Report

An Improved Method for Rapid Sequencing of Support-Bound Peptides by Partial Edman Degradation and Mass Spectrometry

Michael C. Sweeney, and Dehua Pei

J. Comb. Chem., **2003**, 5 (3), 218-222• DOI: 10.1021/cc020113+ • Publication Date (Web): 19 April 2003

An Improved Method for Rapid Sequencing of Support-Bound Peptides by Partial Edman Degradation and Mass Spectrometry

Michael C. Sweeney and Dehua Pei*

Department of Chemistry and Ohio State Biochemistry Program, The Ohio State University, 100 West 18th Avenue, Columbus, Ohio 43210

Received December 12, 2002

Peptide libraries have found widespread applications in defining the substrate specificity of enzymes,[1] identifying specific ligands of protein modules or receptors,[2] the design of enzyme inhibitors,[3] and the development of new catalysts.[4] Frequently, peptide libraries are synthesized and screened on the solid phase in the one-bead-one-peptide format.[5] This results in a large number of positive beads that need to be individually characterized, requiring a rapid and inexpensive method for high-throughput sequencing of library derived peptides. Several methods have been used for this purpose, including conventional Edman sequencing,[1c,e,f,2a,c,e] the use of encoding tags,[6,7] and tandem mass spectrometry.[8] Each of these methods has some disadvantages,[9] compromising their utility in the routine high-throughput sequencing of library-derived peptides. We[9] recently reported a partial Edman degradation method that converts a support-bound peptide into a series of sequence-related truncation products (a peptide ladder) by repeated treatment of the support-bound peptide with a 10:1 mixture of phenyl isothiocyanate (PITC) and phenyl isocyanate (PIC).[10] Subsequent analysis of the peptide ladder by matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) reveals the sequence of the original peptide. This method is rapid, sensitive, and inexpensive. However, during the subsequent application of this method to other peptide libraries, we and others have frequently experienced low success rates in obtaining the full-length sequences. We have since undertaken a systematic study of the method with the intent of finding better reagents and reaction conditions to increase the success rate. Herein we report an improved method that has resulted from this study.

In the original method,[9] positive beads selected from a combinatorial library (typically 50−100 beads) are pooled and treated with a 10:1 (v/v) mixture of PITC and PIC in pyridine/water; both reagents react with the N-terminal amine of the support-bound peptides. After washing to remove the excess reagents, the beads are treated with trifluoroacetic acid (TFA). For ∼90% of the peptides that react with PITC, TFA treatment results in the cleavage of the N-terminal amino acid (Edman degradation). For the other ∼10% peptides that react with PIC, no cleavage is possible under the conditions used, and these peptides become permanently N-blocked by a phenylcarbamoyl group (N-capping). Repetition of the partial Edman degradation reaction for *n* cycles (*n* is the number of residues to be sequenced) results in a series of sequence-related truncation products. Next, the treated beads are individually treated with CNBr to cleave the peptides off the resin and are analyzed by MALDI-TOF mass spectrometry.

At first glance, the Edman degradation/N-capping ratio of 9:1 seemed to be caused by the 10:1 ratio of PITC/PIC, as originally employed by Chait et al.[10] However, because PIC is much more reactive toward nucleophiles than PITC, a much higher percentage of N-capping would have been expected during each cycle of degradation. We have found that the reaction solvent, which typically contained 50% pyridine in water, can dramatically affect the product ratio by competing with the N-terminal amine of peptides for reaction with PIC and PITC. Thus, the amount of N-capping is dependent on not only the PITC/PIC ratio, but also a host of other factors, including the solvent composition, the nature and number of amines present in a peptide, and the reaction temperature. Consequently, the optimal degradation condition, which should produce 5−10% N-capping at each cycle, may differ from one peptide to another, or for the same peptide, from one position to the next. Failure to obtain the optimal degradation condition results in either too much N-capping, thus, the absence of low-molecular-weight fragments, or too little N-capping and the absence of high-molecular-weight species. To arrive at the optimal conditions, multiple trials have to be conducted with different PITC/PIC ratios as well as solvent compositions. Once the optimal conditions are found, a success rate of >90% has been achieved within the same batch of beads (usually 50−100 beads are degraded in one pot).[9] However, when applied to a different batch of beads from the same library, these "optimal" conditions may fail to produce MS spectra of sufficient quality for unambiguous sequence determination.

To improve the reliability of the technique, we reasoned that a better capping agent must be found. An ideal capping agent should have the following properties. First, it should react only with the N-terminal amine of peptides but not with the solvent (e.g., pyridine or water). Second, it should not react with any amino acid side chains, or if it does, it should result in complete modification of that side chain. Finally, it should have a reactivity toward peptidylamines that is similar to that of PITC. With such a capping agent, the fraction of N-capping should generally be determined by the ratio of PITC and the capping agent. Among the many reagents that have been tested, we found that *N*-hydroxysuccinimidyl benzoate (Bz-NHS) meets most of the above requirements. At near neutral pH, NHS esters react readily with free amines at a rate comparable to PITC, but only slowly with water, and no side chain modifications were detected, with the exception of lysine. Last, the molecular weight of the added benzoyl cap (104 mass units) figures nicely with the *N*-Ac-Gly and *N*-Ac-Ala (99 and 113,

* Corresponding author. Phone: (614) 688-4068. Fax: (614) 292-1532. E-mail: pei.3@osu.edu.

10.1021/cc020113+ CCC: $25.00
Published on Web 04/19/2003

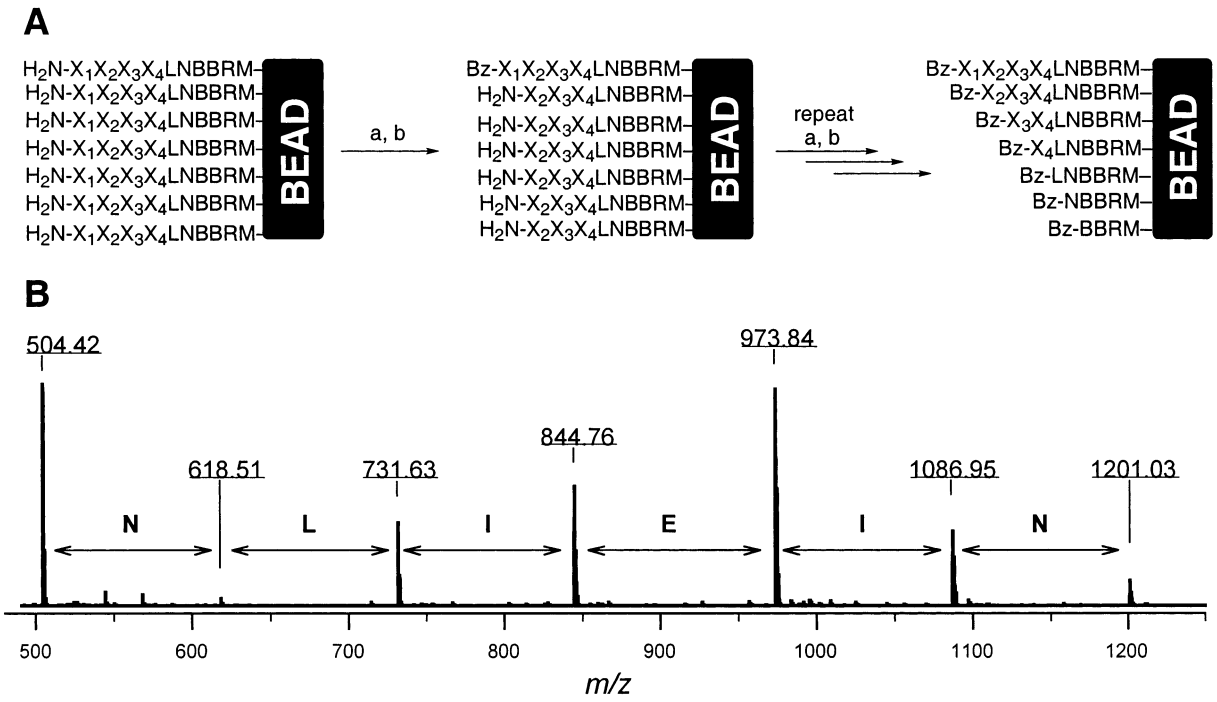

Figure 1. Sequencing of support-bound peptides by partial Edman degradation and mass spectrometry. (A) Reactions involved in partial Edman degradation. (a) 3:2 PITC/Bz-NHS in 4:1 pyridine/water; (b) TFA. Bz, benzoyl. (B) MALDI mass spectrum of a peptide and its truncation products obtained with 7% of the material from a single 130-μm bead. The sequence of the unknown peptide is NH_2-NIEILNBBRM. M, methionine prior to CNBr cleavage and homoserine lactone after CNBr cleavage.

respectively) capping agents to produce characteristic doublets used to distinguish the degenerate amino acids (vide infra).

The modified method was first applied to sequence a peptide library containing four randomized positions, H_2N-XXXXLNBBRM-resin (X = norleucine or any of the 18 α-amino acids, except for cysteine and methionine; B = β-alanine), synthesized on TentaGel S-NH_2 resin (130 μm) by the split-synthesis method.[11] Partial Edman degradation reactions were typically carried out with ~50 beads in a manner similar to that described above, except that the reagents contained a 3:2 (mol/mol) mixture of PITC and Bz-NHS dissolved in 4:1 (v/v) pyridine/water. Repetition of the degradation/capping procedure for six additional cycles permitted the sequence determination of the four randomized residues as well as the first two defined residues, Leu-Asn (Figure 1a). Since each bead carried a unique full-length peptide and six truncation products, MALDI analysis of the cleavage mixture generated a peptide ladder containing seven individual peaks. Figure 1b shows a typical MALDI spectrum, obtained with 7% of the peptide mixture isolated from a single resin bead. The protonated full-length peptide gave a peak at *m*/*z* 1201.03. The truncation products produced four peaks at *m*/*z* 1086.95, 973.84, 844.76, 731.63, 618.51, and 504.42. The mass difference between the full-length peptide and the first truncation product (*m*/*z* 1086.95) is 114.08, indicating that the N-terminal residue of the peptide is Asn. Likewise, the mass difference between the first and second truncation products is 113.11, which is the residue weight of Leu, Ile, or Nle. The absence of a second peak near the *m*/*z* 973.84 peak indicates that the penultimate residue must be Ile, because Leu and Nle would appear as doublets separated by −5 or +9 mass units, respectively, as a result of the addition of encoding tags during the synthesis of Leu and Nle.[9] On the basis of the mass differences between adjacent peaks in the ladder, the rest of the sequence was determined as Glu, Ile, Leu, and Asn, respectively.

To test the reliability of the modified method, we performed the same degradation reactions on five batches of beads from the above library in five separate experiments (~50 beads in each batch). Unambiguous sequences were obtained for >92% of the beads in each batch, and the overall success rate was 95% (Table 1). For the 5% of the beads that did not yield complete sequences, some had spectra devoid of any significant signals. This suggests that the cause of failure was likely due to poor ionization during the automated MALDI analysis or failure during library synthesis. Indeed, we were able to reanalyze many of the beads to obtain complete sequences (in the case of failed ionization). In another experiment, the new method was applied to sequence a phosphotyrosine (pY)-containing peptide library, H_2N-XXpYXXXLNBBRM-resin. Out of the 24 beads analyzed by MALDI MS, 18 (75%) produced high-quality spectra and complete sequences for the six-residue randomized region. The incomplete spectra of the six remaining peptides missed part or all of the pY-containing fragments, likely because of the lower ionization efficiency of pY peptides in the positive ion mode. We later found by serendipity that the use of a nicotinoyl group (instead of benzoyl) can significantly improve the quality of the MALDI MS spectra. For yet unknown reasons, the N-terminally nicotinoylated peptides produced much "cleaner" spectra and higher signal-to-noise ratio. Thus, by performing partial Edman degradation with *N*-hydroxysuccinimidyl nicotinate (Nic-NHS) in place of Bz-NHS, we were able to increase the signal-to-noise ratio of the pY-containing fragments in

Table 1. Success Rate for Sequencing Support-Bound Peptides by Partial Edman Degradation

peptide library	trial	no. of beads analyzed by MS	no. of complete sequences obtained (%)
H_2N-XXXXLNBBRM-resin	1[a]	48	44 (92)[c]
	2[a]	45	42 (93)[c]
	3[a]	46	46 (100)[c]
	4[a]	48	45 (94)[c]
	5[a]	48	46 (96)[c]
H_2N-XXpYXXXLNBBRM-resin	6[a]	24	18 (75)
	7[b]	24	22 (92)

[a] Bz-NHS was used as the capping agent. [b] Nic-NHS was used as the capping agent. [c] Complete sequences for the four-residue randomized region only.

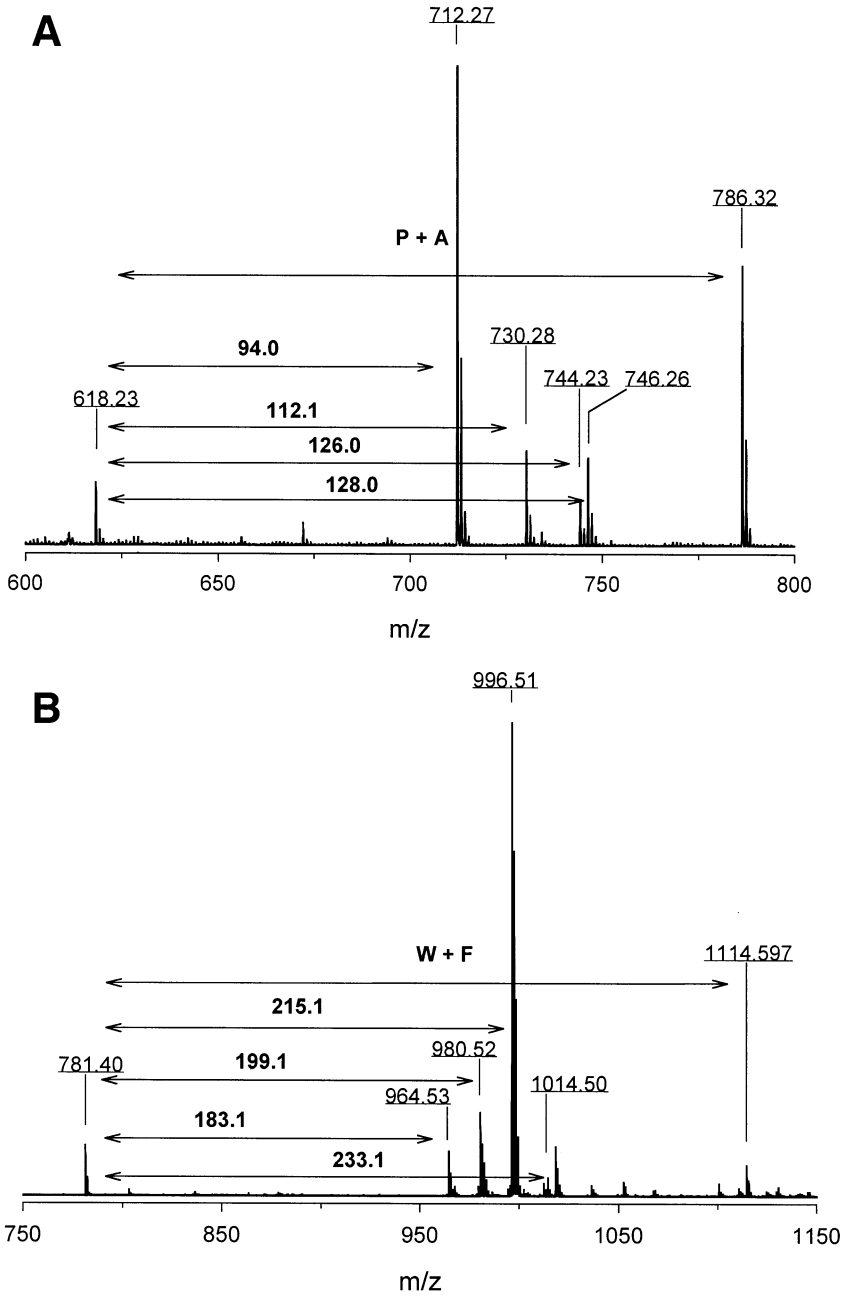

Figure 2. MALDI MS spectra of the peptide ladders derived from proline- (A) and tryptophan-containing peptides (B). In spectrum A, instead of a single peak at *m*/*z* 715, proline produces a group of four peaks at 94, 112, 126, 128 mass units from the *m*/*z* 618 peak. In spectrum B, tryptophan produces three peaks at 183, 199, and 215 mass units from the *m*/*z* 781 peak, instead of the expect peak at *m*/*z* 967.

the MALDI spectra and, consequently, the sequencing success rate to 92% (Table 1). Finally, we performed degradation reactions with a reduced (10) or an increased number of beads (~100) in the vessel under identical conditions to test the scalability of the method. Similar success rates (>90%) were obtained in each case. All together, we have used the new reagents to sequence nearly a dozen batches of beads (~500 beads total) and have

succeeded in every batch. These results demonstrate that the modified method is highly reliable for sequencing anywhere from a few to a few hundred beads each time.

It is worth noting that proline- or tryptophan-containing peptides produce some unusual features in the spectra. When the degradation reached the proline or tryptophan, the corresponding N-acylated peptide (*m*/*z* 97 and 186 from the previous peak, respectively) was rarely observed. Instead, they each produced a unique set of peaks. With Bz-NHS as the capping agent, a proline always produced a group of four peaks at *m*/*z* 94, 112, 126, and 128 from the previous peak in the spectrum; whereas a tryptophan gave three peaks at 199, 215, and 233 mass units from the previous peak (Figure 2). To ascertain that the observed patterns were caused by the prolines and tryptophans in the sequences, we synthesized peptides DFWYLNBBRM and FRAPLNBBRM individually and subjected them to the same degradation procedure. The same distinct patterns were observed. In the case of prolyl residues, the +128 peak appears to be the uncyclized *N*-phenylthiourea formed between PITC and the *N*-prolyl peptide [136 (PITC) + 97 (proline) − 105 (benzoyl) = 128 Da]. The other species of the lower masses are likely the degradation products of the thiourea (e.g., substitution of sulfur by oxygen or loss of water). The peaks associated with tryptophan residues are likely formed by a similar mechanism. Further experiments are necessary to firmly establish the identity of the species and the mechanism of their formation. However, their presence in the spectra is easily identifiable and, thus, does not affect the sequence determination. Another residue that was frequently modified is lysine. In most peptides, the lysine side chain was partially or completely acylated by a benzoyl (when Bz-NHS was the capping agent) or nicotinoyl group (when Nic-NHS was the capping agent). This resulted in a peak at 232 (benzoyl) or 233 mass units (nicotinoyl) from the previous peak in the spectra. Again, the side chain modification does not adversely affect sequence determination. Rather, it is quite helpful in differentiating Lys from Gln, which have the same nominal mass.

In conclusion, the NHS esters (Nic-NHS, in particular) are superior to PIC as capping agents for partial Edman degradation. This provides a highly reliable, sensitive, inexpensive, and rapid method for sequencing support-bound peptides derived from combinatorial libraries. We expect this sequencing strategy to become the method of choice in the sequence determination of combinatorial peptide libraries in the future.

Acknowledgment. We thank Mr. Peng Wang of this laboratory for assistance in the synthesis of the peptide libraries. This work was supported by the National Institutes of Health (AI40575 and GM62820). M.S. was supported by an NIH Chemistry/Biology Interface training grant (T32GM08512).

Supporting Information Available. Experimental Section. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) For examples, see: (a) Matthews, D. J.; Wells, J. A. *Science* **1993**, *260*, 1113−1117. (b) Meldal, M.; Svendsen, I.; Breddam, K.; Auzanneau, F.-I. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3314−3318. (c) Peterson, J. J.; Meares, C. F. *Bioconjugate Chem*. **1998**, *9*, 618−626. (d) Hu, Y.-J.; Wei, Y.; Zhou, Y.; Rajagopalan, P. T. R.; Pei, D. *Biochemistry* **1999**, *38*, 643−650. (e) Wu, J.; Ma, Q. N.; Lam, K. S. *Biochemistry* **1994**, *33*, 14825−14833. (f) Cheung, Y. W.; Abell, C.; Balasubramanian, S. *J. Am. Chem. Soc*. **1997**, *119*, 9568−9569. (g) Harris, J.; Backes, B.; Leonetti, F.; Mahrus, S.; Ellman, J.; Craik, C. *Proc. Natl. Acad. Sci. U.S.A*. **2000**, *97*, 7754−7759. (i) Wang, P.; Fu, H.; Snavley, D. F.; Freitas, M. A.; Pei, D. *Biochemistry* **2002**, *41*, 6202−6210.

(2) For examples, see: (a) Chen, J. K.; Lane, W. S.; Brauer, A. W.; Tanaka, A.; Schreiber, S. L. *J. Am. Chem. Soc.* **1993**, *115*, 12591−12592. (b) Lam. K. S.; Hruby, V. J.; Lebl, M.; Knapp, R. J.; Kazmierski, W. M.; Hersh, E. M.; Salmon, S. E. *Bioorg. Med. Chem. Lett*. **1993**, *3*, 419−424. (c) Songyang, Z.; Shoelson, S. E.; Chaudhuri, M.; Gish, G.; Pawson, T.; Haser, W. G.; King, F.; Roberts, T.; Ratnofsky, S.; Lechleider, R. J.; Neel, B. G.; Birge, R. B.; Fajardo, J. E.; Chou, M. M.; Hanafusa, H.; Schaffhausen, B.; Cantley, L. C. *Cell* **1993**, *72*, 767−778. (d) Strickler, N.; Christopherson, K.; Yi, B.; Schatz, P.; Raab, R.; Dawes, G.; Bassett, D.; Bredt, D.; Li, M. *Nat. Biotechnol*. **1997**, *15*, 336−342. (e) Muller, K.; Gombert, F. O.; Manning, U.; Grossmuller, F.; Graff, P.; Zaegel, H.; Zuber, J. F.; Freuler, F.; Tschopp, C.; Baumann, G. *J. Biol. Chem.* **1996**, *271*, 16500−16505. (f) Sparks, A. B.; Adey, N. B.; Quiliam, L. A.; Thorn, J. M.; Kay, B. K. *Methods Enzymol*. **1995**, *255*, 498−509. (g) Dooley, C. T.; Ny, P.; Bidlack, J. M.; Houghten, R. A. *J. Biol. Chem*. **1998**, *273*, 18848−18856. (h) Fuh, G.; Pisabarro, M. T.; Li, Y.; Quan, C.; Lasky, L. A.; Sidhu, S. S. *J. Biol. Chem*. **2000**, *275*, 21486−21491. (i) Beebe, K. D.; Wang, P.; Arabaci, G.; Pei, D. *Biochemistry* **2000**, *39*, 13251−13260.

(3) For examples, see: (a) Eichler, J.; Houghten, R. A. *Biochemistry* **1993**, *32*, 11035−11041. (b) Wang, C.-I.; Yang, Q.; Craik, C. S. *J. Biol. Chem*. **1995**, *270*, 12250−12256. (c) Lutzke, R. A. P.; Eppens, N. A.; Weber, P. A.; Houghten, R. A.; Plasterk, R. H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11456−11460. (d) Gao, J.; Cheng, X.; Chen, R.; Sigal, G. B.; Bruce, J. E.; Schwartz, B. L.; Hofstadler, S. A.; Anderson, G. A.; Smith, R. D.; Whitesides, G. M. *J. Med. Chem.* **1996**, *39*, 1949−1955. (e) Smith, H. K.; Bradley, M. *J. Comb. Chem*. **1999**, *1*, 326−332.

(4) (a) Miller, S. J.; Copeland, G. T.; Papaioannou, N.; Horstmann, T. E.; Ruel, E. M. *J. Am. Chem. Soc.* **1998**, *120*, 1629−1630. (b) Copeland, G. T.; Jarvo, E. R.; Miller, S. J. *J. Org. Chem*. **1998**, *63*, 6784−6785. (c) Berkessel, A.; Herault, D. A. *Angew. Chem., Int. Ed. Engl*. **1999**, *38*, 102−105. (d) Francis, M. B.; Jamison, T. F.; Jacobsen, E. N. *Curr. Opin. Chem. Biol*. **1998**, *2*, 422−428. (e) Hoveyda, A. H. *Chem. Biol*. **1998**, *5*, R187−R191. (f) Copeland, G. T.; Miller, S. J. *J. Am. Chem. Soc.* **2001**, *123*, 6496−6502.

(5) Lam, K. S.; Lebl, M.; Krchnak, V. *Chem. Rev.* **1997**, *97*, 411−448.

(6) (a) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5381−5383. (b) Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10700−10704. (c) Nikolaiev, V.; Stierandova, A.; Krchnak, V.; Seligmann, B.; Lam, K. S.; Salmon, S. E. *Pept. Res*. **1993**, *6*, 161−170. (d) Kerr, J. M.; Banville, S. C.; Zuckermann, R. N. *J. Am. Chem. Soc*. **1993**, *115*, 2529−2531. (e) Ohlmeyer, M. H. J.; Swanson, R. N.; Dillard, L. W.; Reader, J. C.; Asouline, G.; Kobayashi, R.; Wigler, M.; Still, W. C. *Proc. Natl. Acad. Sci. U.S.A*. **1993**, *90*, 10922−10926. (f) Moran, E. J.; Sarshar, S.; Cargill, J. F.; Shahbaz,

M. M.; Lio, A.; Mjalli, A. M.; Armstrong, R. W. *J. Am. Chem. Soc*. **1995**, *117*, 10787−10788. (g) Nicolaou, K. C.; Xiao, X.-Y.; Parandoosh, Z.; Senyei, A.; Nova, M. P. *Angew. Chem., Int. Ed. Engl*. **1995**, *34*, 2289−2291. (h) Battersby, B. J.; Bryant, D.; Meutermans, W.; Matthews, D.; Smythe, M. L.; Trau, M. *J. Am. Chem. Soc*. **2000**, *122*, 2138−2139.

(7) (a) Youngquist, R. S.; Fuentes, G. R.; Lacey, M. P.; Keough, T. *J. Am. Chem. Soc*. **1995**, *117*, 3900−3906. (b) Cho, C. Y.; Youngquist, R. S.; Paikoff, S. J.; Beresini, M. H.; Hebert, A. R.; Berleau, L. T.; Liu, C. W.; Wemmer, D. E.; Keough, T.; Schultz, P. G. *J. Am. Chem. Soc*. **1998**, *120*, 7706−7718. (c) Hilaire, P. M. S.; Lowary, T. L.; Maldal, M.; Bock, K. *J. Am. Chem. Soc*. **1998**, *120*, 13312−13320.

(8) Biederman, K. J.; Lee, H.; Haney, C. A.; Kaczmarek, M.; Buettner, J. A. *J. Pept. Res*. **1999**, *53*, 234−243.

(9) Wang, P.; Arabaci, G.; Pei, D. *J. Comb. Chem*. **2001**, *3*, 251−254.

(10) Chait, B. T.; Wang, R.; Beavis, R. C.; Kent, S. B. H. *Science* **1993**, *262*, 89−92.

(11) (a) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991,** *354*, 82−84. (b) Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. *Int. J. Pept. Protein Res.* **1991,** *37*, 487−493. (c) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature* **1991**, *354*, 84−86.

CC020113+