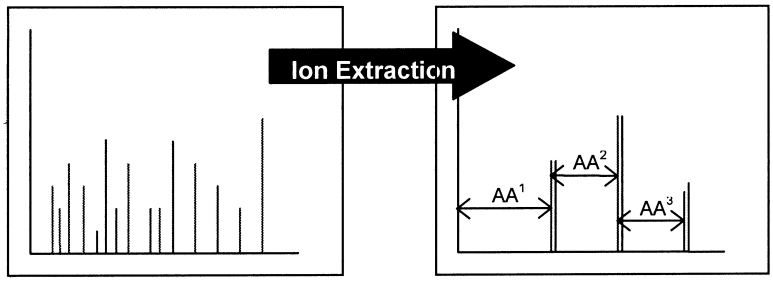
JOURNAL OF combinatorial CHEMISTRY

Article

Ion-Extraction Ladder Sequencing from Bead-Based Libraries

Jose J. Pastor, Iain Lingard, Gurdip Bhalay, and Mark Bradley

J. Comb. Chem., **2003**, 5 (2), 85-90• DOI: 10.1021/cc0200437 • Publication Date (Web): 25 January 2003



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 3 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

Journal of Combinatorial Chemistry is published by the American Chemical Society.
1155 Sixteenth Street N.W., Washington, DC 20036

Articles

Ion-Extraction Ladder Sequencing from Bead-Based Libraries

Jose J. Pastor,[†] Iain Lingard,[‡] Gurdip Bhalay,[§] and Mark Bradley*[,‡]

Departament de Química Orgànica, Universitat de Barcelona, Martí i Franquès, 1-11, 08028 Barcelona, Spain, Department of Chemistry, University of Southampton, Southampton, SO17 1BJ, U.K., and Novartis Pharmaceuticals, Wimblehurst Road, Horsham, West Sussex, RH12 5AB, U.K.

Received June 15, 2002

Ion-extraction mass spectrometry of ladders of mixtures of isotopically labeled compounds from single beads allows the unambiguous sequencing of bead-based peptides and offers significant advantages over traditional methods of library analysis.

Introduction

Since the introduction of combinatorial "split-and-mix" solid-phase synthesis by Furka et. al. in 1991,[1] peptide libraries have been important for the discovery of active peptides for a variety of biological applications. Early developments included the discovery of ligands to a monoclonal antibody against β-endorphin[2] (although resin accessibility issues have always been an issue of contention) and peptide based antimicrobials.[3] After the synthesis of a peptide library and selection of beads with the desired biological activity or binding affinity, it is necessary to obtain the sequence of the resin bound peptide. Edman sequencing is widely used and can be routinely performed using commercially available instruments. This technique is not without its drawbacks, however; it is time-consuming and expensive, limiting the number of samples that can be conveniently analyzed. In addition, it is a requirement of the technique that the N terminus of the peptide is free, while only peptides containing conventional α-amino acid residues can be sequenced, seriously limiting the complexity of the libraries that can be prepared.

An elegant solution to these problems was reported by Youngquist and Keough[4,5] with the introduction of the concept of ladder sequencing. In this method, a small portion of the growing library was capped (in the original paper with *N*-acetylalanine) to give at the end of the synthesis a family of peptides with the mass difference between the family members equating to the mass of the amino acid component. Once active beads have been selected and picked from the library, the peptide is cleaved from the resin along with the capped fragments and the mixture analyzed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF MS), allowing the full peptide sequence to be read from the N terminus by the mass difference between adjacent peaks from high mass to low mass. Modification of this technique has been reported by several groups and has included replacement of the capping step by the partial incorporation of a methionine residue at each coupling stage (cleavable by cyanogen bromide),[6] while a recent paper by Griesinger et. al. describes the use of dual capping groups and analysis of molecular ion redundancy to enable direct structure elucidation by MS-based methods.[7]

In this paper, a new analysis method for the sequencing of compounds selected from "split-and-mix" combinatorial libraries is described and is a major improvement on Youngquist's original method. The analysis method is adopted from that reported by Lane and Pipe,[8] who used a combination of stable isotope labeling and mass spectrometry cluster analysis. Data reduction software is in essence used to select mass spectrometry signals containing only the signature corresponding to the incorporated isotope mixture. This process selects only compounds from the cleavage mixture that have been tagged and therefore gives an effectively "infinite signal-to-noise ratio". It also allows the analysis of very small quantities of compound from samples containing large amounts of impurities, because only the compounds of interest are shown in the processed spectrum. This technology is ideal for the application to mass spectrometry-based peptide ladder sequencing to allow the selection of the ladder signals from the background noise produced by cleavage reagents, protecting group fragments, scavengers, and the full length peptide.

To develop the methodology, the peptides were synthesized using a methionine linker with either a bis- or tris-6-aminohexanoic acid spacer to separate the peptide from the resin and increase the overall mass of the cleaved species to ensure that the smaller peptide ladder fragments have a mass greater than the "chemical noise". During library generation, after each round of peptide coupling, the growing peptide

* To whom correspondence should be addressed. E-mail: mb14@soton.ac.uk.
[†] Universitat de Barcelona.
[‡] University of Southampton.
[§] Novartis Pharmaceuticals.

10.1021/cc0200437 CCC: $25.00
Published on Web 01/25/2003

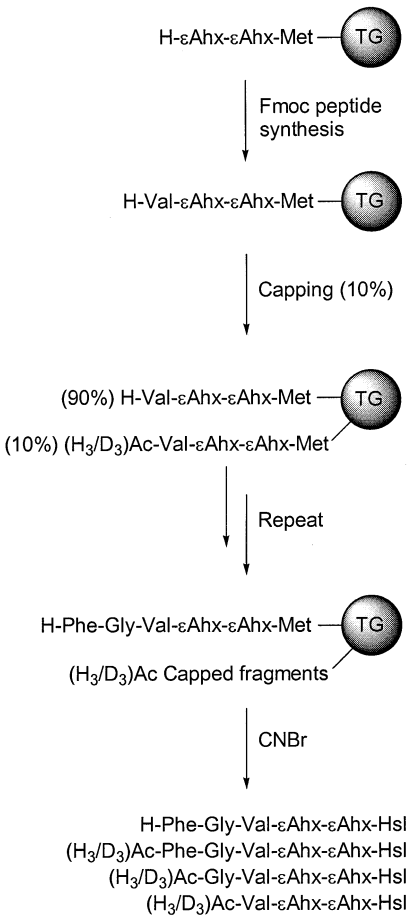


Figure 1. Synthesis and cleavage of the test peptide and its family of capped ladder fragments.

chain was partially terminated by reaction with an isotopically labeled capping group (a 1:1 mixture of acetic acid and D_3-acetic acid), simultaneously generating the peptide ladder fragments and providing the isotope labels required for software-based signal selection. Following cleavage from the resin, MALDI-TOF mass spectrometry and cluster-analysis-based data reduction lead to a mass spectrum in which only the peaks corresponding to the peptide ladder are visible, allowing simple determination of the peptide sequence.

Results and Discussion

Generation of Test Peptide Libraries. To validate the methodology, the peptide H-Phe-Gly-Val- was synthesized using Fmoc peptide chemistry on the linker spacer construct ϵAhx-ϵAhx-Met on TentaGel S resin (Figure 1), capping as described above. It was observed that the use of DIC/HOBt as the activating agent gave unsatisfactory incorporation of the capping agent, as did the anhydrides in the presence of triethylamine. Capping was facilitated by the use of PyBOP,[9] which could be used in excess, since PyBOP does not form the N-terminal guanadinium side products observed with carbodiimides and uronium-based reagents.[10] After peptide synthesis, methionine cleavage was accomplished by treatment of single beads with cyanogen bromide,[11] generating a homoserine lactone (Hsl) and releasing the peptides from the resin.

MALDI Sequencing of the Test Peptide. Following cleavage from the resin, samples were analyzed by MALDI-TOF mass spectrometry. A typical spectrum and its analysis are shown in Figure 2a,b. The initial spectrum (Figure 2a) shows the unprocessed spectrum and clearly shows little useful information about the peptide structure, whereas in Figure 2b, the spectrum has been reprocessed using a cluster analysis algorithm, and only peaks that have a partner peak of approximately equal intensity 3.02 Da apart (mass of 3D − 3H = 3.018) are displayed. It can be clearly seen that this process completely removes the noise from the spectrum, leaving only the peaks due to the capped peptide ladder fragments. The sequence of the peptide can therefore be identified from the mass difference between the adjacent pairs of peaks, from the N terminus at highest mass to the amino acid attached to the spacer−linker construct at low mass.

Synthesis and Sequencing of Libraries. Three libraries were synthesized to test the methodology. The linker−spacer construct was in these cases extended by an extra ϵAhx residue to ϵAhx-ϵAhx-ϵAhx-Met to shift the spectral envelope further from the background noise we had observed in the low-mass region of the mass spectrum.

(1) Library 1. A library of tripeptides was prepared. In the first cycle, the amino acids used were Val, Phe, and Ala; in the second cycle, Asn, Ser(O^{t}Bu) or Gln; and in the third, Asp(O^{t}Bu), Tyr(O^{t}Bu), and Trp(Boc) to give 27 compounds. Single beads were randomly selected from each library, cleaved, and sequenced to give the data shown in Figure 3.

(2) Library 2. The more difficult residues Arg(Pmc), Trp(Boc), or Asn were incorporated at each of the three coupling cycles, generating a 27-member library constructed solely from these three amino acids. Again, beads were randomly selected from the library, cleaved, and sequenced (Figure 4).

(3) Library 3. Capping with two labeled capping groups was investigated to look at distinguishing isobaric residues. Therefore, the sequence X-Gly-Val-ϵAhx-ϵAhx-ϵAhx-Met-TentaGel was synthesized, where X was one of the following amino acids: Lys(Boc), Gln, Leu, Ile, Asp(O^{t}Bu), or Asn. After coupling one member of each isobaric pair, (Lys, Leu, or Asp) or (Ile, Gln, or Asn), 10% capping was carried out with a 1:1 mixture of either acetic/D_3-acetic acid or benzoic/D_5-benzoic acid. Thus, on sequencing members of this library, two separate cluster analyses were run, one as before with a mass difference of 3.02 Da and the second with a mass difference of 5.03 Da (5D − 5H = 5.03). If the first cluster analysis identifies only two of the amino acids, the third must be identified from the second analysis. Single beads were randomly selected from the library, cleaved, and sequenced (Figure 5).

Details of the results obtained from the sequencing of the three libraries are shown in Table 1. Every sample attempted from library 1 was successfully sequenced, including one sample that had contained the cleavage products from two beads (the two different ladders were easily deconvoluted). Library 2 was sequenced with an 85% success rate. The four failed samples gave partial sequences, but it was not possible to unequivocally denote the sequence. All samples taken from library 3 were successfully sequenced.

Several peptide modifications were observed during the screening. Arg and Ser were often observed in both the

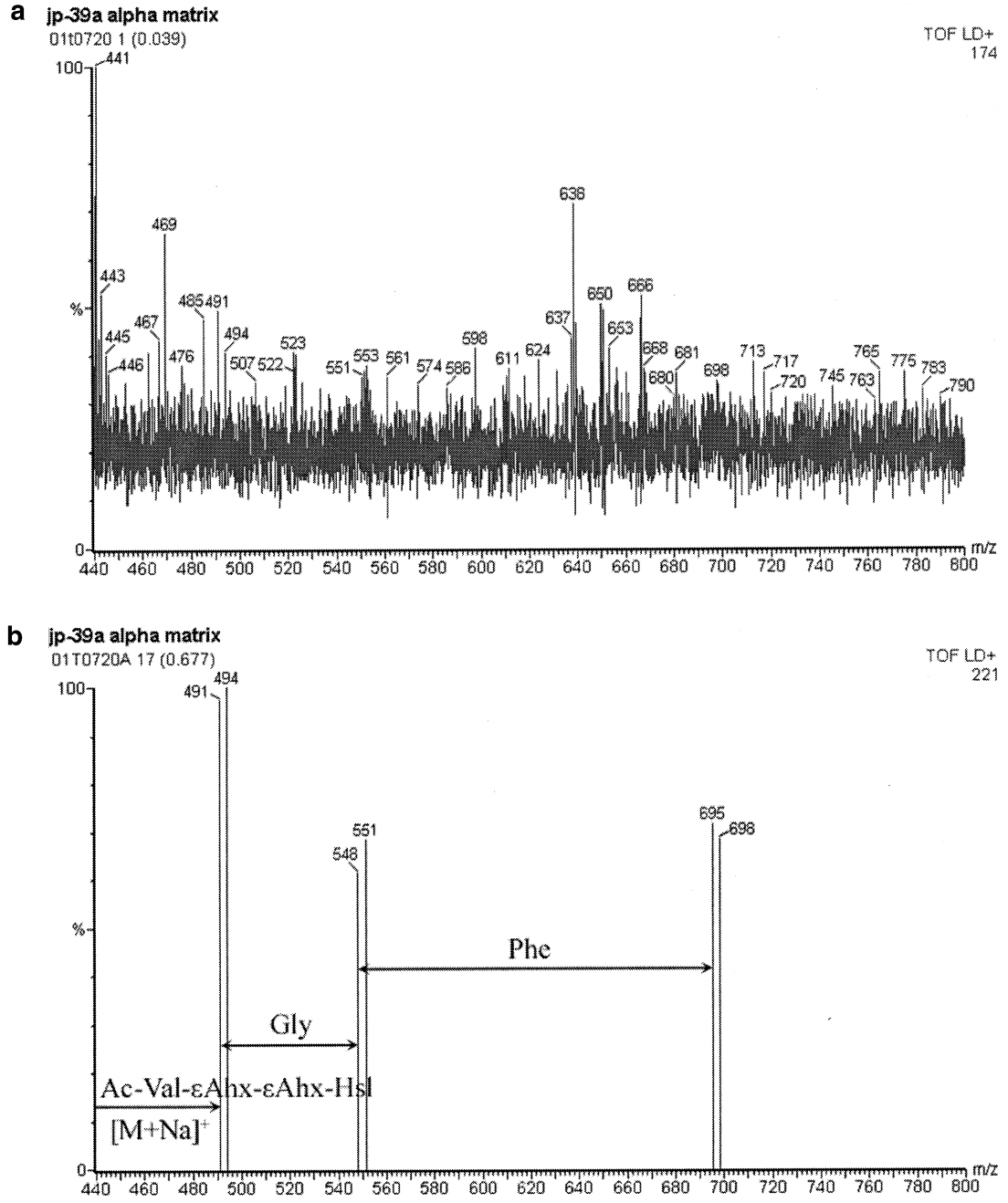


Figure 2. (a) Raw MALDI-TOF trace of cleaved test peptide and its capped ladder fragments. (b) Reprocessed MALDI-TOF spectrum of cleaved test peptide showing the ladder peaks for each amino acid. The peaks observed are 695 (Ac-Phe-Gly-Val-ϵAhx-ϵAhx-Hsl [M + Na]), 548 (Ac-Gly-Val-ϵAhx-ϵAhx-Hsl [M + Na]), and 491 (Ac-Val-ϵAhx-ϵAhx-Hsl [M + Na]).

protected and unprotected forms, and Trp was often observed as the *N*-oxide. This oxidation of tryptophan is a commonly observed modification during MALDI analysis, as reported by Youngquist et al.[5]

Conclusions

An isotope-labeling, chain-termination strategy generates peptide ladders that are suitable for MALDI-TOF-MS sequencing analysis. These data, in association with ion extraction and cluster analysis, give molecular ion information with an excellent signal-to-noise ratio in a highly efficient manner. This allows the rapid sequencing of peptides from beads selected from a split-and-mix peptide library. The advantages of this technique over other methods of peptide sequencing are principally those of time, cost, and reliability. Using a 100-well MALDI sample plate to introduce samples into the mass spectrometer allows the samples to be analyzed at a rate of ~2 min/sample. Data reduction and ladder analysis is easy and quick, leading to a total sequencing time of ~5 min/sample. In comparison, Edman sequencing of peptides takes typically 1−2 h/residue and is expensive, an issue when many beads have to be analyzed, and other MS-based methods can often be problematic as a result of noise. No specialist equipment is required for this protocol, which can be performed on a standard laboratory MALDI-TOF spectrometer with industry standard MS analysis software, and the method is also highly reliable. Peptides in this series of experiments were capped

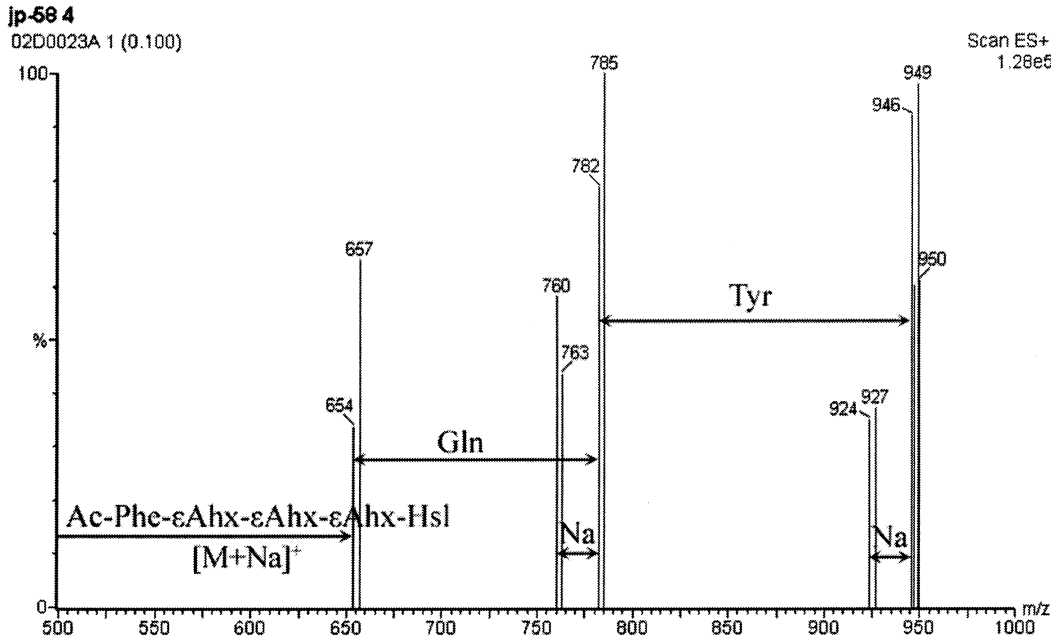


Figure 3. Sample bead from library 1, processed, and assigned. The sequence of the peptide is Tyr-Gln-Phe.

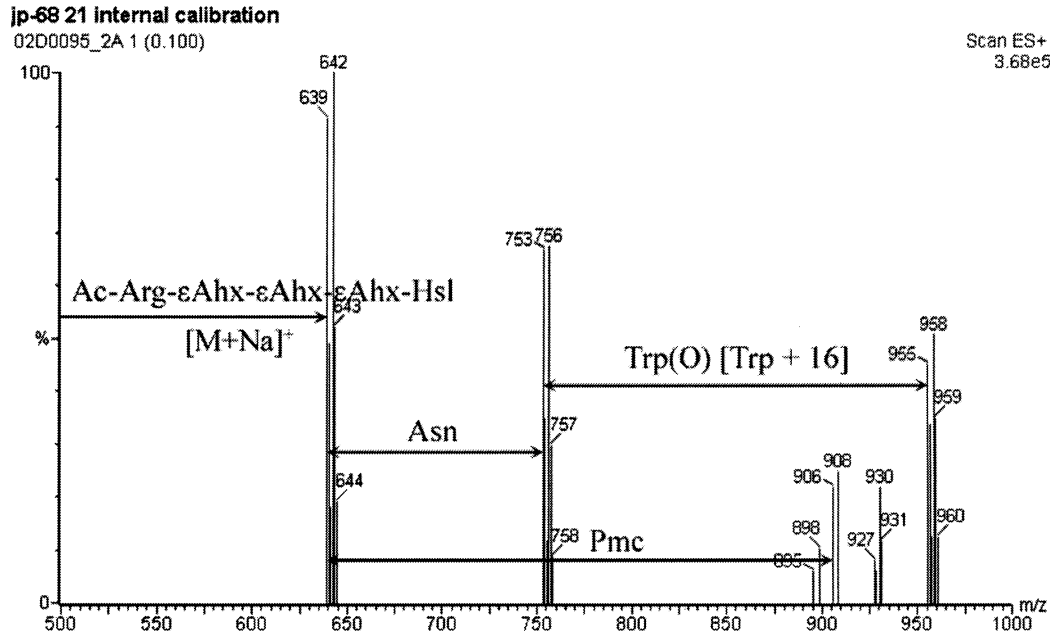


Figure 4. Sample bead taken from library 2, processed, and assigned. The sequence is Trp-Asn-Arg. In addition to the peaks due to the amino acids, peaks due to incomplete removal of the Pmc group from Arg and oxidation of the Trp are also observed.

at a level of 10%. This is a level that would be clearly unsuitable for longer peptides. However, MALDI-TOF mass spectrometry is a highly sensitive technique that allows the detection of compounds at the fmol level, and a single bead may contain ~400 pmol of sites. Thus, capping at a level of 5% would give 10 pmol of each capped fragment pair, which is well within the detection limits of the technique and would allow a hexamer to be prepared and capped with the loss of only 30% of the total sites. That reduced levels of capping are possible can be seen from the excellent signal-to-noise ratio in Figures 3 and 5. In addition, Figure 4 shows not only the fully deprotected peptide series but also related peptides from this series with side-chain protection intact (Pmc) and tryptophan oxidation. Thus, the 10% capping has been spread among a series of molecular ions, but all were assignable and had good signal-to-noise. The level of tagging, we believe, could be reduced to 1−2%. The major limitations are not really MS-based but are more experimental with respect to the accurate capping of resin sites. This MS method has, we believe, the potential to reduce the effort and errors associated with current single-bead MS analysis.

Experimental Section

Peptide Synthesis: Materials and Methods. Peptides were synthesized manually using standard Fmoc chemistry[12] using *N*-hydroxybenzotriazole (HOBt) and *N*,*N*′-diisopropylcarbodiimide (DIC). Capping was realized using a 1:1 mixture of acetic acid/D_3-acetic acid or benzoic acid/D_5-benzoic acid, as described below, using benzotriazole-1-yl-

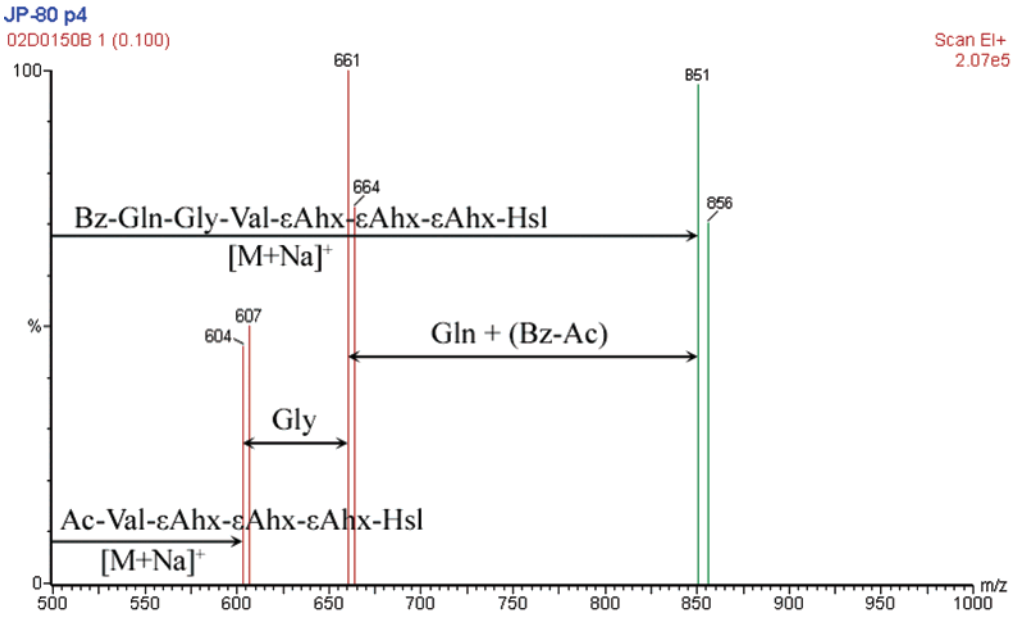


Figure 5. Sample bead from library 3, combination of two cluster analyses. Red peaks show analysis with a mass difference of 3.02. Only the first two amino acids are seen in the processed spectrum. The terminal amino acid, therefore, must be labeled with benzoic acid. Green peaks show analysis with a mass difference of 5.03. This gives the identity of the terminal amino acid as Gln; therefore, the total sequence is Gln-Gly-Val.

Table 1. Results obtained in the sequencing of the three peptide libraries

	library		
	1	2	3
library size	27	27	6
samples analyzed	30	27	8
sequences obtained	30	23	8
% sequencing success	100	85	100

oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBop) and *N*,*N*-diisopropylethylamine (DIPEA).

Synthesis of Test Peptide 1. TentaGel S NH_2 resin (0.27 mmol/g, 500 mg, 0.135 mmol) was used as the solid support for the synthesis. The spacer linker construct ϵAhx-ϵAhx-Met was synthesized by the addition of a 5-fold excess of Fmoc-amino acid in DMF (5 mL), preactivated for 10 min by treatment with DIC and HOBt (5 equiv of each), to the resin that had been preswollen in DCM (5 mL), followed by shaking for 2 h. After confirmation of reaction completion by the Kaiser ninhydrin test,[13] and thorough washing, the Fmoc group was removed by treatment with 20% piperidine in DMF (5 mL, 5 min repeated for 10 min), followed by thorough washing. The amino acids Fmoc-Val, Fmoc-Gly, and Fmoc-Phe were then coupled sequentially using the same method. After each coupling−deprotection step, partial capping was realized by shaking the resin with 0.1 equiv of a 1:1 mixture of AcOH/D_3-AcOH, followed by the addition of 0.5 equiv of PyBop and DIPEA for 45 min. The resin was then thoroughly washed with DMF (3 × 5 mL), DCM (5 × 5 mL), and methanol (3 × 5 mL), dried under vacuum, and stored at −10 °C.

Synthesis of Peptide Library 1. The resin was substituted with the spacer−linker construct ϵAhx-ϵAhx-ϵAhx-Met as described above. Split-and-mix synthesis was then used to generate a library of 27 tripeptides using the coupling−deprotection−capping strategy described above. The amino acid used in the first coupling was Fmoc-Val, Fmoc-Phe, or Fmoc-Ala; in the second position, Fmoc-Asn, Fmoc-Ser(O^{t}Bu), or Fmoc-Gln; and in the third position, Fmoc-Asp(O^{t}Bu), Fmoc-Tyr(O^{t}Bu), or Fmoc-Trp(Boc). Side-chain protecting groups were removed by treatment for 1 h with a mixture of TFA (1.6 mL), H_2O (0.2 mL), TIPS (0.1 mL), and phenol (100 mg). The resin was then thoroughly washed with DCM (5 × 5 mL) and methanol (3 × 5 mL), dried under vacuum, and stored at −10 °C.

Synthesis of Peptide Library 2. A 27-member tripeptide library was generated on the spacer-linker construct ϵAhx-ϵAhx-ϵAhx-Met described above using the three Fmoc amino acids Arg(Pmc), Trp(Boc), or Asn and capped as above. Side-chain deprotection was realized as above with a 5-h treatment.

Synthesis of Peptide Library 3. The resin was substituted with the peptide−spacer−linker construct Gly-Val-ϵAhx-ϵAhx-ϵAhx-Met partially capped as described above, then split into six portions, and each was coupled with one of the following Fmoc amino acids: Lys(Boc), Leu, Asp(O^{t}Bu), Ile, Gln, or Asn. The portions substituted with Lys, Leu, or Asp were then partially capped with AcOH/D_3-AcOH as above; the portions substituted with Ile, Gln, or Asn were capped by the same method with a 1:1 mixture of benzoic acid/D_5 benzoic acid. The portions were then mixed, and side-chain deprotection was carried out as above.

Cleavage of Peptides from Single Beads. Single beads were picked randomly with a micropipet under 50× magnification and placed into 100-μL capacity glass vial inserts inside 500-μL microcentrifuge tubes. A 20-μL portion of a 20 mg/mL solution of cyanogen bromide in 1:1 H_2O/TFA was added, and the tubes were sealed, sonicated for 5 min, and then allowed to react in the dark for 24 h. The samples were then centrifuged under vacuum (50 °C) until dry. A 10-μL portion of HPLC grade MeCN was added, and the

samples were sonicated for 5 min, then centrifuged at 5000 rpm for 5 min.

MALDI Sample Preparation. A 2 μL portion of each sample was applied to the MALDI sample plate and allowed to dry. A further 1 μL of sample and 1 μL of a saturated solution of α-cyano-4-hydroxycinnamic acid in 1:1 MeCN/H_2O containing 0.1% TFA was then applied on top, thoroughly mixed, and allowed to dry.

MALDI Mass Spectrometry. All spectra were recorded in the positive ion mode on a Dynamo MALDI-TOF spectrometer (Thermo BioAnalysis, Hemel Hempstead, U.K.). Spectra were obtained by summing multiple laser shots. Terfenadine (*m*/*z* 472.32 [M + H]) and Renin substrate tetradecapeptide (Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-His-Leu-Leu-Val-Tyr-Ser (*m*/*z* 1760 [M + H])) were used as external and internal calibrants, respectively. Peptide ladder fragments were generally observed as $[M + Na]^+$ adducts of the lactone.

Data Analysis. Sequences were obtained by analysis of raw spectra using MassLynx v3.2 (Micromass Ltd., Manchester, U.K.). The cluster analysis function was used with the following parameters: First mass difference, 3.02 or 5.03; first ratio, 1:1; mass tolerance, 0.1; ratio tolerance, 80%; threshold, 0.1−1%.

Acknowledgment. We thank Dr. John Langley and Miss Julie Herniman of the University of Southampton for their advice and assistance with the mass spectrometry. This research was supported in part by an industrial CASE studentship from Novartis the EPSRC, the JIF initiative, and also by an FI-FPI grant (CIRIT) from the Generalitat de Catalunya (J.P.).

References and Notes

(1) Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487−493.
(2) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354*, 84−86.
(3) Houghten, R. A.; Dinh, K. T.; Burchin, D. E.; Blondell, S. E. In *Techniques in Protein Chemistry IV*; Amgeletti, R. H., Ed.; Academic: San Diego, CA, 1993; pp 249−256.
(4) Youngquist, R. S.; Fuentes, G. R.; Lacey, M. P.; Keough, T. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 77−81.
(5) Youngquist, R. S.; Fuentes, G. R.; Lacey, M. P.; Keough, T. *J. Am. Chem. Soc.* **1995**, *117*, 3900−3906.
(6) Davies, M.; Bradley, M. *Tetrahedron Lett.* **1997**, *38*, 8565−8568.
(7) Hoffmann, C.; Blechschmidt, D.; Krüger, R.; Karas, M.; Griesinger, C. *J. Comb. Chem.* **2002**, *4*, 79−86.
(8) Lane, S. J.; Pipe, A. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 782−793.
(9) Coste, J.; Le-Nguyen, D.; Castro, B. *Tetrahedron Lett.* **1990**, *31*, 205−208.
(10) Albericio, F.; Cases, M.; Alsina, J.; Triolo, S. A.; Carpino, L. A.; Kates, S. A. *Tetrahedron Lett.* **1997**, *38*, 4853−4856.
(11) Hancock, W. S.; Marshall, G. R. *J. Am. Chem. Soc.* **1975**, *97*, 7488−7489.
(12) Chan, W.; White, P. *Solid-Phase Peptide Synthesis*; Oxford University Press: Oxford, U.K., 1999.
(13) Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I. *Anal. Biochem.* **1970**, *34*, 595.

CC0200437