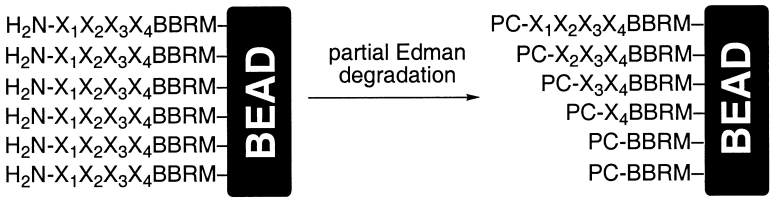
JOURNAL OF combinatorial CHEMISTRY

Report

Rapid Sequencing of Library-Derived Peptides by Partial Edman Degradation and Mass Spectrometry

Peng Wang, Gulnur Arabaci, and Dehua Pei

J. Comb. Chem., **2001**, 3 (3), 251-254• DOI: 10.1021/cc000102l • Publication Date (Web): 30 March 2001



More About This Article

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

- Supporting Information
- Links to the 2 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

Reports

Rapid Sequencing of Library-Derived Peptides by Partial Edman Degradation and Mass Spectrometry

Peng Wang, Gulnur Arabaci, and Dehua Pei*

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

Received November 28, 2000

The well-established peptide chemistry on the solid phase permits the ready access to large combinatorial peptide libraries via split-and-mix synthesis.[1] Consequently, screening of peptide libraries has been widely practiced to define the substrate specificity of enzymes,[2] to develop enzyme inhibitors,[3] to search for ligands of protein modules or receptors,[4] and to identify new catalysts.[5] For resin-bound libraries, high-throughput screening often generates a large number of positive beads that need to be individually characterized. This has created the demand for a rapid and inexpensive method for high-throughput sequencing of library-derived peptides. Several methods have been used for this purpose. The first method is to directly sequence the resin-bound peptides by Edman degradation.[1a,2d,g,h,4a] However, Edman sequencing is expensive and time-consuming, making it impractical for the routine sequencing of large numbers of peptides (>100). A variation of the Edman method involves pooling all of the selected peptides and sequencing them as a mixture.[4c,e] While this method is simple, it loses valuable information about the individual sequences and only gives an overall consensus, which may not necessarily represent the most active sequence. A second method utilizes various types of tags to encode the library, and the structure of an active compound is deduced by examining the tags present on the same bead that carries the active compound.[6] The drawbacks of the coding strategy include requirement of additional, compatible chemical steps in the library synthesis, potential interference with screening by the coding structure, and limited amount information that can be coded by the tags. A third method involves direct sequencing of a library-derived peptide by tandem mass spectrometry.[7] Although conceptually straightforward, this technique has not yet matured into a general, reliable method that can be practiced in an unspecialized laboratory. Finally, Youngquist et al. have developed a peptide ladder sequencing strategy, which encodes the peptide on a bead by generating a series of sequence-specific, partially terminated products during the synthesis of the library.[8] The sequence of the full-length peptide is determined by analyzing the peptide ladder formed by these termination products in a mass spectrum. This rapid and inexpensive method has been successfully applied to both peptide and peptidomimetic libraries.[2e,4i,8,9] A drawback of this method is that, because of the different reactivities of 20 amino acids (or other library building blocks), the amount of chain termination varies at each position in an unpredictable manner, resulting in heterogeneous amounts of full-length peptides on all library beads. This biases the screening against those peptides that have higher percentages of termination products (thus, less full-length peptides). The use of the same amino acid as capping agent (e.g., use of Boc-Ala as the capping agent during the addition of Fmoc-Ala) has only partially solved this problem.[10] Another potential problem is that the termination products may interfere with library screening.

In this report, we have adapted a peptide ladder sequencing method pioneered by Chait et al.[11] to rapidly sequence support-bound peptides derived from combinatorial libraries. In this method, a peptide library is synthesized on solid support without partial chain termination (except for Ile and Nle) and, therefore, all of the library beads carry an equal amount of full-length peptides. After screening against a molecular target (e.g., a protein), the peptide on a selected bead is subjected to multiple cycles of partial Edman degradation to generate a peptide ladder and its sequence is determined by matrix-assisted laser desorption ionization (MALDI) mass spectrometry.

The validity of this method was demonstrated by the synthesis and screening of a peptide library to identify the optimal substrates of *E. coli* peptide deformylase (PDF).[12] An N-formylated tetrapeptide library, f-XXXXBBRM-resin (X = Ala, Asn, Asp, Gln, Glu, Gly, His, Ile, Leu, Nle, Phe, Pro, Ser, Thr, Trp, Tyr, or Val; B = β-alanine), was synthesized on TentaGel S resin (100 μm) via the split-synthesis method.[1] A common linker (BBRM) was added to facilitate enzymatic reaction and mass spectrometric analysis,[13] whereas Lys, Arg, Cys, and Met (replaced with Nle) were excluded from the randomized region to simplify the screening process.[2e] The resulting one-bead−one-compound library (theoretical diversity is 83 521) was screened against Co(II)-substituted *E. coli* PDF as previously described.[2e] Briefly, enzymatic removal of the N-terminal formyl group by PDF exposes a free NH_2 group at the N-terminus of a peptide. The exposed amino group was selectively derivatized with a biotin by incubating the library with *N*-hydroxysuccinimidobiotin. The resin was then incubated with a streptavidin−alkaline phosphatase conjugate, which is subsequently recruited to the reacted beads by the bound biotin. Hydrolysis of 5-bromo-4-chloro-3-indolyl

* To whom correspondence should be addressed. Phone: (614) 688-4068. Fax: (614) 292-1532. E-mail: pei.3@osu.edu

10.1021/cc000102l CCC: $20.00
Published on Web 03/30/2001

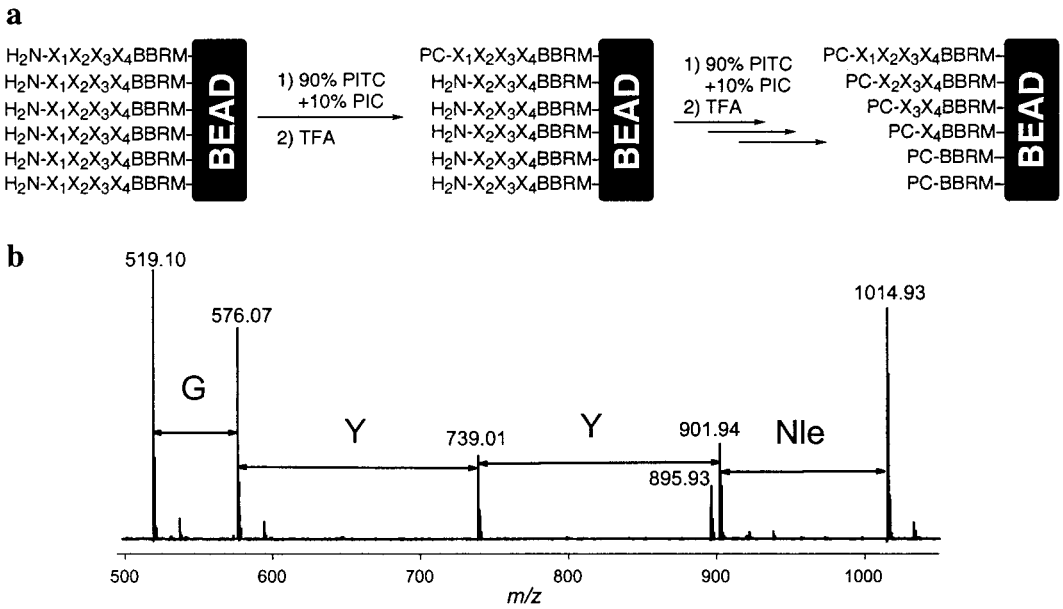


Figure 1. Sequencing of support-bound peptides by partial Edman degradation and mass spectrometry: (a) reactions involved in partial Edman degradation (PITC, phenyl isothiocyanate; PIC, phenyl isocyanate; PC, phenylcarbamoyl); (b) MALDI mass spectrum of a peptide and its truncation products, obtained with 20% of the material from a single bead. The sequence of the unknown peptide is f-(Nle)-YYGBBRM. M is methionine prior to CNBr cleavage and is homoserine lactone after CNBr cleavage.

phosphate (BCIP) by the bound phosphatase resulted in an intense turquoise color at the surface of a positive bead. The colored beads were readily identified under a dissecting microscope and manually removed from the library.

Screening of 70 mg of resin (~200 000 beads) resulted in ~100 colored beads, and the 72 most-colored ones were removed from the library for sequence analysis. All 72 beads were placed in a 0.5 mL column fitted with a 0.22 μm filter at one end. The beads were incubated in a MeOH solution containing 7% hydrochloric acid for 60 h to remove the N-terminal formyl group.[14] The resulting beads were subjected to partial Edman degradation to generate a peptide ladder according to the method originally developed by Chait et al.[11] Specifically, the beads were treated with a 10:1 (v/v) mixture of phenyl isothiocyanate (PITC) and phenyl isocyanate (PIC) in pyridine (250 μL), both of which reacted with the N-terminal amine of the support-bound peptides (Figure 1a). After being washed with CH_2Cl_2 to remove the excess reagents, the beads were treated with trifluoroacetic acid (TFA). For ~90% of the peptides that had reacted with PITC, TFA treatment resulted in the cleavage of the N-terminal amino acid (Edman degradation). For the rest of the ~10% peptides that had reacted with PIC, no cleavage was possible under the conditions used, and these peptides became permanently N-blocked by a phenylcarbamoyl (PC) group. The resulting beads, still placed in the same column, were washed with CH_2Cl_2 to remove the phenylthiohydantoin (PTH)-amino acid derivatives and subjected to the next cycle of partial Edman degradation. Repetition of the partial Edman degradation reaction for $n + 1$ cycles (n is the number of random positions in a library and equal to 4 in this work) resulted in a series of sequence-specific truncation products. Next, the treated beads were individually picked, placed in 72 separate microcentrifuge tubes, and treated with CNBr. This cleaved the peptides off the support at their C-terminal methionine, generating the corresponding peptide homoserine lactones. After lyophilization to remove the excess reagents, the peptides released from each bead were analyzed by MALDI-TOF mass spectrometry as previously described.[2e] Since each bead carried a unique full-length peptide and four truncation products, MALDI analysis of the cleavage mixture generated a peptide ladder containing five individual peaks. Figure 1b shows a typical MALDI spectrum, obtained with 20% of the peptide mixture isolated from a single resin bead. The protonated full-length peptide gave a peak at *m*/*z* 1014.93. The truncation products produced four peaks at *m*/*z* 901.94, 739.01, 576.07, and 519.10. The mass difference between the full-length peptide and the first truncation product (*m*/*z* 901.94) is 113.0, indicating that the N-terminal residue of the peptide is Nle, Leu, or Ile. Likewise, the rest of the residues were determined as Tyr, Tyr, and Gly, respectively, based on the mass differences between adjacent peaks in the ladder. To differentiate norleucine (used as a surrogate for methionine) from the isobaric leucine and isoleucine, a small amount of chain-termination products were generated by adding 10% *N*-acetylalanine into the coupling reaction of Nle and 10% *N*-acetylglycine into the coupling reaction of Ile during library synthesis.[8] Thus, if a peptide contains a Nle at a given position, its mass spectrum would show at that position a doublet separated by 6 mass units, due to the presence of both the chain-termination product derived from library synthesis and the truncation product derived from postscreening partial Edman degradation. Similarly, a peptide containing an Ile should produce a doublet separated by 20 mass units at the corresponding position. The presence of a doublet at *m*/*z* 895.93 and 901.94 in Figure 1b therefore indicates that the N-terminal residue is a norleucine and the unknown peptide has the sequence of NH_2-Nle-Tyr-Tyr-Gly.

Out of the 72 positive beads analyzed by this method, 65 (90%) produced high-quality spectra, which allowed unambiguous sequence assignment at all randomized positions

Table 1. Selected Sequences for *E. coli* PDF[a]

MFYY	MVVL	MLGP	EMYI
MGYY	MFVI	MYSH	EEYV
MGYP	MLVP	MYSE	EFLP
MPYV	MLVY	MEHT	EYMV
MHYP	MYVF	MTHL	HMHY
MMYV	MEAV	MFHH	HATL
MYYV	MFAT	MMFE	FSPY
MMYE	MVAE	MSFG	FPHY
MEYL	MAAA	MMEY	AVLL
MPYF	MPAH	MPPL	AVFF
MYYG	MLLY	MMPV	SHYH
MTYS	MILY	MVTT	LETP
MPVH	MLLA	YFLF	GFLL
MHVH	MLLP	YVYG	PDLG
MVVL	MLLS	YPGY	
MVVL	MYGV	YYGP	
MVVL	MVGV	YAMY	

[a] M, norleucine.

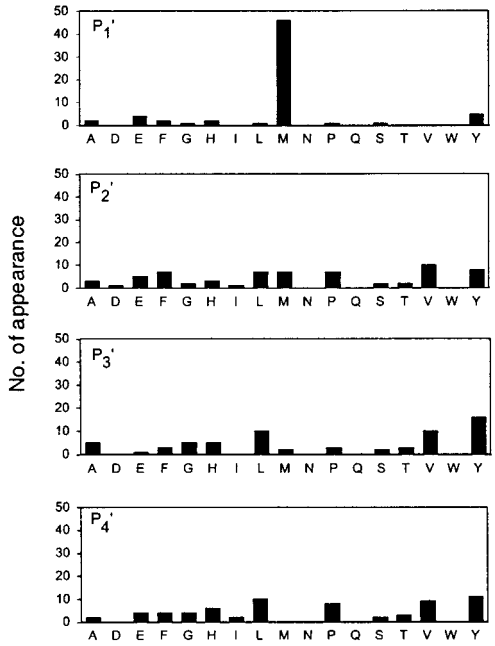


Figure 2. Substrate specificity of *E. coli* Co(II) PDF. Displayed are the amino acids identified at each of the four N-terminal positions (P_1' is the N-terminal residue). Number of appearance on the *y* axis represents the number of selected sequences that contain a particular amino acid at a certain position (out of a total of 65 sequences). M is norleucine.

(Table 1). The other 7 beads gave spectra that missed one or a few peaks, and therefore, complete sequence assignment was not possible. It should be noted that no special effort has been made to optimize the degradation reaction or the MALDI MS conditions at this point, and we believe that a still higher success rate is possible after optimizing these conditions. Analysis of the 65 selected sequences revealed that norleucine is the most preferred amino acid at the P_1' position (46 sequences had norleucine at this position) (Table 1 and Figure 2). This is consistent with earlier studies[2e,15] and the fact that methionine is always the N-terminal residue of a physiological substrate of PDF. Consistent with these earlier studies, PDF has little selectivity at the P_2' position. At the P_3' and P_4' positions, PDF clearly prefers hydrophobic residues and tyrosine in particular (Figure 2). This is again similar to our previous screening results, although the extent of preference for tyrosine is weaker in this work. A number of factors may have contributed to the slight difference observed at the P_3' and P_4' positions, including a subtle difference in screening stringency, the use of Co−PDF (this work) vs Fe−PDF (earlier work), and bias against peptides containing slow-coupling amino acids (e.g., Val and Thr) in the earlier work. A few sequences that contain other N-terminal amino acids (e.g., Tyr, Phe, Glu, Ala) were also obtained, as observed previously.[2e] We have previously shown that the *N*-formylphenylalanyl and *N*-formyltyrosyl peptides are indeed excellent PDF substrates, whereas the peptides with the other N-terminal amino acids (e.g., Gly, Ala) are poor substrates. Preliminary studies seem to suggest that the false positives were a result of background hydrolysis of their N-formyl group during the process of library screening.

The above results demonstrate that the partial Edman degradation strategy is an effective method for sequence determination of library-derived peptides. We have indeed recently applied this method to sequence over a hundred phosphothreonyl peptides that bound to a forkhead-associated (FHA) domain.[16] This method offers several advantages over existing methods. First, it generates a peptide ladder only after the screening step so that libraries can be prepared with few chain-termination events (except for the homopolymers of Nle and Ile), and therefore, there is minimal variation in the quantity of full-length peptides on different beads. It thus avoids the sequence bias problem associated with the Youngquist method. It also simplifies library synthesis and largely eliminates the potential interference with screening by an encoding tag (e.g., a chain-termination product). Second, since all peptides are support-bound, the operation is very simple; no peptide isolation is necessary after each cycle of Edman degradation and excess reagents, and reaction byproducts are removed by simply washing the beads with solvents. Third, the spatial separation of different peptides on individual beads permits simultaneous degradation of all selected beads (anywhere from one to a few million) in one pot. This tremendously reduces the amount of labor required for sequencing a large number of beads. Finally, this method is highly sensitive, rapid, and yet inexpensive. On a Bruker Reflex III MALDI instrument, we have been able to routinely sequence 20−30 beads in an hour and at a cost of a few U.S. dollars per peptide. Only <20% of the peptides isolated from a single 100 μm bead is needed for each MS experiment. This permits multiple MS experiments for each bead, if necessary, to ensure that the sequence is unambiguously determined. The only limitation of this method is that a peptide to be sequenced must contain only α-amino acids and a free N terminus or an N-terminal structure (e.g., *N*-formyl- or *N*-Fmoc-protected) that can be converted into a free N terminus prior to Edman degradation.

In conclusion, a rapid, sensitive, and inexpensive method has been developed to sequence peptides derived from combinatorial libraries. Although the concept of partial Edman degradation has previously been demonstrated in the solution phase,[11] to our knowledge, this work represents the first application of the method to the sequence determination of library-derived peptides on the solid phase. We believe

that this sequencing strategy will further expand the utility of combinatorial peptide libraries in biomedical research.

Acknowledgment. We thank Dr. Kari Green-Church for her technical assistance in the MALDI MS experiments. This work was supported by the National Institutes of Health (Grant AI40575).

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

References and Notes

(1) (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.

(2) Examples: (a) Matthews, D. J.; Wells, J. A. *Science* **1993**, *260*, 1113−1117. (b) Smith, M. M.; Shi, L.; Navre, M. *J. Biol. Chem.* **1995**, *270*, 6440−6449. (c) Meldal, M.; Svendsen, I.; Breddam, K.; Auzanneau, F.-I. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3314−3318. (d) Peterson, J. J.; Meares, C. F. *Bioconjugate Chem.* **1998**, *9*, 618−626. (e) Hu, Y.-J.; Wei, Y.; Zhou, Y.; Rajagopalan, P. T. R.; Pei, D. *Biochemistry* **1999**, *38*, 643−650. (f) Coombs, G. S.; Bergstrom, R. C.; Pellequer, J.-L.; Baker, S. I.; Navre, M.; Smith, M. M.; Tainer, J. A.; Madison, E. L.; Corey, D. R. *Chem. Biol.* **1998**, *5*, 475−488. (g) Wu, J.; Ma, Q. N.; Lam, K. S. *Biochemistry* **1994**, *33*, 14825−14833. (h) Cheung, Y. W.; Abell, C.; Balasubramanian, S. *J. Am. Chem. Soc.* **1997**, *119*, 9568−9569.

(3) Examples: (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) Examples: (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.

(5) (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.* **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.

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

(8) Youngquist, R. S.; Fuentes, G. R.; Lacey, M. P.; Keough, T. *J. Am. Chem. Soc.* **1995**, *117*, 3900−3906.

(9) 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.

(10) Hilaire, P. M. S.; Lowary, T. L.; Maldal, M.; Bock, K. *J. Am. Chem. Soc.* **1998**, *120*, 13312−13320.

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

(12) Giglione, C.; Pierre, M.; Meinnel, T. *Mol. Microbiol.* **2000**, *36*, 1197−1205.

(13) Yu, Z.; Chu, Y.-H. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 95−98.

(14) Sheehan, J. C.; Young, D. H. *J. Am. Chem. Soc.* **1958**, *80*, 1154−1158.

(15) Ragasu, S.; Mouchet, P.; Lazennec, C.; Dive, V.; Meinnel, T. *J. Mol. Biol.* **1999**, *289*, 1445−1457.

(16) Liao, H.; Yuan, C.; Su, M.-I.; Yongkiettrakul, S.; Qin, D.; Li, H.; Byeon, I.-J.; Pei, D.; Tsai, M.-D. *J. Mol. Biol.* **2000**, *304*, 941−951.

CC000102L