Generation and Screening of Combinatorial Peptide Libraries Designed for Rapid Sequencing by Mass Spectrometry†

R. Scott Youngquist,* Gary R. Fuentes, Martin P. Lacey, and Thomas Keough*

Contribution from The Procter & Gamble Company, Miami Valley Laboratories, P.O. Box 538707, Cincinnati, Ohio 45253-8707

Received June 3, 1994®

Abstract: We have developed a method for rapidly sequencing compounds isolated from support-bound combinatorial libraries. During the synthesis of the library, a capping reagent is used to effect partial termination at each coupling step. In this way, each of the resin beads carries not only the full-length product but also small amounts of sequence-specific termination products. The sequence of the full-length product is determined with matrix-assisted laser desorption ionization mass spectrometry, which is used to measure mass differences between adjacent members of the termination series. The observed mass differences identify the corresponding monomers. With mixtures of terminating reagents, monomers of identical mass can also be distinguished. This method is fast, accurate, and reliable, and it is compatible with a wide range of unnatural synthetic libraries. Because our sequencing method directly analyzes the compounds used for screening, synthetic byproducts can be easily identified and, with this information, reaction conditions can be optimized to minimize them. To demonstrate the utility of the method, we have prepared combinatorial libraries of acetylated and non-acetylated pentapeptides on polystyrene beads. These libraries were used to identify ligands to an HIV-neutralizing antibody and to streptavidin.

Introduction

Screening of chemicals and fermentation broths has long been used to identify novel lead compounds for drug discovery. This approach has the advantage of being empirical, requiring no knowledge of the target protein or its ligands. The difficulty has been identification of the active compounds from the complex mixtures. Recently, new methods have been developed for generating and screening combinatorial libraries of synthetic compounds. They combine the empirical advantage of traditional screening with a systematic means for rapid identification of the active materials. There are many demonstrations of the utility of peptide libraries. Houghten et al. have used soluble libraries to identify peptide antimicrobials with activity equal to or greater than naturally occurring peptides such as magainins and cecropins.[1] They have also identified novel ligands to opioid receptors.[2] Lam and co-workers used support-bound libraries to identify ligands to a monoclonal antibody against β-endorphin[3] and to GPIIb/IIIa.[4] This strategy is being extended to non-peptide libraries including *N*-substituted glycines,[5] carbamates,[6] and benzodiazapines.[7]

With resin-bound libraries, identification of the active compound on a given bead is a critical step. Conventional Edman sequencing is widely available and has been used for this purpose;[3] however, it has significant limitations. It requires natural amino acids as monomers, works only with free *N*-termini, and is time-consuming. To avoid these limitations, coding strategies have been developed that enable blocked *N*-termini and unnatural recognition elements to be identified in chemical libraries. These strategies use a branched linker with the recognition arm synthesized in parallel with an arm that encodes the recognition polymer's sequence. The identity of the recognition polymer is then determined by sequencing the coding arm. Both peptides and DNA have been used as the encoding polymer.[8,9] Unfortunately, use of a coding polymer can greatly increase the number of chemical steps needed for library synthesis. It also requires that each bead carry a large amount of coding polymer that could affect screening. Ohlemeyer et al.[10] have reported a different strategy that involves labeling the growing peptide chain with halobenzene tags that encode the recognition sequence. The tags are cleaved from the bead and read using gas chromatography. This system is effective, but is limited in the amount of information that the binary code can carry. A common limitation of all of the coding systems is that they are indirect. They don't allow direct determination of the quality of the compounds in a chemical library. Synthetic problems such as deletion peptides, side reaction products, or incomplete side chain deprotection would not be observed, but could have a significant impact on library quality.

We describe a mass spectrometry method that enables rapid determination of the sequences of biologically active compounds

* To whom correspondence should be addressed.

† Abbreviations: MALDI, matrix-assisted laser desorption ionization; Ac-Ala, *N*-acetyl-D,L-alanine.

® Abstract published in *Advance ACS Abstracts,* April 1, 1995.

(1) Houghten, R. A.; Dinh, K. T.; Burchin, D. E.; Blondell, S. E. In *Techniques in Protein Chemistry IV*; Angeletti, R. H., Ed.; Academic: San Diego, CA, 1993; pp 249-256.

(2) Dooley, C. T.; Chung, N. N.; Schiller, P. W.; Houghten, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10811-10815.

(3) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature (London)* **1991**, *354*, 82-84.

(4) Salmon, S. E.; Lam, K. S.; Lebl, M.; Kandola, A.; Khattri, P. S.; Wade, S.; Patek, M.; Kocis, P.; Krchnak, V.; Thorpe, D.; Felder, S. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11708-1712.

(5) Simon, R. J.; Kania, R. S.; Zuckerman, R. N.; Huebner, V. D.; Jewel, A. A.; Banville, S.; Ng, S.; Wang, L.; Rosenberg, A.; Marlowe, C. K.; Spellmeyer, D. C.; Tan, R.; Franmel, A. D.; Santi, D. V.; Cohen, F. E.; Bartlett, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9637-371.

(6) Cho, C. Y.; Moran, E. J.; Cherry, S. R.; Stephans, J. C.; Fodor, S. P. A.; Adams, C. L.; Sundaram, A.; Jacobs, J. W.; Schultz, P. G. *Science* **1993**, *261*, 1303-1305.

(7) Bunin, B. A.; Plunkett, M. J.; Ellman, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4708-712.

(8) 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.

(9) Nikolaiev, V.; Stierandova, A.; Krchnak, V.; Seligmann, B.; Lam, K. S.; Salmon, S. E. *Peptide Res.* **1993**, *6*, 161-170.

(10) 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.

0002-7863/95/1517-3900$09.00/0

isolated from support-bound combinatorial libraries.[11] Our strategy is to preserve the sequence information available as the compounds are being synthesized and use that information to determine the sequences of the final products. This approach differs from other mass spectrometry sequencing methods which rely on degradation on the full-length compound, either chemically or through fragmentation, to generate the information needed for sequencing.[12-16] During the synthesis of the library, a small percentage of the growing peptide chains are capped at each step to generate a series of sequence-specific termination products on each bead. These sequence-specific termination products provide the information needed to determine the sequence of the full-length product. After the library is screened *in vitro*, positive beads are isolated and each full-length product, with its family of termination products, is chemically removed from each bead. The mass spectrum of each of these mixtures is recorded using matrix-assisted laser desorption ionization (MALDI) mass spectrometry. The sequence is determined from the mass differences between termination products. Since the masses of the monomers in a library are generally different, this sequencing system can distinguish a very large number of monomers at each position in the library. Because this method analyzes the actual library members, unlike the indirect encoding schemes, it allows a direct assessment of the quality of the compounds in a synthetic library. Products resulting from deletions, side reactions, or incomplete deprotection are readily identified by their effect on the masses of the parent compound and the termination products. Direct analysis of the ligand on each bead eliminates any confusion about the identity of the compound being recognized by the target protein.

To demonstrate the utility of this method, we have prepared a series of peptide libraries and screened them against two protein targets. Peptide libraries are a good starting point because they have well-established synthetic chemistry and most of the amino acid monomers are uniquely defined by their masses. Isobaric amino acids such as leucine and isoleucine or stereoisomers can also be differentiated with this method, by simply using mixtures of different capping reagents. In this paper we report the synthesis of acetylated and non-acetylated peptide libraries containing up to 10^6 members, and the screening of these libraries to identify ligands to an anti-HIV-1 gp120 monoclonal antibody and to streptavidin.

Results and Discussion

Generation of the Support-Bound Libraries. The support-bound libraries were generated by the split-synthesis method described by Lam and co-workers[3] and originally reported by Furka.[17] This method results in the synthesis of many copies of a single peptide on each resin bead. TentaGel S resin was used as the solid support. A peptide linker, H_2N-(β-Ala)-XXXRM, was synthesized on the beads to allow the peptides to be efficiently released and to improve their analysis by MALDI. The methionine at the *C*-terminus enabled the peptides to be removed from the resin by cyanogen bromide digestion. Because a fixed positive charge should both improve MALDI sensitivity and ensure reasonably uniform response for each member of the termination synthesis family, we included an arginine in the linker. Another important role for the linker is to increase the molecular mass of the compounds in the library beyond 500 Da. This ensures that each member of the peptide ladder has a mass greater than the "chemical noise" produced by desorption of the UV-absorbing matrix. Various combinations of amino acids have been used at the remaining three positions of the linker (XXX), but in the libraries reported here, proline was used. A flexible amino acid, β-alanine, separated the linker from the randomized region of the library.

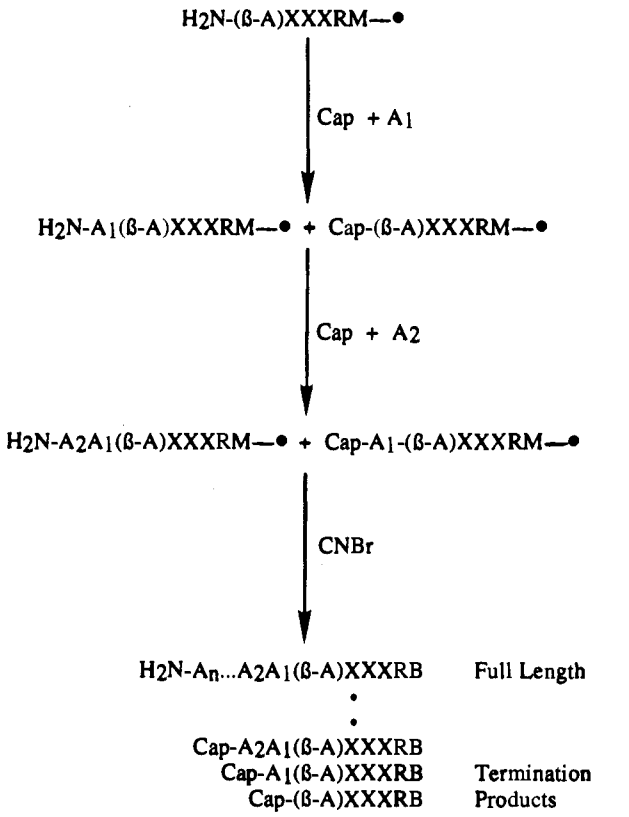

Figure 1. Termination synthesis method used to produce sequence-specific families of peptides on each resin bead in the library. After screening, active peptides were isolated and the peptide products were released from the bead by cyanogen bromide digestion (M is converted into homoserine lactone, B).

Two sets of libraries were prepared with either a free amine at the N-terminus or with *N*-terminal acetylation. We prepared a set of sublibraries where each position in the randomized region comprised 7 amino acids, Ala, Arg, Glu, Gly, His, Phe, and Pro, and a set of full libraries where each position in the randomized region included 16 amino acids, Ala, Arg, Asn, Asp, Gln, Glu, Gly, His, Leu, Lys, Phe, Pro, Ser, Thr, Trp, and Tyr. During the coupling of each position in the randomized region, a small proportion (10%) of a capping agent, *N*-acetyl-D,L-alanine (Ac-Ala), was added to each reaction vessel along with the individual amino acid (90%) (Figure 1). This terminated the synthesis at a small percentage of the growing peptide chains as each monomer was added. These terminated peptides provide a record of the synthesis on each resin bead. At the end of the synthesis, each resin bead contains a full-length peptide as the major product as well as the corresponding family of termination products.

MALDI Sequencing of Peptides. Beads with peptides to be sequenced are removed from the library, placed in separate microcentrifuge tubes, and treated with cyanogen bromide. This

(11) This work was initially communicated in Youngquist, R. S.; Fuentes, G. R.; Lacey, M. P.; Keough, T. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 77-81.

(12) Wang, R.; Chait, B. T.; Kent, S. B. H. *Techniques in Protein Chemistry IV*; 1993; pp 471-478.

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

(14) Wang, R.; Chait, B. T.; Kent, S. B. H. *Techniques in Protein Chemistry V*; 1994; pp 19-26.

(15) Brummel, C. L.; Lee, N. W.; Zhou, Y.; Benkovic, S. J.; Winograd, N. *Science* **1994**, *264*, 399-402.

(16) Sepetov, N. F.; Issakova, O. L.; Lebl, M.; Swiderek, K.; Stahl, D. C.; Lee, T. D. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 58-62.

(17) Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. *Int. J. Peptide Protein Res.* **1991**, *37*, 487-493.

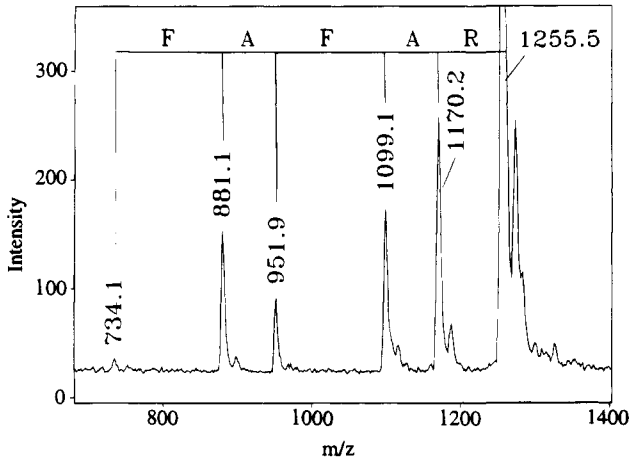

Figure 2. MALDI mass spectrum of a peptide and its termination products using 5% of the material from a single bead. Each of the peaks had a smaller peak at higher mass (+18) that resulted from hydrolysis of the homoserine lactone.

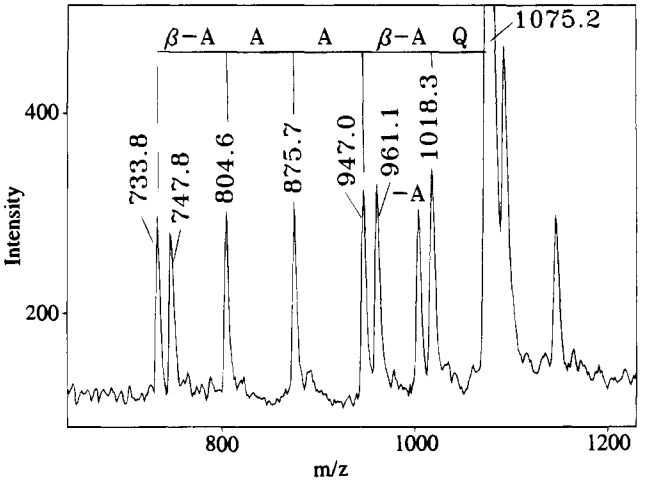

Figure 3. MALDI mass spectrum of a coded peptide containing both alanine and β-alanine. β-Ala is coded with two terminating reagents to produce a doublet in the MALDI spectrum. The peak at 1005.5 was the result of an alanine deletion during synthesis.

cleaves the peptide from the bead at the methionine residue, generating a homoserine lactone and releasing the peptide from the bead. Each of the peptide mixtures is then analyzed by MALDI. A typical MALDI spectrum is shown in Figure 2. Only 5% of the peptide mixture isolated from a single resin bead from a pentamer library was used, demonstrating the high sensitivity of the technique. The spectrum shows an excellent signal to noise ratio and contains ions corresponding to the protonated full-length peptide Ac-RAFAF(β-Ala)PPPRB (acetylated *N*-terminus, B = homoserine lactone) at *m/z* 1255.5, and ions for protonated members of the termination synthesis family (Ac-Ala capping group) at *m/z* 1170.2, 1099.1, 951.9, 881.1, and 734.1. The amino acid identities are established from the mass differences observed between adjacent members of the peptide ladder. For example, the *N*-terminal residue is identified by taking the difference in mass between that of the full-length peptide and that of the highest mass termination product (1255.5 − 1170.2 + 71 = 156.3). An additional 71 Da is added to this result because the full-length peptide is acetylated, while the termination products are capped with Ac-Ala (a difference of 71 Da). The calculated mass difference of 156.3 Da is in close agreement with the actual mass of R (156.2 Da). Simple mass differences are used to assign the remaining members of the termination synthesis family. Reading from the full-length product toward lower mass, the reading direction is from the *N*-terminus to the *C*-terminus and the sequence of this pentamer is read as Ac-RAFAF-Linker.

Capping with Ac-Ala works well for cases where monomer masses differ by at least 1 Da. We have also used MALDI sequencing with structural isomers and stereoisomers where monomers have degenerate masses. Isobaric monomers are differentiated with mixtures of capping reagents. This is illustrated in the MALDI spectrum of the peptide Ac-Q(β-Ala)-AA(β-Ala)-Linker (Figure 3). In this peptide, the two isomers of alanine are distinguished by capping β-Ala with a mixture of two capping reagents, *N*-acetylalanine and *N*-propionylalanine. This produces a doublet in the MALDI spectrum, separated by 14 mass units, each time β-Ala is incorporated into the peptide. The ion for the full-length peptide, Ac-Q(β-Ala)AA(β-Ala)(βAla)PPPRB, is observed at 1075.2 Da, and the ions for the protonated members of the Ac-Ala termination synthesis family are 1018.3, 947.0, 875.7, 804.6, and 733.8 Da. Using the mass differences, the sequence is assigned as Ac-Gln followed by four alanines. Additional ions at 747.8 and 961.1 Da result from *N*-propionylalanine capping and mark the second and fifth residues as β-Ala. The sequence of this pentamer is then read as Ac-Q(β-Ala)AA(β-Ala)-Linker. In the large library, this coding strategy was used to differentiate Asn and Asp and to differentiate Gln, Glu, and Lys. The members in these sets of residues differ in mass by 1 Da or less.

Table 1. Test of MALDI Peptide Sequencing

	sequence		
	reported	actual	accuracy
1	Ac-QAGEP	Ac-QAGEP	5/5
2	Ac-VERHV	Ac-VERHV	5/5
3	Ac-RHPQF	Ac-RHPQF	5/5
4	Ac-PGHQA Ac-PGHAQ	Ac-PGHQA	5/5
5	Ac-QGWSA	Ac-QWSAG	3/5
6	no signal	Ac-RVAYS	fully protected[a]
7	Ac-FVPRH	Ac-FVPRH	5/5
8	Ac-GKFAH	Ac-GKFAH	5/5
9	Ac-YAWSQ	Ac-YAWSQ	5/5
10	YPYPY	YPYPY	5/5
11	Ac-QAKRV Ac-QAKVR	Ac-QAKRV	5/5
12	GPGRAF	GPGRAF	6/6
13	Ac-RHPQF	Ac-RHPQF	5/5
14	Ac-PGRAF	Ac-PGRAF	5/5
15	Ac-PGdRAF	Ac-PGdRAF	5/5
		total	69/71

[a] Not a valid test of the accuracy of MALDI sequencing.

To test our ability to sequence peptides using the termination synthesis method, we synthesized a series of 15 peptides and sequenced them as unknowns (Table 1). This series contained 13 different amino acids and included a set of stereoisomers, D- and L-Arg. The stereoisomers of Arg were differentiated with a mixture of capping reagents. Most of these peptides had blocked *N*-termini and could not have been sequenced by classical methods. Of the 71 amino acid residues that were identified, 69 were assigned correctly. This is an accuracy of about 97%. In addition to correctly sequencing nearly all of the peptides, the method also correctly differentiated peptides containing Arg with D and L stereochemistry.

During this exercise, we encountered three sequencing problems. All of these problems resulted from synthetic problems that have subsequently been corrected. In one sample, the correct amino acids were identified, but they were incorrectly ordered (Table 1, peptide 5). A major portion of the material from this bead was a Gly deletion product. This peak was

Table 2. Byproducts Observed in Support-Bound Combinatorial Peptide Libraries

mass	modification	site
+16	O	Trp
+42	CH_3CO	*N*-terminal Pro
+56	C_4H_9	Lys
+79	Br	Trp, Tyr
+96	CF_3CO	Lys, *N*-terminal amine
+158	di-Br	Trp
+266	Pmc	Arg

misinterpreted as part of the sequencing ladder, and this led to the incorrect ordering of the amino acids in the peptide. The second problem, which occurred in two peptides (Table 1, peptides 4 and 11), prevented the order of two residues from being unambiguously determined. This was caused by a missing peak in the sequencing ladder. Because the observed mass between two peaks in the sequencing ladder was larger than the mass of any single amino acid, it was clear that a peak was missing. The pair of amino acids that summed to the measured mass could be easily determined; however, the order of the two residues could not be assigned. We considered this a successful sequencing because the problem was clear and, despite this ambiguity, the method narrowed the number of possible peptides from hundreds of thousands to two. Finally, on one bead no signal was observed by MALDI (Table 1, peptide 6). Due to an instrument problem during TFA deprotection, this peptide still had almost all of its side chain protecting groups. Protected peptides give little or no signal under our measurement conditions. Correcting the instrument problem and optimizing the synthesis to prevent deletions and provide more even capping have corrected the problems that led to these sequencing problems.

Identification of Byproducts from Peptide Synthesis. Problems can arise during library syntheses that lead to the formation of undesired compounds. They include modification of the peptide by incomplete removal of side chain protecting groups or by reaction with the deprotection reagents. We have been able to identify these problems by observing their effects on the masses of the peptide products (Table 2). Some of these byproducts do not affect the quality of the library that is screened because they are formed after the screening step. For example, the oxidation of tryptophan and the bromination of tryptophan and tyrosine both occur during preparation of the sample for analysis. The oxidation of tryptophan occurs when the sample is applied to the probe tip used for MALDI analysis. We verified this by analysis of a Trp-containing peptide with ion spray mass spectrometry, a mass spectrometry method that did not show sample oxidation. Furthermore, if a larger amount of sample was loaded onto the MALDI probe tip, the mass for the unoxidized material was also observed. Bromination of Trp and Tyr is well-known during cyanogen bromide digestion,[18] the procedure used to release the peptides from the beads. Some modifications, on the other hand, do affect the peptides that are being screened. We observe trifluoroacetylation of primary amines on Lys and on the *N*-terminus of the peptide. However, these byproducts, which are formed during side chain deprotection, can be reversed by treatment of the peptide resin with piperidine.[19] Two other byproducts listed in Table 2 result from problems with the amino acid side chain protecting groups, the incomplete removal of the PMC protecting group from Arg,

(18) Fontana, A.; Grandi, C.; Dalzoppo, D.; Zambonin, M. In *Progress in Tryptophan Serotonin Research, Proceedings—Meeting of the International Study Group*; Schlossberger, H. G., Ed.; de Gruyter: Berlin, Germany, 1984; 839–844.

(19) Atherton, E.; Wolley, V.; Sheppard, R. C. *J. Chem. Soc., Chem. Commun.* **1980**, 970–971.

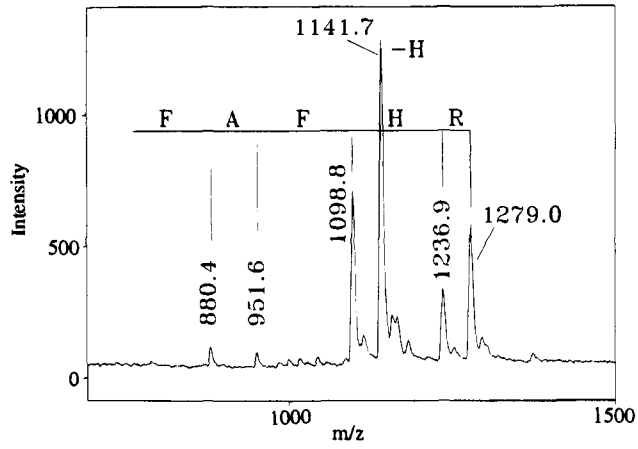

Figure 4. MALDI mass spectrum of a deletion peptide from a pentapeptide library. The major compound on the bead was a tetrapeptide, RFAF, and not the full-length product, RHFAF, which was present to a small extent.

and the alkylation of the ϵ-amino group of Lys with a *tert*-butyl group. These problems can be controlled by increasing the amount of scavengers used during deprotection and by using longer reaction times.

The most confusing modification we observed was a partial acetylation of *N*-terminal prolines in libraries with *N*-terminal amines. We could find no explanation for this. Acetylation is specific for proline, so it must not arise from general contamination of the synthesizer or any of the reagents. Because acetylation is only seen at *N*-terminal proline, and not Pro at other positions, the FMOC-Pro we used cannot be responsible. The only source of acetyl groups in the synthesis is the acetylalanine used for partial termination of the reactions, but it is not clear how those acetyl groups could transfer from the alanine to the proline. Furthermore, there is no evidence in the MALDI spectra for the intermolecular transfer of the acetyl group from acetylalanine to Pro.

In addition to byproducts resulting from modification of the peptides, we have seen products resulting from amino acid deletions. These deletion peptides occur when an amino acid does not couple completely to the growing peptide chain. In general, they can be minimized by repeated couplings and by the use of large excesses of amino acids. Examples of deletion peptides are shown in Figures 3 (A deletion) and 4. In the latter example, the deletion peptide RFAF is the major product on the bead. The full-length peptide, RHFAF, is present to a small extent. Significantly, this bead is from a combinatorial library in which no other major deletions had been observed, so a problem would not have been suspected.

By directly analyzing the ligands used for screening, we have been able to identify byproducts that arise during the synthesis and workup of the samples. This information has guided our optimization of reaction conditions. Unfortunately, all byproducts are not the result of systematic problems and, therefore, cannot be eliminated. By directly analyzing the peptides used for screening, we can catch these problems when they arise and avoid being misled as to the identity of the compound being presented on a bead.

Identifying Peptide Ligands to Streptavidin. Two sets of acetylated and free *N*-terminal pentapeptide libraries were constructed; sublibraries using 7 amino acids (A, R, Q, G, H, F, and P) and full libraries using 16 amino acids (A, R, N, D, Q, E, G, H, L, K, F, P, S, T, W, and Y). The sublibraries contain 7^5 or 16 807 members while the full libraries contain 16^5 or 1 048 576 members. An aliquot of each of the sublibraries (100 mg) was screened against streptavidin–alkaline phosphatase at

Table 3. Peptide Sequences Identified by Screening Acetylated Libraries with Streptavidin

	Sublibrary							
150 ng/mL streptavidin			Ac	H	P	Q	F	P	(3)[a]
			Ac	H	P	Q	F	F	(6)
			Ac	H	P	Q	F	G	(2)
			Ac	H	P	Q	F	H	
			Ac	H	P	Q	F	A	
			Ac	F	P	Q	F	P	
		Ac	H	H	P	Q	F		
		Ac	F	H	P	Q	F		(3)
consensus			Ac	H	P	Q	F		
	Full Library							
80 ng/mL streptavidin			Ac	H	P	Q	F	N	
			Ac	H	P	Q	F	R	
			Ac	H	P	Q	F	F	
		Ac	K	H	P	Q	F		
		Ac	S	H	P	Q	F		
		Ac	S	H	P	Q	N		
			Ac	H	P	Q	N	N	
	Ac	R	G	H	P	Q			
consensus			Ac	H	P	Q	F		

[a] For sequences observed more than once, the number of occurences is given in parentheses.

Table 4. Peptide Sequences Identified by Screening Amine Libraries with a gp120 Antibody

R	I	Q	R	G	P	G	R	A	F	V	T	I	G	K[a]
				Sublibrary										
60 ng/mL Ab				+[b]	H	G	R	A	F					
				+	P	G	R	A	F	(2)[c]				
				+	Q	G	R	A	F	(2)				
				+	F	G	R	A	F					
				+	P	G	A	A	F					
6 ng/mL Ab				+	H	G	R	A	F	(3)				
				+	P	G	R	A	F	(2)				
				+	Q	G	R	A	F	(2)				
				+	F	G	R	A	F	(2)				
				+	P	G	A	A	F	(2)				
consensus				+	X	G	R	A	F					
				Full Library										
1.8 ng/mL Ab				+	Y	G	R	A	F					
				+	Q	G	R	A	F	(2)				
				+	T	G	K	A	F					
consensus				+	X	G	R/K	A	F					

[a] The peptide sequence against which the monoclonal antibody was generated. [b] The + represents the positive charge at the amine on the *N*-terminus. [c] For sequences observed more than once, the number of occurences is given in parentheses.

150 ng/mL. Since each 100 mg of library contains about 60 000 beads, most sequences in the sublibrary should be represented several times. With the acetylated sublibrary, 18 beads were found to be positive and were sequenced (Table 3). A 500 mg sample of the full acetylated library was screened with streptavidin–alkaline phosphatase at 80 ng/mL, and 18 stained beads were isolated; however, one of these beads did not give an interpretable sequence (Table 3). This amount of resin contains about 30% of the sequences in the full library. The amine sublibrary required nearly 10-fold higher concentration of streptavidin–alkaline phosphatase, 1.2 μg/mL, for the beads to begin staining. Because a much higher concentration of target protein was required for staining, we did not screen the amine libraries for streptavidin-binding peptides.

Streptavidin has been used as a target for screening using libraries of peptides on phage[20] and peptides on beads.[3] Both of these methods have identified HPQ as the streptavidin recognition sequence. With our libraries, 17 of the 18 beads isolated from the acetylated sublibrary carried the tetrapeptide sequence HPQF, in agreement with HPQ as the streptavidin recognition sequence and clearly pointing to a *C*-terminal Phe as an additional key residue for binding. This is consistent with a recent report of HPQF as a better streptavidin recognition sequence.[21] It is interesting that in 13 of the 17 HPQF sequences the HPQF is at the *N*-terminus of the peptide presented as acetyl-HPQF. This alignment is consistent with the *N*-terminal acetyl group constituting an additional recognition element for peptide binding to streptavidin.

With the full library, the sequences of 8 of the 17 peptides contained HPQ and 5 of them had HPQF (Table 3). The remaining eight peptides had no significant homology except that they contained a large number of Trp residues. In addition, recognition of streptavidin–alkaline phosphatase by these peptides is not affected by competition with biotin (data not shown). This indicates that they are binding a site on streptavidin–alkaline phosphatase other than the biotin binding site. Because there is no sequence consensus, these Trp-containing peptides could result from nonspecific hydrophobic binding that is not blocked by our buffer. In summary, the data from the acetylated sublibrary and full library clearly show HPQF as the optimal peptide recognition sequence for streptavidin. This result is consistent with literature reports, and demonstrates the utility of the novel sequencing method reported here.

Identifying Peptide Ligands to a Monoclonal Antibody. Both sets of acetylated and free *N*-terminal pentapeptide libraries were also screened against a monoclonal antibody to the HIV-1 protein gp120. This antibody was generated against a 15 amino acid peptide, and its epitope has not been fully defined.[22] A 100 mg aliquot of each of the acetylated and *N*-terminal amine sublibraries was screened against the anti-gp120 antibody at both 60 and 6 ng/mL. This amount of resin should contain 3–4 copies of each sequence on average. The full libraries were screened using 500 mg of resin with an antibody concentration of 1.8 ng/mL. This will present about 30% of the sequences in the full library. Positive beads were selected, and then each of these beads was digested with cyanogen bromide and the released peptides were analyzed by MALDI. The resulting sequence assignments are summarized in Tables 4 and 5.

The results from screening the pentapeptide amine sublibrary are given at the top of Table 4. We sequenced 18 beads, and 15 were found to have GRAF at the *C*-terminus. This is an exact match to four residues of the antigenic sequence. The other three peptides were all PGAAF, which also exactly matches four residues of the antigenic sequence. Additionally, all eighteen sequences have the *N*-terminal amine positioned one residue from the Gly in the consensus sequence. With the full pentapeptide library, five beads were sequenced. One of them gave an ambiguous sequence. The other four match the results from the sublibrary, affording a consensus of G(R/K)-AF at the *C*-terminus (Table 4).

From the pentapeptide acetylated libraries, a different consensus sequence was observed (Table 5). For the sublibrary, 12 of the 16 sequences have a consensus of Ac-RAFxF or Ac-xAFRF where x is a position that did not have a strong consensus. All six of the peptides identified at the lowest antibody concentration match these sequences. Only 1 of these

(20) Devlin, J. J.; Panganiban, L. C.; Devlin, P. E. *Science* **1990**, *249*, 404–406.

(21) Schmidt, T. G. M.; Skerra, A. *Protein Eng.* **1993**, *6*, 109–122.

(22) Durda, P. J.; Bacheler, L.; Clapham, P.; Jenoski, A. M.; Leece, B.; Matthews, T. J.; McKnight, A.; Pomerantz, R.; Rayner, M.; Weinhold, K. J. *AIDS Res. Hum. Retroviruses* **1990**, *6*, 1115–1123.

Table 5. Peptide Sequences Identified by Screening Acetylated Libraries with a gp120 Antibody

	R	I	Q	R	G	P	G	R	A	F	V	T	I	G	K[a]
								Sublibrary							
60 ng/mL Ab					Ac	P'	G	R	A	F			(2)[b]		
					Ac	A	G	R	A	F					
						Ac	G	R	A	F	H				
							Ac	R	A	F	Q	F			
							Ac	R	A	F	F	F			
							Ac	A	A	F	H	F			
							Ac	A	A	F	R	F			
							Ac	Q	A	F	R	F			
							Ac	(G	A)[c]	F	R	F			
6 ng/mL Ab							Ac	R	A	F	Q	F			
							Ac	R	A	F	A	F			
							Ac	R	A	F	H	F			
							Ac	A	A	F	R	F			
							Ac	Q	A	F	R	F	(2)		
consensus							Ac	R/X	A	F	X/R	F			
								Full Library							
1.8 ng/mL Ab							Ac	H	A	F	N	F			
							Ac	H	A	F	R	F	(2)		
							Ac	A	A	F	R	F			
							Ac	K	A	F	H	F			
							Ac	K	A	F	W	F			
							Ac	R	A	F	E	F			
							Ac	R	A	F	R	L			
							Ac	R	A	F	H	L			
							Ac	R	A	F	Q	W	(2)		
							Ac	R	A	W	D	F			
							Ac	R	A	W	N	W			
consensus							Ac	R/X	A	F	X/R	F			

[a] The peptide sequence against which the monoclonal antibody was generated. [b] For sequences observed more than once, the number of occurences is given in parentheses. [c] The order of these two amino acids could not be defined.

12 peptides did not contain an arginine; however, that peptide had a histidine which would be protonated and could make interactions similar to the positively charged arginine. The four peptides that did not fit the Ac-(R/x)AF(x/R)F consensus match the GRAF consensus observed with the amine library. Of the 14 beads selected from the full library, 13 sequenced well while 1 did not give good sequence information. These 13 sequences also match the Ac-RAFxF or Ac-xAFRF consensus sequence identified from the sublibrary. Some conservative replacement of Arg with Lys and Phe with Trp was observed.

The consensus sequences from the acetylated and amine libraries can be overlapped at the GRAF sequence to produce a heptapeptide sequence, xGRAFxF. This overlap aligns the structurally homologous glycine and acetamide, and is consistent with acetylation not simply changing the charge at the *N*-terminus of the peptide but serving as a recognition element. While the consensus sequences from both libraries overlap, they extend in opposite directions. The consensus from the amine libraries has an additional amino acid at the *N*-terminus. The protonated amine on this residue gives the peptide an *N*-terminal positive charge. The consensus sequence from the libraries of acetylated peptides adds two residues to the *C*-terminus, a nonconserved position and a highly conserved Phe. Neither the positive charge from the amine libraries nor the Phe from the acetylated libraries is present in the antigenic sequence.

To determine how well the peptides identified from the libraries bind the antibody, a series of 12 pentapeptides from both the amine and acetylated libraries were synthesized and their relative affinities for the antibody were measured using an antibody capture ELISA with antigen competition. None of them showed significant binding, indicating that pentapeptides may not be long enough to produce good binding to the antibody in solution. However, a seven amino acid peptide designed from the heptapeptide consensus sequence described above did bind well (Table 6). This peptide, PGRAFQF, had an IC_{50} of 50 nM in the ELISA, comparable to that of the 15 amino acid antigenic peptide (IC_{50} = 11 nM). The corresponding seven amino acid peptide from the antigenic sequence had no significant binding, neither as the acetylated peptide nor as the free amine, demonstrating that the consensus heptapeptide is making interactions that are not being made by this region of the antigenic peptide (Table 6). There are two recognition elements in the consensus heptapeptide that are not present in the antigenic peptide, the *C*-terminal Phe and the *N*-terminal amine. Removing either of these elements abolished or greatly reduced binding of the consensus heptapeptide (Table 6), demonstrating that both are critical for strong binding. In summary, we have been able to combine the results from two different sets of pentapeptide libraries to identify a 7 amino acid recognition sequence for this antibody which has about the same affinity for the antibody as the 15 amino acid antigenic peptide. These results clearly demonstrate the importance of screening peptide libraries with different *N*-termini to identify the highest affinity compounds.

Table 6. Relative Affinities of Synthetic Peptides to the gp120 Antibody

peptide sequence	IC_{50}
RIQRGPGRAFVTIGK[a]	11 nM
PGRAFQF	50 nM
Ac-PGRAFQF	525 nM
PGRAFQ	>1 μM
PGRAFVT	>1 μM
Ac-PGRAFVT	>1 μM

[a] The peptide sequence against which the monoclonal antibody was generated.

Conclusions

The current method, which uses termination synthesis and MALDI detection, provides a fast and efficient means for sequencing peptides from individual resin beads isolated from support-bound combinatorial peptide libraries. We were able to sequence peptides at a rate of 25−30 residues per hour. This is much faster than can be achieved with conventional peptide sequencing. The method is also dependable. Only 3 of the 80 beads from the library screenings we report could not be sequenced, and in only one instance was there ambiguity about a sequence. In this latter case the order of a dipeptide segment could not be determined; however, the number of potential compounds was narrowed from hundreds of thousands to only two possibilities. Sensitivity is excellent. We have been able to use as little as 1% of the material from a single 88 μm bead to get good sequencing data. This high sensitivity has enabled us to use beads as small as 17 μm in diameter, increasing the density of compounds in our libraries and allowing us to automate library sorting with FACS. Problems with byproducts from library synthesis, such as deletions and side chain modifications, are easily identified by their effect on the masses of the compounds in the library. This is in contrast to indirect methods for identifying biologically active compounds which cannot identify synthetic byproducts.

Termination synthesis with MALDI detection is a very general method. Since it is not a degradative method, it is compatible with libraries of any compounds that can be prepared by stepwise synthesis. For instance, it can be used with libraries of poly-*N*-substituted glycines, oligocarbamates, or nucleic acid analogs as well as libraries that combine several different monomers. A demonstration of non-peptide sequencing with MALDI can be found in our previous work with methyl

phosphonate oligodeoxyribonucleotides.[23] Finally, we have demonstrated that this sequencing method is practical for combinatorial libraries by screening four peptide libraries with two different proteins, streptavidin and a monoclonal antibody, and efficiently identifying ligands to each of the protein targets.

Experimental Section

Peptide Synthesis: Material and Methods. Peptides and peptide libraries were synthesized using an ACT396 automated peptide synthesizer (Advanced ChemTech, Louisville, KY) and standard FMOC chemistry[24] with *N,N'*-diisopropylcarbodiimide and *N*-hydroxybenzotriazole as the activating system. *N*-Terminal acetylation was carried out with 10% acetic anhydride in DMF. TentaGel S resin was purchased from Rapp Polymere (Tübingen, Germany). *N*-Propionyl-D,L-alanine was prepared as described in the literature.[25] Protected amino acids and Rink amide resin were purchased from Advanced ChemTech. HPLC was performed on a Beckman system with a Vydac 300 Å, 5 μm C18 column. Solvent A was 0.1% TFA in water, and solvent B was 0.1% TFA in acetonitrile. The gradient was 0−40% B over 40 min. TFA was purchased from Pierce Chemicals (Rockford, IL). Solvents were purchased from J.T. Baker (Phillipsburg, NJ) and were used without further purification. All other synthesis reagents were purchased from Aldrich Chemical (Milwaukee, WI).

Peptide Library Synthesis. TentaGel S resin (88 μm beads, 0.25 mequiv/g loading) was used as the solid support for the peptide libraries. Reactions were run on a 25 μmol scale. For the peptide libraries, the six amino acid linker peptide (β-ala)PPPRM was first synthesized on the resin. Randomized positions in the library were generated using a split-synthesis approach.[17] The resin was divided such that there was one reaction vessel for each amino acid to be incorporated at that position in the library. The resin was deprotected with 25% piperidine in DMF for 1 min and then 15 min. After washing the resin five times with NMP, a different amino acid was coupled to the resin in each of the reaction vessels using a 10-fold excess of amino acid. After each coupling, all of the FMOC-protected resin was removed from the synthesizer, mixed, and redistributed to the reaction vessels. This process was repeated to produce pentapeptide libraries. Partial termination was effected by adding 10% *N*-acetyl-D,L-alanine to the amino acid solutions used for coupling at the randomized positions. When peptide synthesis was complete, the resin was washed with methanol (1.5 mL, 5 × 1 min) and dried under high vacuum. The side chains were deprotected for 2 h using 1 mL of the following mixture per 100 mg of peptidyl resin: 10 mL of TFA, 0.75 g of phenol, 0.5 mL of mercaptoacetic acid, 0.5 mL of thioanisole, and 0.5 mL of water. After side chain deprotection, the resin was washed with 90% TFA (1 mL, 2 × 5 min), 10% *N,N*-diisopropylethylamine in DMF (1.5 mL, 1 and 10 min), and DMF (1.5 mL, 5 × 1 min) and stored at 4 °C.

Soluble Peptides. Rink amide resin was used as the support for the soluble peptides. Reactions were run on a 50 μmol scale using a 5-fold excess of amino acid. After synthesis, the resins were washed twice with 2 mL of methanol and dried. The peptides were cleaved from the resin and deprotected for 2 h with 2 mL of the following mixture: 10 mL of TFA, 0.75 g of phenol, and 0.25 mL of ethanedithiol, 0.5 mL of thioanisole, and 0.25 mL of water. The samples were concentrated under nitrogen, precipitated with diethyl ether, taken up in 15 mL of 1% acetic acid, and lyophilized. The lyophilized peptides were then taken up in 5 mL of 1% acetic acid and extracted twice with 50 mL of diethyl ether. The samples were lyophilized, and peptides were purified in portions by HPLC. All peptides gave a single peak by HPLC and the expected mass by ion spray mass spectrometry.

Library Screening. For the sublibraries, 100 mg of beads was mixed with 1 mL of TBS blocking buffer (25 mM Tris, 150 mM NaCl, 0.1% TWEEN 20, 1% casein, pH 7.4) for 1 h to block nonspecific protein binding sites. The buffer was removed and replaced with 1 mL of TBS blocking buffer containing 150 ng/mL streptavidin−alkaline phosphatase conjugate (Pierce Chemicals, Rockford, IL) or either 60 or 6 ng of anti-HIV-1 gp120 monoclonal antibody (NEA-9205, Du Pont NEN Research Products, Boston, MA). For the antibody screening, after 1 h the solution was removed and the beads were washed three times with TBS washing buffer (100 mM Tris, 276 mM NaCl, 5.4 mM KCl, 0.1% TWEEN 20, pH 7.2). Following the washing, 1 mL of TBS blocking buffer containing 35 ng/mL of goat antimouse IgG alkaline phosphatase conjugate (Pierce Chemicals, Rockford, IL) was added to the library and the beads were mixed for 1 h. The final step for screening with both streptavidin and the monoclonal antibody was to wash the beads five times with TBS washing buffer and to transfer them to a filtration apparatus fitted with a 0.22 μm cellulose acetate filter. A standard alkaline phosphatase substrate (4 mL), 5-bromo-4-chloro-3'-indolyl phosphate *p*-toluidine salt and nitro-blue tetrazolium chloride, was added to each set of beads, and the staining reaction was allowed to proceed for 0.5−1 h. For the antibody screening, the libraries were prescreened with the goat antimouse IgG alkaline phosphatase conjugate to ensure that it only bound to the beads through the monoclonal antibody. Staining was terminated by removing the substrate solution with vacuum and washing once with deionized water. Positive beads were identified by their dark blue color and were removed under low magnification.

The full libraries were screened using 500 mg aliquots of resin and the general procedure outlined above except that reaction volumes were increased to 5 mL. The streptavidin screening was run with a protein concentration of 80 ng/mL and the antibody screening with an antibody concentration of 1.8 ng/mL.

Release of Peptides from Isolated Beads and Sample Preparation Prior to MALDI. Individual stained beads were placed into 500 μL Eppendorf tubes, and 15 μL of 20 mg/mL cyanogen bromide in 0.1 N HCl was added. The reactions were run at 25 °C for 16 h in the dark. The reactions were stopped by freezing and lyophilizing to dryness. The resulting peptide mixtures were redissolved in 20 μL of aqueous 0.1% TFA. A 1 μL aliquot of the TFA solution was mixed with an equal volume of a saturated solution of α-cyano-4-hydroxycinnamic acid,[26] which was prepared by dissolving 10 mg in 1 mL of aqueous 30% acetonitrile (v/v) containing 0.1% TFA. A 1 μL aliquot of that solution was analyzed by MALDI after thorough water washing to remove any alkali salt contaminants.

Mass Spectrometry. All MALDI mass spectra were obtained in the positive ion mode on a Vestec (Houston, TX) VT2000 linear time-of-flight mass spectrometer as previously described.[23,27] Peptide samples, mixed in the UV-absorbing matrix, were irradiated with the output of a frequency-tripled Q-switched Nd:YAG laser. Spectra obtained from 50 laser pulses were acquired in 10 s and added to produce the MALDI results used for sequencing. Mass measurement accuracy of ±0.3 Da was achieved by use of small peptides of known mass as standards for internal calibration.[12,13] Ion spray MS was run on a Sciex API-III (PE Sciex, Thornhill, Canada).

Competition ELISA. Test peptides (5−3000 nM) and the anti-HIV-1 gp120 monoclonal antibody (1.5 μg/mL) were added to microtiter wells coated with the peptide to which the antibody was raised, RIQRGPGRAFVTIGK, using a standard format for antibody capture ELISA with antigen competition.[28] IC_{50} values were calculated for inhibition of antibody binding to the support-bound antigenic peptide.

Acknowledgment. We thank Y. Sun and M. D. Bauer who performed the ion spray measurements. We also thank C. Oppenheimer for assistance with the ELISA assay and for critical review of the paper.

JA9417411

(23) Keough, T.; Baker, T. R.; Dobson, R. L. M.; Lacey, M. P.; Riley, T. A.; Hasselfield, J. A.; Hesselberth, P. E. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 195−200.

(24) Aetherton, E.; Sheppard, R. C. *Solid Phase Peptide Synthesis: A Practical Approach*; IRL Press: Oxford, 1989.

(25) Otani, T. T.; Briley, M. R. *J. Pharm. Sci.* **1978**, *67*, 520−526.

(26) Beavis, R. C.; Chaudhary, T.; Chait, B. T. *Org. Mass Spectrom.* **1992**, *27*, 156−158.

(27) Keough, T.; Takigiku, R.; Lacey, M. P.; Purdon, M. *Anal. Chem.* **1993**, *64*, 1594−1600.

(28) Harlow, E.; Lane, D. *Antibodies: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Plainview, NY, 1988; pp 570−573.