Protein Ladder Sequencing: A Conceptually Novel Approach To Protein Sequencing Using Cycling Chemical Degradation and One-Step Readout by Matrix-Assisted Laser Desorption Mass Spectrometry

Rong Wang, Brian T. Chait
The Rockefeller University, New York, NY 10021

Stephen B.H. Kent
The Scripps Research Institute, La Jolla, CA 92037

I. Introduction

Stepwise degradation from the amino terminus and analysis of the released derivatives has been used as a long-standing technique to determine the sequence of amino acids in a polypeptide chain. Almost all highly sensitive protein and peptide sequencing is currently done by automated Edman degradation (1,2) whereby practical analysis of 10-100 picomole amounts of protein has become routine. There is, however, a great current need for more rapid, highly sensitive protein sequencing methods, which give data that are easy to interpret. The most efficient and accurate techniques for sequencing of biopolymers are those that involve the readout of the entire sequence-defining data set in one operation. Thus, optimally, all members of a sequence-defining set of fragments of the biopolymer, each differing by one monomer unit, are simultaneously examined by a high resolution read-out method. The sequence of the parent polymer is deduced from the set of fragmentation products. Such a data set contains mutually interdependent information that determines the identity and order of monomer units in the parent molecule.

TECHNIQUES IN PROTEIN CHEMISTRY IV
Copyright © 1993 by Academic Press, Inc.
All rights of reproduction in any form reserved.

The advent of matrix-assisted laser desorption mass spectrometry (3) together with the development of new matrix materials (4,5) have provided a powerful new tool for the accurate measurement of the mass of intact polypeptide chains containing up to hundreds of amino acids. Strong, clean mass spectrometric data are obtained in the order of 1-3 minutes from picomole or subpicomole amounts of peptide or protein samples (5).

In this paper, we describe a completely new principle in protein sequencing: controlled chemical generation of a family of sequence-defining fragments from a polypeptide chain, followed by read-out of the complete data set as a *protein sequencing ladder* in a single operation, using matrix-assisted laser desorption mass spectrometry. The sequence-defining fragments are generated by chemical degradation of the peptide chain in the presence of a terminating agent which blocks a small fraction of the peptide chain at each residue. In one variation of the approach, a peptide was subjected to repeated cycles of the Edman degradation using 5% phenylisocyanate as a terminating agent, without separation and analysis of the low molecular weight reaction products. The final unfractionated product mixture was subjected to mass spectral analysis, to give a "protein sequencing ladder" data set. Mass differences between adjacent peaks identify each amino acid, and the position in the data set defines the sequence of the original peptide chain.

II. Experimental

A. *Peptide and Chemicals*

[Glu^{1}]-fibrinopeptide B was purchased from Sigma Chemical Co. and used with no further purification. It has an amino acid sequence of $E^{1}GVNDNEEGFFSAR^{14}$. Phenylisothiocyanate (PITC), pyridine, and trifluoroacetic acid (TFA) were purchased from Pierce. Phenylisocyanate (PIC), trimethylamine (TMA), ethyl acetate, and 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP), α-cyano-4-hydroxy-cinnamic acid (4HCCA) were purchased from Aldrich. Heptane was purchased from Burdick & Jackson.

B. *Chemical Ladder Reaction*

A modified manual Edman degradation procedure was employed to generate the desired peptide ladder. Instead of using PITC as coupling agent alone, a mixture of PITC and PIC (20:1, v/v) was used in the stepwise degradation reaction. Here, PITC was used as a

coupling agent and PIC was used as a terminating agent. All of the reaction cycles were carried out in a single 1.5 ml Eppendorf microcentrifuge vial under a stream of dry N_2. Peptide (200 pmoles to 10 nmoles) was dissolved in 20 µl of 25% TMA:Pyridine (1:1). 20 µl of the coupling agent containing PITC:PIC:Pyridine:HFIP (20:1:76:4, v/v) was added to the reaction vial. The coupling reaction was allowed to proceed at 50 °C for 3 minutes. The coupling agent, and side-reaction products were extracted using heptane and ethyl acetate solvents. 400 µl of heptane:ethyl acetate (10:1, v/v) was added to the reaction vial, gently vortexed and centrifuged to clear phases. The upper phase was aspirated and discarded. The above washing procedure was repeated, followed by washing twice with heptane:ethyl acetate (2:1, v/v). The remaining solution in the reaction vial was dried in a Speed-Vac centrifuge. The cleavage step was carried out by adding 20 µl of anhydrous TFA to the dry residue in the reaction vial and allowing the reaction to proceed at 50 °C for 5 minutes, followed by centrifuge-vacuum drying. The above-described coupling-cleavage cycle was repeated the desired number of times. The low molecular weight ATZ/PTH derivatives released at each cycle were not separated/analyzed. Washes/extractions were performed after each coupling step because preliminary experiments showed that otherwise an intractable polymeric material was formed. Analysis was performed only once after the completion of all cycles of the degradation chemistry.

C. Amino Acid Sequence Read-Out by Laser Desorption Spectrometry

After the designated number of cycles, the peptide mixture was analyzed by matrix-assisted laser desorption mass spectrometry. An aliquot, 1 µl (~ 50 pmol total peptide), was mixed with 9 µl of a solution of α-cyano-4-hydroxy-cinnamic acid as matrix (5 g/l in 0.1% TFA:acetonitrile, 2:1, v/v) and 0.5 µl of this mixture of total peptides and matrix was applied to the probe tip and dried in a stream of air at room temperature. Mass spectra were acquired in positive ion mode using a laser desorption time-of-flight instrument constructed at The Rockefeller University (6). Spectra obtained from 200 laser pulses were acquired over 80 seconds and added to give a mass spectrum of the sequencing ladder. No peaks from the α-cyano-4-hydroxy-cinnamic acid matrix material used were detected above 500 Da.

III. Results and Discussion

Sequence-defining sets of fragments can potentially be generated in a controlled fashion, from a polypeptide chain in a number of ways. One straightforward way is to carry out stepwise degradation in the presence of a terminating agent. Thus, a protein sequencing ladder can be generated by carrying out rapid stepwise Edman degradation in the presence of a small-amount of terminating agent such as phenylisocyanate (Figure 1). In this way, a small proportion of N-terminally blocked peptide chain is generated at each cycle of the

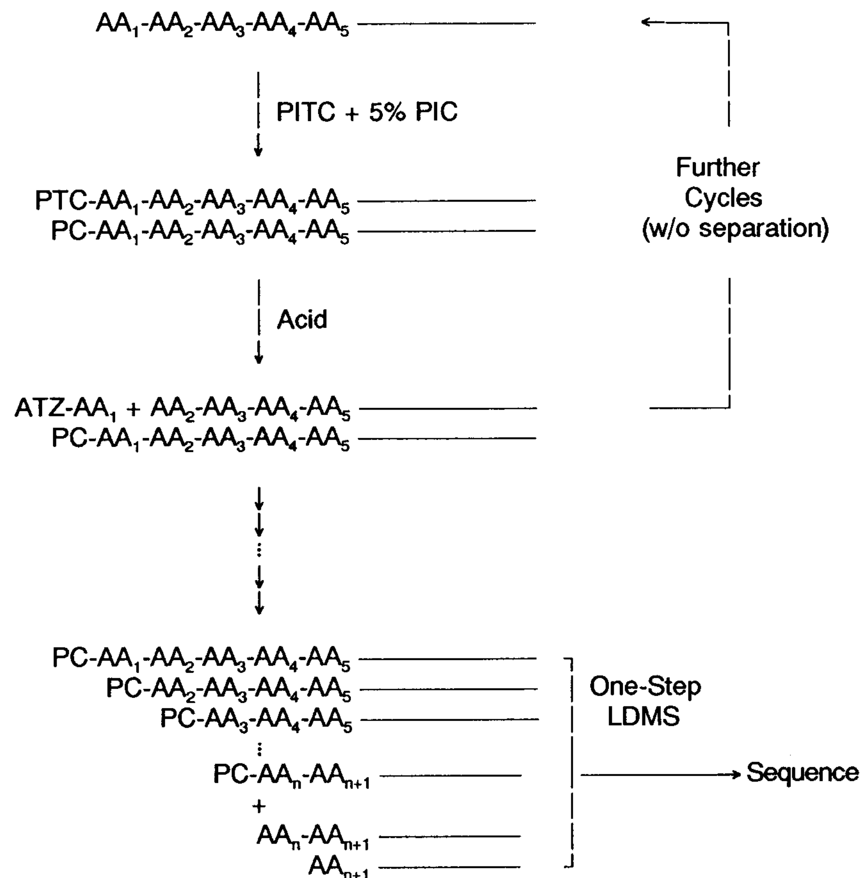

Figure 1. Schematic of the generation of a sequence-defining set of fragments by cycling Edman degradation in the presence of a terminating agent, phenylisocyanate. A small fraction of the peptide is blocked by phenylisocyanate in each cycle. These blocked peptides will not undergo further degradation in the following cycles and will accumulate forming the protein sequencing ladder.

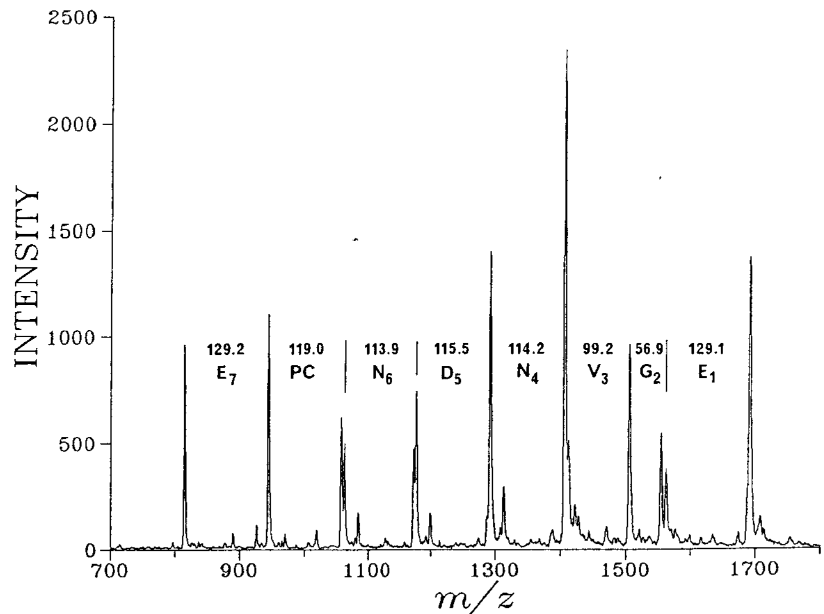

Figure 2. Protein ladder sequencing of [Glu[1]]-fibrinopeptide B. Positive ion matrix-assisted laser desorption mass spectrum was acquired in a single operation after seven cycles of the protein ladder sequencing reaction. Each amino acid is identified by the mass difference between neighboring peaks and the sequence of parent peptide is defined by the position of each amino acid in the same spectrum.

degradation. After the desired number of cycles has been performed, without intermediate separation and analysis of the released amino acid derivatives, the resulting mixture of terminated peptide chains is read out in a single operation by matrix-assisted laser desorption mass spectrometry, to yield a *protein sequencing ladder* (*i.e.*, a mass spectrum consisting of protonated molecules with masses corresponding to each terminated polypeptide species present). The mass differences between consecutive peaks each correspond uniquely to an amino acid (with the exception of Leu/Ile), and their order of occurrence in the data set defines the sequence of amino acids in the original peptide chain.

A sequence-defining set of fragments was generated from [Glu[1]]-fibrinopeptide B by seven cycles of the Edman degradation carried out in the presence of 5% v/v phenylisocyanate, as described in Methods, without separation or analysis of the low molecular weight ATZ/PTH derivatives resulting from each cyclization/cleavage step. The mass spectrum of the sequencing ladder is shown in Figure 2. All peaks present in the mass spectrum were identified (7). The mass differences between each adjacent peak were calculated and compared

with the known mass of each amino acid residue. The interpretation of the data as amino acid sequence is illustrated in Figure 2.

In this example, the ratio of degradation/terminating reagent (PITC/PIC) was arbitrarily selected. The spectrum yields a simple and useful sequencing ladder (Figure 2). No effort was made to optimize the coupling or cleavage yields in the chemical degradation because incomplete reactions of coupling or cleavage in a given cycle, will be continued in following cycles. Thus, the accuracy of protein ladder sequencing is unaffected by yields of individual steps, over a wide range. Success depends only on the ratio of reactions, terminating-to-degradation. Consequently, reactions used in this approach can be simple and fast. This should be contrasted with the stringent yield requirements of the standard stepwise Edman degradation sequencing procedure (8).

The amino acid sequence is simply read off from the protein sequencing ladder. Mass differences between neighboring peaks identify each amino acid, and their position in the data set defines the sequence in the parent peptide. Because the degradation was started from the N-terminus, the amino acid sequence is read out from the high mass end to lower mass. For example, the mass difference between peaks at m/z 1690.9 and 1561.8 is 129.1 Da which corresponds to the residue mass of glutamic acid (calculated residue mass 129.1 Da), at the first position. The next pair of peaks at m/z 1561.8 and 1504.9 gives a mass difference of 56.9 Da which represents the amino acid residue glycine (calculated residue mass 57.1 Da), at the second position. The measured mass differences observed for [Glu^1]-fibrinopeptide B are given in Figure 2. The mass accuracy obtained was sufficient to unambiguously distinguish all of the amino acids. Because the amino acid identification relies on mass differences, less stringent demands are placed on absolute mass measurement. The laser desorption time-of-flight instrument used in these studies routinely gives mass accuracies of 0.01% (1 part in 10,000) (9). Furthermore, because the sequence-defining data set is read-out in one operation, the data is mutually interdependent. Thus, the likelihood of errors in the determination of the identity of an amino acid is reduced.

Because the sequence-defining set of fragments is read-out in one operation, protein ladder sequencing can be very fast. The relatively slow (typically 25-30 min) reverse-phase HPLC analysis of each released PTH amino acid derivative (used in conventional Edman sequencing) is not required. The ladder-generating chemistry lends itself to being performed simultaneously on a large number of samples. Because the laser desorption mass spectrometric read-out is fast (~ 1-3 minutes per sample), by multiplexing the ladder-generating chemistry, this technique can potentially give a throughput of 10-20 residues per minute (~ 1,000 residues per hour), at very low cost per amino acid

residue. Such rapid, inexpensive sequencing could vastly expand the applications and use of protein sequence analysis in biological research.

The protein ladder sequencing is, in principle, applicable to the analysis of peptides and proteins with post-translational modifications. All modified residues that are stable to the chemical conditions used to generate the sequence-defining set of terminated peptides will show up as an additional mass difference in their residue masses. Partly stable modifications may actually yield even more useful data, in that two series of peaks will be seen. The mass difference between the two series will define the mass of the modification. Both the site and identity of the modified amino acid can be determined, and the nature of the modification can in many cases be deduced.

The principle of protein ladder sequencing is also directly applicable to C-terminal sequence analysis using simple variations of existing chemistries, and also can be used for sequencing of mixtures of peptides.

IV. Conclusions

The protein ladder sequencing approach involves the use of controlled degradation chemistry to generate a sequence-defining set of terminated peptides. The protein ladder is read-out in a single operation by matrix-assisted laser desorption mass spectrometry. The protein sequencing ladder data set is generated as a whole, and is straightforward to interpret. The mass difference between neighboring peaks defines the identity of an amino acid residue (with the exception of Leu/Ile since they have the same residue mass), while the position of each mass difference in the data set establishes the sequence of amino acids in the parent peptide chain. The overall process is extremely rapid and has a very low cost per amino acid residue sequenced.

This robust, practical new approach to the rapid determination of protein amino acid sequences has considerable potential for increasing the application and use of protein sequencing in biological research.

Acknowledgements

This work was supported by grants from the National Institutes of Health (RR00862, GM38274), and by funds from the Markey Foundation.

Reference

1. Edman, P. (1950), *Acta Chem. Scand.* **4**, 283-293.
2. Edman, P. and Begg, C. (1967), *Eur. J. Biochem.* **1**, 80-91.
3. Karas, M., Bachmann, D., Bahr, U. & Hillenkamp, F. (1987) *Int. J. Mass Spectrom. Ion Proc.* **78**, 53-68.
4. Beavis, R.C. & Chait, B.T. (1989) *Rapid Commun. Mass Spectrom.* **3**, 432-435.
5. Beavis, R.C., Chaudhary, T. & Chait, B.T. (1992) *Org. Mass Spectrom.* **27**, 156-158.
6. Beavis, R.C. & Chait, B.T. (1989), *Rapid Commun. Mass Spectrom.* **3**, 233-237.
7. The peak at m/z 1553.8 (third peak from the right) corresponds to the protonated [Pyr[1]]-fibrinopeptide B, which is the N-terminal dehydration product of [Glu[1]]-fibrinopeptide B.
8. Laursen, R.A., Lee, T.T., Dixon, J.D. & Liang, S.P. (1991) *Methods Protein Sequence Anal. [Proc. Int. Conf.] 8th.*
9. Beavis, R.C. & Chait, B.T. (1990) *Anal. Chem.* **62**, 1836-1840.