Protein Ladder Sequencing: Towards Automation

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

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

I. Introduction

We have recently developed a new method for sequencing polypeptides: *Protein Ladder Sequencing* (1,2). In this method, a controlled stepwise chemical degradation is performed on one terminus of a polypeptide in the presence of a small amount of terminating reagent. The reaction yields a set of sequence-defining peptides, each differing from the next by a single amino acid residue. This protein sequencing ladder is read out rapidly in an one-step operation by matrix-assisted laser desorption mass spectrometry. Each amino acid is identified from the mass difference between successive peaks in the mass spectrum of the sequencing ladder, and the position in the data set defines the sequence of the original peptide chain. Our experimental results (3) indicate that ladder sequencing has the potential for providing extremely rapid, sensitive sequence analysis of polypeptides with high sample throughput. In addition, the method can be used to directly identify the sites and nature of post-translationally modified amino acid residues in proteins.

Until recently, we have performed the ladder generating chemistry manually. In an effort to improve the sensitivity, speed, and ease of the method, we have reprogrammed a conventional protein sequencer (ABI model 471A) to perform the ladder generating chemistry automatically. The present contribution describes the optimization of these automated chemical manipulations with respect to sensitivity, speed, reaction yield, and minimization of undesired side-reactions. The feasibility of sequencing, in parallel, multiple

TECHNIQUES IN PROTEIN CHEMISTRY V

polypeptide samples was also tested. Sequencing rates of 10 residues/hour have been attained on single peptides (at sensitivity levels <10 pmole of total initial sample) and considerably higher sequencing rates (> 100 residues/hour, including readout) appear to be achievable by carrying out the ladder generating chemistry in parallel on a large number of different polypeptide samples.

II. Experimental

A. Peptide and chemicals

[Glu^1]-fibrinopeptide B (EGVNDNEEGFFSAR) was obtained from the Sigma Chemical Co. (St.Louis, MO) and used with no further purification. 5% phenylisothiocyanate (PITC) in n-heptane, 12.5% trimethylamine (TMA) in water, trifluoroacetic acid (TFA), and ethyl acetate (EA) were sequencing grade obtained from Applied Biosystems Corp. (Foster City, CA). α-cyano-4-hydroxycinnamic acid (4HCCA) was obtained from Aldrich (Milwaukee, WI) and used without further purification. Immobilon-CD membrane was obtained from Millipore Corp (Bedford, MA).

B. Sample preparation for sequencing

Immobilon-CD membrane was cut into pieces having dimensions of 2x2 mm^2 or 5x5 mm^2. The smaller pieces of membrane were used to immobilize quantities of peptide sample in the range 1-10 pmole while the larger pieces were used to immobilize quantities of peptide in the range 50-100 pmole. Prior to application of the peptide, the membranes were pre-wetted with 2 or 5 μL of 12.5% trimethylamine/methanol (1:1, v/v). 1-50 pmole of the peptide sample was absorbed onto the pre-wetted immobilon-CD membrane, which was then dried under nitrogen flow and loaded into the reaction cartridge of the protein sequencer.

C. Polypeptide degradation

To gain a detailed understanding of the degradation chemistry, terminating reagents (2) were omitted from these experiments. An ABI model 471A protein sequencer was used to perform the degradation reactions. The sequencing cycle was modified, varied, and tested with respect to coupling, cleavage, and wash times; reaction temperatures; reagent volumes; and the number of repetitive reagent deliveries.

D. Peptide sample extraction for mass spectrometric analysis

After the degradation reaction, the immobilon-CD membrane was cut into pieces with dimensions of 1x1 mm^2 and the peptide mixture (sequencing ladder) was extracted with 10 µL of 2.5% trifluoroacetic acid in acetonitrile/water (6:4, v/v) by sonication for 5 minutes.

E. Amino acid sequence read-out

The resulting peptide mixture (sequencing ladder) from the degradation reaction was analyzed by matrix-assisted laser desorption mass spectrometry. 1-2 µL of the extraction solution was mixed with 2 µL of α-cyano-4-hydroxycinnamic acid (5 g/L in 0.1% TFA in water/acetonitrile 2:1 v/v) (4). 1 µL of the mixed solution (3-5% of the final peptide sample) was applied on the sample 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 mass spectrometer constructed at The Rockefeller University (5,6). Spectra obtained from 200 laser pulses were acquired and added to give a mass spectrum of the sequencing ladder. The mass accuracy for the determination of the amino acid residues was better than ± 0.4 u.

III. Results and Discussion

Protein ladder sequencing, is performed in two steps. The first step involves the generation of the protein sequencing ladder, i.e. a family of sequence-defining fragments from a polypeptide chain. This protein sequencing ladder is produced by the degradation of the polypeptide chain from one terminus in the presence of a terminating agent which blocks a small fraction of the peptide chain at each residue (1,2). The second step involves readout of the entire sequencing ladder in a single operation by matrix-assisted laser desorption mass spectrometry (7). To gain a detailed understanding of the degradation chemistry, terminating reagents were omitted from the present experiments. In the absence of such terminating reagents, the sequencing ladder of peptides can be generated by incomplete Edman degradation reaction and/or by side-reactions that block a fraction of the amino-termini of the peptides to further sequential degradation reaction. The sequencing ladder produced by incomplete Edman degradation provides a direct measure of the efficiency of the degradation chemistry, while the observation of blocked products provides a measure of the extent of side-reactions.

A. Reduction of the duration of the degradation cycle time

The Edman degradation chemistry was performed on a commercial protein sequencer (ABI Model 471A) using degradation cycles highly modified in comparison to those normally employed for conventional sequencing applications. In particular, the degradation cycle times were greatly reduced over those normally employed and the reaction temperature was varied. In these shortened cycle programs, the coupling time, cleavage time, and the amounts and number of repetitions of chemical delivery were varied. Figure 1 shows a mass spectrum of a peptide ladder generated by one of these shortened programs (see Table 1) with a total cycle time of 5'52" minutes. A total of 50 pmole of [Glu^1] fibrinopeptide B was subjected

Table 1. An example of a reduced duration degradation cycle program for protein ladder sequencing

Step	Function Code	Function	Duration (sec.)	Accumulated Time
1	4	PREPARE R2[a]	5	00:00:05
2	5	DELIVER R2	10	00:00:15
3	1	PREPARE R1[b]	5	00:00:20
4	2	DELIVER R1	2	00:00:22
5	29	DRY CARTRIDGE	5	00:00:27
6	5	DELIVER R2	120	00:02:27
7	29	DRY CARTRIDGE	10	00:02:37
8	13	PREPARE S2[c]	5	00:02:42
9	14	DELIVER S2	8	00:02:50
10	39	WAIT	5	00:02:55
11	29	DRY CARTRIDGE	5	00:03:00
12	14	DELIVER S2	8	00:03:08
13	39	WAIT	5	00:03:13
14	29	DRY CARTRIDGE	5	00:03:18
15	14	DELIVER S2	8	00:03:26
16	39	WAIT	5	00:03:31
17	29	DRY CARTRIDGE	30	00:04:01
18	7	PREPARE R3[d]	5	00:04:06
19	9	LOAD R3	8	00:04:14
20	29	DRY CARTRIDGE	5	00:04:19
21	39	WAIT	40	00:04:59
22	15	LOAD S2	5	00:05:04
23	30	BLOCK FLUSH	5	00:05:09
24	14	DELIVER S2	8	00:05:17
25	39	WAIT	5	00:05:22
26	29	DRY CARTRIDGE	30	00:05:52

Note: [a] R2: 12.5% trimethylamine in water;
[b] R1: 5% phenylisothiocyanate in n-heptane;
[c] S2: Ethyl acetate;
[d] R3: Trifluoroacetic acid.

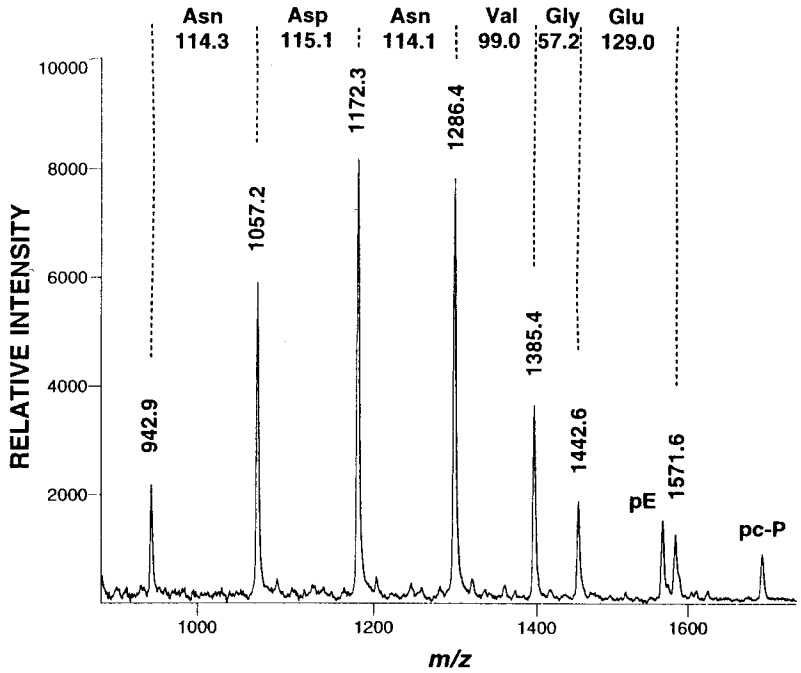

Figure 1. The mass spectrum of the protein sequencing ladder generated from a total of 50 pmole [Glu[1]] fibrinopeptide B (E-G-V-N-D[5]-N-E-E-G-F[10]-F-S-A-R) by 6 cycles of Edman degradation at 53 °C *in the absence of terminating reagent*. 3% of the final sample was used for the mass spectrometric measurement. The duration of the degradation cycle was 5'52" minutes, with a coupling time of 2'37" and a cleavage time of 54" (Table 1). The peaks labeled with pE and pc-P correspond to [Glu[1]] fibrinopeptide B with pyroglutamic acid and phenylcarbamyl blocking groups at the N-terminus. The mass differences between successive peaks and the identities of the corresponding amino acids are also given.

to 6 cycles of degradation at 53 °C. The data demonstrate that useful sequencing ladders can be generated with cycle times as short as 6 minutes. In contrast to conventional sequencing, high reaction yields are not required in order to generate useful sequencing ladders (1,2). In the cycle program used to generate the ladder shown in Fig. 1 (see Table 1), the coupling time was 2'37" and the cleavage time 54". A degradation reaction yield of ~65% was estimated from the relative heights of the peaks in Fig. 1, assuming that the degradation yield is constant from cycle to cycle and that the peak heights in the mass spectrum are representative of the relative amounts of the different peptides present in the mixture. Other experiments (data not shown) indicated that multiple PITC delivery was not necessary to improve the coupling reaction yield but that

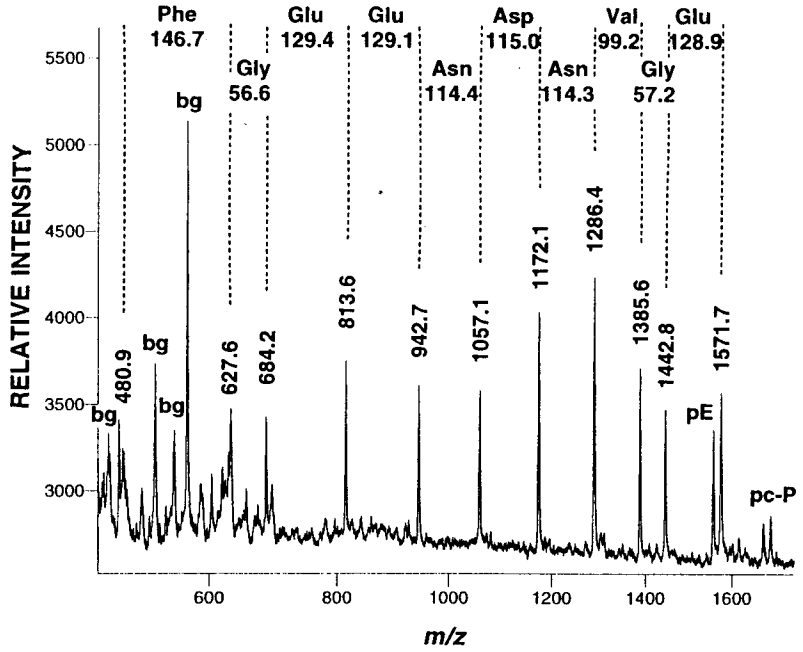

Figure 2. The mass spectrum of the protein sequencing ladder generated from a total of 10 pmole [Glu[1]] fibrinopeptide B by 10 cycles of Edman degradation. 5% of the final sample was used for the mass spectrometric measurement. The duration of the degradation cycle time was 6'02", with a coupling time of 2'37" and a cleavage time of 1'09". The reaction temperature was 59 °C. The peaks labeled with bg are known background peaks from the laser desorption matrix. The peaks labelled pE and pc-P correspond to [Glu[1]] fibrinopeptide B with pyroglutamic acid and phenylcarbamyl blocking groups at the N-terminus. The mass differences between successive peaks and the identities of corresponding amino acids are also given.

scrupulous removal of TFA from previous cleavage reaction cycles was essential to achieve optimal coupling reaction yields. It proved better to use three pulsed washes after the coupling reaction rather than a continuous wash because the pulsed wash was more efficient and produced lower sample loss. Reaction temperatures of 48, 53, 56, 58, 59 °C were tested. Although the higher temperatures gave higher reaction yields, the yield of undesired side-reactions also increased as a function of increased temperature. Because the optimum reaction conditions are directly related to the method of sample immobilization, we believe that further reductions in the cycle time may be possible through the use of improved immobilization methods.

B. Sensitivity

The readout of the sequencing ladder by matrix-assisted laser desorption mass spectrometry is highly sensitive (in the low femtomole range) (1,2). However, because it is necessary to immobilize the sample of interest, perform extensive chemistry on the immobilized sample, and finally extract the resultant peptide mixture from the immobilizing medium, significant losses of sample occur. The sensitivity of ladder sequencing is therefore limited largely by sample handling prior to the mass spectrometric readout. Figure 2 shows a sequencing ladder mass spectrum obtained from a total of 10 pmole of [Glu^1]-fibrinopeptide B after 10 degradation cycles (the experimental conditions are indicated in the figure legend).

We believe that further improvements in sequencing sensitivity will be achieved through additional improvements in sample handling.

C. Parallel degradation of multiple peptides

We have also tested the feasibility of carrying out the ladder-generating chemistry in parallel on multiple polypeptide samples. Results obtained on pairs of peptides inserted into the reaction cell on separate pieces of membrane (data not shown here) yielded good sequencing data without discernible interference or cross contamination between the different peptide ladders. These results suggest that it will be possible to carry out the ladder generating chemistry in parallel on large numbers of different peptides (e.g. all of the peptides generated by trypsin digestion of a typical protein). Such large scale parallel sequence analysis should lead to effective sequencing rates greater than 100 residues/hour, including readout.

IV. Conclusions

Protein ladder sequencing is inherently highly accurate because the one-step readout data contains mutually interdependent information, which determines the identity and order of the amino acid residues in the parent molecule. Post-translational modifications (e.g. phosphorylation (2), glycosylation (3)) can also be directly determined.

Ladder sequencing is rapid compared to conventional Edman sequencing because (i) a high reaction yield is not required and (ii) the one-step readout is very fast (< 1 min.). An Edman degradation rate of 10 residues/hour has been achieved by the use of an automated instrument.

The sample throughput is potentially very high because the ladder-generating chemistry can be carried out simultaneously on a

large number of samples. Taken together with the fast degradation, sequencing rates greater than 100 residues/hour, including readout of the sequencing ladder data, appear feasible.

Ile cannot be distinguished from Leu (since they have the same residue mass); blocked peptides are not amenable to sequence analysis; presently the method is limited to the analysis of peptides shorter than approximately 60 amino acid residues.

It is desirable to further optimize sample handling, reduce interfering side-reactions, and tune the termination chemistry.

Sensitivities of < 10 pmole of total initial sample have been obtained and it is likely that considerably better sensitivities will be achieved in the future.

Acknowledgments

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

References

1 Chait, B.T., Wang, R., Beavis, R.C., and Kent, S.B.H. (1992) *Proceedings of the 40th ASMS Conference on Mass Spectrometry and Allied Topics, Washington, DC,* May 31- June 5.

2 Chait, B.T., Wang, R., Beavis, R.C., and Kent, S.B.H. *Science*, in press.

3 Wang, R., Walker, S., Kent, S.B.H., Beavis, R.C., and Chait, B.T. (1993) *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA*, May 30 - June 4.

4 Beavis, R.C., Chaudhary, T., and Chait, B.T. (1992) *Organic Mass Spectrom.* **27**, 156-158.

5 Beavis, R.C., and Chait, B.T. (1989) *Rapid Commun. Mass Spectrom.* **3**, 233-237.

6. Beavis, R.C. and Chait, B.T. (1990) *Anal. Chem.* **62**, 1836-1840.

7. Hillenkamp, F., Karas, M., Beavis, R.C., and Chait, B.T. (1991) *Anal. Chem.* **63**, 1193A-1203A.