JOURNAL OF combinatorial CHEMISTRY

Subscriber access provided by American Chemical Society

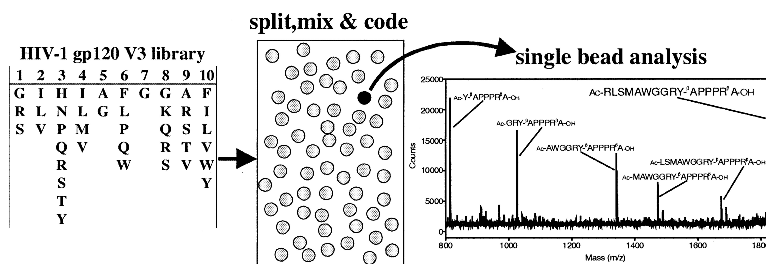
Article

Mass Spectrometric Sequencing of Individual Peptides from Combinatorial Libraries via Specific Generation of Chain-Terminated Sequences

Christian Hoffmann, Dierk Blechschmidt, Ralf Krger, Michael Karas, and Christian Griesinger

J. Comb. Chem., **2002**, 4 (1), 79-86• DOI: 10.1021/cc010057x • Publication Date (Web): 19 December 2001

Downloaded from http://pubs.acs.org on March 20, 2009



More About This Article

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

- Supporting Information
- 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

ACS Publications
High quality. High impact.

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

Mass Spectrometric Sequencing of Individual Peptides from Combinatorial Libraries via Specific Generation of Chain-Terminated Sequences

Christian Hoffmann, Dierk Blechschmidt,† Ralf Krüger,‡ Michael Karas,‡ and Christian Griesinger*

Institute of Organic Chemistry, University of Frankfurt, Marie-Curie-Strasse 11, D-60439 Frankfurt am Main, Germany, and the Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

Received August 15, 2001

Combinatorial peptide libraries are a versatile tool for drug discovery. On-bead assays identify reactive peptides by enzyme-catalyzed staining and, usually, sequencing by Edman degradation. Unfortunately, the latter method is expensive and time-consuming and requires free N termini of the peptides. A method of rapid and unambiguous peptide sequencing by utilizing synthesis-implemented generation of termination sequences with subsequent matrix-assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometric analysis is introduced here. The required capped sequences are determined and optimized for a specific peptide library by a computer algorithm implemented in the program Biblio. A total of 99.7% of the sequences of a heptapeptide library sample could be decoded utilizing a single bead for each spectrum. To synthesize these libraries, an optimized capping approach has been introduced.

Introduction

Solid-phase combinatorial peptide libraries are important since their first applications[1,2] in the early 1990s in lead substance determination and drug discovery.[3] Peptide ligands to a monoclonal antibody against β-endorphin[1] and to GPIIb/IIIa[2] were the first examples using screening techniques in combination with peptide libraries. The broad versatility of the application of resin-bound peptide libraries has been put into evidence by the determination of substrate specifity for the peptide deformylase[4] or the confirmation of cross-reactivity between Lyme antibody H9724 and human heat shock protein 60 by epitope mapping.[5,6] Some essential strategies for a successful setup of combinatorial peptide libraries have been discussed by Chu et al.,[7] such as finding the minimal number of amino acids necessary for antibody recognition or optimizing the linker construct between the bead matrix and the functional peptide.

In addition, the concept of solid-phase-based biological screening has been extended to other solid-phase library types, such as benzodiazepine,[8] glycopeptide,[9] peptide mimetics, peptide nucleic acids (PNA), and oligocarbamate libraries.[3]

For all peptide libraries synthesized on the solid phase, the sequence of a peptide on a specific bead needs to be derived after selection for a specific binding affinity or biological response. One way is the Edman degradation, which can be done on commercially available instrumentation. However, free N termini of the peptides are required and the technique is rather expensive and time-consuming. Therefore, methods for coding of the libraries have been developed, resulting in quick and unambiguous identification of an individual sequence.

One can distinguish three major coding strategies: positional, chemical, and, more sophisticated, electronic encoding.[10] Positional encoding was one of the first applications of this approach, realized by pin synthesis[11] of peptide libraries; today's most frequent application is the "DNA chip technology". Chemical encoding is usually carried out by tagging the individual members of the library either with nucleic acids[12] or with more robust, small organic molecules.[13] Finally, electronic encoding is accomplished by utilization of a radio frequency memory tag.[14] A reaction platform containing an rf chip therefore is tagged in a machine-readable fashion.[10]

Chemical encoding with small molecular tags is most frequently used for the construction and synthesis of combinatorial libraries. However, the chemical synthesis of a library can be complicated because of the necessity to perform the synthesis of the coding tags.[15] While libraries with millions of members have been reported in the literature,[13] most libraries using chemical tagging did not exceed several thousands of members.[12,15] Last, the chemical coding provides only indirect information about the coded

* To whom correspondence should be addressed. E-mail: cigr@mpibpc.gwdg.de.

† Current address: Massachusetts Institute of Technology, Francis Bitter Magnet Laboratory, 170 Albany Street, Cambridge, MA 02139.

‡ Division of Instrumental Analytical Chemistry, University of Frankfurt, Theodor-Stern-Kai 7, D-60590 Frankfurt am Main, Germany.

10.1021/cc010057x CCC: $22.00 © 2002 American Chemical Society
Published on Web 12/19/2001

compound of the library; its identity and purity need to be determined by different means.

The method we describe here advances a concept originally proposed by Youngquist and Keough.[16] It enables the determination of a peptide sequence, which was identified as biologically active, simply by cleaving the peptide off the bead and recording a mass spectrum.

This approach comprises the following steps: A combinatorial peptide library usually is synthesized by the "split, mix, and combine" method by Furka[17] and Lam.[1] The resin quantity is split into *m* portions, where *m* is the number of amino acids at the mentioned library position. The amino acids are coupled to the resin; then the whole quantity is pooled and thoroughly mixed. After removal of the N-terminal protecting group, the reaction cycle is reiterated, and finally one distinct peptide per single bead is provided. To distinguish between the various peptides in the library, barring isobaric amino acids and multiple sequence similarities, sequences that are capped at specific positions are introduced during the synthesis of the library. In addition to the desired peptide sequence, each bead then contains also a set of capped sequences. Synthesis is performed by using mixtures of the desired FMOC-protected amino acid and its permanently N-terminally blocked derivative, for instance, with an acetyl or benzoyl group.

The amount of the capping reagents is 5% with respect to the whole resin occupancy. This is achieved because of the equal coupling kinetics of the FMOC protected amino acid as the capped amino acid in the coupling reaction (data not shown). As shown later, 5% of each capped sequence is sufficient for mass spectrometric analysis and renders the desired peptide to be the main component on the bead because of this optimized capping scheme. Furthermore, our approach utilizing amino acid derivative mixtures makes capping applicable and manageable for large combinatorial peptide libraries up to several thousand members.

In contrast to the original approach[16] with fixed capping steps and capping reagents, we use an adaptive capping strategy that allows us to minimize the necessary number of capping positions and quantity of capped sequences. Moreover, the limited diversity of a library synthesized with a "fixed capping" strategy is overcome by the adapted capping that we propose. Because the capped peptides are as close in structure to that of the main peptide as possible, it is unlikely that the capped peptides exhibit biological activity that is different from that of the main peptide.

Our approach, implemented in the program Biblio, optimizes the capping pattern for a given library with respect to the number of capping steps; e.g., even in large libraries, not every amino acid position has to be capped. Nevertheless, for every position of any peptide of the library, any amino acid can be chosen and later be decoded.

Thus, our approach solves the problem of isobaric amino acids and sequence similarities occurring in large libraries. Previous approaches to solve the problem of isobaric masses included the use of halogenation[18] or incorporation of isotopes.[19] In the chemical encoding approach, isobaric masses are no problem. In our approach, the diversity of peptide libraries to be synthesized is not restricted. As shown below by means of a model peptide sequence, the quantity of the main sequence is maximized.

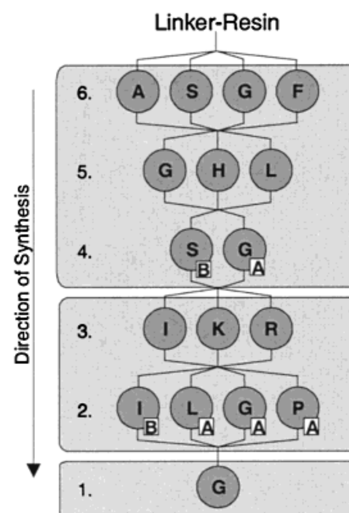


Figure 1. Different steps of coding a sample peptide library using the algorithm in Biblio. The peptide is synthesized on a linker bound to a resin. The first capping position is amino acid 4 because identical masses occur, e.g., for the sequences SGG and GGS. Capping with different capping reagents denoted by A and B resolves this ambiguity. The next capping is done at position 2 because the dipeptides IL and II are isobaric. For this library two capping steps are sufficient.

Results and Discussion

Algorithm of Biblio. Biblio fulfills two tasks. In the first step, it suggests the optimal capping pattern for mass spectrometric analysis for a given peptide library, and in a second step it translates a set of mass peaks into the most probable peptide sequence from the given library. For the first step, the algorithm relies on the fact that every peptide in a specific library is described by its unique set of masses brought about by the capping procedure. One mass in the set is given by the mass of the full-length peptide. The other masses depend on the peptide sequence and the capping pattern of the library. The determination of that pattern with a minimal number of capping steps, namely, the coding of the library, is suggested by Biblio and is illustrated by the library given in Figure 1. First, the possible masses of the monopeptides, i.e., the C-terminal amino acids, are calculated and compared. In the case of isobaric amino acids, capping would be already necessary in the first step. If not, as is the case here, the second amino acid position is added to the calculation. The masses of the possible dipeptides (AH, AL, SG, SH, ...) are calculated and compared. These steps are repeated until similar or identical masses are found. Two masses are similar or identical if they deviate by less than $\pm$1 Da. After the third amino acid positions, there are similar masses in the library (Figure 1, top); for example, the tripeptides with the sequences GXS and SXG have identical masses (X is any amino acid at position 5). Therefore, capping is necessary. The correct pattern of capping reagents at the given amino acid position is determined by permutation of the given set of capping reagents summarized in Table 2. A correct capping pattern will render the masses of all

members of the library different up to step 3. Since all amino acids at position 3 in the library of Figure 1 are capped, the library is determined unambiguously to that point. Thus, this part of the library can be excluded from further considerations when continuing the synthesis. Starting with the fourth position, masses of the monopeptides (I, K, R) are calculated and compared. No similarities are found. Subsequently the fifth position is added. The masses of the possible dipeptides (II, IL, IG, IP, KI, KL, ...) are calculated and compared. Peptides with identical masses are found because L and I have the same masses. After the capping pattern is determined, the library is unambiguously determined up to fifth amino acid position. In that sample library, the sixth position contains a single G only, so the calculation is finished (Figure 1, bottom) and the hexapeptide library requires only two capping steps for unambiguous analysis, i.e., complete coding and decoding. Every full-length (and every non-full-length) peptide sequence is described by its unique set of masses. This algorithm keeps some redundancies and does not always lead to the theoretical minimum number of capping steps, since all amino acids at a specific position are always capped; for example, in the sample library, capping of G and P at the fifth position is not necessary for coding the library unambiguously. However, incomplete capping of amino acid positions would prevent breaking the whole library into subgroups, which would in turn require more computational time to analyze larger libraries. Thus, the proposed adaptive coding scheme is slightly redundant. This ensures that the method tolerates errors of masses in the mass spectroscopic analysis and even sometimes missing peaks.

Decoding of the library is also done with the help of the program Biblio by comparing a given set of masses (from the mass spectrometric analysis) with all the masses the coded library can produce. For every mass the probability of a specific sequence is generated by comparing the observed masses with the theoretically predicted mass pattern. The matching of the masses is measured by a quality factor Q that is defined in the following way. Q equals 1, 0.8, and 0.3 if an experimentally determined mass matches a calculated mass within ±1, ±2, and ±3 Da, respectively. If the mass deviation is bigger than ±3 Da, the quality factor is 0. The sequence rendering the highest sum of quality factors is assumed to be the correct result. Sequence determination using this parametrical approach was found to be robust against artifacts; even if there are more mass peaks present in the spectrum (e.g., those arising from artifacts), the correct mass is determined. The lower quality factors 0.8 and 0.3 are introduced to take smaller random measurement errors into account.

Construction of Encoded Solid-Phase Combinatorial Libraries. The combinatorial peptide libraries have been synthesized on TentaGel S HMB resin by the "split, mix, and combine" method [17] using standard FMOC chemistry.[20] Capped amino acids in the library are introduced by application of mixtures of FMOC-protected amino acids and their permanently blocked ("capped") derivatives to an amount of 5%. Capping is only performed at positions suggested by Biblio. A peptide linker, $^{\beta}$Ala-Pro-Pro-Pro-Arg-$^{\beta}$Ala, between the HMB resin and the functional amino acid sequence was introduced for the following reasons. First, a spacer enlarges the distance between the HMB resin and the peptide sequence, increases its mobility, and enhances the peptide yield and availability for the biological target. We proved our spacer as being biologically inert by on-bead screening with an HIV-positive serum and visualizing the results with BCIP/NBT staining.[16] No differences to negative control experiments have been observed. It is known that short peptidic linkers containing proline, arginine, and $^{\beta}$alanine do not interfere with biological screening results.[5,7,16] Second, for matrix-assisted laser desorption ionization mass spectrometric (MALDI-MS) measurements, a "charged anchor" that can easily be protonated is benefical for uniform ionization behavior of all involved peptides and, eventually, enhanced spectrum quality. Therefore, we use arginine with its guanidine group as "charged anchor".[32] Third, a uniform cleavage off the bead is achieved by inserting $^{\beta}$alanine at the C terminus and using a weak aqueous base. This allows us to abandon the BrCN cleavage of the C-terminal methionine used earlier.[16] Because we use TentaGel S (hydroxymethylbenzoic acid) HMB resin, the peptide is covalently attached to the bead via an ester bond, providing a greater oxidation resistance with respect to the aggressive side chain deprotection conditions than the relatively labile methionine residue. Standard cleavage conditions from the TentaGel resin would be incubation with 1 N aqueous NaOH. However, because single-bead measurements are to be performed, the residual alkali ion content is incompatible with MALDI ionization. For this reason, we cleaved the peptides off via in situ generation of hydroxyl ions by incubating a bead with a solution of 10% aqueous triethylamine overnight. The linker is big enough to increase the mass of even the smallest capped peptide above 600 Da. Therefore, disturbances from the matrix signals in the MALDI mass spectrum even at an abundance of femtomole (10^{-15} mol) of the desired peptide are prevented.

Table 1. Pentapeptide Library Scheme with 144 Species[a]

amino acid positions				
1	2	3	4	5
A	F[2]	A	F	I[2]
G	G[2]	G	Q	L[1]
	I[4]	L	T	
		Y		

[a] The N terminus is the free amine, and $^{\beta}$APPPRM was used as the linker. The capping positions predicted by Biblio are denoted by superscripts. The translation of the superscripts into the N-terminal protecting group is listed in Table 2.

MALDI Sequencing of Selective Beads of Two Model Libraries. The first model library is a pentapeptide library with 144 different species, as denoted in Table 1. The linker used here is $^{\beta}$Ala-Pro-Pro-Pro-Arg-Met. The N terminus of every full-length peptide of this library is a free amine. Before synthesis, Biblio was fed the library data: number of amino acid positions, desired amino acids in each position, and selection of a mass variance interval. The program then suggested only two capping steps (at positions 2 and 5) requiring a mass measurement accuracy of ±1 Da for each MALDI signal. It uses acetyl, benzoyl, 2-naphthoyl, and

Table 2. N-Terminal Modifications Used for Blocking the Amino Termini of the Capping Amino Acids[a]

group	modification	abbreviation	mass
1	acetyl-	Ac-	+42.0
2	benzoyl-	Bz-	+104.0
3	2-naphthoyl-	Naph-	+154.2
4	propionyl-	Prop-	+56.0

[a] The mass is denoted relative to the unprotected amino acid residue.

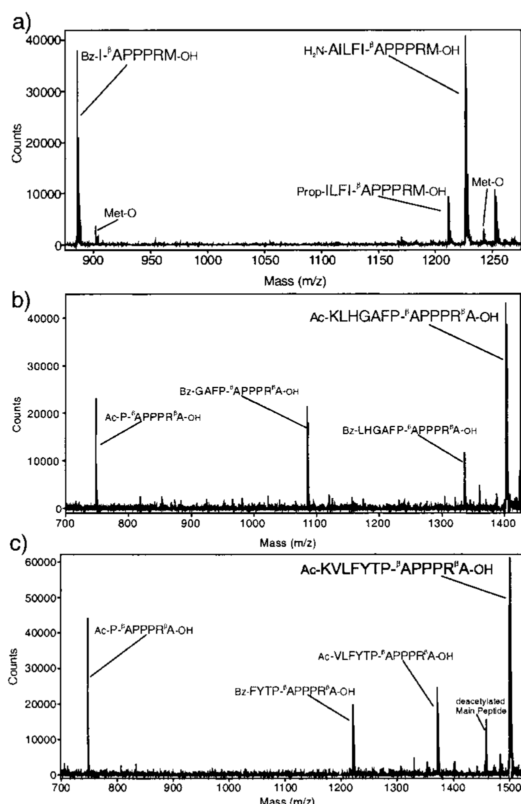


Figure 2. (a) MALDI mass spectrum of the coded pentapeptide AILFI and its associated two coding subsequences. Since methionine is used for the linker, additional peaks resulting from methionine oxidation are observed. (b) MALDI mass spectrum of the heptapeptide KLHGAFP and its three coding peptides. The assignment of the peaks was calculated using Biblio. (c) MALDI mass spectrum of the peptide KVLFYTP and its three coding peptides. The assignment of the peaks was also calculated using Biblio. The major byproduct displays a peak resulting from deacetylation of the main peptide.

propionyl as capping agents (Table 2). If a fixed capping strategy had been used, five capping steps would have been necessary.

Several beads have been picked with a needle. After the peptide was cleaved off the bead, the peptides of individual beads were analyzed by MALDI mass spectrometry. A typical spectrum is shown in Figure 2a.

Two signals of the capping peptides and one of the main peptide had to be expected and are indeed observed with an excellent signal-to-noise ratio. Because of the variable capping reagents, the sequence cannot directly be read from the spectrum. The three observed masses were fed to Biblio, which unambiguously determined Ala-Ile-Leu-Phe-Ile as the sequence of the peptide. The peak multiplicity arising from isotopic resolution in the peptide signals is discussed later.

Table 3. Heptapeptide Library Scheme with 5184 Species[a]

amino acid positions						
1	2	3	4	5	6	7
A	I^1	A	F^2	A	F	I^2
G	L^2	H	G^2	G	Q	L^1
F	P^1	L	I^4	L	T	P^1
K	V^1			Y		

[a] The N terminus is acteylated, and $^{\beta}$APPPR$^{\beta}$A was used as the linker. The capping positions predicted by Biblio are denoted by superscripts. The translation of the superscripts into the N-terminal protecting group is listed in Table 2.

Because of the use of methionine in the linker, artifacts of methionine oxidation byproducts are observed. Another artifact is emerging in all pentapeptide library mass spectra. A peak with a mass difference of +26 Da with respect to the main peptide of the spectrum occurs with significant signal intensity. Unfortunately, the origin of this peak, probably an N-terminal modification of the main peptide, remains unresolved; neither MALDI time-of-flight postsource decay (PSD) fragmentation analysis[21] nor nano-electrospray-ionization (ESI)-mass-spectrometry spectra recorded in addition provided conclusive data.

To expand the diversity of the library, a heptapeptide library (Table 3) consisting of 5184 sequences was synthesized after calculating the optimal capping pattern with Biblio. Only three positions of this library instead of seven had to be capped (positions 2, 4, and 7 in Table 2), requiring again a mass measurement accuracy of ±1 Da. The linker used from now on is $^{\beta}$Ala-Pro-Pro-Pro-Arg-$^{\beta}$Ala, avoiding the difficulties of methionine oxidation. Acetylation of the N terminus of the main peptide prevented the occurrence of artifact peaks related to the major signal. Typical spectra are shown in parts b and c of Figure 2. Four signals are expected: three coding peptides and the main sequence. The detected signals in Figure 2b, for instance, are assigned with the corresponding peptide sequence revealed by Biblio; the peaks corresponding to the protonated main peptide show up at m/z = 1401.5 (Ac-KLHGAFP-linker) and the capped peptides at 1335.4 (Bz-LHGAFP-linker), 1085.1 (Bz-FP-linker), and 747.8 (Ac-P-linker). Figure 2c shows decoding of another peptide from a single bead: m/z = 1499.2 for the main peptide Ac-KVLFYTP-linker, and 1370.9 (Ac-VLFYTP-linker), 1220.8 (Bz-FYTP-linker), and 747.5 (Ac-P-linker) for the capped peptides. In both spectra, the main peptide is found to be the most intense signal. Nevertheless, MALDI mass spectrometry does not allow a quantitative analysis of peak intensities; therefore, the molar ratio of main to capped peptides has to be evaluated separately as shown later.

All recorded spectra display an excellent signal-to-noise ratio and resolution to the monoisotopic masses (Figure 3).

Undesired chain-terminated sequences, incomplete cleavage of side chain protection groups, or sequences with missing amino acids are an important problem in peptide synthesis. These byproducts could prevent the decoding process because the signal of such peptides could only be distinguished from the desired capped peptides on the basis of their molar ratio. Only few byproducts appeared in our

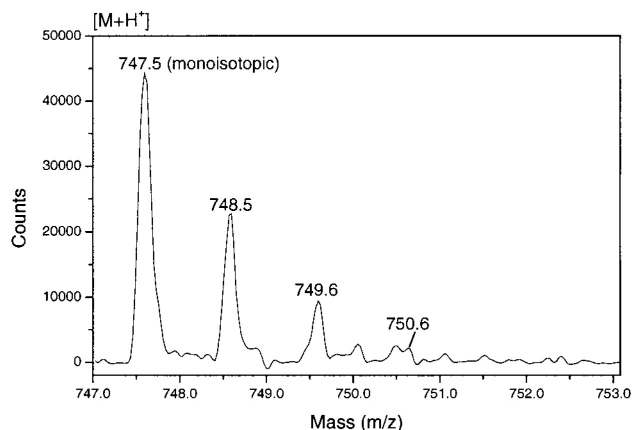


Figure 3. Experimentally found isotopic pattern of Ac-P-$^{\beta}$APPPR$^{\beta}$A-OH. On the basis of the assumption of 90 pmol per single bead, this spectrum was recorded with an amount of 100 fmol (1 fmol = 10^{-15} mol).

Table 4. Byproducts Observed in Single-Bead Sequencing[a]

modification	mass
Na	+22.0
K	+38.0
Ac	+42.0
deacetylation	−42.0
O	+16.0

[a] Oxidation is only detected in peptides with methionine residues.

libraries and are summarized in Table 4. Major peaks arise from adducts of the peptide with sodium or potassium. These signals mainly appeared together with the full-length peptide because of the fact that it has the highest abundance. Furthermore, desacetylation of the main peptide is observed as a well-known side reaction during side chain deprotection.

Incidentally, a signal 42 Da heavier than the unmodified peptide occurs because of a second acetylation site in the peptide (not shown here). Under the acetylation conditions used (10 equiv of acetic anhydride, 50 equiv of triethylamine, 2 h), the guanidine terminus of arginine is additionally acetylated,[34] even though it is protected because of the requirements of peptide synthesis.

Statistical Evaluation of 299 Heptapeptide Library Probes. To ensure the quality of the coding scheme computed by Biblio and to show the reliability of the single-bead decoding and sequencing concept, 300 beads of the heptapeptide library have been picked and the peptides have been analyzed by MALDI mass spectrometry. A total of 299 of the 300 beads could be sequenced successfully; in any case, the desired three plus one signals (three capped peptides plus one main peptide) were observed. Moreover, all spectra could be assigned to a single peptide sequence, unambiguously resolved by Biblio. This is a hit rate of about 99.7%. One bead revealed only one signal of low mass, which could not be related to any full-length peptide but rather to a termination sequence, Bz-GLFI-linker. This sequence is coding for 48 various full-length peptide sequences of the library. Thus, the true sequence of this full-length peptide could not be determined. As shown in Table 5, 249 individual sequences have been identified overall. Of the 249, 206 appeared only once. The sequences occurring more than twice are denoted individually.

Table 5. Result of the Mass Spectrometric Analysis of 299 Beads from the Heptapeptide Library[a]

number of sequences	multiplicity
206	1
37	2
5	3
1	4

[a] Overall, 249 different heptapeptide sequences have been observed in 299 probes. The five sequences appearing three times are FVHIYFP, GVAIYQP, GVHIYQP, KHVIGFP, KVLIYQP. The sequence appearing four times is KVHIYQP.

For statistical analysis of the sequenced peptides we looked for the occurrence of the amino acids used in each position and for the independence of occurrence of these amino acids in these positions. The occurrence of the amino acids in each position should depend on the fractions taken for each coupling step. These fractions were rather equal except for one amino acid in each synthetic step, which was applied to the so-called reserve batch. The reserve batch contained between 47% and 73% of the whole resin, as was observed from the occurrence of the respective amino acid.

Whether the occurrence of specific amino acids is correlated in different positions was tested as well. χ^2 tests of independence were performed for each pair of amino acid positions (21 tests in all). Two of these pairs showed significant dependence at the level of $p = 0.01$: positions 1 and 6, and positions 5 and 6. The other 19 tests did not reveal significant dependence at the 0.01 or the 0.05 level.

As a second test of independence of the amino acid occurrence in the different peptide positions, the number of sequences found to be present k times ($k = 1, 2, 3, ...$) in the experimental sample of 299 beads was compared with that in a computer experiment in which the computer drew 100 000 times the 299 samples from the library of beads, assuming the experimental a priori probabilities for the amino acids in each position and independence of each position. The number of sequences present only once ($k = 1$) was strikingly higher in the *in silico* drawings than in the actual sample; in only 59 of 100 000 permutations, there were as few of these as actually observed. Figure 4 compares the predicted multiplicity distribution by this means with the experimentally determined distribution indicating the variance. The experimental probe result lies within 3σ of the *in silico* probe.

It thus appears that, for reasons as yet unknown, the amino acids in the positions are not independent. Though clearly detectable, this effect appears to be much less important in disturbing the desired equal distribution of sequences than the nonuniformity of amino acids in the individual positions. For further experiments, the fraction of the reserve position has to be minimized, and in experiments yet to be published we obtained a more homogeneous distribution of peptide sequences.

Estimation of Main Peptide/Capping Peptides Ratio by HPLC. For testing in biological systems, it is desirable to keep the amount of the main peptide on a bead maximal. Since MALDI-TOF MS is not a quantitative method, the yield of main peptide versus capped sequences had to be established by another method. Therefore, we synthesized

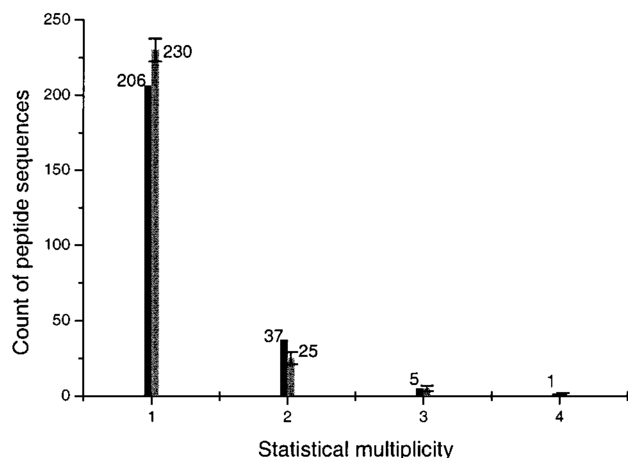


Figure 4. Experimentally determined distribution of 299 decoded peptides from the heptapeptide library with respect to their occurrence multiplicity (denoted in black) in comparison with the theoretically predicted data (denoted in gray). The theoretical data are derived from an *in silico* experiment where the a priori probabilities of each amino acid in each position have been taken into account and a statistical independence of the amino acid positions is assumed. The error bar corresponds to the standard deviation. The result from our drawing lies within 3σ.

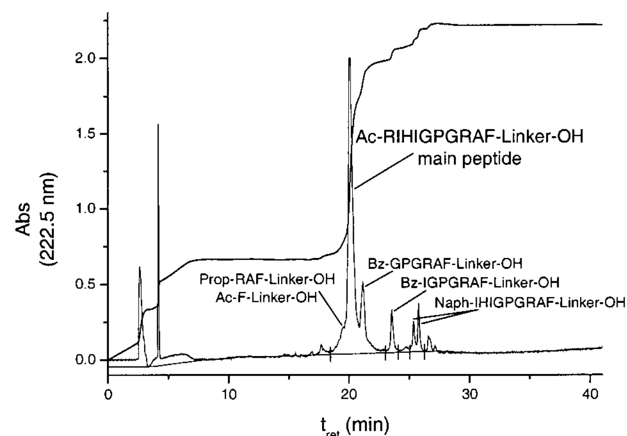


Figure 5. HPLC chromatogram of the peptide mixture of a decapeptide and its corresponding coding subsequences. The column conditions that best resolved the mixture were the following: gradient 0−100% aqueous acetonitrile over 50 min (40 min shown), using a Eurospher RP 18, 4 mm × 250 mm, 5 μm, 100 Å C18 column; detection was performed at 222.5 nm. The peptide region of the chromatogram spans a region of t_r = 18−26 min. The integral within this region containing the main peptide and three capping peptide contains 88.3% of the whole peptide region. The remaining two capping peptides make up for 4.4% and 7.3%, respectively.

the N-terminally acetylated peptide Ac-RIHIGPGRAF-$^{\beta}$APPPR$^{\beta}$A and its capping peptides manually on a larger amount of resin (50 mg) in order to quantify the peptide sequences by HPLC. The main peptide is a sequence found in the V3 region of the HIV-1 clade B gp120 envelope protein.[23] Details and use of the library are described later in the text. This peptide and five capped truncated versions were synthesized, as described above, using 5% of the capping reagent. Thus, the following peptides were present:

main peptide	Ac-RIHIGPGRAF-$^{\beta}$APPPR$^{\beta}$A
capping peptides	Naph-IHIGPGRAF-$^{\beta}$APPPR$^{\beta}$A Bz-IGPGRAF-$^{\beta}$APPPR$^{\beta}$A Bz-GPGRAF-$^{\beta}$APPPR$^{\beta}$A Prop-RAF-$^{\beta}$APPPR$^{\beta}$A Ac-F-$^{\beta}$APPPR$^{\beta}$A

Because five termination sequences are divergently synthesized to an individual amount of 5%, the relative yield of the main peptide should be $(0.95)^5 \approx 77\%$ as opposed to 5% of each capped peptide. This is checked by HPLC, which provides a good estimation of peptide amounts.[33]

The chromatogram resolving the peptide mixture best is shown in Figure 5. The peak containing the main peptide and unresolved three capping peptides, as revealed by MS, has an integral of 88.3%; the residual capping peptides have an integral of 4.4% (Bz-IGPGRAF−linker) and 7.3% (Naph-IHIGPGRAF-linker), respectively. The integration assumes equal molar UV extinction of all peptides.[33] A more sophisticated analysis based on estimation of the UV extinctions of each peptide bond[22] and of the naphthoyl group present in one capped peptide does not change this result appreciably. On the basis of this latter estimation, the expected area of the main peak (main peptide plus three termination sequences) is 81.1% of the total peak area. Additionally, the HPLC peak containing Bz-IGPGRAF-linker is calculated to cover 3.3% of the total area of peptidic HPLC peaks; finally, the peak containing Naph-IHIGPGRAF-linker is estimated to cover 15.6% of the total area of the peptidic HPLC peaks. These values all are within the expected tolerance interval of such a peak area estimation (about 10%).[33] No other functional groups in the peptides show up as UV-active at the considered wavelengthh (222.5 nm).

Table 6. Decapeptide Library Scheme with 432 000 Species Derived from the V3 Loop of HIV-1 clade B (for Details See Text)[a]

amino acid positions									
1	2	3	4	5	6	7	8	9	10
G	I[3]	H	I[2]	A[1]	F	G	G[1]	A	F[1]
R	L[1]	N	L[1]	G[2]	L		K[1]	R	I[1]
S	V[2]	P	M[1]		P		Q[2]	S	L[2]
		Q	V[1]		Q		R[4]	T	V[1]
		R			W		S[2]	V	W[1]
		S							Y[1]
		T							
		Y							

[a] The N terminus is acteylated, and $^{\beta}$APPPR$^{\beta}$A was used as the linker. The capping positions predicted by Biblio are denoted by superscripts. The translation of the superscripts into the N-terminal protecting group is listed in Table 2.

Application of the "Adapted Capping" Method to a Biologically Relevant Decapeptide Library. After establishing the method introduced above, the concept is applied to a biologically relevant decapeptide library with 432 000 peptides derived from the V3 loop that is part of the viral envelope protein gp120 of HIV-1 and is one of its immunodominant regions.[24−27] The decapeptide library is summarized in Table 6. It is derived from HIV-1 gp120 sequence data.[23] This library covers the variety of HIV-1 clade B, which is the most divergent of all known HIV clades. Because the tip of the V3 loop is especially known to be antigenic to linear peptides,[24−29] these sequence positions have been varied in the library. The sequence-coding pattern has been computed using Biblio as described above, sug-

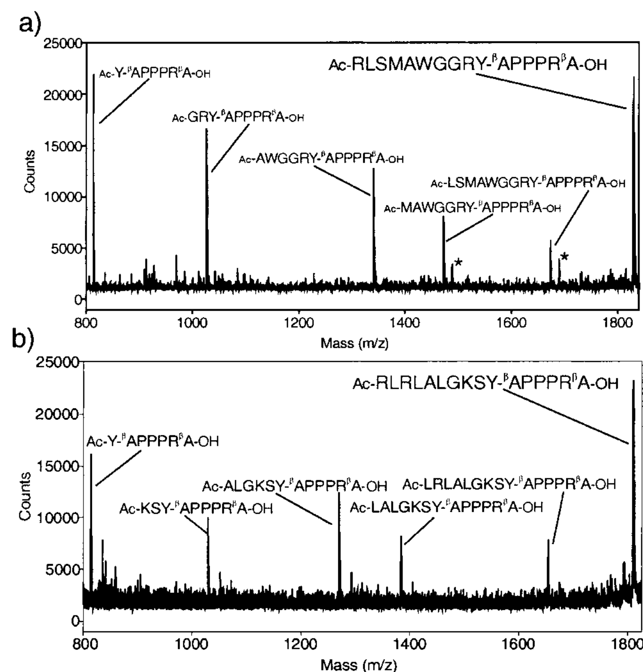


Figure 6. MALDI mass spectrum of the decapeptide sequences (a) Ac-RLSMAWGGRY and (b) Ac-RLRLALGKSY with their corresponding five coding sequences. The two peaks marked with an asterisk are methionine oxidation byproducts.

gesting five capping positions. Two beads have been picked. As demonstrated in Figure 6, all six anticipated peptides, the main peptide plus five capping peptides, have been observed for the two beads successfully decoded with mass spectrometry. Biblio unambiguously determined the decapeptide sequences as denoted in the figure. We have in the meantime applied the approach to another decapeptide library with 80 640 members and have found 37 biologically active peptides of which we could sequence 36.

Conclusions

The modified capping method we describe here is a development of the concept originally presented by Youngquist and Keough.[16] It allows a quick, straightforward, and unambiguous identification of individual sequences of a peptide library by well-established MALDI mass spectroscopy without need for special instrumentation. We showed on a penta-, hepta-, and decapeptide library that the decoding by mass spectroscopy was reliable. A total of 99.7% of the beads in a heptapeptide library sample were successfully analyzed. The program Biblio, which is accessible, suggests a capping pattern with minimal use of capping reagents and a minimized number of capping steps depending on the number of used capping reagents and the expected resolution of the mass spectrometry. Using mixtures of FMOC and acetylated amino acids instead of using acylation agents allowed reliable percentages of capped sequences and reduced the number of synthetic steps. The optimized ratio of the capping sequences and their maximal identity with the main peptide ensure that the results of biological tests are not flawed by introduction of the capped peptides. As revealed by HPLC, the capping peptides have an abundance of approximately 5% and the desired peptide makes up for the remainder. In our syntheses, byproducts occurred only with low yield and never disturbed the decoding. The applicability of our adapted capping method has been demonstrated for peptide libraries up to 432 000 members. The MALDI MS sequencing is cheaper than the Edman degradation and does not have to resort to additional molecules on the bead for encoding as used for GC−MS.

Experimental Section

Peptide Synthesis: Materials and Methods. Peptide and peptide libraries have been synthesized manually by standard FMOC chemistry[20] using *N*-hydroxybenzotriazole and *N,N′*-diisopropylcarbodiimide as activation agents. The optimum coupling time was found to be 4 h. TentaGel S HMB (hydroxymethylbenzoic acid) resin, bead diameter of 90 μm, was provided by Rapp Polymere GmbH, Tübingen. N-terminal acetylation was performed by using 10 equiv of acetic anhydride and 50 equiv of triethylamine (1 mL per 100 mg of peptidyl resin) for 2 h.[34] Where special amino acid derivatives were needed, they have been prepared according to literature protocols[30] or by simply modifying them attached to the solid phase; 2-chlorotrityl chloride resin (Calbiochem Novabiochem, Läufelfingen, Switzerland) was used. HPLC[33] was performed on a Merck-Hitachi L-6200 system with an L-3000 diode array detector and a Eurospher RP 18, 4 mm × 250 mm, 5μm, 100 Å C18 column. Solvent A was 0.1% TFA in water, and solvent B was 0.1% TFA in acetonitrile. The gradient was 0−100% B over 50 min. TFA was purchased from Acros Organics (Geel, Belgium). Solvents were purchased from Merck (Darmstadt, Germany) and were used without further purification. FMOC amino acid derivatives were purchased from Calbiochem Novabiochem (Läufelfingen, Switzerland). All other reagents were purchased from Sigma-Aldrich (Milwaukee, WI).

Synthesis of Peptide Libraries. TentaGel S HMB (hydroxymethylbenzoic acid) resin (bead diameter of 90 μm, capacity 0.26 mmol/g) was used as the solid phase. The pentamer and the heptamer libraries were calculated for the use of 400 mg of resin, and the decapeptide library was calculated for the use of 650 mg of resin. In all libraries, an excess of about one-third of the calculated resin mass was additionally used because of an expected loss of resin during the various splitting and mixing steps. First, a hexapeptide linker was attached to the resin; the resin was occupied with $^{\beta}$alanine as described.[31] Then the remaining five amino acids were attached, resulting in $^{\beta}$Ala-Pro-Pro-Pro-Arg-$^{\beta}$Ala as the linker construct. Library diversity was introduced by Furka's and Lam's "one bead one peptide" approach.[1,17] The bead quantity was divided in *m* portions, whereas *m* is the number of desired amino acids per library position. To introduce capped amino acids during library synthesis, a mixture of an FMOC-protected amino acid and an individually capped amino acid derivative, as predicted by Biblio, was used. The amount of capped amino acid was 5% with regard to the corresponding resin loading. To ensure manageable quantities of both capped and FMOC-protected derivatives, a 12-fold excess of amino acid was used (with 13.2 equiv of *N,N′*-diisopropylcarbodiimide and 18 equiv of *N*-hydroxybenzotriazole). The N-terminal protecting group (FMOC) was removed with 40% piperidine in DMF for 20 min, and this

was repeated three times, with thorough washing for subsequent performance of the next cycle. This process was repeated until the peptide sequence was finished. After intense resin washing, the library was acetylated as described above. Then the side chain protection groups were cleaved in 4 h by using 1 mL of the following mixture per 100 mg of resin: 10 mL of TFA, 750 mg of phenol, 500 μL of mercaptoacetic acid, 500 μL of thioanisole, and 500 μL of water. Thereafter, the resin was washed with 90% aqueous TFA and 10% triethylamine in DMF, several times with DMF, and finally twice with diethyl ether. The resin is stored at 4 °C.

MALDI Sample Preparation. Individual beads were picked randomly with a thin steel needle and placed in an Eppendorf cap containing a solution of 10% (v/v) triethylamine in water. The cleaving reaction was run at room temperature overnight. A 1 μL aliquot of the reaction mixture was pipetted onto the sample plate and dried in a cold airflow. Then 1 μL of a solution of 0.1 mol/L dihydroxybenzoic acid (DHB) in 30% aqueous acetonitrile containing 0.1% TFA was dropped onto this sample and dried in a cold airflow. In all cases, specular crystals have been obtained.

Mass Spectrometry. The MALDI mass spectra[29] were recorded on a PE Biosystems (Framingham, MA) Voyager DE Pro time-of-flight mass spectrometer. All data were acquired in the positive-ion reflector mode with delayed extraction. For each spectrum, 128 laser pulses have been accumulated. A mixture of Angiotensin I and III (10^{-5} M each) was used for external calibration. Since the identification of peptide sequence is based on specific capping resulting in large mass differences, one-time external calibration (for a whole series of experiments) was sufficient, resulting in a mass accuracy of $\pm$1 Da. Spectra were typically recorded at a mass resolution of about 2500 (fwhm). Again, tuning of the instrument for better resolution values was not necessary.

Acknowledgment. This work was supported by the Fonds der Chemischen Industrie, the DFG, and the MPG. We thank U. Dietrich for collaborating on the HIV library and J.B. Ferebee for help with the statistical analysis of the heptapeptide library.

References and Notes

(1) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354*, 82−84.
(2) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature* **1991**, *354*, 84−86.
(3) Jung, G., Ed. *Combinatorial Peptide and Nonpeptide Libraries*; VCH: Weinheim, New York, 1996.
(4) Hu, Y. J.; Wei, Y.; Zhou, Y.; Rajagopalan, P. T. R.; Pei, D. *Biochemistry* **1999**, *38*, 643−650.
(5) Yu, Z.; Chu, Y.-H. *Anal. Chem.* **1997**, *69*, 4515−4518.
(6) Tu, J. T.; Yu, Z.; Chu, Y.-H. *Clin. Chem.* **1998**, *44* (2), 232−238.
(7) Chu, Y.-H.; Yu, Z. *Bioorg. Med. Chem. Lett.* **1997**, *7* (1), 95−98.
(8) Bunin, B. A.; Plunkett, M. J.; Ellman, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4708−4712.
(9) St. Hilaire, P. M.; Lowary, T. L.; Meldal, M.; Bock, K. *J. Am. Chem. Soc.* **1998**, *120*, 13312−13320.
(10) Czarnik, A. W. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12738−12739.
(11) Geysen, M. H.; Meloen, R. N.; Barleling, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 3998−4002.
(12) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5381−5383.
(13) 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.
(14) Czarnik, A. W.; Nova, M. P. *Chem. Br.* **1997**, *33*, 39.
(15) Still, W. C. *Acc. Chem. Res.* **1996**, *29*, 155−163.
(16) Youngquist, R. S.; Fuentes, G. R.; Lacey, M. P.; Keough, T. *J. Am. Chem. Soc.* **1995**, *117*, 3900−3906.
(17) Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487−493.
(18) Luo, Z.; Zhang, Q.; Oderaotoshi, Y.; Curran, D. P. *Science* **2001**, *291*, 1766−1769.
(19) Geysen, H. M.; Wagner, C. D.; Bodnar, W. M.; Markworth, C. J.; et al. *Chem. Biol.* **1996**, *3*, 679−688.
(20) Fields, G. B.; Noble, R. L. *Int. J. Pept. Protein Res.* **199**, *35*, 161−214.
(21) Spengler, B.; Kirsch, D.; Kaufmann, R. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 105−108. Spengler, B. *J. Mass Spectrom.* **1997**, *32*, 1019−1036.
(22) Saidel, L. J.; Lieberman, H. *Arch. Biochem. Biophys.* **1958**, *76*, 401−409.
(23) Korber, B.; Kuiken, C.; Foley, B.; Hahn, B.; McCutchan, F. C.; Mellors, J.; Sodorski, J. *Human Retroviruses and AIDS*; Los Alamos National Laboratory: Los Alamos, NM, 1998.
(24) Goudsmit, J.; Debouck, C.; Meloen, R. H.; Smit, L.; Bakker, M.; Asher, D. M.; Wolff, A. V.; Gibbs, C. J.; Gajdusek, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 4478−4482.
(25) Javaherian, K.; Langlois, A. J.; McDanal, C.; Ross, K. L.; Eckler, L. I.; Jellis, C. L.; Profy, A. T.; Rusche, J. R.; Bolognesi, D. P.; Putney, S. D.; Matthews, T. J. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 6768−6772.
(26) Scott, C. F.; Silver, S.; Profy, A. T.; Putney, S. D.; Langlois, A.; Weinhold, K.; Robinson, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8597.
(27) Moore, J. P.; Nara, P. L. *AIDS* **1991**, *5*, 21.
(28) Zolla-Pazner, S.; Gorny, M. W.; Nyambi, P. N.; VanCott, T. C.; Nádas, A. *J. Virol.* **1999**, *73*, 4042−4051.
(29) Schreiber, M.; Wachsmuth, C.; Muller, H.; Odemuyiwa, S.; Schmitz, H. *J. Virol.* **1997**, *71*, 9198−9205.
(30) Otani, T. T.; Briley, M. R. *J. Pharm. Sci.* **1978**, *67*, 520−526.
(31) Blankemeyer-Menge, B.; Nimitz, M.; Frank, R. *Tetrahedron Lett.* **1990**, *31*, 1701−1704.
(32) Karas, M.; Glückmann, M.; Schäfer, J. *J. Mass Spectrom.* **2000**, *35*, 1−12.
(33) Unger, K. K., Weber, E., Eds. *Handbuch der HPLC* (Handbook of HPLC); GIT Verlag: Darmstadt, Germany, 1995.
(34) Hsieh, K.-H.; Demane, M. M.; Gurusidaiah, S. *Int. J. Pept. Protein Res.* **1996**, *48*, 292−298.

CC010057X