Automated Sequence Analysis of Proteins and Peptides

Hugh D. Niall

Theoretical and practical advances in the chemistry of the phenyl isothiocyanate degradation of proteins by P. Edman and G. Begg (1967) have resulted in the development of an instrument, the protein Sequenator, capable of carrying out the coupling and cleavage reactions at close to 100% efficiency. With this procedure amino-terminal sequences of 40–80 residues have been determined on a number of proteins. Shorter peptides, however, cannot be readily analyzed by this original method, since there is considerable loss of material in the organic solvents used for extraction at each cycle. Recently an automated approach has been devised which allows complete degradations to be carried out on submicromolar amounts of peptides of chain length 2–32 residues. It depends on the use of volatile reagents with flexible programming of solvent extractions to minimize peptide loss as the carboxyl terminus is approached. A complete cycle of degradation may be carried out in less than 60 min.

A formidable array of methods is now available for separation and isolation of proteins; some of the more recent and sophisticated approaches are reviewed elsewhere in this symposium. In general, it is possible to purify almost any protein provided it possesses some specific chemical or biological property which can be quantitated and which is not significantly destroyed in the process of isolation. Until recently, however, successful purification of a protein was a mere prelude to a lengthy and difficult sequence determination. This situation has now been altered, chiefly through the work of Edman and Begg in developing and automating the phenyl isothiocyanate (PITC) degradation (Edman and Begg, 1967). The major portion of a protein sequence may now be completed within a matter of weeks. Automated methods more suitable for shorter peptide chains have subsequently been devised (Niall *et al.*, 1969).

In this review only procedures based upon the Edman degradation will be considered. An excellent description of the principles of the PITC degradation with many important experimental details has recently been published (Edman, 1970). Alternate methods for automated sequence analysis (for example by mass spectrometry) are being actively studied in a number of laboratories. However, at present, these procedures seem far from becoming generally available and are useful only for certain limited applications.

THE EDMAN DEGRADATION

The single most important method for sequence analysis of proteins was discovered in 1950 by Pehr Edman, then at the University of Lund in Sweden (Edman, 1950). The method is based upon the reaction of the reagent PITC with the α-amino group of the protein or peptide to form a phenylthiocarbamyl adduct. Under anhydrous acidic conditions the N-terminal amino acid is then cleaved as an anilinothiazolinone derivative (Edman, 1956). This is separated from the peptide chain by a solvent extraction step and the identity of the cleaved amino acid established by one of a variety of means. The whole procedure is then repeated on the shortened peptide chain, the second amino acid being coupled with PITC and cleaved in the same way (Figure 1). By successive cycles of this degradation, the sequence is determined.

Endocrine Unit, Massachusetts General Hospital, Boston, Mass. 02114

The basic Edman procedure in its original form or modifications of it have been extensively applied in the structural analysis of proteins and peptides over the last two decades. For reasons to be discussed, most degradations were limited to a few cycles, the structure of the proteins being established by compositional analysis and partial sequencing of large numbers of short peptides produced by two or more different cleavage procedures.

More recently, fundamental investigation by Edman of the reaction mechanisms involved has led both to considerable improvement in the results of the degradation as carried out manually and to its successful automation (Edman, 1957, 1960). To illustrate the progress in this field, the problems encountered by early workers trying to apply the method will be considered in some detail. These problems commonly resulted in rapidly falling yields of the thiohydantoin derivatives and incompleteness of reaction.

PROBLEMS IN THE COUPLING REACTION

Insolubility. Both the protein (or the peptide) and the coupling reagent should be soluble in the medium used. Since proteins are largely polar and water-soluble and the coupling reagent (PITC) is water-insoluble, a mixture of an organic solvent (usually pyridine or propanol) with water represents a suitable compromise. Some proteins, particularly those of high molecular weight, are insoluble or poorly soluble in these media. Other proteins, initially soluble, become harder to dissolve as the degradation proceeds. This is sometimes seen with native proteins containing intact disulfide bridges, and may be due to polymerization *via* disulfide interchange reactions (Edman, 1960). Prior performic acid oxidation or reduction and alkylation of the protein may prevent this difficulty. At times the coupling reaction seems to proceed to completion even though the protein appears to dissolve only partially or not at all in the coupling mixture. Presumably this represents a heterogeneous reaction between the protein in the solid phase and the PITC in solution. The protein must swell sufficiently in the solution to form a gel-like matrix for penetration by the coupling reagent.

It is much more desirable, however, to have the protein in solution. This is aided by dispersal of the protein in a finely divided form. In the manual degradation the protein tends to form an adherent film after the cleavage step. It may be dispersed effectively prior to the addition of coupling buffer for the next cycle by suspending it in a polar solvent (ethyl acetate) and grinding it with a glass rod. In an early modi-

fication of the Edman procedure, the protein was dispersed on a sheet of filter paper for the duration of the degradation (Frankel-Conrat, 1954). In the automated sequenator method, the protein dissolves readily because it is spread as a thin film on the wall of the spinning glass cup, as described below.

Very insoluble proteins (for example, fibrinogen or fibrin derivatives) may be coupled in solutions containing 40% urea (Blomback and Yamashina, 1958). The protein is precipitated and the urea extracted after coupling by aqueous acetone. Solubility is aided by carrying out the reaction at a moderately elevated temperature (40–50° C).

Poor Control of pH. If the pH of the coupling medium is too high (>10), destruction of the PITC becomes the predominant reaction. The coupling is incomplete and the breakdown products of the PITC interfere with identification of the phenylthiohydantoin derivatives. If the pH falls below 8.5, the α-amino group becomes largely protonated. Since in this form it will not react with the PITC, the coupling reaction proceeds very slowly. The suitable pH range for the reaction (9.0–9.5) may be maintained by periodic additions of alkali (since the reaction consumes alkali) or by use of a base which buffers at this pH. Primary and secondary amines are unsuitable since they will react with and consume the PITC. A variety of tertiary amines may be used. Those which possess an allyl group have been found suitable (Niall and Edman, 1962; Blomback *et al.*, 1966). We have used dimethylallylamine, diethylallylamine, and *N*-allyl piperidine successfully for different applications (Sauer *et al.*, 1970). Recently we have found that 1,4-diazabicyclo[2.2.2]octane is a most effective buffer and may be used without special purification (Niall *et al.*, 1971).

Amino Terminal Blocking Reactions. These fall into two categories. Extrinsic blocking reactions are due to impurities in the reagents. Though there is no doubt that use of unpurified or inadequately purified reagents leads to amino terminal blocking reactions which make the peptide unavailable for further degradation, there is little evidence for the exact nature of these impurities. There are probably several independent mechanisms. Attention has been particularly directed at the possibility of trace quantities of aldehydes forming Schiff bases with the α-amino group (Edman and Begg, 1967). Purification procedures which reduce the level of aldehydes in the reagents are associated with improved repetitive yields. However, reagents which give negative tests for aldehydes may still give poor results, so other mechanisms must also be involved.

A most important side reaction leading to amino terminal blocking is oxidative desulfuration of the phenylthiocarbamyl (coupled) protein to the phenylcarbamyl derivatives (Ilse and Edman, 1963). The latter product, as discovered many years ago, cyclizes only under extremely rigorous conditions to form the phenylhydantoin of the terminal amino acid (Bergmann and Miekeley, 1927). This does not occur under the usual conditions for Edman degradation. The oxidation can be due either to dissolved oxygen in the reagents or to any oxidant present as an impurity. The oxidation may take place during the actual coupling or during the subsequent solvent washes and vacuum stages up to the cleavage. Oxidation after the cleavage may lead to partial or complete breakdown of the thiazolinone or thiohydantoin but will not affect the efficiency of the repetitive removal of amino acids from the protein chain.

Intrinsic blocking reactions are due to alterations in the covalent structure of the polypeptide chain. For example,

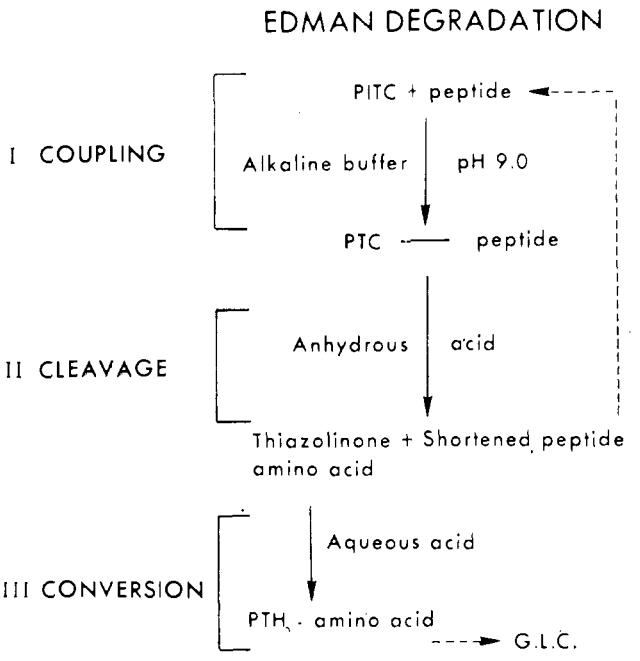

Figure 1. A simplified outline of the three-stage Edman degradation (Edman, 1960): PITC, phenyl isothiocyanate; PTC, phenylthiocarbamyl; PTH, phenylthiohydantoin; glc, gas–liquid chromatography

amino terminal glutamine may cyclize to form a pyroglutamyl ring. N–O acyl shifts involving hydroxy amino acids and α–β aspartyl shifts are other intrinsic blocking mechanisms which may partly or completely halt the progress of the degradation.

Some blocking reactions are potentially reversible. For example, a blocking agent present on the α-amino group during coupling may be labile to the acidic conditions of cleavage. This could cause asynchrony in the degradation, since the proportion of the protein which had been blocked could couple and cleave normally at the next cycle, out of phase with the rest of the material. There is some reason to believe that carbon dioxide may represent such a reversible blocking group (Hirschmann *et al.*, 1967).

In the absence of solubility problems or blocking reactions, and with adequate control of pH, the coupling reaction appears to go to completion within 15 to 20 min at 50° C, in the presence of excess PITC.

PROBLEMS IN THE SOLVENT EXTRACTIONS

Extractions with an organic solvent are necessary after coupling (to remove excess reagent and by-products and to precipitate the protein) and after cleavage (to separate the cleaved thiazolinone from the residual protein). The only major problem at these steps has been extractive loss of material when degradations are attempted on shorter peptides, since these have appreciable solubility in the organic solvents. These losses can be reduced though not eliminated by suitable choice of solvent and by reduction in the volumes used for extraction. Adequate purification of reagents also helps to reduce the need for extraction. Reduction in the scale of the degradation is also helpful; by suitable micromodifications of the Edman procedure, degradations may be carried out on proteins as well as peptides on only a few nanomoles of material (Niall *et al.*, 1971).

PROBLEMS IN THE CLEAVAGE REACTION

The use of fluoroacids for cleavage (Elmore and Toseland, 1956; Edman, 1957) has overcome many earlier difficulties.

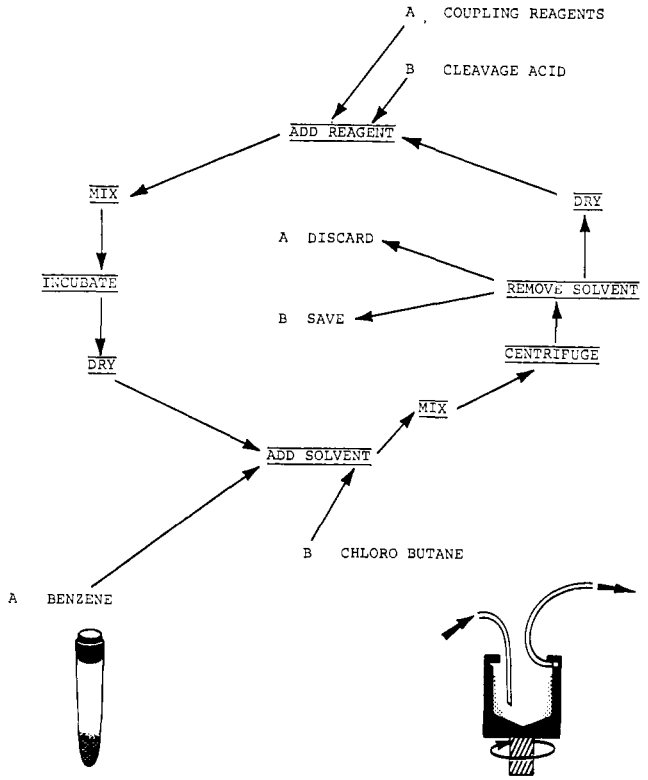

Figure 2. An operational analysis of the Edman degradation. Note that an actual degradation cycle includes two very similar operational cycles; A, coupling; B, cleavage. In the degradation carried out in a test tube (lower left of figure) these operations are diverse. Simplicity and ease of automation results from use of a spinning cup for the reaction (lower right)

Trifluoroacetic acid or its higher analog, heptafluorobutyric acid, are excellent solvents for proteins and cause few side reactions. A small degree of peptide bond cleavage along the chain seems to occur concomitantly with the cleavage of the terminal amino acid (Edman and Begg, 1967). The mechanism for this cleavage, described as nonspecific, is uncertain. There is insufficient evidence at present to decide whether, in fact, it is truly nonspecific, or whether certain peptide bonds are especially labile. In any case the gradual accumulation of false end groups due to this cleavage process is an important factor in termination of the degradation, since the true sequence can no longer be deciphered from the background of other thiohydantoin derivatives.

Incomplete cleavage has been reported for aspartic acid and glutamic acid residues. It has been suggested that an equilibrium exists in the cleavage reaction (Edman and Begg, 1967). Addition of a second cleavage overcomes this problem and reduces asynchrony from this cause to very small proportions.

PROBLEMS IN THE CONVERSION REACTION

Ilse and Edman (1963) described conditions suitable for routine conversion of the thiazolinones of all amino acids to phenylthiohydantoins. These conditions necessarily represented a compromise. The serine and tryptophan derivatives partly decompose to unknown products. Conversion of the glycine derivative is incomplete (about 70%); asparagine and glutamine residues are partially deamidated. However, conditions have been described which minimize these side reactions if it is important to obtain higher yields (Ilse and Edman, 1963; Edman and Lauber, 1956; Ingram, 1953) of these derivatives.

In some modifications of the Edman degradation (Gray and Hartley, 1963; Hirs *et al.*, 1960) direct identification of the phenylthiohydantoin is not employed and hence the conversion step is omitted.

PROBLEMS OF IDENTIFICATION

Identification procedures may be based upon direct analysis of the phenylthiohydantoin derivatives by thin-layer chromatography (Edman and Begg, 1967) or gas–liquid chromatography (Pisano and Bronzert, 1969; Niall and Potts, 1970). Alternatively the thiazolinone or the thiohydantoin derivative may be hydrolyzed in strong acid or alkali to give the free amino acid which is identified and quantitated by standard procedures. Indirect methods of identification (for example the dansylation procedure, or the "subtractive" Edman degradation) are applicable mainly to manual procedures and will probably not find extensive use as automated procedures become more widespread (Gray and Hartley, 1963; Hirs *et al.*, 1960).

The numerous problems which may be encountered during these various identification procedures are beyond the scope of this review. In general if the coupling and cleavage reactions are proceeding quantitatively or nearly so, there is no difficulty in identifying the cleaved residues. Since no single identification approach is free of ambiguities, it is desirable that facilities be available for at least two different identification systems.

AUTOMATION OF THE EDMAN DEGRADATION: THE PROTEIN SEQUENATOR

Prior to about 1960 there would have been little point in attempting to automate the Edman degradation, since the chemistry was poorly understood and repetitive yields were low so that only short degradations were possible. However, by 1960, most of the problems outlined above had been identified and eliminated by Edman and his collaborators and, using the three-stage manual procedure (Edman, 1960), quite extensive degradations (up to 20 cycles) were carried out at 90–95% repetitive yield. At this point automation became a realistic possibility.

The operations required to carry out the manual degradation are rather diverse; they include pipetting, centrifuging, extracting with organic solvents, and vacuum drying, as well as the incubations required for the coupling and cleavage steps (Figure 2). The procedure could probably have been automated or, more correctly, mechanized essentially as it was; however, it would not have been a very simple or appealing system.

Edman and Begg (1967) introduced a most novel concept in their highly successful automated system. The entire degradation is carried out with the protein immobilized as a thin film on the wall of a continuously spinning cylindrical glass vessel (Figure 2). The solvent and reagent additions are made *via* a thin Teflon line which delivers the liquid to the floor of the cup. Centrifugal force spreads the liquid as a film over the surface of the protein for reaction during coupling and cleavage. The contents of the cup are partly or completely reduced to dryness after the coupling and cleavage steps by application of a vacuum. Solvent extractions are carried out by delivering a continuous flow of solvent into the cup. The solvent rises over the surface of the protein film and flows into a horizontal groove in the top of the cup (Figure 2). There it is scooped off into a tangential Teflon line, whence it flows to a fraction collector (for collection of the amino acid derivatives) or to a waste container.

The detailed design and performance of the instrument, called by Edman and Begg the "protein sequenator," are found in their paper (Edman and Begg, 1967). Instruments based essentially upon the design of Edman and Begg are now functioning in many laboratories.

It should be stressed that the engineering and chemical problems involved in this kind of instrumentation are not by any means simple. The toxicity and explosive properties of some of the chemicals require careful attention to safety features in the design. The need to exclude essentially all atmospheric oxygen and to avoid reagent contamination imposes further limitations. Consideration of such engineering details, however, is beyond the scope of this review.

THE PROTEIN SEQUENATOR: APPLICATIONS

In their original paper Edman and Begg described the removal and identification of the amino terminal 60 residues of humpback whale myoglobin. Subsequent applications of their instrument established amino-terminal sequences on human immunoglobulin light chains (Niall and Edman, 1967), human placental lactogen (Catt *et al.*, 1967), cyanogen bromide fragments of streptokinase (Morgan and Henschen, 1969), and scorpion neurotoxins (Rochat *et al.*, 1970). More recently, investigators in other laboratories have studied particularly immunoglobulin sequences.

Immunoglobulin light and heavy chains have been isolated from myeloma proteins, from normal γ-globulin, and from antibody fractions of limited heterogeneity. This kind of study is starting to provide information concerning the molecular basis of the antigen–antibody reaction and, in more detail, concerning the genetic interrelationships of different classes of immunoglobulin (Jaton *et al.*, 1970; Hood *et al.*, 1970; Capra and Kunkel, 1970; Terry and Ohms, 1970). Partial sequences have been obtained on a number of other molecules, including viral proteins, polypeptide hormones, and hemoglobins (Niall *et al.*, 1970; Riggs, 1970). Degradations have varied in length from 20–80 amino acids, depending mainly on the length of the polypeptide chain. The longer the chain, the greater is the opportunity for internal peptide bond cleavages and the more rapid is the buildup of background relative to the yield of the main sequence. So far no large protein has been completely sequenced by automated procedures; the largest polypeptide completely analyzed in the Sequenator is bovine parathyroid hormone, which contains 84 residues.

The amino terminal 54 residues were identified through a degradation on the intact hormone. The five arginine residues in the molecule were found at positions 20, 25, 44, 50, and 52. This suggested an approach for rapid completion of the sequence. The parathyroid hormone was succinylated to block the ϵ-amino groups of lysine and digested with trypsin. Since no trypsin sensitive site remained carboxyl terminal to the arginine at position 52, the fragment 53–84 was easily isolated in high yield. An automated degradation of 31 cycles (using the peptide methodology to be described below) was performed. The carboxyl terminal residue, glutamine, was identified directly by amino acid analysis. Thus the complete sequence was established by two automated degradations (Niall *et al.*, 1970).

This approach illustrates a general strategy for complete sequence analysis of proteins. An initial degradation is performed on the whole molecule. The information obtained, taken with the known amino acid composition, may suggest a specific cleavage procedure suitable for generating a small number of large fragments. Further degradation can be carried out on these isolated fragments. As with earlier approaches, sufficient overlaps must then be obtained through an alternate cleavage procedure to align these fragments.

In summary, three main areas of application of the protein Sequenator have been delineated so far: extensive degradations from the amino terminus of purified isolated proteins as the initial step in total structural analysis (Edman and Begg, 1967); comparative amino terminal sequence studies on a group of closely related proteins to define their genetic or functional relationships, as in the immunoglobulin field (Niall and Edman, 1967); and sequence studies on more or less heterogeneous protein mixtures, difficult or impossible to examine by conventional means (Jaton *et al.*, 1970; Niall and Edman, 1967).

THE PEPTIDE SEQUENATOR

An important potential area of application for automated sequencing equipment is the analysis of smaller peptides (5–50 residues). Though peptide degradations can be carried out in instruments of design similar to the one originally described (Edman and Begg, 1967), difficulties arise with peptides which have a shorter chain length or which consist mainly of hydrophobic amino acids. This limitation is important when one considers that many biologically interesting peptides are of low molecular weight. Again, in completing the amino acid sequence of a protein, the usual chemical and enzymatic cleavage procedures generate small as well as larger fragments. It would be desirable and economical of material and effort to be able to sequence these by automated procedures. The main problem in applying the automated procedures as developed for proteins to degradation of smaller peptides lies in the greater solubility of the peptides in the organic solvents used for the extraction steps. This leads to loss of material at each cycle, rapidly ending the degradation. The extractive loss increases very sharply as the residual peptide chain is reduced to a length corresponding to 10–15 amino acids.

These problems in automated peptide degradation may be approached by several different means. The extractive loss of peptide may be prevented by its covalent attachment to an insoluble polymer (Laursen, 1966). Though a number of difficulties remain, progress has been made in improving this procedure, which has recently been automated (Laursen, 1970). An alternative method (Stark, 1965; Dowling and Stark, 1965) involves the use of a solid phase isothiocyanate as the coupling reagent. After cyclization the amino terminal residue remains attached to the polymer, while the shortened peptide is eluted and an aliquot subjected to amino acid analysis. The main problem with this procedure has been incomplete coupling to the solid phase isothiocyanate due to a competing reaction of uncertain nature. This leads to poor repetitive yields which have so far limited the application of this rather novel approach.

In the application of the Sequenator to automated degradation of peptides, the following considerations must be borne in mind. Most of the need for extraction with organic solvents arises from the use of the polar, nonvolatile organic base *N,N,N',N'*-tetrakis(2-hydroxypropyl)ethylenediamine (Quadrol, Wyandotte Chemicals Corp., Wyandotte, Mich.) in the coupling buffer. The polarity of Quadrol makes it an excellent protein solvent, presumably because its hydroxyl groups are available for forming hydrogen bonds with constituents on the polypeptide chain. Its lack of volatility is a most desirable feature for use in Sequenators of design

similar to that of Edman and Begg. This is because the large vapor space surrounding the reaction cell tends to result in evaporation of volatile reagents from the cup. In the case of the tertiary amine used to buffer during coupling, this evaporation in turn may result in a fall in pH of the coupling mixture and hence in incomplete reaction. Use of volatile tertiary amines in large vapor space Sequenators also may result in troublesome accumulations of amine salts throughout the system. It has been reported (Nolan, 1969) that certain conservative design changes will allow the use of volatile reagents while retaining a relatively large vapor space. Further details of this work will be of great interest.

Our own approach toward automated peptide sequencing was based on the development of an instrumental design which would allow the use of volatile as well as nonvolatile reagents (Niall *et al.*, 1969). This instrument is now represented by the Beckman "Sequencer," Model 890.

It includes a reaction cell module in which the vapor space has been greatly reduced, a magnetic drive, and a heating system based on circulating air instead of the electroconductive layer described by Edman and Begg. In this instrument protein degradations can be carried out using the Quadrol system. In addition we have evaluated a variety of approaches to automated peptide degradation using this system. Detailed descriptions of this work are in preparation. However, the principles involved will be described briefly here.

AUTOMATED PEPTIDE DEGRADATION

Systematic examination of the peptide content of the various organic solvent extractions has been carried out by collecting individual extractions during a degradation and subjecting them to quantitative amino acid analysis. This showed that very little loss of material occurred during benzene extraction. However, appreciable quantities of shorter peptides may be lost in the ethyl acetate extraction after coupling and, much more markedly, in the 1-chlorobutane extraction after cleavage.

Omission of the ethyl acetate extraction is possible if a less polar tertiary amine is substituted for Quadrol. As indicated above, the design of the protein/peptide Sequenator (the Beckman "Sequencer") is such that it is immaterial whether this amine is volatile or nonvolatile. A wide range of amines is potentially useful. These include aliphatic, heterocyclic, and aromatic tertiary monoamines and polyamines; members of each of these groups are available which buffer in the correct pH range. Volatility may vary from trimethylamine (bp 3° C) to certain polyamines which are virtually nonvolatile. In practice amines which boil below 60° C are not useful since the temperature of coupling is usually 50–55° C, and evaporative losses are too great. Selected amines representing a wide range of volatility are shown in Table I.

Removal of the less volatile amines after the coupling step depends upon their solubility in benzene. Amine salts are also poorly volatile, even those derived from amines which are themselves quite volatile. These also must be largely removed by solvent extraction. In theory one would choose an amine carrying a bulky long chain alkyl, heterocyclic, or aromatic substituent to facilitate its extraction by benzene. We have examined several such amines, including *N,N'*-dimethylpiperazine, 1,4-diazabicyclo[2.2.2]octane, and 4-dimethylaminopyridine. These three amines are soluble in benzene and buffer very strongly in the correct pH range for coupling. However, their removal by benzene is less efficient than might be imagined since the matrix of the dried polar peptide is poorly penetrated by nonpolar solvents. This is evidenced by the inefficient extraction by benzene of the thiazolinones of nonpolar amino acids (*e.g.*, valine or leucine) when this solvent is substituted for 1-chlorobutane after cleavage (Niall *et al.*, 1971). Though evaluation of several alternative amines is proceeding, at present our findings suggest that a volatile amine is preferable since most of it can be removed simply by vacuum. Small amounts of residual amine salt are largely extractable by benzene with the qualification mentioned above that the solvent probably does not completely penetrate the substance of the peptide film. With a volatile amine it is also possible to omit the benzene extraction completely for a few cycles. This is helpful for the last few residues of a hydrophobic peptide.

Table I. Volatility of Selected Amines

Amine	Bp (°C)
Ethyldimethylamine	38
Dimethylallylamine	64
Triethylamine	89.5
Diethylallylamine	110–113
Dipropylallylamine	153
1,3-Bis(dimethylamino)butane	160
Heptyldiethylamine	200
1,1,1-Tris(2-dimethylaminoethyl)ethane	112 at 4 mm

Of the amines evaluated we have found dimethylallylamine and diethylallylamine to be the most useful. The former is used as a 1 *M* solution in 1-propanol–water (3:2, v/v) without pH adjustment. The latter is used as a 0.3 *M* solution in the same propanol/water mixture.

The small vapor space surrounding the cup prevents undue evaporation of these tertiary amines; hence pH is maintained in the alkaline range necessary for coupling. After coupling the constituents of the buffer are removed by a vacuum step usually followed by a brief extraction with benzene. Peptide losses are minimal. After a second vacuum drying step, a single cleavage is performed, and the thiazolinone extracted with a small volume of 1-chlorobutane. (The cycle of degradation may be completed in 50–70 min.) A number of changes in the program are necessary to prevent physical displacement of the peptide film up the wall of the cup. With a program similar to this it was possible to carry out a complete (20-cycle) degradation on the insulin "A" chain. The final residue (asparagine) was identified as the free amino acid by amino acid analysis. Degradation of 25–30 cycles has been carried out on the hydrophobic 32 amino acid peptide hormone calcitonin. The complete sequence of ovine calcitonin has been established using these approaches (Sauer *et al.*, 1970). A variety of smaller peptides (4–12 residues) have also been completely sequenced (Niall *et al.*, 1971).

Since peptides show great variability in size and polarity, there is considerable need for a varied and flexible methodology for automated peptide degradation. A series of programs have been developed showing a gradation from those suitable for proteins and longer polypeptides to those suitable for a short hydrophobic tetrapeptide (Niall *et al.*, 1971). Since the properties of the peptide may vary during the actual degradation as its length becomes shorter and it loses various polar or nonpolar amino acids, it is evident that more than one program will often be necessary in the sequencing of a single peptide.

For example, let us consider a hypothetical degradation of a peptide of 50 residues, containing both acidic and basic residues. The acidic residues will be ionized at the pH of coupling; hence during the extractions after coupling there will be several charged groups (α-, β- and γ-carboxyl groups). This means that initial extractive losses into the organic solvents used after coupling will not be significant. After cleavage, the acidic functions will be suppressed by the extremely strong fluoroacid; hence they will be uncharged and of no help in minimizing extractive losses at this step.

The influence of basic groups in the peptide is complex. Lysine residues make the peptide more hydrophobic since after the first coupling cycle the ϵ-amino groups are substituted with bulky phenylthiocarbamyl groups. Hence the peptide becomes more soluble in the extracting solvents and the loss of material is correspondingly increased. If the basic groups are arginine side chains, such losses are very small since the guanidino group does not react with PITC and is so basic that it remains positively charged throughout the degradation. The influence of histidine is harder to predict; the imidazole function probably is capable of reacting with PITC, though direct evidence for this has not been obtained.

In the hypothetical degradation the following kind of reasoning would apply in choosing a program. The peptide contains amino acids which will carry positive or negative charges during the solvent extractions both after coupling and after cleavage. This fact, taken with the size of the peptide (50 residues), means that regardless of the program extractive losses initially will not be great. Hence one should choose a program which will give the most efficient degradation with the minimum of overlap. A reasonable decision would be to use a Quadrol program. In protein programs 1 *M* Quadrol is used. This high concentration is useful because of the protein solvent properties of Quadrol mentioned above and because a high concentration of base is necessary to neutralize acid which becomes bound to the protein during cleavage. However, neither of these reasons usually applies to a peptide of 50 amino acids. Hence more dilute Quadrol may be used, a concentration of 0.1–0.2 *M* being suitable. Since less Quadrol is used, the ethyl acetate and benzene extractions after coupling may be somewhat reduced. A double cleavage should be used in these initial cycles of degradation to reduce overlap (asynchrony).

This program has a number of advantages. Quadrol is available in highly purified form for protein degradations. The ethyl acetate extractions necessary to remove the Quadrol have the additional merit of removing polar impurities (*e.g.*, diphenylthiourea) which result from side reactions of the PITC. The double cleavage keeps overlap to low levels so that when the program is switched to a single cleavage mode as the peptide shortens, the consequent overlap is more easily tolerated.

This Quadrol program cannot, however, be continued indefinitely since peptide losses will become prohibitive. After 20 cycles or so the degradation should be halted until all the residues to that point have been identified. A program can then be chosen to continue the degradation based upon the polarity of the residual peptide assessed from the composition of the remaining amino acids. If, for example, several polar residues, particularly one or more arginines, still remain, it may be possible to continue the Quadrol program for up to another 15–20 cycles. However, when the residual peptide is 10–15 residues long, when yields of the PTH derivatives start to fall or when few charged groups remain (whichever comes sooner), a program change should be made.

The Quadrol coupling buffer should be replaced by the dimethylallylamine–propanol–water buffer described above. The benzene extraction after coupling is reduced. The ethyl acetate extraction may be completely omitted. With these changes loss of peptide is greatly reduced.

Omission of the second cleavage in this program results in some increase in overlap. However, the degradation should be completed before this increases to the extent of causing ambiguity in identification.

Though we have not yet had occasion to degrade peptides as large as 50 amino acids using this kind of approach, recent success with a 32-residue peptide from bovine parathyroid hormone is encouraging. This peptide (about 0.2 μmol) was degraded to the carboxyl terminal residue using a Quadrol program for six initial cycles followed by 25 cycles using the volatile system (Niall *et al.*, 1970).

In summary our general approach towards automated degradation of peptides could be defined as follows. The Quadrol program, or other program suitable for proteins, may be used unmodified until the residual peptide chain is about 30 residues long. The program should include an ethyl acetate extraction after coupling sufficient to remove Quadrol and polar impurities. Two cleavages should be used to minimize the development of overlap.

As the residual peptide becomes shorter due to its increased solubility in the organic solvents at some point in the range of 15–30 residues, it will be necessary to change to one of several alternative peptide programs. The exact point at which this should be done is decided on the basis of repetitive yield data and on the polarity of the residual peptide, known from its amino acid composition.

A series of peptide programs have been devised to include one or more of the following features: (a) more dilute Quadrol with a shorter ethyl acetate extraction; (b) omission of the second cleavage; (c) use of a peptide mode type of cleavage, with less acid and a longer drying time (Sauer *et al.*, 1970); (d) substitution of a volatile amine or a less volatile but benzene soluble amine for Quadrol (*e.g.*, dimethylallylamine or 4-dimethylaminopyridine), together with complete omission of the ethyl acetate extraction; and (e) omission of the benzene extraction.

Since these programs are usually associated with an increase in overlap, their introduction should not be made prematurely. The most useful combinations are (a),(b),(c) and (b),(c),(d). As a rough guide, the (a),(b),(c) program is suitable while the peptide is in the 15–30 residue size range. The (b),(c),(d) program is suitable for the range 4–15 residues. Omission of the benzene extraction can only be tolerated for a few cycles and should therefore not be introduced until the degradation has about reached the carboxyl terminal residue (2–4 residues). It is most convenient to identify the final residue as the free amino acid, thus verifying that the end of the peptide has indeed been reached (Niall *et al.*, 1970).

Variations on this basic approach may be made on the basis of the polarity of the peptide and the quantity available.

SUMMARY: FUTURE POSSIBILITIES

A revolution in protein and peptide sequencing methods is in progress, stimulated by the theoretical and practical advances of Edman and his collaborators (Edman 1950, 1956, 1960; Edman and Begg, 1967). Though it is not yet possible to carry out continuous degradations of longer

than perhaps 100 cycles, it can be anticipated that even high molecular weight proteins may be readily sequenced using a combination of the protein methodology with more recently developed methods for automated degradation of shorter peptides (Niall *et al.*, 1969, 1970; Sauer *et al.*, 1970).

There is still, however, considerable room for further improvement in the methods. A totally automated system incorporating an on-line identification procedure with direct readout of the amino acid sequence and capable of analysis at the nanomolar level of sensitivity is what is ultimately required. This seems quite feasible with only conservative extrapolation from techniques already available. The rate-limiting factor is likely to be the present restriction of funding for biomedically relevant research rather than to any intrinsic difficulty in the approaches to be used.

LITERATURE CITED

Bergmann, M., Miekeley, A., *Justus Liebigs Ann. Chem.* **458,** 40 (1927).
Blomback, B., Yamashina, I., *Ark. Kemi* **12,** 299 (1958).
Blomback, B., Blomback, M., Edman, P., Hessel, B., *Biochim. Biophys. Acta* **115,** 371 (1966).
Capra, J. D., Kunkel, H. G., *Proc. Nat. Acad. Sci.* **67,** 87 (1970).
Catt, K. J., Moffat, B., Niall, H. D., *Science* **157,** 321 (1967).
Dowling, L. M., Stark, G. R., *Fed. Proc.* **24,** 225 (1965).
Edman, P., *Acta Chem. Scand.* **4,** 283 (1950).
Edman, P., *Acta Chem. Scand.* **10,** 761 (1956).
Edman, P., *Ann. N. Y. Acad. Sci.* **88,** 602 (1960).
Edman, P., *Proc. Roy. Aust. Chem. Inst.* **24,** 434 (1957).
Edman, P., in "Protein Sequence Determination," Springer-Verlag, 1970, pp 211–255.
Edman, P., Begg, G., *Eur. J. Biochem.* **1,** 80 (1967).
Edman, P., Lauber, K., *Acta Chem. Scand.* **10,** 466 (1956).
Elmore, D. T., Toseland, P. A., *J. Chem. Soc.* 188 (1956).
Frankel-Conrat, H., *J. Amer. Chem. Soc.* **76,** 3606 (1954).
Gray, W. R., Hartley, B. S., *Biochem. J.* **89,** 379 (1963).
Hirs, C. H. W., Moore, S., Stein, W. H., *J. Biol. Chem.* **235,** 633 (1960).
Hirschmann, R., Strachan, R. G., Schwam, H., Schoenewaldt, E. F., Joshua, H., Barkemeyer, B., Veber, D. F., Palevada, W. J., Jr., Jacob, T. A., Beesley, T. E., Denkewalter, R. G., *J. Org. Chem.* **32,** 3415 (1967).
Hood, L., Lackland, J., Krause, R., *Fed. Proc.* **29,** 438 (1970).
Ilse, D., Edman, P., *Aust. J. Chem.* **16,** 411 (1963).
Ingram, V. M., *J. Chem. Soc.* 3717 (1953).
Jaton, J. C., Waterfield, M. D., Margolies, M. N., Haber, E., *Proc. Nat. Acad. Sci.* **66,** 959 (1970).
Laursen, R. A., *Fed. Proc.* **29,** 727 (1970).
Laursen, R. A., *J. Amer. Chem. Soc.* **88,** 5344 (1966).
Morgan, F. J., Henschen, A., *Biochim. Biophys. Acta* **181,** 93 (1969).
Niall, H. D., Edman, P., *J. Gen. Physiol.* **45,** Suppl. 185 (1962).
Niall, H. D., Edman, P., *Nature (London)* **216,** 262 (1967).
Niall, H. D., Potts, J. T., Jr., In "Peptides, Chemistry and Biochemistry: Proc. 1st Amer. Peptide Symposium," Marcel Dekker, 1970, pp 215–241
Niall, H. D., Sauer, R., Potts, J. T., Jr., unpublished (1971).
Niall, H. D., Penhasi, H., Gilbert, P., Myers, R. C., Williams, F. G., Potts, J. T., Jr., *Fed. Proc.* **28,** 661 (1969).
Niall, H. D., Keutmann, H., Sauer, R., Hogan, M., Dawson, B., Aurbach, G., Potts, J. T., Jr., *Hoppe-Seyler's Z. Physiol. Chem.* **351,** 1586 (1970).
Nolan, C., personal communication (1969).
Pisano, J. J., Bronzert, T. J., Jr., *J. Biol. Chem.* **24,** 5597 (1969).
Riggs, A., personal communication (1970).
Rochat, H., Rochat, C., Kupeyan, C., Miranda, F., Lissitzky, S., Edman, P., *FEBS (Fed. Eur. Biochem. Soc.) Lett.* **10,** 349 (1970).
Sauer, R., Niall, H. D., Potts, J. T., Jr., *Fed. Proc.* **29,** 728 (1970).
Stark, G. R., *Fed. Proc.* **24,** 225 (1965).
Terry, W. D., Ohms, J., *Proc. Nat. Acad. Sci.* **66,** 558 (1970).

Received for review December 24, 1970. Accepted April 2, 1971. Presented at the Division of Agricultural and Food Chemistry, 160th Meeting, ACS, Chicago, Ill., September 1970.