18. Y. Masuho and T. Hara, *Gann* **71**, 759 (1980).
19. V. Raso and T. Griffin, *J. Immunol.* **125**, 2610 (1980).
20. V. Raso, *Immunol. Rev.* **62**, 93 (1982).
21. D. G. Gilliland, Z. Steplewski, R. J. Collier, K. F. Mitchell, T. H. Chang, H. Koprowski, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4539 (1980).
22. V. Raso, J. Ritz, M. Basala, S. F. Schlossman, *Cancer Res.* **42**, 457 (1982).
23. P. E. Thorpe, W. C. J. Ross, A. J. Cumber, C. A. Hinson, D. C. Edwards, A. J. S. Davies, *Nature (London)* **271**, 752 (1978).
24. W. C. J. Ross, P. E. Thorpe, A. J. Cumber, D. C. Edwards, C. A. Hinson, A. J. S. Davies, *Eur. J. Biochem.* **104**, 381 (1980).
25. D. G. Gilliland and R. J. Collier, *Cancer Res.* **40**, 3564 (1980).
26. Y. Masuho, T. Hara, T. Noguchi, *Biochem. Biophys. Res. Commun.* **90**, 320 (1979).
27. P. E. Thorpe, A. J. Cumber, N. Williams, D. C. Edwards, W. C. J. Ross, A. J. S. Davies, *Clin. Exp. Immunol.* **43**, 195 (1981).
28. P. E. Thorpe, A. N. F. Brown, W. C. J. Ross, A. J. Cumber, S. I. Detre, D. C. Edwards, A. J. S. Davies, F. S. Stirpe, *Eur. J. Biochem.* **116**, 447 (1981).
29. B. L. Kagen, A. Finkelstein, M. Colombini, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4950 (1981).
30. E. S. Vitetta, unpublished observations.
31. S. Olsnes and A. Pihl, *Pharmacol. Ther.* **15**, 355 (1982).
32. E. S. Vitetta, K. A. Krolick, J. W. Uhr, *Immunol. Rev.* **62**, 159 (1982).
33. S. M. Fu, R. J. Winchester, H. G. Kunkel, *J. Immunol.* **114**, 250 (1975).
34. D. W. Hough, R. P. Eady, T. J. Hamblin, F. K. Stevenson, G. T. Stevenson, *J. Exp. Med.* **144**, 960 (1977).
35. G. Haughton, L. L. Lanier, G. F. Babcock, M. A. Lynes, *J. Immunol.* **121**, 2358 (1978).
36. J. N. Hurley, S. M. Fu, H. G. Kunkel, G. McKenna, M. D. Scharff, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5706 (1978).
37. R. Levy, R. Warnke, R. F. Dorman, J. Haimovich, *J. Exp. Med.* **145**, 1014 (1977).
38. K. A. Krolick, P. C. Isakson, J. W. Uhr, E. S. Vitetta, *Immunol. Rev.* **48**, 81 (1979).
39. S. Slavin and S. Strober, *Nature (London)* **272**, 624 (1977).
40. M. J. Muirhead, P. C. Isakson, K. A. Krolick, J. W. Uhr, E. S. Vitetta, *Am. J. Pathol.* **105**, 295 (1981).
41. M. J. Muirhead, J. M. Holbert, J. W. Uhr, E. S. Vitetta, *ibid.*, p. 306.
42. E. Vitetta, E. Puré, P. Isakson, L. Buck, J. Uhr, *Immunol. Rev.* **52**, 211 (1980).
43. K. A. Krolick, J. W. Uhr, S. Slavin, E. S. Vitetta, *J. Exp. Med.* **155**, 1797 (1982).
44. R. P. Gale, *J. Am. Med. Assoc.* **243**, 540 (1980).
45. K. A. Krolick, J. W. Uhr, E. S. Vitetta, *Nature (London)* **295**, 604 (1982).
46. J. H. Kersey, H. J. Meuwissen, R. A. Good, *Hum. Pathol.* **2**, 389 (1971).
47. M. E. Tyan, *Transplantation* **15**, 601 (1973).
48. H. V. Rodt, S. Therfelder, M. Eulitz, *Eur. J. Immunol.* **4**, 25 (1974).
49. R. Korngold and J. Sprent, *J. Exp. Med.* **148**, 1687 (1978).
50. D. A. Vallera, C. Soderling, G. Carlson, J. H. Kersey, *Transplantation* **31**, 218 (1981).
51. H. G. Prentice, H. A. Blacklock, G. Janossy, *Lancet* **1982-I**, 700 (1982).
52. D. A. Vallera, R. J. Youle, D. M. Neville, Jr., J. H. Kersey, *J. Exp. Med.* **155**, 949 (1982).
53. S. Slavin, S. Strober, Z. Fuks, H. S. Kaplan, *Transplant. Proc.* **9**, 1001 (1977).
54. S. Slavin, S. Morecki, L. Weiss, *J. Immunol.* **124**, 586 (1980).
55. N. A. Mitchison, *Eur. J. Immunol.* **1**, 68 (1971).
56. D. J. Volkman, A. Ahmad, A. S. Fauci, D. M. Neville, Jr., *J. Exp. Med.* **156**, 634 (1982).
57. J. S. Goodwin, Ed., *Suppressor Cells in Human Disease* (Dekker, New York, 1981).
58. R. Van Furth and Z. A. Cohn, *J. Exp. Med.* **128**, 415 (1968).
59. We thank M. Bagby-Wyatt, C. Bockhold, Y. Chinn, and L. Trahan for expert technical assistance, D. Marcoulides and G. A. Cheek for helpful secretarial assistance and L. Eidels for reviewing the manuscript. Supported by NIH grants AI-11851 and CA28149. K.A.K. was a postdoctoral fellow in the department of microbiology, University of Texas Southwestern Medical School, Dallas, at the time much of this work was conducted.

Protein Sequence Analysis: Automated Microsequencing

Michael W. Hunkapiller and Leroy E. Hood

Summary. The automated microsequencing of proteins can now be carried out at the 5- to 10-picomoles (submicrogram) level on polypeptides obtained directly from one- and two-dimensional gel electrophoresis. The techniques are applicable to polypeptides ranging in size from small peptides (less than 10 residues) to large proteins (more than 1000 residues).

Since its introduction 25 years ago, the Edman degradation has been the most widely used method for the direct determination of the primary structure of proteins and peptides (*1*). However, the introduction of rapid, simple methods of DNA sequencing, by which protein sequences are obtained indirectly, has raised questions about the utility of protein sequencing. Several years ago in one leading scientific journal, an editorial appeared under the title *The Decline and Fall of Protein Chemistry* (*2*). However, the advent of modern micromethods requiring as little as a few picomoles of proteins and peptides for sequence analysis has firmly established the importance of protein sequencing as a tool for biochemistry and molecular biology.

Chemistry

The Edman chemistry is shown in Fig. 1. One cycle, which results in removal of one amino acid from the amino-terminal end of a peptide and generation of a new peptide that is one amino acid shorter, consists of two separate chemical steps. In the first (coupling), phenyl isothiocyanate is coupled under basic conditions to the amino end of the peptide to form a phenylthiocarbamyl peptide. In the second (cleavage), treatment with a strong, anhydrous acid removes the derivatized amino acid as its anilinothiazolinone. The latter is usually converted in a third reaction (conversion) into the more stable phenylthiohydantoin (Pth) for subsequent analysis.

Repetition of this sequence, in theory, allows one to proceed from the amino-terminal to carboxyl-terminal end of a protein to define its primary structure. In practice, side reactions, incomplete reactions, and loss of sample usually limit successful degradations in a single run to 30 to 70 cycles. The sequence of the remainder of the protein is determined after fragmentation by chemical or enzymic methods (or both) to generate a set of overlapping peptides that are individually analyzed by the Edman chemistry.

Instrumentation

Manual. The Edman degradation was originally developed as a manual method. Despite the subsequent popularity and success of automated methods, there remain many applications where the manual approach is suitable, particularly when there is need for rather limited sequence information and there is no access to automated instruments. Improvements in technique (*3*) and chemistry (*4*) make it particularly attractive for initial screening of a large number of small- to medium-sized peptides to select those most suitable for further, extensive sequence analysis by automated methods. Experienced protein chemists can

M. W. Hunkapiller is a Senior Research Fellow and L. E. Hood is Bowles Professor of Biology at the Division of Biology, California Institute of Technology, Pasadena 91125.

usually expect five to ten successful degradations with as little as 1 to 10 nanomoles of peptide, and several peptides can be analyzed simultaneously. Manual methods are not routinely used for long amino-terminal sequence analyses of proteins because of the extent of side reactions inherent in the manual manipulations.

Spinning cup (liquid phase) sequenator. The spinning cup sequenator has been the workhorse of protein sequence analysis since its introduction by Edman and Begg (*5*) in 1967. Its design includes (i) an inert gas pressure supply and distribution system, (ii) a series of reagent and solvent reservoirs pressurized by the inert gas, (iii) a reaction vessel consisting of a spinning (1200 to 3600 revolutions per minute) glass cup contained in a sealable chamber, (iv) delivery valves that direct reagents and solvents into and out of the cup, (v) a vacuum system to remove volatile chemicals from the cup, (vi) a fraction collector into which the amino acid anilinothiazolinones cleaved from the peptide are deposited, and (vii) an electronic programmer that controls instrument operation. Some spinning cup sequenators also include a secondary reaction vessel in which the anilinothiazolinones are converted into the Pth amino acids (*6*).

The key feature of the spinning cup design is the cup itself, since it provides the means of immobilizing the peptide during the many cycles of chemical manipulation. A solution of the sample is added to the cup and dried into a thin film on its interior wall. Reagents are added to dissolve the sample and effect the Edman degradation. Excess reagents, by-products, and the anilinothiazolinones are removed first by evacuation and then by extraction with solvents flowing over the dried peptide film.

Since a considerable mass (a few milligrams) is required to form a stable film on the cup wall, the spinning cup sequenator was first used with several hundred nanomoles of protein. Moreover, the large volume of extraction solvents required to clean the cup initially limited its usefulness in sequencing small peptides because many peptides are somewhat soluble in the solvents. The introduction of a nonprotein carrier, Polybrene (see discussion below), has minimized these problems and greatly extended the capabilities of the spinning cup sequenator (*7*).

Solid phase sequenator. Laursen (*8*) sought to eliminate the main problem with the original Edman sequenator, the loss of sample from the cup, by replacing the cup with a column packed with a solid phase support (derivatized glass or polystyrene beads) onto which the peptide is covalently attached prior to initiating the Edman degradation. The column system, owing to its potential miniaturization, efficient flow properties during solvent extractions, and mechanical simplicity, offers substantial theoretical advantages, especially for microsequencing, compared to the spinning cup system. In practice, these advantages have not been realized because of difficulties in coupling many peptides and most proteins in high yield to the solid support. With proper care, reasonable coupling efficiencies can be obtained in certain cases, however, and sequencing in the low nanomole sample range is possible (*9*).

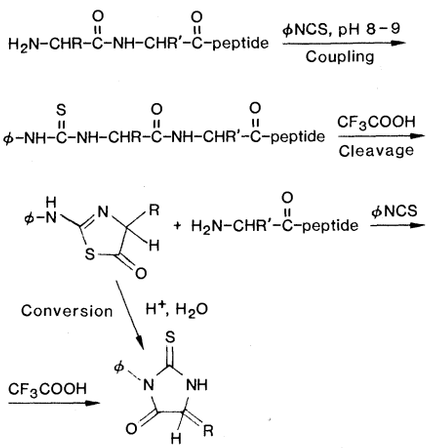

Fig. 1. Edman degradation.

Gas phase sequenator. Recently, a new type of sequenator, one in which gas rather than liquid phase reagents are used at critical points in the Edman degradation, was described (*10*) (Fig. 2). This sequenator, which resembles the solid phase instrument in having a cartridge-type reaction chamber, differs from the solid phase instrument in its use of gas phase base and acid for the coupling and cleavage steps, respectively. If these reagents were delivered as liquids, the sample would be dissolved and carried away unless it was covalently attached to the support. The gas phase delivery eliminates the need for covalent attachment and thus the problems associated with this process. The sample solution is applied to the support, a thin disc of glass filter paper coated with Polybrene, in much the same manner as the sample is applied to the spinning cup. The miniaturization possible with this system (Fig. 3) allows use of minimal

Table 1. Sample of proteins sequenced on gas phase sequenator.

Protein	Molecular size (daltons)	Residues identified	Amount	
			pmole	μg
Angiotensin II	1,000	8	500	0.5
Angiotensin II	1,000	6	50	0.05
Somatostatin	1,600	14	1,200	2.0
Insulin, B chain	3,400	30	260	1.0
Aplysia neuropeptide B	3,800	31	500	2.0
Dynorphin (1–17)	2,000	14	20	0.04
Dynorphin (1–32)	4,000	32	200	0.6
Calliphora polypeptide	4,000	36	75	0.3
Myoglobin	17,500	90	10,000	175
Myoglobin	17,500	22	5	0.09
Drosophila larval cuticle protein*	18,000	55	850	15
Aplysia membrane phosphoprotein†‡	22,000	23	15	0.3
Human histocompatibility antigen, HLA-DR, α chain‡	32,000	49	700	23
Human histocompatibility antigen, HLA-DR, β chain‡	26,000	39	500	13
Mouse immune response antigen E_β^k*†‡	25,000	21	20	0.5
Human erythropoietin§	44,000	40	100	4
Human melanoma cell surface antigen†‡	95,000	13	60	5.5
Eel acetylcholine receptor α subunit†‡	40,000	68	400	16
Calf acetylcholine receptor α subunit†‡	42,000	35	50	2
MOPC-315 heavy chain	55,000	20	20	1
Colony stimulating factor†‡	30,000	25	30	1
Bovine lens gap junction protein, CNBr fragment†‡	14,000	34	120	2
Platelet-derived growth factor subunit†	18,000	20	50	1
Rat transforming growth factor	6,000	42	250	1.5
Transcription regulatory factor, CNBr fragment†	20,000	35	20	0.4
Rat transforming growth factor, lys-C fragment	2,300	20	25	0.06

*Purified by isoelectric focusing in polyacrylamide gels containing urea. †Electrophoretically eluted from Coomassie blue-stained, sodium dodecyl sulfate–polyacrylamide gels. ‡Integral membrane protein. §Sixty percent carbohydrate by weight.

solvent volumes and makes the gas phase instrument suitable for small peptides as well as proteins.

Pth amino acid analysis. Edman originally used thin-layer chromatography to identify Pth amino acids (*5*). Others have used a variety of analytical procedures, including gas chromatography (*11*), mass spectrometry (*12*), and amino acid analysis after hydrolysis of the Pth amino acids (*13*). None of these methods alone provided a quantitative analysis of all common Pth amino acids, and their inadequacy hindered sequencing by any of the methods described above, especially at micro levels. The development of high-performance liquid chromatography (HPLC) has provided the analytical resolution, quantitation, speed, and reliability required for microsequencing (*14*). An example of the quality of HPLC resolution obtainable is shown in Fig. 4, and the sensitivity of this system is such that as little as a few hundred femtomoles (0.1 nanogram) of Pth amino acids can be detected.

Microsequencing

Many interesting biological systems yield very little protein or peptide for study. Hence, the ability to sequence very small amounts of sample (a few micrograms or less) is often crucial. Although the Edman chemistry was originally designed to handle several milligrams of protein (and was at the time of its introduction considered a microsequencing method), the general power of this chemistry is illustrated by the success of efforts to increase the sensitivity of sequencing by several orders of magnitude. The term "microsequencing," although it is arbitrary and changes as technology improves, may now be considered to involve sequencing less than 0.5 nmole of sample. Three microsequencing approaches have been used; two are based on isotopic labeling, primarily to improve sensitivity of the Pth analysis, and the third is based on direct analysis of nonlabeled Pth amino acids.

External label. The first serious attempts at microsequencing involved increasing sensitivity of the phenylhydantoin detection system by introducing an isotopic label via phenylisothiocyanate. The efforts included using [^{3}H]- and [^{14}C]phenylisothiocyanate and [^{35}S]-phenylisothiocyanate as the coupling reagent in spinning cup sequencing (*15, 16*) and solid phase sequencing (*8, 17*). These external labeling procedures suffered from the use of expensive, unstable, and difficult-to-purify radioactive labeling reagents. Nevertheless, they allowed sequencing at the 0.5- to 10-nmole level. Their usefulness has diminished with the advent of HPLC analysis of Pth amino acids and the excellent sensitivity of this type of analysis.

Internal label. A more successful exploitation of the sensitivity of isotopic labeling methods was the use of internal labeling of the protein itself during its synthesis. This has been achieved by short-term tissue culture procedures that readily incorporate several labeled amino acids into newly synthesized polypeptides (*18*), by longer-term tissue culture techniques in which a mixture of labeled amino acids and Krebs cycle intermediates is used (*19*), and by cell-free translation of messenger RNA (mRNA) with labeled amino acids (*20*). In the former, sequenator analysis of the partially labeled proteins results in partial amino acid sequence data with the unlabeled residues being registered as blanks, while in the latter cases complete sequence information is possible. These techniques have drawbacks as routine microsequencing procedures in that they are expensive and they require a source of cells producing protein or active mRNA. However, they can be used with any of the sequencing instruments, and they require protein that is only free of other labeled proteins. The presence of nonlabeled protein, such as antibody used in purification of the labeled protein, does not interfere with the sequencing. The in vivo labeling methods allow sequencing with as little as a few pico-

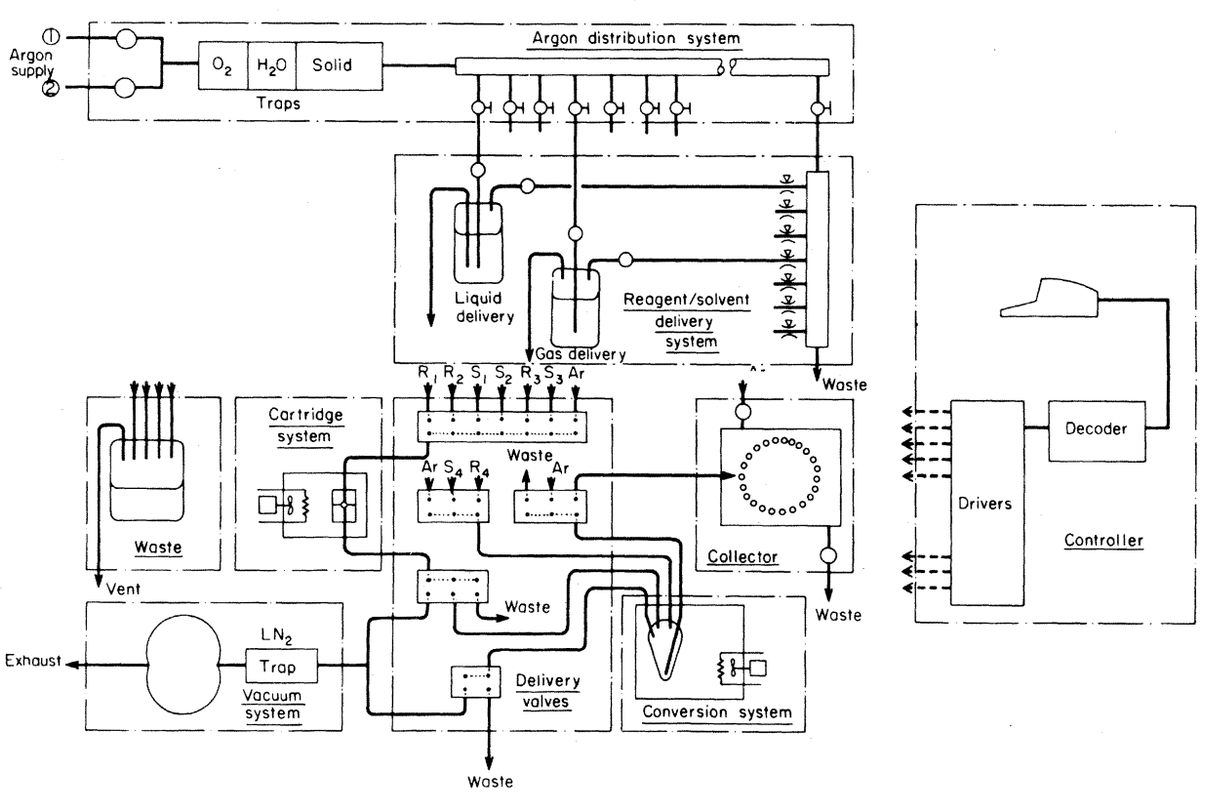

Fig. 2. Schematic diagram of gas phase sequenator (*10*). [Courtesy of the *Journal of Biological Chemistry*]

moles of protein, while the in vitro procedure has been used with as little as 100 fmole of sample. The power of the techniques is illustrated by the structural analysis of the mouse transplantation antigens (*21, 22*).

Microsequencing without isotopic labeling. The development of so-called direct microsequencing, that is with no isotopic labeling, has depended on advances in three areas. The first was the introduction of HPLC analysis of Pth amino acids (*14*). The second was the introduction of a nonprotein carrier to retain the protein in the reaction chamber of the sequenator without the troublesome requirement of covalent attachment to a solid support. The "magic" compound that has revolutionized sequencing is Polybrene, a polymeric quaternary ammonium salt first used by Tarr *et al.* (*7*) to sequence small peptides in the spinning cup sequenator. The Polybrene readily forms a stable film on glass surfaces into which the polypeptide can be embedded and protected from mechanical dislodging by the moving solvent streams. The film is readily penetrated by the Edman reagents, however, and the degradation proceeds smoothly in its presence.

The third area of advancement was in the instrumentation used for performing the Edman degradation—the sequenator. With an HPLC system that can detect less than a picomole of Pth amino acid and with Polybrene to hold picomole quantities of protein in the sequenator, sequencing at the picomole level is possible if the sequenator can produce phenylhydantoins free of significant amounts of ultraviolet-absorbing contaminants (primarily side reaction products of the phenylisothiocyanate) that would interfere with the HPLC analysis. Hunkapiller and Hood (*23, 24*), utilizing many design changes originated by Wittmann-Liebold (*25*), were able to sequence both proteins and peptides at the subnanomole level with a spinning cup sequenator. These improvements involved extensive changes in almost all mechanical components of the instrument, including the reagent and solvent delivery valves, spinning cup reaction chamber, vacuum system, inert gas supply, and reagent-solvent storage assemblies, as well as the addition of a secondary reaction vessel for automated Pth conversion (*4*), extensive purification of the reagents and solvents, and use of a versatile microprocessor controller.

Because of problems in purification of some of the reagents required in the spinning cup sequenator, particularly the quadrol (*5*) used as a coupling buffer, and

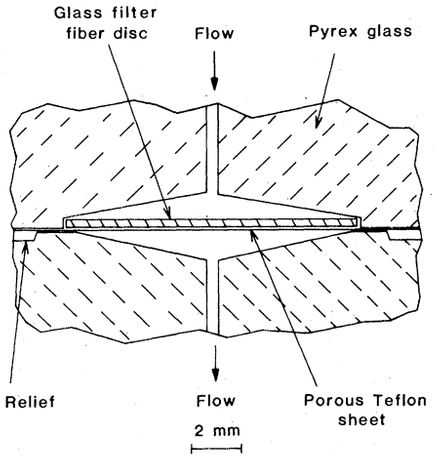

Fig. 3. Enlarged diagram of reaction cartridge for gas phase sequenator (*10*). [Courtesy of the *Journal of Biological Chemistry*]

because the large solvent extractions tended to dislodge some small samples from the spinning cup, sequencing below 100 pmole in the spinning cup is still very difficult to achieve on a routine basis. These problems have largely been overcome in the gas phase sequenator (*10*). The key to the performance of this sequenator in microsequencing is the combination of miniaturization and a flow path that permits use of less reagent than other sequenators and allows more efficient removal of the reagents and their by-products. The low levels of contaminants appearing in the HPLC analysis of fractions from this instrument are shown in Fig. 5.

The gas phase sequenator has been used to analyze a wide variety of proteins and peptides. These samples include glycoproteins, integral membrane proteins, proteins and peptides purified by one- or two-dimensional polyacrylamide gel electrophoresis (PAGE) and isoelectric focusing, short hydrophobic peptides, large proteins (> 90,000 daltons), and peptides containing numerous prolyl residues. A partial list of the samples analyzed on the Caltech gas phase sequenator is shown in Table 1 and illustrates its versatility. It can provide extended runs (> 70 residues) with a few nanomoles of protein, complete to nearly complete sequencing of small- or medium-sized peptides (< 40 residues) in a single run with a few hundred picomoles or less (Fig. 6), and 15 to 30 residues with as little as 5 to 20 pmole of proteins and many peptides (Fig. 5).

Sample Preparation

Crucial to any microsequencing method is the purification and handling of small quantities of polypeptide without contamination that would interfere with the Edman degradation. The failure of microsequencing efforts can, in fact, be traced most often to loss of sample at the final purification step (for example, by its adherence to the walls of test tubes) or by oxidation reactions that block the amino-terminal residues, or contamination by reagents (or impurities in them) used in the purification.

Two methods of preparation are suitable: (i) HPLC for peptides and small proteins and (ii) PAGE for long peptides and proteins. HPLC is the faster and simpler method, and it causes little damage to the samples. The development of volatile elution systems with the use of trifluoroacetic acid (*26*) makes handling peptides particularly easy. Using larger pore packings and better ion pairing agents such as $NaClO_4$ (*27*), one can extend the HPLC method to many large peptides and small proteins. Polyacryl-

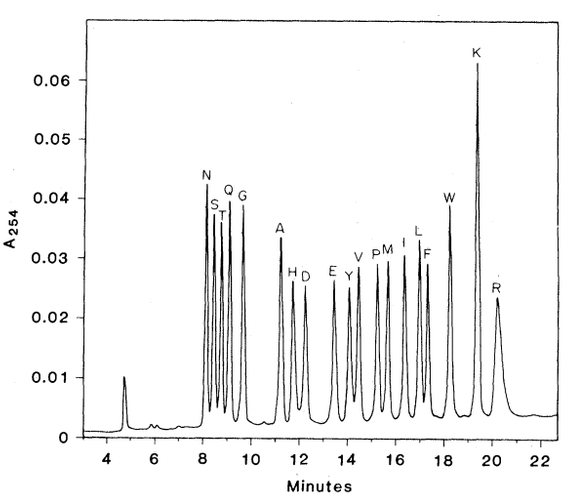

Fig. 4. Separation of Pth amino acids on a cyano column (IBM). The HPLC parameters are 32°C; flow rate, 1 milliliter per minute; column dimensions, 4.6 millimeters (inside diameter) by 25 centimeters in length; guard column, Du Pont Permaphase ETH, 4.6 millimeters (inside diameter) by 5 centimeters in length; aqueous phase 0.015*M* sodium acetate buffer, *p*H 5.8; hydrophobic phase (B), acetonitrile and methanol (4:1); gradient, 18 to 35 percent B in 1 minute, 35 to 55 percent B in 10.2 minutes, 55 to 60 percent B in 1.9 minutes, 60 to 18 percent B in 3.7 minutes. Sample: 10 microliters of methanol containing 0.5 nanomole of each of 19 Pth amino acids.

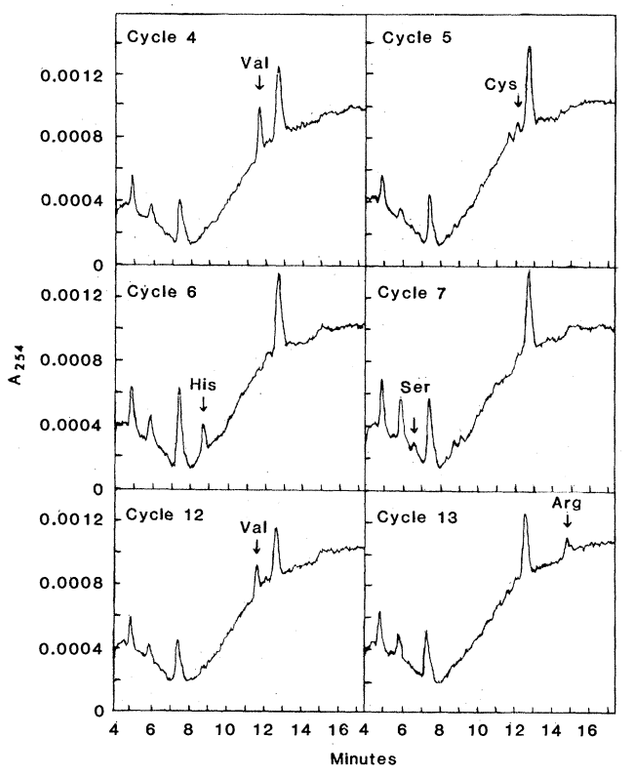

Fig. 5. Representative HPLC traces from gas phase sequenator analysis of 15 pmole of a tridecapeptide. The HPLC conditions were as described (Fig. 4). The traces are not corrected for background from either the HPLC system (the sloping rise and fall of the baseline) or the sequenator (peaks at 4.8 and 5.8 minutes from dithiothreitol, at 6.3 minutes from phenylthiocarbamyl dimethylamine, and 12.4 minutes from diphenylthiourea). The phenylthiocarbamyl dimethylamine came from slight contamination of the particular batch of trimethylamine used as the coupling buffer for this sequenator run.

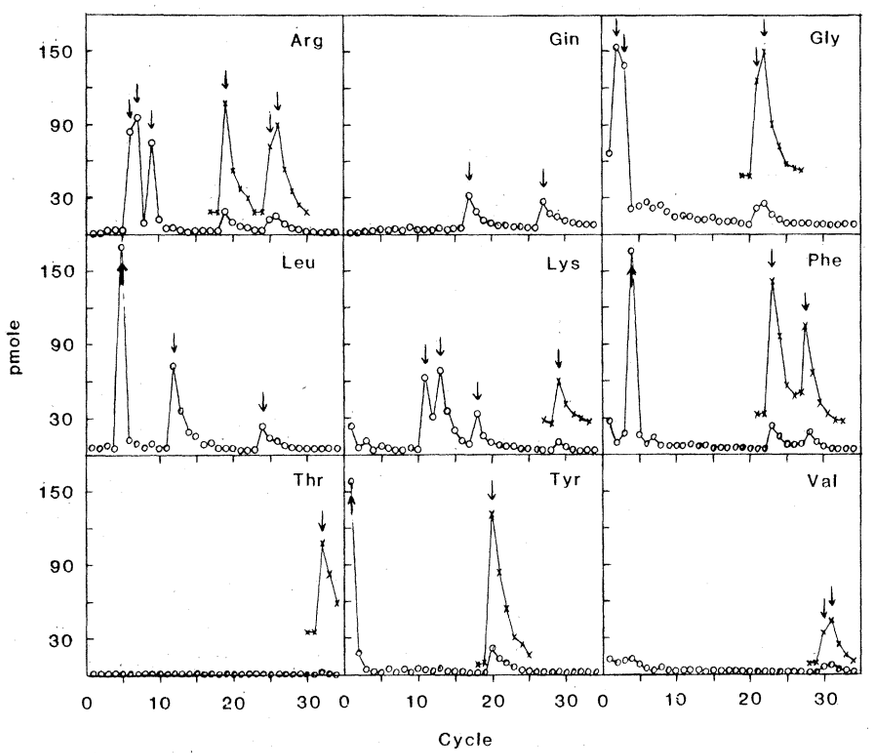

Fig. 6. Yields of selected Pth from analysis of 200 pmole of dynorphin (1–32) on the gas phase sequenator. Portions of the flask extract from each cycle were analyzed by HPLC, peaks were identified and quantified by comparison of peak positions and heights with values for a standard Pth amino acid mixture, and yields were normalized to 100 percent injection. The preparation of the sample and the sequenator analysis were reported by Fischli *et al.* (*37*). Abbreviations: Arg, arginine; Gln, glutamine; Gly, glycine; Leu, leucine; Lys, lysine; Phe, phenylalanine; Thr, threonine, Tyr, Tyrosine; Val, valine.

amide gel electrophoresis requires more effort to prevent damage to or loss of the sample, but electroelution of proteins, even after staining with Coomassie blue, can be used to prepare as little as 1 to 2 micrograms of protein for sequence analysis (*28*). Since many sequencing experiments (for example, those with proteins containing blocked amino-terminal residues) require generation and separation of peptides, the HPLC and gel techniques may be used several times on a given sample before the desired sequence is obtained. This can increase the amount of starting samples required for analysis by a factor of 10 to 20.

Applications

Complementary DNA cloning. With the advent of rapid DNA sequencing, the role of protein sequencing by Edman degradation has changed. In most cases, complete sequence analysis of large proteins can be obtained much more efficiently by DNA sequence analysis if the DNA encoding the protein can be cloned and identified. With genes that produce small quantities of mRNA (rare-message genes) it is often difficult to obtain complementary DNA (cDNA) probes for subsequent gene cloning by the standard techniques used for frequent-message genes such as antibodies and hemoglobins. An alternative approach is to determine the partial amino acid sequence of a small amount of the desired gene product and then to search through these protein sequence data for stretches of five to seven contiguous amino acids that have minimal ambiguity when translated by the genetic code dictionary into DNA language (*29*). With recent improvements in DNA synthesis (*30, 31*), one can generate mixtures of DNA probes consisting of 14 to 20 bases that cover all possibilities for a five– to seven–amino acid sequence obtained by microsequence analysis. These synthetic DNA probes can then be used as probes to screen cDNA libraries directly or as primers to generate specific probes. Even with a mixture of as many as 32 related oligonucleotides (15 bases in length), one can achieve specific hybridization of the correct oligonucleotide to the desired DNA sequence (*32*). Thus, protein sequencing can provide a powerful approach to the cloning of rare-message genes.

Synthetic polypeptides. An emerging tool of molecular biology is the use of synthetic peptides based on protein or DNA sequences to generate antibodies that recognize the parent protein (*33,*

34). The work of Bittle, Lerner, and their colleagues with peptides from foot and mouth disease virus (*35*) illustrates the power of the technique and points to the potential development of important vaccines. A second use of fundamental importance will be the identification of biologically active polypeptides that must be purified 10^4-fold or more. After sequence analysis, one must always ask whether the sequence is derived from the peptide with appropriate biological activity, or whether it represents a copurifying contaminant. After microsequencing, peptides can be synthesized and used as immunogens, and the resulting antibodies can be used to verify the biological activity of the protein and, if appropriate, to purify larger amounts of the protein.

Other applications. Although DNA sequencing can give the linear peptide sequence of the primary product of translation, it cannot answer questions of posttranslation proteolytic processing or side chain modification (that is, removal of signal peptides, carbohydrate attachment, and phosphorylation). Subtle changes in protein structure, such as the attachment of methionine to the amino terminus of a protein produced by recombinant DNA technology, are often almost impossible to detect and quantify by any method other than protein sequence analysis. Direct protein sequencing also can answer questions of stoichiometry of multisubunit protein complexes that cannot be addressed by DNA sequencing. It may be the method of choice for such studies because of the minimum number of assumptions and manipulations required in data analysis (*36*). Protein sequencing also is becoming an important method for identifying protein or peptide fractions and, perhaps more important, assessing their purity.

The Future

In a variety of ways protein sequencing can provide useful information about the structures of functional proteins and peptides. With only an additional slight increase in the sensitivity of current microsequencing methods to the subpicomole level, one could employ one of the most powerful analytical separation procedures in modern biochemistry, two-dimensional gel electrophoresis, as a single-step preparative method for obtaining proteins available in very small quantities. Direct microsequencing at this subpicomole level would provide even more striking new opportunities to analyze new systems in developmental biology and neurobiology.

References and Notes

1. P. Edman, *Acta Chem. Scand.* **10**, 761 (1956).
2. Anonymous, *Nature (London)* **275**, 90 (1978).
3. G. Tarr, in *Methods in Protein Sequence Analysis*, M. Elzinga, Ed. (Humana Press, Clifton, N.J., 1982), p. 223.
4. J. Y. Chang, D. Brauer, B. Wittmann-Liebold, *FEBS Lett.* **93**, 205 (1978).
5. P. Edman and G. Begg, *Eur. J. Biochem.* **1**, 80 (1967).
6. B. Wittmann-Liebold, H. Graffunder, H. Kohls, *Anal. Biochem.* **75**, 621 (1976).
7. G. E. Tarr, J. F. Beecher, M. Bell, D. J. McKean, *ibid.* **84**, 622 (1978).
8. R. A. Laursen, *Eur. J. Biochem.* **20**, 89 (1971).
9. J. E. Walker *et al.*, *ibid.* **123**, 253 (1982).
10. R. M. Hewick, M. W. Hunkapiller, L. E. Hood, W. J. Dreyer, *J. Biol. Chem.* **256**, 7990 (1981).
11. J. J. Pisano, T. J. Bronzert, H. B. Brewer, Jr., *Anal. Biochem.* **45**, 43 (1972).
12. B. Wittmann-Liebold, A. W. Geissler, E. Marzinzig, *J. Supramol. Struct.* **3**, 426 (1975).
13. O. Smithies, D. M. Gibson, E. M. Fanning, R. M. Goodfliesch, J. G. Gilman, D. L. Ballantyne, *Biochemistry* **10**, 4912 (1971); E. Mendez and C. Y. Lai, *Anal. Biochem.* **68**, 47 (1975).
14. C. L. Zimmerman, E. Appella, J. J. Pisano, *ibid.* **77**, 569 (1977).
15. S. Oroszlan, T. Copeland, M. Summers, G. Smythers, in *Solid Phase Methods in Protein Sequence Analysis*, R. A. Laursen, Ed. (Pierce Chemical Company Press, Rockford, Ill., 1975), p. 179.
16. J. W. Jacobs and H. D. Niall, *J. Biol. Chem.* **250**, 3629 (1975).
17. J. Bridgen, *FEBS Lett.* **50**, 159 (1975).
18. J. Silver and L. Hood, *Contemp. Top. Mol. Immunol.* **5**, 35 (1976).
19. B. Ballou, D. J. McKean, E. F. Freedlender, O. Smithies, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 4487 (1976).
20. Y. Burstein and I. Schecter, *ibid.* **74**, 716 (1977).
21. J. Silver and L. Hood, *ibid.* **73**, 599 (1976).
22. J. E. Coligan, T. J. Kindt, H. Uehara, J. Martinko, S. G. Nathenson, *Nature (London)* **291**, 35 (1981).
23. M. W. Hunkapiller and L. E. Hood, *Biochemistry* **17**, 2124 (1978).
24. ———, *Science* **207**, 523 (1980).
25. B. Wittmann-Liebold, *Hoppe-Seyler's Z. Physiol. Chem.* **354**, 1415 (1973).
26. W. C. Mahoney and M. A. Hermodson, *J. Biol. Chem.* **255**, 1199 (1980).
27. K. J. Wilson, M. W. Berchtold, P. Zumstein, S. Klauser, G. J. Hughes, in *Methods in Protein Sequence Analysis*, M. Elzinga, Ed. (Humana Press, Clifton, N.J., 1982), p. 401.
28. M. W. Hunkapiller, E. Lujan, F. Ostrander, L. E. Hood, *Methods Enzymol.*, in press.
29. B. Noyes, M. Mevarech, R. Stein, K. Agarwal, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1770 (1979).
30. M. D. Mattucci and M. H. Caruthers, *J. Am. Chem. Soc.* **103**, 3185 (1981).
31. S. L. Beaucage and M. H. Caruthers, *Tetrahedron Lett.* **22**, 1859 (1981).
32. M. Steinmetz *et al.*, *Nature (London)* **300**, 35 (1982).
33. G. Walter, K.-H. Scheidtmann, A. Carbone, A. P. Laudano, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5197 (1980).
34. J. G. Sutcliffe, T. M. Shinnick, N. Green, F.-T. Liu, H. L. Niman, R. A. Lerner, *Nature (London)* **287**, 801 (1980).
35. J. L. Bittle, R. A. Houghten, H. Alexander, T. M. Shinnick, J. G. Sutcliffe, R. A. Learner, D. J. Rowlands, F. Brown, *ibid.* **298**, 30 (1982).
36. M. A. Raftery, M. W. Hunkapiller, C. D. Strader, L. E. Hood, *Science* **208**, 1454 (1980).
37. W. Fischli, A. Goldstein, M. W. Hunkapiller, L. E. Hood, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5435 (1982).

38. This work was supported by the Weingart Foundation, National Science Foundation Grant PCM 80-05999, and National Institutes of Health Grant GM 06965.