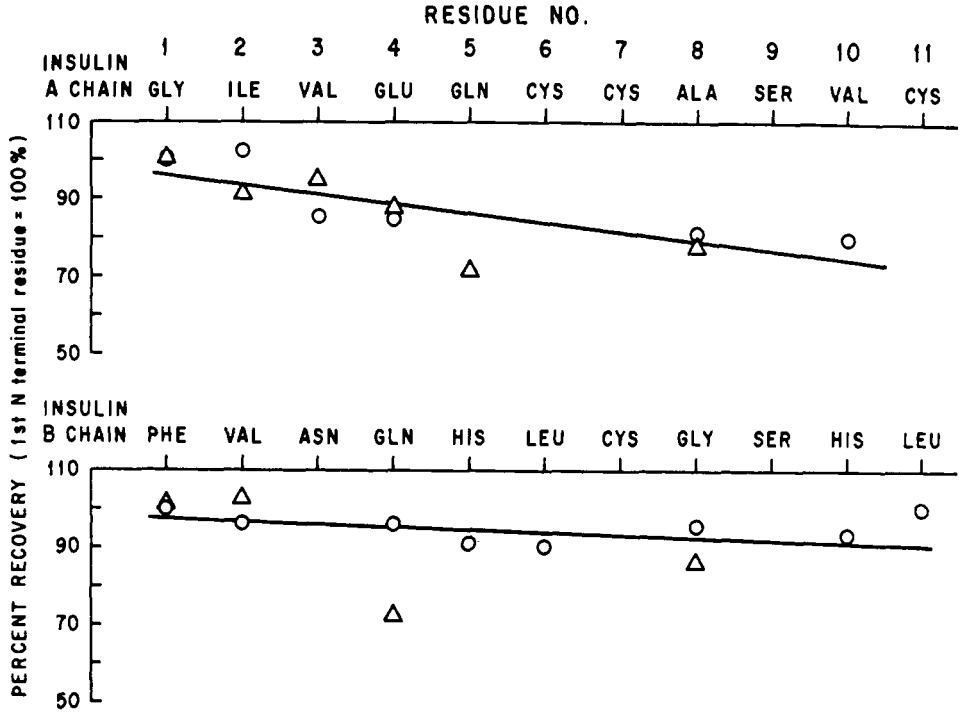

FIGURE 3: Recovery of methylthiohydantoin amino acids on simultaneous sequential degradation of the A and B chains of insulin: (—Δ—) recovery values using ^{15}N methylthiohydantoin derivatives in the isotope dilution procedure; (—O—) recovery values using ^{15}N amino acids. Average loss in reactive N-terminal residues equals 2.5% per residue (A chain); 0.9% per residue (B chain).

(25% of the total number of moles of protein present) which is unambiguous except for residue 9. Here either alanine or serine might fit. Alanine was chosen despite the fact that the value seems high, because the serine values then fit more closely to the extrapolated values for the remaining two sequences. There is however real ambiguity at this locus. *S*-methylcysteine residues at residues 5 and 6 and the histidine residue at locus 10 were not quantitated. Nevertheless inspection shows that it is possible to assign *S*-methylcysteine unambiguously to the correct polypeptide chains.

When the second sequence is subtracted, it is clear that the remainder (section C) corresponds to two further sequences present in approximately equimolar amounts each totaling about 14% of the protein present. Since the same quantities of each sequence are present it is not possible to assign each residue to either one chain or the other. The histidine residue at locus 10 clearly belongs to one of these sequences.

Thus from a single run it is possible to determine that there are four polypeptides present, the relative molar quantities of each; the amino acid sequence of the two major polypeptides up to residue 10 and the amino acids present at each locus in the remaining two sequences. The labor is somewhat less than the determination of the sequence analysis of a single polypeptide using conventional thin-layer chromatography detection.

Subunit Structure of Yeast Enolase. Early work by Malmström *et al.* (1959) had suggested that yeast enolase is a single polypeptide, mol wt 67,000. Brewer and Weber (1968) and Gavronski and Westhead (1968) demonstrated that the enzyme contained two subunits of equal size. Recent studies suggest that of the two bound magnesium ions, one is more tightly bound, and that its binding induces conformational change in the protein. The second, more loosely bound magnesium ion, is probably involved in the catalytic functions of the protein (Brewer and Weber, 1966; Hanlon and Westhead, 1969). The question raised by these findings about the identity or nonidentity of the subunits have been examined using tryptic digestion, cyanogen bromide fractionation, and quantitative N-terminal degradation by means of isotope dilution and mass spectrometry (Brewer *et al.*, 1970).

When 0.64 μmole of yeast enolase was degraded, 1.13 μmoles of N-terminal alanine methylthiohydantoin (mean value) was recovered. Repetitive experiments gave a standard deviation of ±0.11 μmole. This corresponds to 1.76 ± 0.19 moles of alanine per mole of protein. The figure for the second residue was 0.96 μmole of glycine as the sole amino acid present. This represents 1.37 moles of glycine per mole of protein originally present. With protein loss correction this increased to 1.74 moles/mole of protein. Further degradation showed that an unique sequence Ala-Gly-Lys-Val-Gly-Asp-Thr-Glu(NH_2) was present. Residues 3–8 showed the presence of only a single amino acid in quantities greater than 5×10^{-9} mole at each locus, making quantitation unnecessary. This evidence suggests that at least as far as residues 1–8 are concerned, the subunits of yeast enolase are identical.

Sequential Degradation Reaction. Figure 3 gives the per cent yield of methylthiohydantoin amino acid derivatives (in terms of the N-terminal residue) at each cycle of the simultaneous N-terminal degradation of the A and B chains of insulin. The reaction has been carried out using both ^{15}N methylthiohydantoin derivatives and ^{15}N amino acids in the isotope dilution procedure.

The experiment indicates the following points. (1) There is somewhat greater scatter when ^{15}N methylthiohydantoin amino acid derivatives are used for quantitation than when the ^{15}N amino acids are employed. (2) There is no evidence that the recoveries of ^{15}N methylthiohydantoin amino acids are either substantially higher or lower when the ^{15}N methylthiohydantoin derivatives are employed in the isotope dilution procedure. This suggests that the reaction rates for MeSNC addition and cyclizations do not differ to a large extent when the amino acid is present as the free acid, or present as the N-terminal residue of a protein. These data also suggest that it is unlikely that the methylthiazolinone derivatives are converted into compounds other than the amino acid methylthiohydantoin derivatives to a degree which is quantitatively important. Had substantial losses of methylthiazolinone derivatives occurred, the recovery values using the ^{15}N methylthiohydantoin amino acids should have been systematically below the values in experiments where ^{15}N amino acids were used, since the loss would have occurred prior to the isotope dilution step. (3) The rate of loss of reactive N-terminal groups was approximately 2.5% per residue for the A chain and 0.9% per residue for the B chain. These loss rates may vary with conditions of the extraction procedure, with the presence of other peptides, the quantities of peptides being sequenced and should be established for each experiment. However, the rate of loss for any one polypeptide chain is approximately linear in all the experimental conditions we have so far tried. This makes possible the extrapolation of values for "missing residues" in which the amino acid has been identified but not quantitated. It is also very useful where the same amino acid occurs in the same locus in more than one of the chains being sequenced simultaneously. Table II illustrates that in this protein mixture experiment, in the first ten residues there were five loci in which a single amino acid appeared in more than one chain (1, 2, 6, 7, 9). In the same experiment there were seven residues not quantitated. Despite this only a single ambiguity due to lack of quantitation occurred in position 9.

In the protein and peptides tested there has been no consistent pattern suggesting a sudden drop in yield at any particular amino acid residue. Where such a drop has occurred, it could not be demonstrated at similar residues in other peptides.

Discussion

There are several methods available which are useful for the quantitative degradation of proteins and peptides. The method of Stark and Smyth (1963) using the cyanate ion gives good quantitative results for the N-terminal amino acid only. A nondestructive quantitative method has been reported by Barrett (1967). A further theoretical method has also been proposed by Gray (1968). We selected the Edman method because there was good evidence that under optimal conditions the loss of protein per residue degraded could not be in excess of 2%. In single runs up to 50 residues of a protein has been sequenced (Edman and Begg, 1967). A theoretical objection to the use of the Edman method for quantitation lies in the presence of the 5-thiazolinone intermediate. It may be possible for this compound to fail to rearrange to the methylthiohydantoin derivative, or to convert into substances other than the methylthiohydantoin derivative. This means that the methylthiohydantoin derivatives formed from thiazolinones is the sum of at least two equilibria. If the methylthiohydantoin derivative, like its phenyl analog is unstable (Fraenkel-Conrat *et al.*, 1955) a further difficulty is introduced. It is possible to correct for losses, in this reaction, by isotope dilution (Callewaert and Vernon, 1968; Laursen, 1969). The data in Figure 1 suggest that such losses are not important in this reaction. Isotope dilution gives the added benefit that after addition of isotope, all further handling steps in the identification of the thiohydantoins need no longer be carried out in a quantitative fashion since only the ratio of ^{14}N to ^{15}N isotope is measured. The ability to resolve sequences from quantitative amino acid data at successive loci is a complex function. It depends principally on the number of sequences present and on the frequency distribution of the amino acids within each sequence. In order to establish the number of polypeptides present there must be loci at which the chains which comprise the mixture have no common amino acids. There should be at least two or three and preferably more such loci, so that the rate of loss of reactive N-terminal amino groups for each chain may be established (see Figure 3). In the four component mixture analyzed there were five such loci in the first ten residues (Table II). As the number of chains increases, these fully resolved loci become less frequent and the number of residues which will have to be sequenced in order to obtain a minimum of three fully resolved loci will increase. We do not know yet what the practical limits of resolution are, but it seems possible that we should be able eventually to sequence perhaps seven or eight polypeptides at once. Automatic methods of sequential degradation which give more reproducible results should increase the power of resolution of the method. Assignment of residues to chains cannot be performed if the chains are present in equimolar amounts (see the insulin chains in Table II). However, protein or mixtures, obtained in enzyme purification, or peptides obtained during protein degradation rarely contain equimolecular mixtures, and more commonly have predominant and minor sequences present.

The method has uses for establishing the homogeneity of protein subunits and one such example is the analysis of yeast enolase given here. Preliminary studies on H_4folate synthetase have shown that four N-terminal histidines are present and that subsequent residues contain only a single amino acid at each locus. A major effort to determine the number of major sequences present in antihapten antibodies elicited both from randomly 2,4-dinitrophenyl-substituted bovine globulin and from defined sequence peptides containing the 2,4-dinitrophenyl determinant at regular intervals (Richards *et al.*, 1967; Richards *et al.*, 1969) is in progress.

Initial experiments with phenyl isothiocyanate demonstrated that the yield of M^+ ions was poor for some amino acid phenylthiohydantoin derivatives. Methyl isothiocyanate gave mostly excellent M^+ ion yields (Figure 2). Because it is volatile it does not require a solvent extraction step for the reagent.

The accuracy of the method is limited by the chemistry of the Edman reaction. It would have been useful to have a second reliable criterion of protein loss at each stage of degradation. Automatic amino acid analysis in our hands gives accuracy in the range of $\mp 3\%$ which is insufficient

to detect losses of reactive amino-terminal groups of the order of 1–4% per residue sequenced. Amino acid analyses at the beginning and end of sequenced runs have given results consistent with the isotope dilution results. At present our best estimates for the amount of reactive amino-terminal groups lost per residue sequenced is 2.7%, for ribonuclease and 3.2% for lysozyme. These losses have decreased as we have become more familiar with the method. Experiments are in progress aimed at transferring the chemistry of the reaction to a modified Bio-Cal ES 300 Protein Sequenator. We hope that this, together with linkage of peptides to insoluble supports will decrease losses still further. The fact that the rate of loss of reactive N-terminal groups was higher in insulin A chain than (2.5%) in the B chain (0.9%) suggests that the observed losses may result from a greater solubility of the A chain in the H_4furan used for extracting the methylthiohydantoin amino acid derivatives from the proteins.

The mass spectrometer assay of the methylthiohydantoin amino acid derivatives is straightforward and presents no particular difficulty. It is entirely compatible with on-line computer handling of data. Programs for data reduction and handling have been written by one of us (R. E. L.).

Certain ^{15}N amino acids, notably cysteine, histidine, and methionine, are either expensive or are difficult to obtain commercially. An occasional unquantitated residue has not proved to be a handicap; it is surprisingly easy to place these residues into their correct sequence when they have been identified at a locus (see Table II). In the mixed-protein analysis reported here no ambiguities arose from a failure to quantitate cysteine or histidine. Methionine did not occur in the polypeptides reported here. We are now preparing the missing ^{15}N-labeled amino acids.

The assay method requires a single-focusing mass spectrometer, an instrument which is in the same cost range as the more expensive amino acid analyzers. The staggered analysis procedure referred to in the Experimental Section greatly reduced the amount of ^{15}N amino acids required and brings the cost down to levels within the financial reach of most protein chemistry laboratories. The assay method for the released thiohydantoin derivative is much more rapid than any of the presently employed assay methods. It does not require conversion of the cyclized derivatives back to their parent amino acids and it will identify unequivocally all the amino acids present, in a single 12-min run on the mass spectrometer. It will sequence a number of peptides or proteins at the same time giving quantitative as well as qualitative information and unlike gas chromatography it does not require substitution of certain unvolatile residues such as arginine, glutamic, and aspartic acid.

Acknowledgments

The authors thank Mrs. Christine Attardo for her synthesis of the peptides used in these experiments. Our thanks are also due to Mr. David Tatem for his skillful assistance.

References

Barrett, G. C. (1967), *Chem. Commun.*, 487.

Biemann, K. (1962), Mass Spectrometry, Organic Chemical Applications, New York, N. Y., McGraw-Hill, p 204.

Brenneman, L., and Singer, S. J. (1969), *Proc. Nat. Acad. Sci. U. S. 60*, 258.

Brewer, J. M., Fairwell, T., Travis, J., and Lovins, R. E. (1970), *Biochemistry 9*, 1011.

Brewer, J. M., and Weber, G. (1966), *J. Biol. Chem. 241*, 2550.

Brewer, J. M., and Weber, G. (1968), *Proc. Nat. Acad. Sci. U. S. 59*, 216.

Callewaert, G. L., and Vernon, C. A. (1968), *Biochem. J. 107*, 728.

Dixon, G. H. (1966), *Essays Biochem. 2*, 147.

Edelman, G. M., and Gall, W. E. (1969), *Annu. Rev. Biochem. 38*, 415.

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

Eisen, H. N., and Siskind, G. W. (1964), *Biochemistry 3*, 996.

Fraenkel-Conrat, H., Harris, J. I., and Levy, A. L. (1955), *Methods Biochem. Anal. 2*, 360.

Gavronski, T. H., and Westhead, E. W. (1968), *Fed. Proc. 27*, 522.

Gray, W. R. (1968), *Nature 220*, 1300.

Hanlon, D. P., and Westhead, E. W. (1969), *Biochemistry 8*, 4247, 4255.

Hirs, C. H. W. (1967), *Methods Enzymol. 11*, 199.

Laursen, R. A. (1969), *Biochem. Biophys. Res. Commun. 37*, 663.

Malmström, B. C., Kimmel, J. R., and Smith, E. L. (1959), *J. Biol. Chem. 234*, 1108.

Merrifield, R. B. (1964), *Biochemistry 3*, 1385.

Miller, E. J., Osterland, C. K., Davie, J. M., and Krause, R. M. (1967), *J. Immunol. 98*, 710.

Petra, P. H., and Neurath, H. (1969), *Biochemistry 8*, 5029.

Richards, F. F., Barnes, W. T., Lovins, R. E., Salomone, R., and Waterfield, M. D. (1968), *Nature 221*, 1241.

Richards, F. F., and Haber, E. (1967), *Biochim. Biophys. Acta 140*, 558.

Richards, F. F., Pincus, J. H., Bloch, K. J., Barnes, W. T., and Haber, E. (1969), *Biochemistry 8*, 1377.

Richards, F. F., Sloane, R. W., Jr., and Haber, E. (1967), *Biochemistry 6*, 476.

Stark, G. R., and Smyth, D. G. (1963), *J. Biol. Chem. 238*, 214.

Stepanov, V. M., and Krivtsov, V. F. (1965), *J. Gen. Chem. (USSR) 35*, 53, 556, and 988.

Zuckerkandl, E. (1968), *in* Structural Chemistry and Molecular Biology, Rich, A., and Davidson, N., Ed., San Francisco, Calif., W. H. Freeman, p 256.