Peptide Sequencing by Edman Degradation

John Bryan Smith, *Celltech Chiroscience plc, Slough, UK*

Secondary article

Article Contents

Edman chemistry is a long-established approach to determination of the primary structure, or sequence, of polypeptides.

Introduction

By the middle of the twentieth century it was known that proteins were composed of amino acids, but not how these were joined together. Were they arranged in blocks of similar residues? or randomly mixed together? or in repeated patterns? At about that time, methods were devised to provide an answer to these questions: namely, that each individual protein has its constituent amino acid residues arranged linearly in a unique order, or sequence. The strategy of Sanger and colleagues for the sequencing of insulin was to characterize series of small overlapping peptides produced by cleavage of the parent molecule. Determination of the overall amino acid content and the identity of the amino- (N-)terminal residue for each peptide allowed deduction of the sequence of the whole molecule (Sanger, 1959). An alternative approach was that described by Pehr Edman (1950). This allowed determination of extended sequences of peptides or whole proteins, and has been used widely up to the present day. The method employs a series of chemical reactions to remove and identify the amino acid residue that is at the N-terminus of the polypeptide chain, i.e. the residue with a free α-amino group. At the same time, the next residue in the sequence is made available and subjected to the same round of chemical reactions. Reiteration of this process reveals the sequence of the polypeptide.

Originally described as a manual method, it was later partly automated (Edman and Begg, 1967) in an instrument termed a 'sequenator', the design of which was used in the commercially available 'spinning cup' sequenator marketed by Beckman in the late 1960s. In 1971 Laursen described a different design of automated peptide 'sequencer' that used the same chemistry but with the sample covalently attached to a solid support that minimized the loss that can occur during steps involving extraction of reaction products from solution (Laursen, 1971). This 'solid-phase' sequencing approach was useful for sequencing of short peptides that were especially easily lost in extraction steps. A significant advance came with the 'gas-phase' peptide sequenator (Hewick *et al*., 1981), so-called because some reagents were delivered as vapour. First commercialized as the Applied Biosystems model 470A, the design included automation of the process of conversion (see the next section), and could undertake solid-phase sequencing as well. Later developments were automated ('on-line') phenylthiohydantoin-amino acid identification by high-performance liquid chromatography, and later still, more than one reaction cartridge so that one sequence run could automatically be followed by another. Today, manual sequencing is rarely done. Automation has brought savings in resources and time and, with minimization of interference by oxygen and water and improved reagents and reaction conditions, a marked increase has been achieved in the efficiency of the whole process. In this way the amount of sample required to obtain a significant sequence is now orders of magnitude less than it was, at low or subpicomolar levels.

The Basis of the Method

Peptide sequencing by Edman chemistry may be divided into steps as follows, and as illustrated in **Figure 1**.

Coupling

Phenyl isothiocyanate (PITC) reacts with an α-amino group (or in the case of prolyl residue with an imino group) at the N-terminal end of the polypeptide chain, to form a phenylthiocarbamyl derivative of the terminal residue. Basic conditions are required for this reaction. Edman originally used pyridine to generate a pH of 8.6. Alternatives used since include 'Quadrol', trimethylamine and *N*-methylpiperidine. Clearly, a free α-amino group is required for this reaction to occur. If the amino group has become modified and will no longer react with PITC, the polypeptide is said to be 'blocked'. Many proteins are blocked by acetyl, formyl or pyroglutamyl groups, for instance. Various remedial techniques have been described to deblock such molecules (Hirano *et al*., 1977). Blockage may also occur artificially during the handling of the sample in the laboratory. Trace contaminants in commercial preparations of non-ionic detergents are known to cause blockage, for instance. Again, a peptide may not have an *N*-terminus at all: cyclosporin is just one example of a natural peptide that is cyclic. In cases where there is no

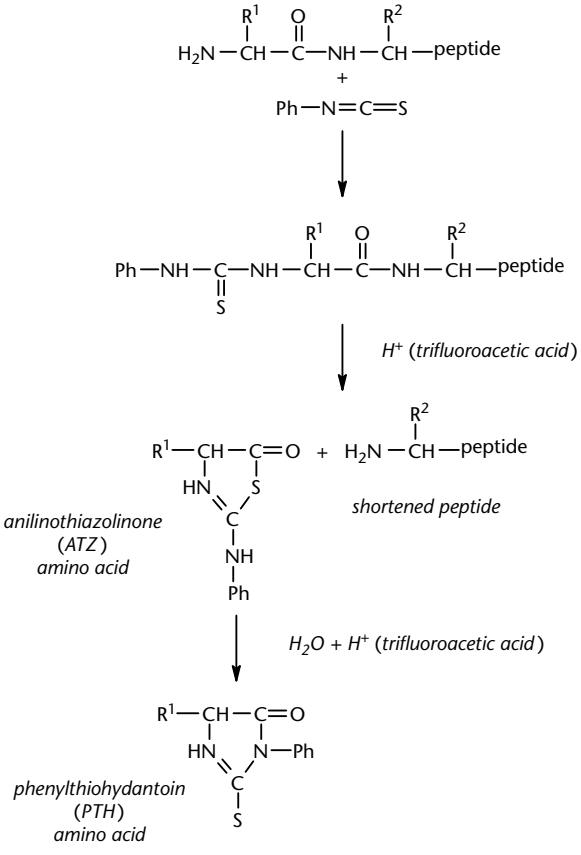

Figure 1 Edman chemistry for N-terminal sequencing of polypeptides.

free N-terminus, the polypeptide must be cleaved by chemical or enzymatic means to yield fragments that do have free N-termini. On a practical note, it is necessary to avoid contamination of the sample with amine-containing nonpeptidic species, such as trizma base, since these, too, may react with PITC and generate products that interfere with subsequent analysis. Reaction of PITC also occurs with traces of water and other molecules that are difficult to exclude completely from the reaction, so byproducts such as diphenylthiourea are commonly found. Minimization of side reactions such as these helps to improve the efficiency of the chemistry and subsequent analysis, and the ability to do this is one of the advantages of automated sequencing methods over older manual methods.

Cleavage

In the presence of strong acid, cleavage occurs at the first peptide bond, giving the peptide (minus the first residue) and the liberated first residue as the anilinothiazolinone (ATZ) form. Once other reactants and products have been washed away, the shortened polypeptide can be taken through another round of coupling and cleavage to release the second residue, and so on in a cyclical fashion. Currently, trifluoroacetic acid (TFA) is used for this cleavage reaction. Conditions are as anhydrous as is practically possible in order to minimize acid hydrolysis at points within the polypeptide chain. Where this happens, a new N-terminus is generated and becomes subject to sequencing. This then contributes to background 'noise' in the subsequent analysis. Minimization of this acid cleavage leads to longer and clearer sequence runs.

Conversion

The ATZ residue is separated from the peptide by extraction in organic solvent (ethyl acetate or chlorobutane), and is then converted to a more stable form to allow better analysis. Conversion to the more stable phenylthiohydantoin (PTH) form is done in aqueous acid (25% TFA, v/v in water). Some modified amino acid residues (such as glycosylated asparaginyl) may be poorly soluble in organic solvent and so give a blank at their corresponding point in the sequence. In solid-phase sequencing, where the peptide is covalently linked to a solid support, alternative means of extraction of the ATZ residue may be tried, in the knowledge that the remaining peptide will not also be extracted and lost (giving large drops in yield).

Analysis of PTH residues

The PTH residue generated by each cycle of Edman chemistry is typically identified by chromatography, originally thin-layer chromatography and latterly reversed-phase high-performance liquid chromatography. The PTH amino acid residue derived from each cycle in turn is identified and quantified by comparison with standards, and the sequence is described by the order of residues from the N- to the C-terminus. If radioactive amino acid residues are present they may be detected by their activity at this stage.

The efficiency of sequencing

The different amino acid residues, being structurally different, react at each stage with different degrees of efficiency. The overall efficiency ('repetitive yield') is less than 100% (usually of the order of 95%), so over the course of a number of cycles the yield of sequence declines, and the degree of stagger, or 'lag', gradually increases. At the same time, the amount of background noise increases. When the sequence signal reaches the level of the background, the sequence becomes uninterpretable. The number of cycles at which this occurs may be small or large (50 or more), depending to a great extent on the size and amino acid content of the polypeptide itself, since this dictates the degree of random acid hydrolysis and other side reactions.

The time taken to sequence a sample is considerably less than it used to be – the time for one cycle of Edman

chemistry has been reduced to 20 minutes. Current sequencers may have more than one reaction cartridge, so the machines can be used for a full 24 hours each day and generate much information in a short time.

The amount of sample required currently to provide a sequence of a few to 20 or so residues is of the order of 1 picomole or less. Automated sequencers can accept samples that are covalently bound to solid supports, such as synthesized peptides remaining on the supports on which they were made, or samples that are noncovalently bound to glass fibre or polyvinylidene difluoride (PVDF). The latter may be used to trap proteins transferred from a polyacrylamide gel, a rapid method for resolution of complex mixtures, and this approach to sample preparation is very common.

The simplest case is one where there is just one polypeptide in the sample, but it is common to have differently processed forms of the same protein or other proteins present as contaminants. These mixtures can be resolved into individual sequences if they are in significantly different quantities, if one of the sequences is known and may be subtracted from the other(s), or if, in the case of a variously processed protein, the frameshift or delayed repeat in the sequence can be spotted. Again, mixtures of sequences can be identified by screening against the rapidly growing database sequences.

The Role of Peptide Sequencing with Edman Chemistry

The purpose of peptide sequencing has changed over the decades. Originally the aim was to determine the sequence of a protein, wholly or in part, in order to better understand its structure and function. The advent of molecular biology provided an alternative and quicker way to do this, via the sequencing of the respective gene. Partial protein sequence was required for design of the oligonucleotide probes used in the process of gene cloning and for confirmation that isolated clones were indeed the relevant ones. As sequencing of whole genomes progresses, we approach the time when all gene (and therefore all protein) sequences will be known. Currently, however, knowledge of the sequence of a gene does not give us complete information on matters that are significant in the function of a protein, such as inter- and intramolecular disulfide bonding patterns or modification and processing events. Determination of protein sequence has a role in providing this, and in identification of proteins that are found to correlate with particular states of development or disease. For this purpose, a short sequence just a few residues long is all that is necessary for screening against databases of known gene sequences, when used in conjunction with other properties such as molecular weight, pI or masses of peptides. Peptide sequencing also has an important role in quality assurance of 'biopharmaceuticals', recombinant therapeutic proteins that are the result of the molecular biology revolution.

Edman chemistry is now a standard method for peptide sequencing. In recent years it has been complemented by a variety of mass spectrometric methods that have been, and continue to be, refined such that they can determine masses of proteins and of peptides derived from them, and can determine peptide sequence from the patterns of fragmentation from peptide into individual amino acid residues. While the methods may complement each other today, it could be that in the next decade or so the mass spectrometry method will replace the Edman chemistry approach. Nevertheless, it remains a credit to Edman's chemistry that it has proved so robust and given such valuable service over decades of rapid technological change in the field of biochemistry.

References

Edman P (1950) Method for determination of the amino acid sequence in peptides. *Acta Chimica Scandinavica* **4**: 283–293.

Edman P and Begg G (1967) A protein sequenator. *European Journal of Biochemistry* **1**: 80–91.

Hewick RM, Hunkapillar MW, Hood LE and Dreyer WJ (1981) A gas–liquid solid phase peptide and protein sequenator. *Journal of Biological Chemistry* **256**: 7990–7997.

Hirano H, Komatsu S and Tsunasawa S (1977) On membrane deblocking of proteins. In: Smith BJ (ed.) *Methods in Molecular Biology*, vol. 64; Protein Sequencing Protocols, pp. 285–292. Totowa, NJ: Humana Press.

Laursen R (1971) Solid-phase Edman degradation. An automatic peptide sequencer. *European Journal of Biochemistry* **20**: 89–102.

Sanger F (1959) Chemistry of insulin. *Science* **129**: 1340–1344.

Further Reading

Reim DF and Speicher DW (1995) N-terminal analysis of proteins and peptides. In: Coligan JE, Dunn BM, Ploegh HL, Speicher DW and Wingfield PT (eds) *Current Protocols in Protein Science*, vol. 2, section 11.10. Chichester: Wiley.

Smith BJ (ed.) (1997) *Protein Sequencing Protocols*. Methods in Molecular Biology, vol. 64. Totowa, NJ: Humana Press.

Smith BJ and Chapman JR (1998) Protein sequencing. In: Rapley R and Walker JM (eds) *Molecular Biomethods Handbook*, pp. 503–525. Totowa, NJ: Humana Press.