A General Method for Designing Combinatorial Peptide Libraries Decodable by Amino Acid Analysis

Jacob Kofoed and Jean-Louis Reymond*

Department of Chemistry and Biochemistry, University of Berne, Freiestrasse 3, CH-3012 Berne, Switzerland

Received July 17, 2007

Herein we describe an algorithm for designing combinatorial peptide libraries for split-and-mix synthesis on solid support that are decodable by amino acid analysis (AAA) of the beads. AAA is a standard service analysis available in most biochemical laboratories, and it allows one to control the quality of the peptide on each bead, an important feature that is missing from most library decoding protocols. In the algorithm, each AA is assigned to two variable positions in the sequence grouped in a "unique pair". This arrangement limits sequence design because both the number of unique pairs U (setting the maximum number of variable AA) and the maximum number S of different AA per variable position depend on the peptide length N ($U = N(N-1)/2$), $S = N-1$). The method is therefore only suitable for focused libraries. An application example is shown for the selection of peptides with *N*-terminal proline or hydroxyproline catalyzing an aldol reaction from a combinatorial library of 65536 octapeptides. A simple enumeration program is available to help design combinatorial libraries decodable by amino acid analysis. The method applies to linear and cyclic peptides, can be used for nonnatural building blocks, including β-amino acids, and should help to explore the vast chemistry of linear and cyclic peptide for catalysis and bioactivity.

Introduction

Split-and-mix peptide synthesis on solid support is one of the founding methods of combinatorial chemistry.[1–3] In each step, the support is split in M equal portions, and M different amino acids are coupled in M separate reaction vessels. The portions are then mixed together and split again for the next coupling round. The protocol realizes an exponential gain in synthetic efficiency over classical synthesis because only $M \times N$ operations are necessary to realize M^N sequences of a peptide of N amino acids in length, typically hundred thousands of compounds in tens of operations. Most remarkably, each bead of the polymeric support carries only a single sequence because it can only be present in one of the M reaction vessels in each coupling step, such that one-bead-one-compound (OBOC) libraries are produced.

The price to pay for this remarkable efficiency is the necessity to deduce the structure of the compound on any bead testing positive in an assay. In the case of peptides, the beads can be analyzed by Edman microsequencing;[2] however, this analysis is limited to linear α-peptides with a free *N*-terminus, and is quite expensive. Mass spectrometry can be used to analyze both linear (by MS ladder sequencing)[4] and cyclic peptide libraries.[5] Alternative methods for compound identification in combinatorial libraries[6] include library deconvolution by positional scanning libraries,[3] bead tagging using binary tags,[7] oligonucleotide tags[8] or radio frequency tags,[9] spatial arraying of the library on microarrays[10] or as SPOT synthesis,[11] isotope encoded[12] analysis, or image analysis of encoding optical[13] or shape markers.[14] All of these methods are quite complex in their implementation. Moreover, with the exception of Edman microsequencing, none of these methods provides a quantitative analysis of the material of the synthesis beads, such that the quality of the library cannot be ascertained, leading to a high risk of misinterpretation of screening data.

We recently investigated combinatorial libraries of peptide dendrimers obtained by using a branching diamino acid at every 3rd or 4th position in a peptide sequence.[15] We used amino acid analysis (AAA) of individual polymer beads as a simple tool to determine the sequence of the peptide dendrimer, because the relative amount of any amino acid (e.g., 1, 2, 4, or 8 copies) indicated its position in the corresponding branch of the dendrimer. Amino acid analysis (AAA) is a routine procedure in protein analytical laboratories, which consists in total acidic hydrolysis in strong acid and elevated temperature, derivatization of the released amino acids by phenyl isothiocyanate, and separation and quantification of each amino acid derivative by analytical HPLC. Using AAA for dendrimer library decoding proved robust and reliable in a variety of settings, including the search for dendritic enzyme models[16,17] and artificial glycoproteins.[18] The method was particularly attractive because it did not require any encoding operation and could be carried out at low-cost on many beads. Most importantly, the method provided a quality control of the material on the bead, with a typical fall-out rate of 20% due to empty beads or incomplete sequences, thus avoiding artifacts. However, the method for dendrimers was not utilizable for linear or cyclic peptides where all positions have the same multiplicity.

* Corresponding author e-mail:jean-louis.reymond@ioc.unibe.ch.

10.1021/cc7001155 CCC: $37.00
Published on Web 10/09/2007

Table 1. Comparison of a "Unique Pair" Library of Tripeptides versus a Full Combinatorial Library

combinatorial library			unique pair library		
X^3	X^2	X^1	X^3	X^2	X^1
A	A	A	A	A	
B	B	B	B		B
C	C	C		C	C

amino acid analysis	parent sequence(s)	parent sequences
ABC	ABC, ACB, BAC, BCA, CAB, CBA	ACB, BAC
A_2B	AAB, ABA, BBA	AAB
A_2C	AAC, ACA, CAA	AAC
B_2A	BBA, BAB, ABB	BAB
B_2C	BBC, BCB, CBB	BCB
C_2A	CCA, CAC, ACC	ACC
C_2B	CCB, CBC, BCC	BCC
A_3, B_3, C_3	AAA, BBB or CCC	
total	27 sequences, 10 profiles	8 sequences, 7 profiles
degeneracy	$D = 2.7$	$D = 1.14$

Table 2. Comparison of Different Unique Pair Libraries

no.	N^a	U^b	split^c	library size (T)	AAA^d	AAA with 1 sequence	AAA with 2 sequences	AAA with 4 sequences	D^e
1	2	1	1	1	1	1	0	0	1.00
2	3	3	2	8	7	6	1	0	1.14
3	4	6	3	81	66	51	15	0	1.23
4	5	10	4	1024	792	560	232	0	1.30
5	6	15	5	15625	11590	7575	4005	10	1.35
6	7	21	6	279936	200469	122052	77892	525	1.40
7	8	28	7	5,764801	4,004490	2,285353	1,698550	20587	1.44
8	8	16^b	4	65536	54730	43960	10752	72	1.20

[a] N is the length of the peptide. [b] U is the number of unique pairs used, which is $U = N(N-1)/2$, except for the last table entry no 8. [c] The split size is calculated as: $s = N - 1$. [d] Number of different amino acid analyses possible for the library. [e] The degeneracy (D) is defined as the total library size divided by the number of hydrolysis profiles ($D = T / P$).

Herein we report a library design algorithm that allows one to use AAA to decode libraries of linear and cyclic peptides. We show that by placing each amino acid at only two positions in the sequence following the principle of "unique pairs", one obtains libraries in which each possible AAA corresponds to only one, two, or rarely four possible peptides in the library. Although "unique pair" libraries are focused and much smaller than complete libraries in which every amino acid is used at each variable position, they reach a useful size in the perspective of split-and-mix combinatorial synthesis. Furthermore, these libraries take full advantage of AAA in terms of simplificty, low price, and bead quality control, which render the combinatorial experiment particularly straightforward. The method is illustrated by the preparation of a 65536 members "unique pairs" octapeptide library, and the functional selection of peptides catalyzing an aldol reaction. A program is presented to help design AAA-decodable combinatorial peptide libraries by complete enumeration.

Results and Discussion

Library Design by Unique Pairs. In a linear peptide all positions have the same multiplicity of one, and the relative amount of an amino acid does not carry any positional information. Although libraries of tetrapeptides, up to 4096 compounds, have been analyzed by AAA, no systematic design concept was presented that would allow its generalization.[19] In testing various design combinations, we have found a general principle to produce combinatorial peptide libraries for AAA decoding which consists in positioning amino acids along the sequence by "unique pairs", as described below.

The principle of "unique pairs" is illustrated for the case of a tripeptide library with three different amino acid building blocks A, B, and C (Table 1). The "full" combinatorial library comprises 27 different peptides. However there are only 10 possible amino acid analysis results, such that the analysis allows unique sequence assignment in only three cases corresponding to the homotripeptides AAA, BBB, and CCC. Six analyses give three possible solutions, e.g. AAB, and one analysis gives six possible sequences, e.g. ABC. In this case, the degeneracy of the information in the amino acid analysis is $D = 27/10 = 2.7$.

One can now restrict the use of each amino acid to two positions only, such that a different pair of positions is used for each of the three building blocks A, B, and C. This results in a smaller library of eight peptides; however, there are now only seven possible amino acid analyses, corresponding to a much lower degeneracy of $8/7 = 1.14$. In this library, only a single analysis, ABC, corresponds to two possible peptides: ACB and BAC. All other peptides are uniquely identified from their amino acid analysis.

We define a "unique pair" as the association of two different positions in a peptide. The number of unique pairs in a peptide of length N is given by the triangular number $U(N) = N(N - 1)/2$.[20] We can form combinatorial peptide libraries by assigning M different building blocks to M different unique pairs (Table 2). This design keeps the degeneracy of the amino acid analysis low, but delivers a sizable number of compounds. For example, using all 15

Table 3. Design of a "Unique Pair" Octapeptide Library

pair no.[a]	AA[b]	X^8	X^7	X^6	X^5	X^4	X^3	X^2	X^1
1		0	0						
2		0		0					
3		0			0				
4	Pro	1				1			
5	Hyp	1					1		
6	His	1						1	
7	Thr	1							1
8			0	0					
9			0		0				
10	Phe		1			1			
11	Met		1				1		
12	Val		1					1	
13	Asp		1						1
14				0	0				
15	Glu			1		1			
16	Ile			1			1		
17	Tyr			1				1	
18	Gly			1					1
19	Arg				1	1			
20	Ser				1		1		
21	Ala				1			1	
22	Leu				1				1
23						0	0		
24						0		0	
25						0			0
26							0	0	
27							0		0
28								0	0
	SUM[c]	4	4	4	4	4	4	4	4

[a] All possible unique pairs are listed by identification of the positions as "0" ("not used here") or "1" (assigned to the indicated amino acid). [b] The indicated amino acids were used at the indicated positions in the sequence. [c] The sum of used amino acids per position gives the split size at each position. The split size can vary depending on the choice of unique pairs. The library size is given by the product of all sums: $4^8 = 65536$ members.

unique pairs with 15 different amino acids in a hexapeptide library $X^6X^5X^4X^3X^2X^1$, and five different amino acids at each position, results in $5^6 = 15625$ different peptides, which can be prepared in $5 \times 6 = 30$ coupling operations. In this combinatorial library there are 11590 possible amino acid analyses, corresponding to either one (7575 profiles, 48.5% of the beads), two (4005 profiles, 51.3% of the beads), or four peptide sequences (10 profiles, 0.26% of the beads). Such a small degeneracy ($D = 1.35$) is an acceptable uncertainty. By comparison, the "complete" combinatorial library corresponding to the hexapeptide discussed above using all 15 variable amino acids at each position would have a size of $15^6 = 11390625$ different peptides. There are 38760 possible amino acid analyses in this library, with an average of around 300 possible sequences for each analysis ($D = 294$). This degeneracy would be too high to provide useful sequencing information.

Application To an Octapeptide Library for Aldol Catalysis. The total number of sequences available in "unique pairs" libraries matches well with what is realistically realizable in a typical split-and-mix synthesis, where the library size is limited by the number of polymer beads, typically 1 million beads per gram, with a requirement for at least 10-fold coverage, and the use of typically 50 mg resin (50000 beads) per screening experiment. The "unique pair" design was tested in such a setting for an octapeptide library $X^8X^7X^6X^5X^4X^3X^2X^1$ using 16 of the 28 possible unique pairs (Table 3). The unique pairs were chosen such as to obtain an even split size $s = 4$ at each position to form a

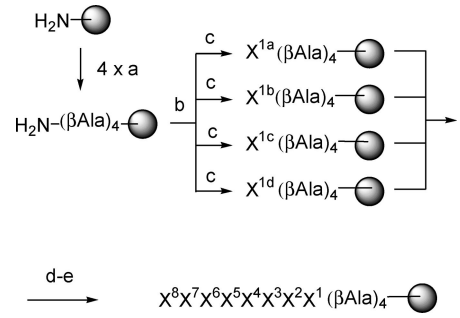


Figure 1. Split-and-mix synthesis of library **L** on tentagel resin (0.63 mmol g^{-1}). Conditions: (a) 2.5 eq. Fmoc-βAla-OH, 2.5 eq. DIPCDI/HOBt, DMF, 25 °C, 1 h, then 20% piperidine in DMF, 25 °C, 30 min.; (b) split in four portions; (c) 2.5 eq. Fmoc-Xi-OH, 2.5 eq. DIPCDI/HOBt, DMF, 25 °C, 1 h, then 20% piperidine in DMF, 25 °C, 30 min.; (d) mix four portions as suspension, then split in four equal portions; (e) repeat b–d seven times with X^{2a-d}–X^{8a-d}, after last cycle TFA/TIS/H_2O (95:2.5:2.5, *v*/*v*/*v*), 25 °C, 4 h. DIPCDI: diisopropylcarbodiimide; HOBt: 1-hydroxybenzotriazole; Fmoc: 9-fluorenylmethoxycarbonyl; DMF: *N*,*N*-dimethylforamide; TFA: trifluoroacetic acid; TIS: triisopropylsilane.

library of 65536 different peptides. The partial use of unique pairs results in lower degeneracy D compared to libraries using all possible unique pairs (Table 2).

The library was used to probe aqueous aldol catalysis by peptides with free *N*-termini, which has been reported for both linear and dendritic peptides.[21,17] Proline and hydroxy proline, which are known to provide active peptide catalysts, were placed at the critical *N*-terminal position X^8, together with threonine and histidine which display hydrogen-bonding groups that might assist aldolization catalysis.[22] The other amino acids were distributed along positions X^7–X^1, including glutamate and aspartate occupying position X^7, X^6, X^4, and X^1 to test various placements of catalytic carboxyl groups in the peptide.

The library was synthesized by Fmoc solid phase split-and-mix peptide synthesis on a 1.0 g batch of TentaGel HL (Rapp Polymere) resin (0.63 mmol/g, ca. 1.0 million beads per gram) ensuring good library coverage (Figure 1). The synthesis was initiated by a $(\beta\text{-Ala})_4$ tetrapeptide spacer. β-alanine served as an internal control for amino acid analysis of the beads. After removal of the last Fmoc protecting group, the side-chain protecting groups were removed by acidic treatment. The quality of the library was checked by amino acid analysis of 20 randomly picked beads. Seventeen beads (85%) gave AAA from which sequences could be assigned, with the expected distribution of amino acids at each position. The remaining three beads (15%) were empty or contained truncated sequences, indicating low loading or a failed synthesis.

The beads were screened for binding to the dye-labeled diketone **1** in aqueous buffer pH 8.5 with DMSO cosolvent to ensure good solubility of the diketone and good swelling of the beads (Figure 2). 1,3-Diketones such as **1** react with aldolase reactive amino groups to form an enaminone, and have been used as reactive haptens to select for aldolase catalytic antibodies.[23] We have used diketone **1** previously to identify aldolase in combinatorial libraries of peptide dendrimers.[16] While there was no staining of the library with

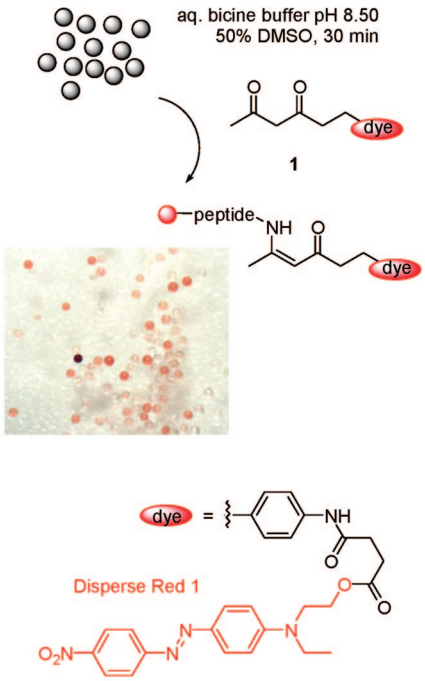


Figure 2. On-bead assay of enaminone formation. Screening conditions: Beads shaken in dimethylsulfoxide-bicine buffer pH 8.5 (1,1, *v*/*v*) with 50 μM diketone **1**. Hits are red to deep-red.

Disperse Red 1 alone (10 mM), beads were selectively stained in the presence of **1** (50 μM). Darkly red colored beads were picked and subjected to AAA. From 10 picked beads, nine analyses returned a readable sequence (Table 4).

The hit sequences in the octapeptide library showed a consensus for the Pro-Phe or Hyp-Phe as *N*-terminal dyad X^8X^7, and all but one sequence had at least one carboxylate side chain (Asp or Glu). Five peptides were synthesized corresponding to three unique and one degenerate amino acid analysis. The aldolase activity of these peptides was confirmed by investigating their activity for the aldol reaction of nitrobenzaldehyde **3** with cyclohexanone **2a**, a reference reaction for aqueous catalysis by pyrrolidine-type catalysts (Table 5).[24] While L-proline was essentially inactive as reported elsewhere,[24] all five peptides tested showed good activity for the reaction. Peptide **8a** was significantly more active than peptide **8b** representing the second possible sequence assigned by decoding, which suggests that **8a** rather than **8b** was present on the selected bead. The most active catalyst was peptide **5**, which could be used with a lower catalyst loading of 1 mol %. The highest stereoselectivity in the series was achieved with peptide **8a** (84% *de* and 70% *ee* in favor of the *anti-R* aldol **4a**).

Limitations of the Unique Pair Design. The design by unique pairs limits sequence coverage since the number of amino acids at each position and their relative placement is constrained. The limits of the unique pair design are fixed by the relationship between the peptide length N and the number of unique pairs U, which sets the maximum number of variable AA to $U = N(N-1)/2$, and the maximum number of different AA per variable position to $S = N - 1$. The method therefore does not replace a complete peptide library, and is most useful for focused libraries with a certain level of design, as shown above for the aldol catalysis library featuring *N*-terminal proline and hydroxyprolines.

An interesting option to extend the concept consists in using certain AA only once, which is possible without any placement constraints for these AA. This option limits the attainable library size per AA, but is interesting if more AA are analyzable by the AAA than the number of available unique pairs U for the given peptide length. For example 15 AA (20 AA) could serve for a pentapeptide (hexapeptide) library using 10 AA (15 AA) for the unique pairs and 5 AA to be used once with even distribution, resulting in a library of size $5^5 = 3125$ ($5 \cdot 6^5 = 38880$) members, which is 3 times (2.5 times) larger than the parent unique pair 1024-member pentapeptide library (15625-member hexapeptide library), with the same level of degeneracy. Concentrating the five additional AA not used for unique pairs to a single position results in library sizes of $9.4^4 = 2304$ ($10.5^5 = 31250$) pentapeptides (hexapeptides), doubling library size compared to the reference unique pair libraries.

A Enumeration Program for Designing and Using AAA-Decodable Combinatorial Libraries. As discussed above the use of "unique pairs" for library design leads to a remarkably low degeneracy in the AAA information for sequence assignment, but imposes limitations in the diversity that can be surveyed in any given library since each amino acid is used only twice in the sequence. At the other extreme a "complete" library using all variable amino acids at each position results in an impractically high degeneracy in AAA data. A large number of designs are conceivable where the use of each amino acid is adjusted to a selected number of positions, with a consequently variable level of degeneracy in the AAA information. The design options include the possibility to select different sets of unique pairs in cases where the number of variable amino acids is lower that the number of possible unique pairs, as in the application example above based on the selection of 16 of the 28 possible unique pairs (Table 3), or to use some of the AA only once as discussed above. Overall, the method will be limited by the number of building blocks that can be identified and quantified during the amino acid analysis, which depends on the resolving power of the HPLC analysis.

To facilitate library design for AAA-decoding, we have written an enumeration program called TAGSFREE. This program takes a library design table as input, consisting in the positional assignment of each variable amino acid along a given sequence (e.g., Table 3). The output of the program is a text file with each line listing an AAA formulated as a string of numbers giving the relative abundance of each amino acid in the order chosen for the input, followed by the peptide sequence(s) corresponding to the AAA. The text file can be used as a help for decoding by matching any measured AAA data to one or more peptide sequences. The program can handle "unique pair" libraries or any other design, and is freely available.

Conclusion

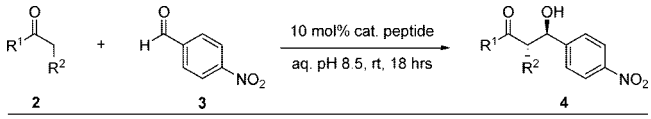
The design principle by "unique pairs" presented here offers unprecedented simplicity due to the absence of encoding operations and the ease and reliability of amino

Table 4. Peptide Sequences Identified by AAA of Red-Colored Beads after Enaminone Formation Assay with **1**

bead no.	AAA[a]	cpd.	X^8	X^7	X^6	X^5	X^4	X^3	X^2	X^1
1	1100201100100001	**5**	Pro	Phe	Tyr	Leu	Phe	Hyp	Val	Asp
2	0020110020000011	**6**	His	Phe	Glu	Ala	Glu	Met	His	Leu
3	2000101011010001	**7**	Pro	Phe	Glu	Leu	Pro	Ile	Val	Gly
4	1100100011110001	**8a**	Hyp	Phe	Glu	Leu	Pro	Ile	Tyr	Gly
		8b	Pro	Phe	Ile	Leu	Glu	Hyp	Tyr	Gly
5	2000100101000111		Pro	Phe	Ile	Leu	Pro	Ser	Ala	Asp
6	1111200001000010		Pro	Phe	Ile	Ala	Phe	Hyp	His	Thr
7	0110100020100011		His	Phe	Glu	Ala	Glu	Hyp	Tyr	Leu
8	0100100110111100		Hyp	Phe	Gly	Arg	Glu	Ser	Tyr	Asp
			Hyp	Asp	Glu	Arg	Phe	Ser	Tyr	Gly
9	1011100010010110		Pro	Phe	Gly	Ala	Glu	Ser	His	Thr

[a] The AAA indicates the relative amounts of amino acids deduced from the HPLC analysis in the order Pro-Hyp-His-Thr-Phe-Met-Val-Asp-Glu-Ile-Tyr-Gly Arg-Ser-Ala-Leu. See the Supporting Information for original HPLC traces and exact integration values.

Table 5. Aldol Reaction Catalyzed by Octapeptides Selected from a "Unique Pair" Combinatorial Library[a]



2 3 4

entry	catalyst	R^1	R^2	time (hours)	yield (%)[b]	isomeric ratio[c]
1		–$(CH_2)_4$–		18	7	na
2	L-proline	–$(CH_2)_4$–		18	9	na
3	PFYLFXVD **5**	–$(CH_2)_4$–		18	>95	69:11:13:7
4	1 mol % **5**	–$(CH_2)_4$–		36	>95	64:11:16:9
5	PFYLFXVD **5**	Me	H	18	94	65:35[e]
6	HFEAEMHL **6**	–$(CH_2)_4$–		18	23	47:19:20:13
7	PFELPIVG **7**	–$(CH_2)_4$–		18	66	74:12:10:4
8	XFELPIYG **8a**	–$(CH_2)_4$–		18	63	78:14:6:2
9	PFILGXYG **8b**	–$(CH_2)_4$–		18	28	52:15:22:11

[a] Conditions: 100 mM aldehyde **3**, 10 mM peptide in cyclohexanone **2a**-aq. bicine buffer pH 8.5 (1:1, *v*/*v*), 18 h, room temperature. [b] Conversion and diastereomeric excess were measured by RP-HPLC on a Waters Atlantis dC18 column (100 mm × 4.6 mm) at 268 nm. [c] Isomeric ratio [(*S*,2*R*)-**4a**: (*R*,2*S*)-**4a**: *syn*-**4a** (major): *syn*-**4a** (minor)] of the individual diastereoisomers of **4a** was measured by chiral phase HPLC on a ChiralPak OD-H column (300 mm × 4.6 mm) at 268 nm and is reported in the following order. The *anti*-**4a** and *syn*-**4a** aldol products were separated by semipreparative RP-HPLC followed by injection of the lyophilized products on chiral HPLC. By comparison with the literature the major product in all case was assigned to (*S*)-2-((*R*)-hydroxy(4-nitrophenyl)methyl)cyclohexanone. [d] 100 mM aldehyde **3**, 10 mM peptide in acetone **2b**-aq. bicine buffer pH 8.5 (1:1, *v*/*v*). [e] 100 mM aldehyde **3**, 10 mM peptide in acetone **2b**-octanol-aq. bicine buffer pH 8.5 (1:1:2, *v*/*v*). The major enantiomer of **4b** obtained is (*R*)-4-hydroxy-4-(4-nitrophenyl)butan-2-one as measured by chiral phase HPLC on a ChiralPak AS column (300 mm 300 mm × 4.6 mm) at 254 nm.

acid analysis. Although very large libraries are possible, libraries of less than 100000 members such as the octapeptide library above are best suited for solid-phase synthesis experiments due to the limited number of beads involved in a typical synthesis (ca. 1 million beads/gram).[25] The amino acid analysis used for decoding provides a useful quality control helping to avoid false positives, and is amenable to high-throughput bead decoding due to its low cost.

Despite of the limitations imposed by the unique pair design as discussed above, the method has advantages in terms of flexibility. For instance, the design is independent of topology and can also be used for libraries of cyclic peptides. One can also incorporate unusual amino acids, such as β-alanine and hydroxyproline as used here, provided that these can be identified (by a specific retention time) and quantified in the amino acid analysis. Furthermore it is possible to use D-amino acids at specific positions. The method should allow many nonspecialized laboratories to prepare and handle combinatorial split-and-mix peptide libraries with ease, and might be extendable to any combinatorial assembly where building blocks can be analytically quantified. The method should help explore the vast chemistry of linear and cyclic peptide for catalysis and bioactivity.

Experimental Section

General. Reagents were purchased in the highest quality available from Fluka, Sigma, Bachem, Novabiochem, NeoMPS, or Aldrich. All solvents used in reactions were bought in p.a. quality or distilled and dried prior to use. Solvents for extractions were distilled from technical quality. Sensitive reactions were carried out under nitrogen or argon, the glassware being heated under high vacuum. Preparative RP-HPLC (flow rate 100 mL min^{-1}) was performed with a Waters Delta Prep 4000 system with a Waters Prepak Cartridge (500 g) as column and Waters 486 Tunable Absorbance Detector. Semipreparative RP-HPLC (flow rate 4 mL min^{-1}) was performed with a Waters 510 pump operated with a Waters Automated Gradient Controller and Jasco UV-2075 Plus Detector on a Vydac 218 TP (1.0 cm × 25 cm) column. Analytical RP-HPLC (flow rate 4 mL min^{-1}) was performed on Waters 600E systems with Waters Atlantis (4.6 mm × 100 mm, dC18, 5 mm) column, UV detection with Waters 996 photodiode array detector). Data recording and processing was done with Waters Empower2 software. Eluents for all systems were (A) water and 0.1% TFA; (D) acetonitrile, water, and TFA (3/2/0.1%). Analytes were quantified using external standards. MS and Hrms analyses were provided by the mass spectrometry service of the Department of Chemistry and Biochemistry, University of Berne. ^1H and ^{13}C NMR spectra were recorded on Bruker AC 300 (300 MHz) and DRX 500 or Avance 500 (500 MHz) instruments. Chemical shifts (*d*) are given in ppm referring to solvent residual peak, coupling constants (*J*) in Hertz (Hz). Solid phase peptide chemistry was performed in polypropylene syringes fitted with a polyethylene frit and a teflon stopcock and stopper.

Fmoc-Synthesis. The resin was washed and swollen inside the reactor with DCM (2 × 5 mL) and DMF (1 × 5 mL). The appropriate resin (library: TentaGel HL (0.65 mmol g^{-1}), single sequences: resin was acylated with 2.5 equivalents of *N*-Fmoc amino acid in the presence of 2.5 equivalents of DIPCDI and 2.5 equivalents of HOBt in DMF. After 1 h, the resin was washed (3× each) with DMF, DCM, and

MeOH, and controlled with the TNBS (trinitrobenzenesulfonic acid) or chloranil test followed by acetylation. The Fmoc protecting group was removed with 5 mL of a solution of DMF-piperidine (4:1, *v*/*v*) for 10 min. After filtration, the procedure was repeated and then washed (3× each) with DMF, DCM, and MeOH. For acetylation in case of incomplete coupling, the resin was treated with a solution of acetic acid anhydride-DCM (1:1, *v*/*v*) for 10 min. After filtration, the procedure was repeated and then washed (3× each) with DMF, DCM, and MeOH. The final cleavage was carried out using TFA-H_2O-TIS (triisopropylsilane) as a (95:2.5:2.5, *v*/*v*) solution for 6 h.

Resin Mixing and Splitting. The resin was suspended in DMF-DCM (2:1, *v*/*v*), and mixed via nitrogen bubbling for 15 min, and then distributed in four equal portions.

On-Bead Assay with Diketone 1. Fifty mg of library resin was swollen overnight in 20 mM DMSO-bicine buffer pH 8.5 (1:1, *v*/*v*). The swelling mixture was removed by filtration and 1 mL of 50 mM solution of **1** in 20 mM DMSO-bicine buffer pH 8.5 (1:1, *v*/*v*) was added. The resin was shaken for 30 min and washed extensively with bicine buffer, DMSO, DMF, MeOH, DCM, MeOH, DMF and finally with bicine buffer again (3× each). A suspension of the resin in DMF was transferred to a silica gel plate and the beads were observed under a microscope. Single red colored beads were transferred via a syringe needle to amino acid analysis vials.

Bead Analysis. After the on-bead assays single peptide-containing resin beads were transferred via a sterile syringe needle to amino acid analysis vial and hydrolyzed with aqueous HCl (6 M) at 110 °C for 22 h, and their amino acid composition was determined quantitatively by HPLC after derivatization with phenyl isothiocyanate. The hydrolysis profile is calculated as the relative amounts of amino acids.

Sequence Assignment from Amino Acid Analysis. The TAGSFREE program is available from the author's Web site (http://www.dcb.unibe.ch/groups/reymond/). The user provides an input file <name>.csv describing the library as follows: Each line lists a different amino acid (in three letter code) followed by the usage vector as a series of 1s and 0s for each position up to the planned number of variable positions all separated by semicolons. The program handles "unique pair" designs or any other design. An excel file listing unique pairs is provided as a help for library design. Running the TAGSFREE program on the input file <name>.csv generates an output file <name>LIB.txt in which all possible AAA and the corresponding sequences are listed in the order of AAA and an output file <name>STAT.txt which contains the analysis of the generated library. The observed composition of the bead is written as an AAA in form of a series of integers (0, 1, 2) indicating the relative amounts of each building block in the order used in the input file. Decoding is carried out by typing an AAA in the search window of a text editor capable of handling the large library file. The search function automatically goes to the line(s) containing the profile, which shows the corresponding sequence(s).

Peptide Synthesis. The peptides were resynthesized on Rink amide NovaGel™ (high-loading PEG-PS resin 0.63 mmol g^{-1}) using the same coupling conditions as described for the linear library synthesis. After TFA cleavage, the peptides were precipitated with methyl *tert*-butyl ether and dissolved in water–acetonitrile mixture.

2-[Hydroxy-(4-nitrophenyl)methyl]cyclohexanone 4a.[26] A mixture of 50 μL of a 200 mM solution of 4-nitrobenzaldehyde in cyclohexanone and 50 μL a 20 mM peptide catalyst solution in 100 mM aqueous bicine buffer pH 8.5 was shaken in an Eppendorf PP-tube for 18 h. Final concentrations: 100 mM aldehyde and 10 mM catalyst. Ten μL aliquots of the reaction mixture were diluted with 100 μL A and injected on analytical RP-HPLC. The aldol product was purified by diluting the reaction mixture with water and injecting on semipreparative RP-HPLC separating the diastereomers. The pure aldol products were injected on chiral phase HPLC separating the enantiomers.

anti-**4**: IR (neat) = 3504, 1688, 1604, 1531, 1508, 1342, 1131, 1044, 855, 842, 800, 702 cm^{-1}. ^1H NMR (300 MHz, $CDCl_3$) δ = 8.22 (d, J = 8.8 Hz, 2H), 7.51 (d, J = 8.7 Hz, 2H), 4.95 (d, J = 8.3 Hz, 1H), 4.09 (br, 1H), 2.32–2.71 (m, 3H), 2.12–2.29 (m, 1H), 1.31–1.87 (m, 5H) ppm. ^{13}C NMR (300 MHz, $CDCl_3$) δ = 214.7, 145.6, 145.2, 127.8, 123.5, 74.0, 57.2, 42.7, 30.8, 27.6, 24.7 ppm. ESI MS(+): calcd for $C_{13}H_{15}NNaO_4$ 272.3, found 272.5. Anal. HPLC (UV 268 nm; gradient 90% A, 10% D to 10% A, 90% D in 15 min): t_R = 11.3 min. Anal. chiral HPLC [Daicel Chiralpak OD-H, *i*-PrOH-hexane (10:90, v/v), UV 268 nm, flow rate 1.5 mL min^{-1}]: t_R = 16.1 min (major); t_R = 22.5 min (minor).

syn-**4**: IR (neat) = 3491, 1693, 1602, 1508, 1447, 1343, 1186, 1131, 1091, 852, 796, 703 cm^{-1}. ^1H NMR (300 MHz, $CDCl_3$) δ = 8.22 (d, J = 8.7 Hz, 2H), 7.49 (d, J = 8.9 Hz, 2H), 5.45 (d, J = 2.1 Hz, 1H), 3.05 (br, 1H), 2.51{2.61 (m, 1H), 2.29–2.49 (m, 2H), 1.95–2.11 (m, 1H), 1.41–1.82 (m, 5H) ppm. ^{13}C NMR (300 MHz, $CDCl_3$) δ = 214.0, 144.6, 143.8, 126.6, 123.5, 70.2, 56.8, 42.6, 27.8, 25.9, 24.8 ppm. ESI MS(+): calcd for $C_{13}H_{15}NNaO_4$ 272.3; found 272.4. Anal. HPLC (268 nm; gradient 90% A, 10% D to 10% A, 90% D in 15 min): t_R = 11.7 min. Anal. chiral HPLC [Daicel Chiralpak OD-H, *i*-PrOH-hexane (10:90, v/v), UV 268 nm, flow rate 1.5 mL min^{-1}]: t_R = 13.5 min (minor); t_R = 15.1 min (major).

H-Pro-Phe-Tyr-Leu-Phe-Hyp-Val-Asp-NH_2 5. From NovaGel (160 mg, 0.63 mmol g^{-1}), **5** was obtained as colorless foamy solid after preparative HPLC purification (71 mg, 62%, as TFA-salt); anal. RP-HPLC (90% A, 10% D to 10% A, 90% D in 15 min): t_R = 9.91 min; ESI MS(+): calcd for $C_{52}H_{69}N_9O_{12}$: 1011.51, found: 1011.50.

H-His-Phe-Glu-Ala-Glu-Met-His-Leu-NH_2 6. From NovaGel (160 mg, 0.63 mmol g^{-1}), **6** was obtained as colorless foamy solid after preparative HPLC purification (32 mg, 35%, as TFA-salt); anal. RP-HPLC (90% A, 10% D to 10% A, 90% D in 15 min): t_R = 5.68 min; ESI MS(+): calcd for $C_{45}H_{65}N_{13}O_{12}S$: 1011.46, found: 1011.75.

H-Pro-Phe-Glu-Leu-Pro-Ile-Val-Gly NH_2 7. From NovaGel (160 mg, 0.63 mmol g^{-1}), **7** was obtained as colorless foamy solid after preparative HPLC purification (55 mg, 55%, as TFA-salt); anal. RP-HPLC (90% A, 10% D to 10% A, 90% D in 15 min): t_R = 9.91 min; ESI MS(+): calcd for $C_{43}H_{67}N_9O_{10}$: 869.50, found: 869.50.

H-Hyp-Phe-Glu-Leu-Pro-Ile-Tyr-Gly NH_2 8a. From NovaGel (160 mg, 0.63 mmol g^{-1}), **8a** was obtained as colorless foamy solid after preparative HPLC purification (66 mg, 61%, as TFA-salt); anal. RP-HPLC (90% A, 10% D to 10% A, 90% D in 15 min): $t_R = 9.43$ min; ESI MS(+): calcd for $C_{47}H_{67}N_9O_{10}$: 949.49, found: 949.38.

H-Pro-Phe-Ile-Leu-Glu-Hyp-Tyr-Gly NH_2 8b. From NovaGel (160 mg, 0.63 mmol g^{-1}), **8b** was obtained as colorless foamy solid after preparative HPLC purification (60 mg, 56%, as TFA-salt); anal. RP-HPLC (90% A, 10% D to 10% A, 90% D in 15 min): $t_R = 8.30$ min; ESI MS(+): calcd for $C_{47}H_{67}N_9O_{10}$: 949.49, found: 949.50.

Acknowledgment. This work was supported by the University of Berne, the Swiss National Science Foundation, and the Marie Curie Training Network IBAAC.

Supporting Information Available. Compound characterization (MS) and amino acid analyses of resin beads from on bead assay. This information is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Furka, A.; Sebestyén, F.; Asgedom, M.; Dibó, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487–493.

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

(3) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature* **1991**, *354*, 84–86.

(4) (a) Chait, B. T.; Wang, R.; Beavis, R. C.; Kent, S. B. H. *Science* **1993**, *262*, 89–92. (b) Youngquist, R. S.; Fuentes, G. R.; Lacey, M. P.; Keough, T. *J. Am. Chem. Soc.* **1995**, *117*, 3900–3906. (c) Wang, X.; Zhang, J.; Song, A.; Lebrialla, C. B.; Lam, K. S. *J. Am. Chem. Soc.* **2004**, *126*, 5740–5749.

(5) Redman, J. E.; Wilcoxen, K. M.; Ghadiri, M. R. *J. Comb. Chem.* **2003**, *5*, 33–40.

(6) (a) For reviews of encoding techniques in the period 1997–2003 see Czarnik, A. W. *Curr. Opin. Chem. Biol.* **1997**, *1*, 60–66. (b) Czarnik, A. W. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 12738–12739. (c) Lam, K. S.; Lebl, M.; Krchnak, V. *Chem. Rev.* **1997**, *97*, 411–448. (d) Barnes, C.; Balasubramanian, S. *Curr. Opin. Chem. Biol.* **2000**, *4*, 346–350. (e) Guillier, F.; Orain, D.; Bradley, M. *Chem. Rev.* **2000**, *100*, 2091–2158. (f) Affleck, R. L. *Curr. Opin. Chem. Biol.* **2001**, *5*, 257–263. (g) Ede, N. J.; Wu, Z. *Curr. Opin. Chem. Biol.* **2003**, *7*, 374–379.

(7) (a) Ohlmeyer, M.; Swanson, R.; Dillard, L.; Reader, J.; Asouline, G.; Kobayashi, R.; Wigler, M.; Still, W. C. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 10922–10926. (b) Nestler, H. P.; Bartlett, P. A.; Still, W. C. *J. Org. Chem.* **1994**, *59*, 4723–4724. (c) Still, W. C. *Acc. Chem. Res.* **1996**, *29*, 155–163.

(8) Nielsen, J.; Brenner, S.; Janda, K. D. *J. Am. Chem. Soc.* **1993**, *115*, 9812–9813.

(9) Herpin, T. F.; Van Kirk, K. G.; Salvino, J. M.; Yu, S. T.; Labaudiniére, R. F *J. Comb. Chem.* **2000**, *2*, 513–521.

(10) (a) Fodor, S. P. A.; Read, J. L.; Pirrung, M. C.; Stryer, L.; Lu, A. T.; Solas, D. *Science* **1991**, *251*, 767–773. (b) Macbeath, G.; Koehler, A. N.; Schreiber, S. L. *J. Am. Chem. Soc.* **1999**, *121*, 7967–7968.

(11) Frank, R. *Tetrahedron* **1992**, *48*, 9217–9232.

(12) Geysen, H. M.; Wagner, C. D.; Bodnar, W. M.; Markworth, C. J.; Parke, G. J.; Schoenen, F. J.; Wanger, D. S.; Kinder, D. S. *Chem. Biol.* **1996**, *3*, 679–688.

(13) Nicolau, K. C.; Pfefferkorn, J. A.; Mithcell, H. J.; Roecker, A. J.; Barluenga, S.; Cao, G. Q.; Affleck, R. L.; Lillig, J. E. *J. Am. Chem. Soc.* **2000**, *122*, 9954–9967.

(14) Vaino, A. R.; Janda, K. D. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 7692–7696.

(15) (a) Clouet, A.; Darbre, T.; Reymond, J.-L. *Angew. Chem. Int. Ed.* **2004**, *43*, 4612–4615. (b) Clouet, A.; Darbre, T.; Reymond, J.-L. *Biopolymers* **2006**, *84*, 114–123.

(16) Darbre, T.; Reymond, J.-L. *Acc. Chem. Res.* **2006**, *39*, 925–934.

(17) Kofoed, J.; Darbre, T.; Reymond, J.-L. *Org. Biomol. Chem.* **2006**, *4*, 3268–3281.

(18) (a) Kolomiets, E.; Johansson, E. M. V.; Renaudet, O.; Darbre, T.; Reymond, J.-L. *Org. Lett.* **2007**, *9*, 1465–1468. (b) Johansson, E. M. V.; Kolomiets, E.; Rosenau, F.; Jäger, K.-E.; Darbre, T.; Reymond, J.-L. *New. J. Chem.* **2007**, *31*, 1291–1299.

(19) Weinberger, H.; Lichte, E.; Griesinger, C.; Kutscher, B. *Arch. Pharm.* **1997**, *330*, 109–111.

(20) Sloane, N. J. A. The On-Line Encyclopedia of Integer Sequences **2006**, The Triangular numbers: http://www.research.att.com/~njas/sequences/A000217.

(21) (a) Kofoed, J.; Nielsen, J.; Reymond, J.-L. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2445–2447. (b) Martin, H. J.; List, B. *SYNLETT* **2003**, 1901–1902. (c) Krattiger, P.; Kovasy, R.; Revell, J. D.; Ivan, S.; Wennemers, H. *Org. Lett.* **2005**, *7*, 1101–1103.

(22) Tsogoeva, S.; Wei, S. *Tetrahedron: Asymmetry* **2005**, *16*, 1947–1951.

(23) Wagner, J.; Lerner, R. A.; Barbas III, C. F. *Science* **1995**, *270*, 1797–1800.

(24) (a) Mase, N.; Nakai, Y.; Ohara, N.; Yoda, H.; Takabe, K.; Tanaka, F.; Barbas III, C. F. *J. Am. Chem. Soc.* **2006**, *128*, 734–735. (b) Hayashi, Y.; Sumiya, T.; Takahasi, J.; Gotoh, H.; Urushima, T.; Shoji, M. *Angew. Chem. Int. Ed.* **2006**, *45*, 958–961.

(25) A 10- to 15-fold excess of beads over sequences ensures good library coverage. Screening itself is often carried out with 50 mg batches of resin, so that only 50000 beads are actually tested in an assay. This corresponds to assaying only approximately 53% of the sequences contained in a library of 65536 sequences. Indeed, the probability of any sequence of not being selected for the assay when taking 50000 beads from the library is $p = (65535/65536)^{50000} = 0.47$.

(26) Sakthievel, K.; Notz, W.; Bui, T.; Barbas III, C. F. *J. Am. Chem. Soc.* **2001**, *123*, 5260–5267.

CC7001155