128 by 128; pixel size, 0.75 arc sec; effective resolution, 2 arc sec.

31. Observed with the FAST camera on the European South Observatory–Max Planck Institute 2.2-m telescope at La Silla, Chile (27). In Sb array, 62 by 58; pixel size, 0.78 arc sec; effective resolution, 2 arc sec.

32. Obtained with the Hat Creek interferometer (12). Effective resolution, 7.5 arc sec. The absolute position of these maps is accurate to 0.15 arc sec.

33. T. R. Geballe, A. G. G. M. Tielens, L. J. Allamandola, A. Moorhouse, P. W. J. L. Brand, Astrophys. J. 341, 278 (1989).

34. M. Haas, D. Hollenbach, E. F. Erickson, Astrophys. J. 301, L57 (1986).

Protein Ladder Sequencing

Brian T. Chait, Rong Wang, Ronald C. Beavis, Stephen B. H. Kent*

A new approach to protein sequencing is described. It consists of two steps: (i) ladder-generating chemistry, the controlled generation from a polypeptide chain by wet chemistry of a family of sequence-defining peptide fragments, each differing from the next by one amino acid; and (ii) data readout, a one-step readout of the resulting protein sequencing ladder by matrix-assisted laser-desorption mass spectrometry. Each amino acid was identified from the mass difference between successive peaks, and the position in the data set defined the sequence of the original peptide chain. This method was used to directly locate a phosphoserine residue in a phosphopeptide. The protein ladder sequencing method lends itself to very high sample throughput at very low per cycle cost.

Direct experimental determination of the amino acid sequence of a polypeptide chain usually gives partial sequence data only. Partial amino acid sequence data may be used to identify isolated proteins (1), and are useful in cloning genes (2). The complete amino acid sequence of a protein is most often determined by nucleic acid sequencing at the cDNA level. However, posttranslational modifications (3) must be characterized at the polypeptide level.

Most direct sequence determination of peptides and proteins is done by automated Edman degradation (4), in which a two-part chemical reaction is used to remove one amino acid at a time from the amino terminal. After release, each amino acid derivative is converted to a stable form and is then identified by analytical reverse-phase high-performance liquid chromatography. Currently such sequencing is limited to less than ~50 residues per day (5). Also, most posttranslational modifications are not identified. Thus, there is a great need for more rapid and versatile protein sequencing methods (6).

The recent advent of matrix-assisted laser-desorption mass spectrometry (LDMS) (7) and the development of improved matrix materials (8) has facilitated the accurate measurement of the mass of intact polypeptide chains. Subpicomole amounts of total sample can be analyzed in seconds with a mass accuracy of up to 1 part in 10,000 (9). Thus the polypeptide itself can be analyzed more readily, with greater speed, sensitivity, and precision, than the amino acid derivative released by stepwise sequencing (10).

We describe a new principle in protein sequencing that combines multiple steps of wet degradation chemistry with a final, single-step mass spectrometric (MS) read-out of the amino acid sequence. First, a sequence-defining concatenated set of peptide fragments, each differing from the next by a single residue, is chemically generated in a controlled fashion. Second, matrix-assisted LDMS is used to read out the complete fragment set in a single operation, as a "protein sequencing ladder" data set.

A concatenated set of peptide fragments can be generated in a controlled fashion (11) by carrying out rapid stepwise degradation in the presence of a small amount of terminating agent, a procedure we call "ladder-generating chemistry" (Fig. 1). A small proportion of peptide chain blocked at the amino terminus is generated at each cycle. A predetermined number of cycles is performed without intermediate separation or analysis of the released amino acid derivatives. The resulting mixture is read out in a single operation by matrix-assisted LDMS (12). The mass spectrum contains molecule ions corresponding to each terminated polypeptide species present. The mass differences between consecutive peaks each correspond to an amino acid residue (13), and their order of occurrence in the data set defines

the sequence of amino acids in the original peptide chain (14).

We sequenced the 14-residue peptide [Glu^1]fibrinopeptide B (15) to illustrate the new method. Eight cycles of manual ladder-generating chemistry were carried out (16), and the resulting product mixture of terminated peptides read out (17) by matrix-assisted LDMS (Fig. 2). All the major components present in the mass spectrum were readily identified, and the data could be simply interpreted to give the sequence of the eight amino-terminal residues of the peptide. The two consecutive peaks with the highest mass differ by 129.1 daltons, identifying the amino-terminal amino acid as a Glu residue (calculated residue mass 129.1). The identities of the next seven residues were read off in a similar fashion (18).

Several features of the protein ladder sequencing experiment are immediately apparent. The mass accuracy obtained (9) was sufficient to unambiguously distinguish Asp [calculated residue mass 115.1] (13) and Asn (calculated residue mass 114.1); Glu [calculated residue mass 129.1] was also identified with sufficient accuracy to distinguish it from Gln [calculated residue mass 128.1]. The arbitrary ratio of degradation-to-terminating reagents and the minimal
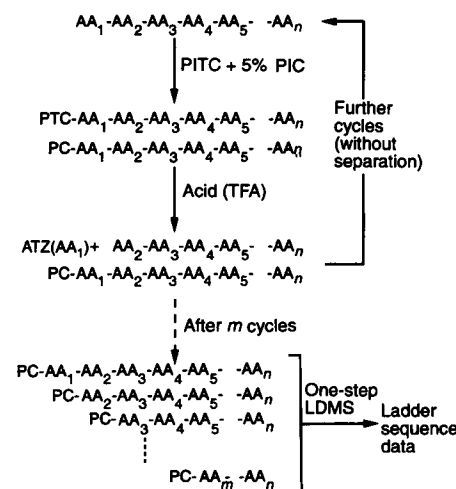
B. T. Chait and R. Wang, The Rockefeller University, New York, NY 10021.
R. C. Beavis, Department of Physics, Memorial University of Newfoundland, St. John's, Newfoundland, Canada A1B 3X7.
S. B. H. Kent, The Scripps Research Institute, La Jolla, CA 92037.

*To whom correspondence should be addressed.

Fig. 1. Protein ladder sequencing principle exemplified by the generation of a set of sequence-determining fragments from an intact peptide chain with controlled ladder-generating chemistry. A stepwise degradation (32) is carried out with a small amount of terminating agent present in the coupling step. In this case, 5% phenylisocyanate (PIC) was added to the phenylisothiocyanate (PITC). The phenylcarbamyl (PC) peptides formed are stable to the trifluoroacetic acid (TFA) used to cyclize and cleave the terminal amino acid (AA) from the phenylthiocarbamyl (PTC) peptide. Successive cycles of ladder-generating chemistry are performed without intermediate isolation or analysis of released amino acid derivatives. Finally, the mixture of PC peptides is read out in one step by matrix-assisted LDMS.

89

reaction conditions employed have yielded a simple, useful sequencing ladder. No effort was made to optimize coupling or cleavage yields in the chemical degradation because the accuracy of protein ladder sequencing is unaffected by the relative abundance, over a wide range, of individual terminated fragments. Obtaining high reaction yields is not critical, and the degradation protocols can be simple and fast. In contrast, extreme (prolonged and forcing) reaction conditions are used in the standard stepwise Edman degradation (19).

A second example illustrates the ladder sequence analysis of both phosphorylated and unphosphorylated forms of a 16-residue peptide containing a Ser residue (20). After 10 cycles of ladder-generating chemistry on each form of the peptide (21), the two separate sequence-defining fragment mixtures were each read out in a single matrix-assisted LDMS experiment (Fig. 3). The protein ladder sequencing method directly identified and located a Ser(Pi) at position five in the peptide (22). There was no detectable loss of phosphate from the phos-

phoserine residue, which has been regarded as the most sensitive and unstable of the phosphorylated amino acids (23).

The inability to directly identify, locate, and quantify phosphorylated residues is a major shortcoming of standard sequencing methods and has imposed major limitations on currently important areas of biological research, such as mechanisms of signal transduction. Protein ladder sequencing has general application to the direct identification of posttranslational modifications present in a peptide chain being sequenced. A modified amino acid residue that is stable (23) to the conditions used in the ladder-generating chemistry reveals itself as an additional mass difference at the site of the covalent modification. Frequently, this will lead to unambiguous identification of the chemical nature of the posttranslational modification (3). The utility of protein ladder sequencing in this regard would apply even to large modifying entities, such as carbohydrate moieties in glycopeptides.

To explore the capabilities and limitations of the ladder sequencing readout by matrix-assisted LDMS, measurements were carried out on sets of sequence-defining unblocked synthetic peptides. This set of peptides was obtained during the course of a total chemical synthesis of the 99–amino

Fig. 2. Protein ladder sequencing of [Glu^1]fibrinopeptide B (15). The peptide, of sequence Glu^1-Gly-Val-Asn-Asp^5-Asn-Glu-Glu-Gly-Phe^{10}-Phe-Ser-Ala-Arg^{14}, was subjected to eight cycles of ladder-generating chemistry (Fig. 1) (16). The matrix-assisted LDMS readout (17) of the resulting sequence-defining set of fragments is shown in two forms: A standard intensity versus mass (33) plot; the data is plotted from high to low mass, so that the amino acid sequence reads from the amino terminal. The upper horizontal lines show the different lengths blocked peptide species present and their relation to the MS data.
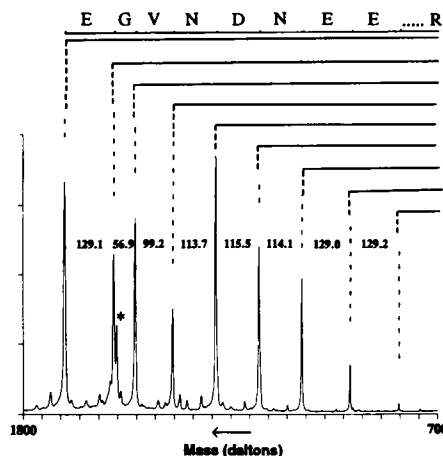

Fig. 3. (left) Protein ladder sequencing of the 16-residue synthetic peptide: Leu-Arg-Arg-Ala-Ser(P$_i$)-Gly-Leu-Ile-Tyr-Asn-Asn-Pro-Leu-Met-Ala-Arg.amide. (A) Phosphorylated peptide. (B) Unphosphorylated peptide. Each peptide sample was subjected to 10 cycles of ladder-generating chemistry. Data defining the 11 amino-terminal residues (21) are shown. The Ser(P$_i$) residue was characterized by a mass difference of 166.7 daltons (Ser, calculated residue mass 87.1; Ser(P$_i$) calculated residue mass 167.1) observed in position five. There is no evidence for loss of phosphate (35).
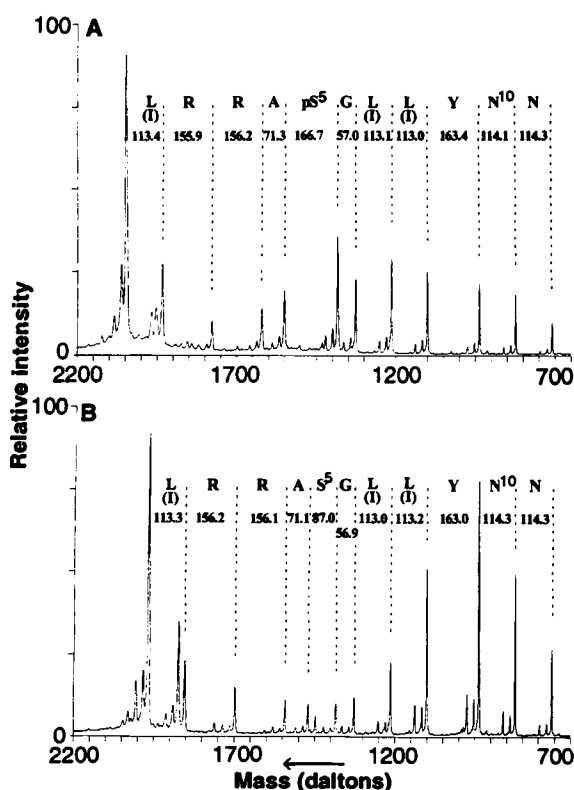
Fig. 4. (right) Extended MS readout of sequence-defining sets of polypeptide fragments. Consecutive samples, after each amino acid addition, were taken during stepwise solid-phase assembly of the 99-residue monomer sequence of HIV-1 protease (24). After release from the solid support and deprotection, pooled peptide samples corresponding to residues 67 to 99 and 33 to 66 were analyzed by matrix-assisted LDMS
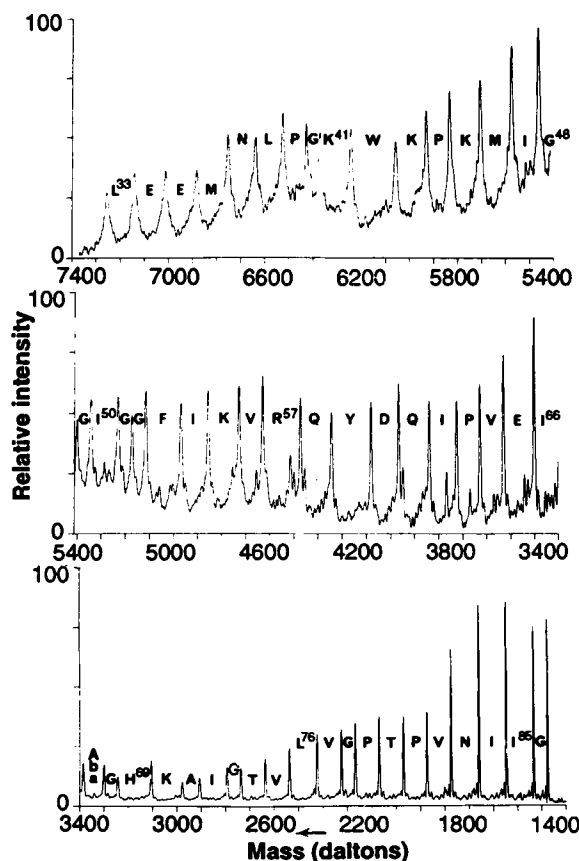
(17). Observed mass differences for each amino acid residue are given in Table 1.

acid monomer polypeptide chain of the human immunodeficiency virus–1 (HIV-1) protease (24). The target sequence was assembled by solid-phase synthesis in step-wise fashion from the resin-bound carboxyl-terminal residue Phe[99]. Samples of peptide resin were taken after addition of each amino acid, from residue 98 to residue 33. The different length peptide resins were pooled in two batches of more than 30 consecutive samples, and the two mixtures were separately deprotected and cleaved (25). The resulting sets of sequence-defining fragments with masses up to 7400 dal-

tons were read out by matrix-assisted LDMS (26) (Fig. 4 and Table 1).

The impact of the average uncertainty in measured mass becomes more significant as the molecular mass increases. Below 3500 daltons, the mass deviation is less than ±0.3 dalton, and there is no ambiguity in distinguishing even the most closely related pairs of amino acids [Leu/Ile have identical mass (13)]. However, above 3500 daltons, uncertainties of 0.4 to 0.9 dalton introduce certain ambiguities in the identification of amino acids of closely similar residue masses (13).

These results illustrate the potential for extended sequence determination with the protein ladder approach with existing matrix-assisted LDMS readout. Out to more than 30 residues, the simple mass differences directly define the amino acid sequence. However, for peptides approximately 35 to 65 residues in length, current instrumental mass accuracy of up to 1 part in 10,000 is not sufficient to unambiguously identify every amino acid residue based on simple mass differences (27). Direct applicability to peptide chains of less than 60 residues is currently a limitation of protein ladder sequencing and means that sequence data cannot be directly obtained from intact proteins larger than ~6500 daltons. However, extensive sequence data can be obtained from larger proteins by the commonly used tactic of chemical or enzymatic cleavage combined with (protein ladder) sequencing of the resulting fragments, by analogy with existing methods (28).

Most current protein sequence determination is carried out with 10 to 100 pmol of sample (5, 19) and extended automated Edman degradation has been demonstrated on <10 pmol samples (29). We can read out a protein sequencing ladder data set containing 2 to 5 fmol of individual components (Fig. 5). We have adapted the protein ladder method to obtain sequence data from low picomole total amounts of peptide samples (30). Thus, the demon-

strated sensitivity of protein ladder sequencing is comparable to that of existing Edman methods, with potential for far greater sensitivity.

Protein ladder sequencing lends itself to very high sample throughput. If the ladder-generating chemistry can be carried out in parallel on multiple samples (30), then a total throughput of >100 residues per hour could potentially be achieved at very low per cycle cost. Such a rapid, inexpensive sequencing technology of enhanced accuracy and generality could vastly expand the applications and use of protein sequence determination in biological research (31).
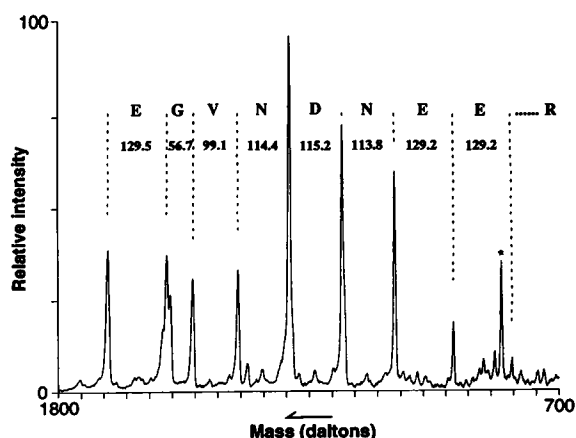
Table 1. Measured mass differences between adjacent peaks of the protein sequencing ladders shown in Fig. 4. The deviation from the calculated value is given in parentheses; Aba, α-amino-n-butyric acid.

Amino acid	Δ Mass (daltons)	Amino acid	Δ Mass (daltons)
Leu[33]	113.3 (0.1)	Asp[60]	114.8 (−0.3)
Glu[34]	129.7 (0.6)	Gln[61]	128.7 (0.6)
Glu[35]	129.5 (0.4)	Ile[62]	113.2 (0.0)
Met[36]	130.8 (−0.4)	Pro[63]	97.0 (−0.1)
Asn[37]	115.0 (0.9)	Val[64]	99.4 (0.3)
Leu[38]	112.4 (−0.8)	Glu[65]	128.6 (−0.5)
Pro[39]	97.9 (0.8)	Ile[66]	113.3 (0.1)
Gly[40]	56.1 (−0.9)	Aba[67]	84.9 (−0.2)
Lys[41]	128.1 (0.0)	Gly[68]	57.0 (0.0)
Trp[42]	186.4 (0.2)	His[69]	137.3 (0.2)
Lys[43]	128.2 (0.0)	Lys[70]	127.8 (−0.4)
Pro[44]	97.1 (0.0)	Ala[71]	71.4 (0.3)
Lys[45]	128.0 (−0.2)	Ile[72]	113.4 (0.2)
Met[46]	131.9 (0.7)	Gly[73]	56.8 (−0.2)
Ile[47]	112.6 (−0.6)	Thr[74]	101.1 (0.0)
Gly[48]	57.9 (0.9)	Val[75]	99.2 (0.1)
Gly[49]	56.3 (−0.7)	Leu[76]	113.1 (−0.1)
Ile[50]	112.4 (−0.8)	Val[77]	99.1 (0.0)
Gly[51]	57.6 (0.6)	Gly[78]	57.1 (0.1)
Gly[52]	57.5 (0.5)	Pro[79]	97.2 (0.1)
Phe[53]	147.3 (0.1)	Thr[80]	101.1 (0.0)
Ile[54]	112.5 (−0.7)	Pro[81]	97.1 (0.0)
Lys[55]	128.9 (0.8)	Val[82]	99.2 (0.1)
Val[56]	99.0 (−0.1)	Asn[83]	113.8 (−0.3)
Arg[57]	156.2 (0.0)	Ile[84]	113.4 (0.2)
Gln[58]	128.4 (0.3)	Ile[85]	113.1 (0.0)
Tyr[59]	162.6 (−0.6)	Gly[86]	57.1 (0.0)

Fig. 5. High-sensitivity protein ladder sequencing readout demonstrated by serial dilution (1 to 1000) of the sample used in Fig. 2. No more than ~25 fmol total peptide was present in the mass spectrometer, that is, <~5 fmol per component. Data were obtained as described (17). Note that only a slight diminution of signal-to-noise ratio was observed compared with readout at 25-pmol total peptide amounts (see Fig. 2) [* = impurity].

E . G . V . N . D . N . E . E .----- R
129.5 . 56.7 . 99.1 . 114.4 . 115.2 . 113.8 . 129.2 . 129.2

Relative Intensity

Mass (daltons)

REFERENCES AND NOTES

1. R. Aebersold and J. Leavitt, *Electrophoresis* **11**, 517 (1990).
2. B. Oesch *et al.*, *Cell* **40**, 735 (1985).
3. S. C. B. Yan, B. W. Grinnell, F. Wold, *Trends Biochem. Sci.* **14**, 264 (1989).
4. P. Edman and G. Begg, *Eur. J. Biochem.* **1**, 80 (1967); R. M. Hewick *et al.*, *J. Biol. Chem.* **256**, 7990 (1981).
5. N. F. Totty, M. D. Waterfield, J. J. Hsuan, *Protein Sci.* **1**, 1215 (1992).
6. In recent years, sequencing of peptides by tandem MS has shown utility for special cases not amenable to the Edman degradation, most notably for peptide chains blocked at the amino terminal and for posttranslational modifications [K. Biemann, *Annu. Rev. Biochem.* **61**, 977 (1992)].
7. M. Karas *et al.*, *Int. J. Mass Spectrom. Ion Processes* **78**, 53 (1987); R. C. Beavis and B. T. Chait, *Rapid Commun. Mass Spectrom.* **3**, 432 (1989); F. Hillenkamp *et al.*, *Anal. Chem.* **63**, 1193A (1991); B. T. Chait and S. B. H. Kent, *Science* **257**, 1885 (1992).
8. R. C. Beavis, T. Chaudhary, B. T. Chait, *Org. Mass Spectrom.* **27**, 156 (1992).
9. R. C. Beavis and B. T. Chait, *Anal. Chem.* **62**, 1836 (1990).
10. For example, a sample of the residual polypeptide chain could be taken after each cycle of Edman degradation, and examined by MS [I. Katakuse *et al.*, *Biomed. Mass Spectrom.* **9**, 64 (1982)]. The mass loss at each cycle identifies the amino acid at that position in the chain. This procedure uses a separate analysis step for each amino acid residue determined, and is a high-sensitivity modern version of "subtractive Edman" sequencing [W. Konigsberg, *Methods Enzymol.* **25B**, 326 (1972)].
11. Previous methods that generated "ragged end" peptides, including exopeptidase digestion [R. Self and A. Parente, *Biomed. Mass Spectrom.* **10**, 78 (1983); L. A. Smith and R. M. Caprioli, *ibid.*, p. 98; B. T. Chait, T. Chaudhary, F. H. Field, *Methods in Protein Sequence Analysis 1986*, K. A. Walsh, Ed. (Humana, Clifton, NJ, 1987), pp. 483–492] and uncontrolled chemical degradation [A. Tsugita, K. Takamoto, M. Kamo, H. Iwadate, *Eur. J. Biochem.* **206**, 691 (1992)], yielded less predictable results.
12. A preliminary report of this method was presented at the 40th American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics, Washington, DC, 31 May to 5 June 1992; at The Protein Society Meeting, San Diego, 20 to 25 July 1992; and at the Ninth International Conference on Methods of Protein Sequence Analysis, Otsu, Japan, 20 to 24 September 1992.
13. Residue masses and one-letter codes of the 20 genetically encoded amino acids are as follows (average isotope composition): Ala (A), 71.1; Arg (R), 156.2; Asn (N), 114.1; Asp (D), 115.1; Cys (C), 103.1; Gln (Q), 128.1; Glu (E), 129.1; Gly (G), 57.1; His (H), 137.1; Ile (I), 113.2; Leu (L), 113.2; Lys (K),

128.2; Met (M), 131.2; Phe (F), 147.2; Pro (P), 97.1; Ser (S), 87.1; Thr (T), 101.1; Trp (W), 186.2; Tyr (Y), 163.2; and Val (V), 99.1. The isomeric residues Leu and Ile have identical mass and cannot be directly distinguished. Lys and Gln are readily distinguished by the modified Lys side-chain ε-amino group formed in the chemistry steps.

14. The most efficient and accurate techniques for biopolymer sequencing are those that involve the readout of a sequence-defining data set in a single experimental operation. The data set can be examined as a whole, and anomalies can be detected and resolved. For example, as originally practiced, DNA sequencing by either (dideoxy) chain-termination [F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977)] or chain-fragmentation [A. M. Maxam and W. Gilbert, *ibid.*, p. 560] involved one-step readout followed by simultaneous inspection of the complete sequence-defining data set.

15. [Glu¹]Fibrinopeptide B was purchased from Sigma (St. Louis, MO). The reported sequence was: Glu¹-Gly-Val-Asn-Asp⁵-Asn-Glu-Glu-Gly-Phe¹⁰-Phe-Ser-Ala-Arg¹⁴. Matrix-assisted LDMS gave a mass of 1570.6 daltons (calculated, 1570.8 dalton) and showed high purity of the starting peptide.

16. A mixture of phenylisothiocyanate (PITC) plus 5% v/v phenylisocyanate (PIC) was used in the coupling step. Phenylisocyanate reacts with the αNH₂ group of a polypeptide chain to yield an Nα-phenylcarbamyl-peptide, which is stable to the conditions of degradation. A variation of manual Edman chemistry was used [G. E. Tarr, *Methods Enzymol.* **47**, 335 (1977)]. All reactions were carried out in the same 0.5-ml polypropylene microfuge tube under a blanket of dry nitrogen. Peptide (200 pmol to 10 nmol) was dissolved in 20 μl of pyridine/water (1:1 v/v; pH 10.1); 20 μl of coupling reagent containing PITC/PIC/pyridine/hexafluoroisopropanol [HFIP] (20:1:76:4 v/v) was added to the reaction vial. After reaction at 50°C for 3 min, the coupling reagents and nonpeptide coproducts were extracted by adding 300 μl of heptane:ethyl acetate (10:1 v/v) and gentle vortexing. The phases were separated by centrifugation, and the upper phase was aspirated and discarded. This washing procedure was repeated once, followed by washing twice with heptane/ethyl acetate (2:1 v/v). The remaining solution containing the peptide products was dried on a vacuum centrifuge. The cleavage step was carried out by addition of 20 μl of anhydrous trifluoroacetic acid (TFA) to the dry residue in the reaction vial and reaction at 50°C for 2 min, followed by drying on a vacuum centrifuge. Coupling-wash-cleavage steps were repeated for a predetermined number of cycles. The low molecular weight derivatives released at each cycle were not separated and analyzed. Finally, the total product mixture was subjected to an additional treatment with PIC to convert any remaining unblocked peptides to their phenylcarbamyl derivatives. The sample was dissolved in 20 μl of trimethylamine/water (25% wt/wt) in pyridine (1:1 v/v); 20 μl of PIC/pyridine/HFIP (1:76:4 v/v) was added to the reaction vial. The coupling reaction was carried out at 50°C for 5 min. The reagents were extracted as described above.

17. The product mixture was dissolved in 0.1% aqueous TFA/acetonitrile (2:1, v/v). A 1-μl aliquot (~250 pmol total peptide, assuming no losses) was mixed with 9 μl of α-cyano-4-hydroxycinnamic acid (5 g/liter in 0.1% TFA/acetonitrile, 2:1 v/v), and 1.0 μl of this mixture of total peptide products (25 pmol) 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 with a time-of-flight LDMS instrument constructed at The Rockefeller University [R. C. Beavis and B. T. Chait, *Rapid Commun. Mass Spectrom.* **3**, 233 (1989); *Anal. Chem.* **62**, 1836 (1990)]. The spectra resulting from 200 15-mJ pulses at a wavelength of 355 nm were acquired over 80 s and added to give a mass spectrum of the protein sequencing ladder. Masses were calculated with matrix peaks of known mass as calibrants.

18. Residue assignment was made by computer in an interactive fashion. First, the intact molecule ion was selected by the user. The program then searched the data for a lower mass peak corresponding to the removal of a single residue. The mass differences between the adjacent peaks were calculated and compared with a look-up table of known residue masses. If the mass difference was within set tolerances, the residue was assigned from the table, otherwise the user was asked to label the peak as an unknown, or reject it.

19. M. W. Hunkapiller, K. Granlund-Moyer, N. W. Whiteley, in *Methods of Protein Microcharacterization. A Practical Handbook*, J. E. Shively, Ed. (Humana, Clifton, NJ, 1986), pp. 223–247; J. E. Shively, R. J. Paxton, T. D. Lee, *Trends Biochem. Sci.* **14**, 246 (1989).

20. The peptide was prepared by highly optimized peptide synthesis [M. Schnölzer, P. Alewood, A. Jones, D. Alewood, S. B. H. Kent, *Int. J. Peptide Protein Res.* **40**, 180 (1992)]. The sequence was: LRRASGLIYNNPLMAR.amide. Matrix-assisted LDMS gave a mass of 1844.3 daltons (calculated, 1844.2 daltons) and showed high purity of the starting peptide. The phosphorylated form was prepared by enzymatic reaction with 3',5'-cyclic AMP–dependent protein kinase. The phosphopeptide had a mass of 1924.2 daltons (calculated, 1924.2 daltons) and showed high purity.

21. Although only 10 cycles of ladder-generating chemistry were performed, sequence-defining fragments corresponding to 11 residues were observed, apparently because of a small amount of premature cleavage [W. A. Schroeder, *Methods Enzymol.* **25**, 298 (1972); G. E. Tarr, *ibid.* **47**, 335 (1977)]. This side reaction, a potential problem for standard Edman methods, has no deleterious effect on the ladder sequencing approach.

22. Serine has a residue mass of 87.1 daltons; addition of −PO₃H₂ in place of a proton results in an additional mass increment of 80.0 daltons, for a Ser(P₁) residue mass of 167.1 daltons.

23. The ladder-generating chemistry used here has no conversion step, and is therefore considerably milder than the Edman (degradation + conversion) chemistry. The serine phosphate within the peptide chain is stable to Edman chemistry [D. B. Rylatt and P. Cohen, *FEBS Lett.* **98**, 71 (1979)]. However, the conversion step typically involves 1 M HCl in methanol or 25% trifluoroacetic acid in water for 10 min at 55° to 65°C, conditions that cause extensive decomposition of Ser(Pi) [C. G. Proud, D. B. Rylatt, S. J. Yeaman, P. Cohen, *ibid.* **80**, 435 (1977)] and other acid-sensitive residues [R. E. H. Wettenhall, R. Aebersold, L. E. Hood, *Methods Enzymol.* **201**, 186 (1991)]. Furthermore, standard extraction techniques used for existing Edman methods do not recover the polar phosphorylated amino acid derivatives for analysis, even where they are stable to the chemistry used [R. Aebersold, J. D. Watts, H. D. Morrison, J. E. Bures, *Anal. Biochem.* **199**, 51 (1991)].

24. R. C. deL. Milton, S. C. F. Milton, S. B. H. Kent, *Science* **256**, 1445 (1992).

25. Stepwise solid-phase peptide synthesis was carried out as described on a modified Applied Biosystems 430A instrument [M. Schnölzer, P. Alewood, A. Jones, D. Alewood, S. B. H. Kent, *Int. J. Peptide Protein Res.* **40**, 180 (1992)]. Samples (~1 μmol each) of butyloxycarbonyl (Boc)–peptide-resins were taken under instrument control after each coupling step. These samples were pooled in batches corresponding to residues 67 to 99, and 33 to 66. The pooled Boc-peptide-resins were deprotected and cleaved as described, extracted into aqueous acetic acid, and lyophilized. Aliquots were used for readout experiments.

26. Measurements were also obtained for peptide fragments between 7500 to 11,000 daltons (corresponding to residues 32 to 1), but the quality of the data was not sufficiently high for extended stretches of unambiguous sequence determination.

27. Although higher mass accuracy is desirable and may be achieved in the future, in many cases it is already possible to resolve ambiguities in the 3500- to 7500-dalton range by looking at the mass differences on either side of the residue in question. For example, residue 40 has a measured mass of 56.1 daltons and residue 39 has a measured mass of 97.9 daltons. This latter value has an uncertainty of roughly ±1 dalton, and thus could correspond to either Pro (97.1 daltons) or Val (99.1 daltons). There is no such ambiguity in the identity of residue 40, which must correspond to Gly (calculated mass 57.1 daltons). Correcting the measured mass difference for Gly⁴⁰ by +1.0 leads to a corrected mass difference for residue 39 of 97.9 − 1.0 = 96.9. This corresponds closely to the mass of a Pro residue. Similar corrections can be applied to the pairs Val⁶⁴/Glu⁶⁵, Gln⁵⁸/Tyr⁵⁹, Ile⁵⁴/Lys⁵⁵, and Asn³⁷/Leu³⁸. In this way, the accessible length of peptide can be extended to more than 60 residues.

28. Some limitations of the protein ladder sequencing method are shared with existing chemical sequencing methods. For example, amino-terminal–blocked samples will not be amenable to the ladder-generating chemistries.

29. P. Tempst and L. Riviere, *Anal. Biochem.* **183**, 290 (1989); S. C. Wong, C. Grimley, A. Padua, J. H. Bourell, W. J. Henzel, in *Techniques in Protein Chemistry IV*, R. H. Angeletti, Ed. (Academic Press, New York, 1993), pp. 371–378.

30. The utility of a variety of other chemical reaction systems to generate sets of sequence-defining peptide fragments has been investigated. In particular, low-level side reactions or chronic incomplete reaction in the stepwise degradation chemistry can be used to generate useful ladder data sets. Further, sequencing of samples immobilized on a solid support reduces handling losses and increases recoveries, and is particularly useful for small amounts of total sample. In this way, data have been obtained from total sample amounts as little as 10 pmol. The sequence-defining fragment sets were generated by accelerated automated Edman degradation, carried out on an ABI 471A, of a peptide sample immobilized on an ion exchange membrane. Multiple samples have been processed simultaneously. After up to 10 cycles of degradation, the peptide products were extracted from the membrane, and an aliquot (5%) used for MS readout (R. Wang, in preparation).

31. A potential application of the protein ladder principle includes carboxyl-terminal sequence analysis, where existing stepwise chemistries [A. S. Inglis, *Anal. Biochem.* **195**, 183 (1991)] have proven inadequate because of low reaction yields that lead to confusing overlaps within the first few cycles of degradation.

32. P. Edman, *Arch. Biochem.* **22**, 475 (1949); *Acta Chem. Scand.* **4**, 283 (1950).

33. Strictly, mass-to-charge ratio. However, under the conditions used, all significant components were singly charged species. The additional intense peak * at 1552.6 daltons corresponds to the blocked peptide [(pyrrolidonecarboxylic acid)¹] fibrinopeptide B, an artifact formed from cyclization of the amino-terminal Glu. Such side reactions are readily identifiable and do not interfere with the sequence determination.

34. The digitized time-intensity data were converted to this graphical form with an 8-bit gray scale.

35. Loss of phosphate by hydrolysis (−80 daltons) or by elimination to form dehydro-alanine (−98 daltons) was *not* detected. A low-level side reaction, unrelated to the modified amino acid, was observed in these experiments; this gave rise to a series of peaks at −93 daltons relative to the main series. The origin of this artifact is under investigation.

36. Supported by grants from the National Institutes of Health [NIH RR00862 and GM38274), and by funds from the Markey Foundation. We thank R. deL. Milton and S. Milton for the HIV-1 PR-derived peptide-resin samples, and S. Walker for synthesis of the unphosphorylated peptide used in Fig. 3B.

11 May 1993; accepted 27 July 1993