 J. Med. Chem. **1998,** *41,* 3804−3811

Design of Self-Coded Combinatorial Libraries To Facilitate Direct Analysis of Ligands by Mass Spectrometry

Ian Hughes

SmithKline Beecham Pharmaceuticals, New Frontiers Science Park, Third Avenue, Harlow, Essex, CM19 5AW, U.K.

Received February 11, 1998

The direct analysis of selected components from combinatorial libraries by sensitive methods such as mass spectrometry is potentially more efficient than deconvolution and tagging strategies since additional steps of resynthesis or introduction of molecular tags are avoided. A substituent selection procedure is described that eliminates the mass degeneracy commonly observed in libraries prepared by "split-and-mix" methods, without recourse to high-resolution mass measurements. A set of simple rules guides the choice of substituents such that all components of the library have unique nominal masses. Additional rules extend the scope by ensuring that characteristic isotopic mass patterns distinguish isobaric components. The method is applicable to libraries having from two to four varying substituent groups and can encode from a few hundred to several thousand components. No restrictions are imposed on the manner in which the "self-coded" library is synthesized or screened.

Introduction

In recent years, the combinatorial synthesis of chemical libraries by the split-and-mix strategy[1] has been adopted as an efficient means of generating, in relatively few synthetic steps, large numbers of compounds (ligands) for biological evaluation.[2] Several techniques have been developed to identify active members of such combinatorial libraries. Positional scanning libraries[3] and iterative deconvolution strategies[4] require multiple syntheses of the same components in subsets of the final library and hence detract in some ways from the efficiency of the combinatorial synthesis. Other methods rely on the selection of active components from mixtures of compounds, either free in solution or tethered to synthesis beads, and their subsequent identification by various means.

Extraction and identification of active components from soluble libraries has been achieved by a number of affinity selection techniques[5] followed by mass spectroscopic analysis. If a library is prepared on solid phase, with a single component on each bead, spatial separation of the beads prior to screening in effect allows screening of the individual components, whether attached to or subsequently released from the beads.[2] To facilitate product identification, elegant encoding methods have been developed which introduce readable tags onto either the synthesis beads[6] or the released ligands themselves.[7] The use of dyes to color-code beads has also been described.[8] These methods have the potential to code very large numbers of components, but inevitably require extra steps to introduce tags, and also limit to varying degrees the scope of the chemistry available for ligand synthesis due to the required orthogonality with tagging chemistry, protection, and cleavage strategies. Recently, the use of radio frequency tags as a nonchemical coding strategy has been reported.[9]

Provided that sufficient material is available, direct analysis of selected ligands obviates the need for tagging, and in the case of peptides and oligonucleotides this can be achieved by sensitive sequencing techniques. For smaller, nonoligomeric compounds, mass spectrometry (MS) provides the most generally applicable technique for analysis of the small quantities (typically in the picomole range) of material available and has been applied to affinity selection methods[5d,e] as well as to the identification of material released from single beads.[10]

While methods such as high-resolution mass measurement, or MS-MS techniques have been used for library analysis[11] and are potentially very powerful for the detailed analysis of modest number of compounds, the need for calibration and interpretation render them less appropriate when large numbers of structurally diverse compounds are involved. The ability to distinguish library components by their nominal (integer) masses would allow the use of simpler instrumentation more suited to running in a high-throughput mode. However, if the masses of the ligands are to be maintained within the range 200−600 Da most likely to give acceptable pharmacokinetics,[12] the library size is restricted to a few hundred components if nominal masses are used. A more significant limitation arises from the nature of the split-and-mix method which rarely results in combinatorial libraries with an even mass distribution throughout the range. Libraries as small as 100 components will typically contain many sets of isobaric components which complicate analysis at the nominal mass level.[13] A genetic algorithm approach has recently been applied to address this issue.[14]

This paper presents a flexible stategy to overcome both limitations.[15] A rule-based procedure for substituent selection allows the design of libraries in which all components have unique nominal molecular weights. Further rules allow the use of isotope patterns to extend the final library size beyond the limits imposed by the desired molecular weight range. Since each component can be uniquely identified by its molecular weight and isotope pattern, such self-coded combinatorial libraries can, in principle, be applied to any screening approach

S0022-2623(98)00095-8 CCC: $15.00 © 1998 American Chemical Society
Published on Web 08/29/1998

Table 1. Chart of Consecutive Integers Arranged in Periodicity 10 with Two Sets Highlighted in Italic and Bold (See Text)

1	2	*3*	4	5	6	7	8	**9**	10
11	12	*13*	14	15	16	**17**	18	19	20
21	22	***23***	24	25	26	27	**28**	29	30
31	**32**	*33*	34	35	36	37	38	39	40
41	42	*43*	44	45	46	47	48	49	50
51	52	*53*	54	55	56	57	58	59	60
61	62	*63*	**64**	65	66	67	68	69	70
71	72	*73*	74	75	76	77	78	79	80
81	82	*83*	84	85	**86**	87	88	89	**90**
91	92	*93*	94	**95**	96	97	98	99	100

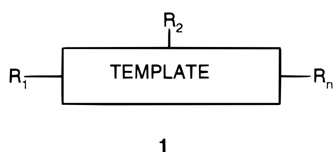
Table 2. Addition Matrix for Set 1 (Italic) and Set 2 (Bold) Chosen from Table 1

	set 2									
set 1	**31**	**32**	**23**	**64**	**95**	**86**	**17**	**28**	**9**	**90**
3	34	35	26	67	98	89	20	31	12	93
13	44	45	36	77	108	99	30	41	22	103
23	54	55	46	87	118	109	40	51	32	113
33	64	65	56	97	128	119	50	61	42	123
43	74	75	66	107	138	129	60	71	52	133
53	84	85	76	117	148	139	70	81	62	143
63	94	95	86	127	158	149	80	91	72	153
73	104	105	96	137	168	159	90	101	82	163
83	114	115	106	147	178	169	100	111	92	173
93	124	125	116	157	188	179	110	121	102	183

where sufficient material is available for MS analysis. This strategy imposes no restriction on the synthetic methods used to prepare the library, which is synthesized once only, requires no additional steps for introduction of tags, and therefore offers an attractive alternative approach to combinatorial library design.

Methods

In general terms, a combinatorial library can be defined as two or more sets of substituents (represented as R-groups) attached to a common core, or template, **1**. Since the template contributes a constant mass to



1

each library member, only variations in the substituent masses need to be considered further.

It can be demonstrated (see below) that pairwise addition of two sets of integers will always give unique sums if all members of the first set have the same remainder when divided by an arbitrary constant (herein referred to as the periodicity) and members of the second set each have different remainders when divided by the same periodicity constant. This can be represented graphically by means of a chart having a number of columns equal to the periodicity (see Table 1, periodicity = 10). Integers with the same remainder on division by the periodicity appear in the same column of the chart, and hence integers in different columns must have different remainders. Two sets of integers having either the same or different remainders on division by 10 are highlighted, in italic and bold, respectively, and the uniqueness of their sums may be seen by inspection of Table 2.

If these sets of integers were to correspond to the molecular weights of two sets of substituents defining a combinatorial library, then all members of the library would have unique nominal masses.

The above can be applied to the design of self-coded combinatorial libraries as follows. The nominal mass of each available substituent is calculated, using only the lower mass for elements having two major natural isotopes (e.g., Cl = 35, Br = 79). The substituents are arranged in chart format such that those having masses that give the same remainder after division by the chosen periodicity appear in the same column. The process is exemplified in Table 3 in which a selection of commonly encountered substituent groups have been arranged, under their molecular weights, into a chart of periodicity 10. For brevity, where multiple isomeric forms of a substituent are possible (e.g., propyl or isopropyl; imidazolyl or pyrazolyl; 2-, 3-, or 4-chlorophenyl, etc.) only a single form is indicated. An alternative approach would be to limit the substituent set to those available (commercially or otherwise) for a particular reagent class (e.g., amines, carboxylic acids, etc.) and to create separate charts for each reagent class to be used in a library synthesis.

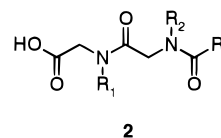
Application of the following rules ensures the selection of two sets of substituents that will combine to give a library with no isobaric components.

Rule 1: Set 1 and set 2 should be chosen from charts of the same periodicity.

Rule 2: Members of set 1 should all be chosen from the same column and be of different masses.

Rule 3: Members of set 2 should all be chosen from different columns.

For example, consider the design of a library of N-acylated dipeptoids **2** where R_1 and R_2 are derived from primary amines[16] and R_3 from carboxylic acids or



2

acid chlorides. The simplest self-coding strategy for this library would be to ensure that all combinations of R_1 and R_2 give unique masses and to retain as separate sublibraries the final pools each containing different, defined members of R_3. Using this approach, a library with n variable groups requires only the first ($n - 1$) R-groups to be coded. Applying rules 2 and 3 to Table 3 reveals numerous possible combinations of groups that might be selected for R_1 and R_2; indeed this is one of the attractions of this coding method. Two possible sets of substituents are indicated in Table 4. This selection is for illustration only and does not necessarily reflect the ready availability and suitable reactivity of all of the reagents. On the basis of this selection each sublibrary would have 150 components (15×10) with unique nominal masses. Assuming 20 acylating agents were chosen for R_3, a total library size of 3000 readily distinguishable components could be prepared.

Use of Isotopes To Extend the Scope. Isobaric compounds containing different numbers and combinations of Br and Cl atoms can be distinguished by the pattern of isotope masses above the nominal mass peak.

Table 3. Chart of Commonly Encountered Substituent Groups (Periodicity 10)

1	72	43	44	15	66	27	58	29	30
Hydrogen	Carbamoylethyl	Propyl	Aminoethyl	Methyl	Pyrrolyl	Vinyl	Aminopropyl	Ethyl	Aminoethyl
31	Aminobutyl	73	84	45	86	57	Carbamoylmethyl	39	80
Hydroxymethyl	Dimethylaminoethyl	Hydroxybutyl	Thiazolyl	Hydroxyethyl	Aminopentyl	Butyl	68	Propynyl	Methylpyrrolyl
41	82	Carboxyethyl	114	55	Carbamoylpropyl	67	Oxazolyl	59	Triazinyl
Propenyl	Methyloxazolyl	83	Carbamoylpentyl	Cyclobutyl	Dimethylaminopropyl	Furyl	Triazolyl	Hydroxypropyl	100
Cyclopropyl	92	Cyclohexyl	124	Butenyl	106	Imidazolyl	78	Carboxymethyl	Carboxamidobutyl
61	Aminophenyl	Thienyl	Methylthiopyridyl	75	Methylaminophenyl	77	Pyridyl	69	Chloropyrrolyl
Methylthiomethyl	Methylpyridyl	93	134	Methylthioethyl	116	Phenyl	98	Cyclopentyl	120
71	102	Hydroxyphenyl	Propylaminophenyl	85	Indolyl	87	Methylthiazolyl	Trifluoromethyl	Dimethylaminophenyl
Pentyl	Cyanophenyl	103	Dimethylaminobenzyl	Hexyl	Cyanobenzyl	Hydroxypentyl	108	Tetrazolyl	Ethylaminophenyl
81	Chloroisoxazolyl	Methylthiobutyl	Benzothiazolyl	Thiadiazolyl	136	Carboxypropyl	Methoxypyridyl	79	Carbamoylphenyl
Methylfuryl	112	Styryl	144	95	Nitrobenzyl	Methoxycarbonylethyl	118	Pyrimidinyl	130
Methylimidazolyl	Nitrofuryl	113	Bromopyrrolyl	Fluorophenyl	146	97	Chlorothiazolyl	89	Methylindolyl
91	Chloropyridyl	Octyl	184	Norbornyl	Bromooxazolyl	Cycloheptyl	Benzotriazolyl	Methylthiopropyl	150
Methylphenyl	Nitroimidazolyl	Difluorophenyl	Benzyloxypyridyl	105	Dichloropyridyl	Methylthienyl	128	99	Chloroindolyl
Benzyl	122	Nitrooxazolyl	194	Ethylphenyl	156	107	Quinolyl	Heptyl	
101	Nitrophenyl	123	Bromoindolyl	Phenethyl	Bromopyridyl	Methoxyphenyl	148	109	
Carboxybutyl	152	Nitropyridyl		Methylbenzyl		Hydroxybenzyl	Diethylaminophenyl	Dihydroxyphenyl	
Chlorofuryl	Chlorobenzoxazolyl	Methylthiophenyl		115		117	Butylaminophenyl	Fluorobenzyl	
Chloroimidazolyl	162	133		Methoxycarbonylbutyl		Trichloromethyl	Nitrostyryl	119	
Methoxycarbonylpropyl	Chloroquinolyl	Butylphenyl		125		Chlorothienyl	168	Propylphenyl	
111	Bromothiazolyl	Benzothienyl		Chlorobenzyl		Benzofuryl	Phenylaminophenyl	Acetylphenyl	
Cyclooctyl	172	143		135		Phenylcyclopropyl		129	
Chlorophenyl	Nitronaphthyl	Methoxycarbonylhexyl		Propyloxyphenyl		Benzimidazolyl		Quinazolinyl	
121		153		Methoxycarbonylphenyl		127		Methoxycarbonylpentyl	
Ethoxyphenyl		Biphenylyl		Adamantyl		Nonyl		139	
Carboxyphenyl		163		Benzothiadiazolyl		Naphthyl		Chlorophenethyl	
Methoxybenzyl		Pentyloxyphenyl		145		137		149	
Methylenedioxyphenyl		173		Dichlorophenyl		Dimethoxyphenyl		Butyloxyphenyl	
Carbamoylpyridyl		Nitroquinolyl		Bromofuryl		Chlorostyryl		Tetrafluorophenyl	
131		183		Bromoimidazolyl		Bornyl		159	
Trifluorophenyl		Benzyloxyphenyl		Trifluoromethylphenyl		147		Dichlorobenzyl	
Tetrahydronaphthyl				155		Pentylphenyl		Trifluoromethylbenzyl	
141				Undecyl		Methylbenzothienyl		169	
Decyl				Bromophenyl		157		Dodecyl	
Naphthylmethyl				165		Bromopyrimidinyl		Phenoxyphenyl	
151				Diethoxyphenyl		167		Bromobenzyl	
Chlorobenzofuranyl				175		Dinitrophenyl		179	
Dimethoxybenzyl				Heptylphenyl		Pentafluorophenyl		Trichlorophenyl	
161				195		Benzylphenyl		189	
Hexylphenyl				Bromobenzofuryl		Chlorobenzothienyl		Octylphenyl	
Bromothienyl						Trimethoxyphenyl		Bromochlorophenyl	
Chloronaphthyl						177			
171						Hexyloxyphenyl			
Trifluoromethylstyryl						Anthryl			
181									
Trimethoxybenzyl									
191									
Heptyloxyphenyl									

Table 4. Substituent Sets Selected from Table 3 According to Rules 2 and 3

set 1 (rule 2)		set 2 (rule 3)	
MW	substituent	MW	substituent
1	hydrogen	71	pentyl
31	hydroxymethyl	82	methyloxazolyl
41	cyclopropyl	93	hydroxyphenyl
71	pentyl	44	aminoethyl
81	methylimidazolyl	15	methyl
91	benzyl	116	indolyl
101	carboxybutyl	77	phenyl
111	cyclooctyl	78	pyridyl
121	ethoxyphenyl	109	fluorobenzyl
131	trifluorophenyl	120	dimethylaminophenyl
141	decyl		
151	dimethoxybenzyl		
161	hexylphenyl		
171	trifluoromethylstyryl		
191	heptyloxyphenyl		

Table 5. Additional Substituents for Set 1 Selected from Table 3 by Rule 5

MW	substituent
101	chlorofuryl
111	chlorophenyl
151	chlorobenzofuranyl
161	bromothienyl
161	chloronaphthyl

Table 6. Additional Substituents for Set 2 Selected from Table 3 by Rule 6

MW	substituent	MW	substituent
111	chlorophenyl	156	bromopyridyl
161	bromothienyl	117	trichloromethyl
152	chlorobenzoxazolyl	137	chlorostyryl
162	bromothiazolyl	157	bromopyrimidinyl
144	bromopyrrolyl	139	chlorophenethyl
125	chlorobenzyl	159	dichlorobenzyl
145	dichlorophenyl	179	trichlorophenyl
146	dichloropyridyl		

On the basis of this observation, further selection rules can be stated which allow the design of expanded self-coded libraries in which each component is uniquely identified by a combination of its nominal mass and isotope pattern.

Rule 4: Either, but not both, of rules 2 and 3 may be replaced by rules 5 and 6, respectively.

Rule 5: Members of set 1 should be chosen from the same column and may have the same mass provided that the isotope pattern (e.g., Br, Cl count) is different.

Rule 6: Members of set 2 can be chosen from any column provided that all members chosen from an individual column have different isotope patterns (e.g., different Br, Cl counts).

For example, the library defined in Table 4 could be extended by applying rule 5 to include five more substituents (Table 5), giving sublibraries of 200 components (20 × 10), or by applying rule 6 to introduce 15 additional substituents to set 2 (Table 6), giving sublibraries of 375 distinguishable components (15 × 25). Assuming once again 20 values for R_3, total self-coded library sizes of 4000 and 7500, respectively, could be achieved.

The foregoing discussion has assumed coding of two sets of substituents and that the identity of a third substituent would be derived from knowledge of its sublibrary. However, if coding is required for three substituent sets (as would be necessary for a library with four R-groups), this can be accommodated by using rule 7, which reserves isotope patterns exclusively to code the third R-group.

Table 7. Substituent Set 3 Having Different Isotope Patterns Selected by Rule 7

MW	substituent
71	pentyl
137	chlorostyryl
159	dichlorobenzyl
117	trichloromethyl
156	bromopyridyl

Rule 7: Exclude groups containing Br and Cl from sets 1 and 2. Members of set 3 can be selected freely, provided that each member has a different, distinguishable isotope pattern (e.g., different Br, Cl count).

Applying rule 7 to the dipeptoid example would require using the original selections defined in Table 4 for R_1 and R_2, which gave unique combinations without recourse to isotope codes, and then selecting isotopically distinguishable substituents for R_3, for example, the five groups shown in Table 7. This would give a dipeptoid library having 750 easily distinguishable components (15 × 10 × 5). If there were a fourth R-group (R_4) having 20 values, self-coded libraries of 15 000 components could be generated in this manner.

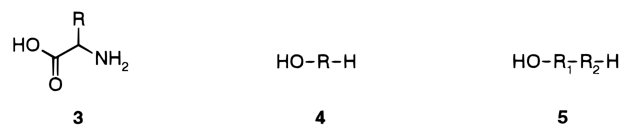
The use of isotopes to augment the substituent sets need not be restricted to bromine and chlorine. Useful additional reagent sets may be created by artificially enriching reagents with other isotopic elements. For example, combining benzoic acid and pentadeuteriobenzoic acid in equimolar proportions gives a reagent with a characteristic M+5 ion.

Since addition is a commutative property, there is no requirement to equate R_1 with set 1 and R_2 with set 2; the reverse assignment is equally valid. Likewise when three sets of substituents are coded, rule 7 may be applied to either R_1, R_2, or R_3—whichever best suits the library design.

Discussion

In the following discussion, several important features of the method will be illustrated by reference to libraries of dipeptides derived from the 20 natural amino acids. Although the method is not particularly suited to peptide libraries, the use of a very restricted set of building blocks will simplify the analysis of the scope and flexibility of the approach.

First, the method is not restricted to simple "monovalent" substituent groups of the type discussed so far. For example, while nineteen of the natural amino acids can be defined as substituted glycine molecules, **3**, the cyclic amino acid proline cannot, so alternative representations of an amino acid **4** and dipeptide **5** must be



used. This is not a problem provided that all substituents in a reagent set are treated similarly. Table 8 shows the 20 natural amino acids arranged in a chart of periodicity 10. Note that, despite the different

Table 8. Chart of Amino Acids (Periodicity 10)

71	**103**	**114**	**115**	**156**	**57**	**128**	**99**
Ala	Cys	Asn	Asp	Arg	Gly	Gln	Val
101	**113**			**186**	**87**	Lys	**129**
Thr	Leu			Trp	Ser		Glu
131	Ile				**97**		
Met	**163**				Pro		
	Tyr				**137**		
					His		
					147		
					Phe		

Table 9. Chart of Amino Acids (Periodicity 11)

57	**113**	**103**	**71**	**128**	**129**	**97**	**87**	**99**
Gly	Leu	Cys	Ala	Gln	Glu	Pro	Ser	Val
101	Ile	**114**	**115**	Lys		**163**	**131**	
Thr		Asn	Asp			Tyr	Met	
156		**147**	**137**				**186**	
Arg		Phe	His				Trp	

Table 10. Chart of Amino Acids (Periodicity 14)

57	**114**	**87**	**186**	**103**	**147**	**163**	**137**	**97**
Gly	Asn	Ser	Trp	Cys	Phe	Tyr	His	Pro
71	**128**	**101**		**131**				
Ala	Gln	Thr		Met				
99	Lys	**115**						
Val	**156**	Asp						
113	Arg	**129**						
Leu		Glu						
Ile								

Table 11. Chart of Amino Acids (Periodicity 16)

97	**114**	**99**	**101**	**71**	**57**	**186**	**156**	**128**
Pro	Asn	Val	Thr	Ala	Gly	Trp	Arg	Gln
113		**115**		**87**	**137**			Lys
Leu		Asp		Ser	His			
Ile		**131**		**103**				
129		Met		Cys				
Glu		**147**						
		Phe						
		163						
		Tyr						

definition of what comprises a substituent, the same groups fall into alignment within a column as would be expected from Table 3 (i.e., Gly, Ser, His, and Phe; cf. hydrogen, hydroxymethyl, imidazolylmethyl, benzyl), albeit not necessarily the exactly corresponding columns.

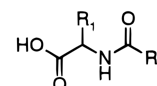
Effect of Periodicity. The choice of periodicity is entirely arbitrary provided that rule 1 is followed. Tables 8–11 show the amino acids arranged in charts of periodicities 10, 11, 14, and 16, respectively, with empty columns omitted. The most obvious and important effect of altering chart periodicity is that different substituents are brought into alignment within columns. For example, periodicity 10 (Table 8) aligns Gly with Ser, Pro, His, and Phe, while periodicity 11 (Table 9) aligns Gly with Thr and Arg. Since entries within columns of a periodicity 14 chart differ by 14 mass units, corresponding to a methylene group, homologous series are brought into alignment as in Table 10 where Gly, Ala, Val, and Leu are in the same column, as are Ser and Thr, Asn, and Gln, etc. Hence judicious variation of periodicity allows considerable flexibility in the diversity of substituent sets that can be selected.

Flexibility. Although the method imposes no restrictions on synthesis or screening strategies, it must, by virtue of being a selection method, restrict to some extent the choice of substituents available. The limitations are most noticeable for the set chosen from the same column by rule 2. This is particularly so when the number of substituents from which to select is low, such as the charts in Tables 8–11. While in such cases the structural diversity within a column can be adjusted by changing the chart periodicity, the actual number of choices available will inevitably remain small. However, even for limited substituent availabilities, the number of options for selecting a set by rule 3 (different columns) is always considerable.

Consider using the periodicity 10 chart of Table 8 to select substituents sets for a dipeptide library. Using rule 2 (same column) for set 1 allows a maximum set size of five (Gly, Ser, Pro, His, Phe). If only three amino acids were required in this set, a total of 13 possible combinations would be possible (e.g., Gly, Pro, Phe; or Cys, Leu, Tyr). Applying rule 3 gives 480 ($3 \times 4 \times 1 \times 1 \times 2 \times 5 \times 2 \times 2$) potential choices of eight amino acids for set 2.

As the number of substituents available increases (e.g., reagent sets such as carboxylic acids, amines, and alcohols), each column becomes increasingly populated, more masses are represented, and each will include many isobaric alternatives, including, of course, isomeric forms (e.g., Table 3), greatly reducing the limitations described above.

If selecting substituents from two different reagent sets, it is therefore apparent that the optimum application of the rules is generally achieved by using rule 2 (same column) to select from the most populated reagent set and reserving rule 3 (different columns) for the smaller, more restrictive set. For example, a simple acylated amino acid library, **6**, would best be designed



6

by applying rule 2 (same column) to the acylating agent (e.g., the 15 substituents in Table 4, set 1, selected from Table 3) while reserving rule 3 (different columns) to select 8 amino acids from Table 8 giving a 120 component library.

Scope. Consider a theoretical set of substituents fully populating a mass range m and arranged in a chart of periodicity p. The maximum number of nonisobaric product masses is m/p (from rule 2) $\times$ p (from rule 3) = m. So for a typical reagent mass range of 200, a maximum of 200 distinguishable products would be possible, spread over a total mass range of 400 mass units. Indeed, the library of 150 components generated from the substituent sets defined in Table 4 is fairly typical of the situation for well-populated reagent sets.

When the number of substituents available becomes more restricted, as in the dipeptide example described above, there is a corresponding decrease in the number of components that can be uniquely encoded. By arranging the amino acid side chains in charts of periodicities from 2 to 20, it can be shown that periodicity 16 (Table 11) maximizes the number of nonisobaric dipeptides available by this method at 45 (five from rule 2, nine from rule 3). To put this into perspective, however,

it should be noted that only 108 different nominal masses are represented among the 400 dipeptides theoretically available from the 20 amino acid building blocks. Of these, only eight are represented once only. It is in fact impossible to generate a dipeptide library in a split-and-mix fashion that accesses each of the 108 masses once only.

As has been demonstrated, the appropriate use of isotopes increases the number of readily distinguishable components severalfold. The simple device of maintaining separate sublibrary pools after introduction of the final R-group substituent set increases the scope by an order of magnitude, and self-coded libraries of several thousand components are not unreasonable.

Mathematical Description of the Method

To demonstrate that pairwise addition of sets chosen according to rules 1–3 generates unique sums, consider the general case of Table 1, periodicity p. The value in row i, column j is given by $V(i,j) = (i-1)p + j$. The qth element of set 1 chosen from column y by rule 2 (same column) is $S1(q) = (q-1)p + y$. If $x(1)$, $x(2)$, $x(3)$, ..., $x(p)$ are the row numbers of elements of set 2 chosen from columns 1, 2, 3, ..., p by rule 3 (different columns), then the rth element of set 2 is $S2(r) = (x(r) - 1)p + r$. Hence the sum of any element q from set 1 and any element r from set 2 is given by eq 1

$$S(q,r) = (q + x(r) - 2)p + (y + r) \quad (1)$$

In order for the sums of two pairs of elements, m and a from set 1 and n and b from set 2, to be equal requires that $S(m,n) = S(a,b)$. Substituting eq 1 gives $(m + x(n) - 2)p + (y + n) = (a + x(b) - 2)p + (y + b)$, hence

$$(m - a + x(n) - x(b))p = b - n \quad (2)$$

Since both b and n are integers, then $b - n$ is either 0 or an integer. If $b - n = 0$, then $b = n$. Substituting into eq 2 gives $(m - a) = 0$, hence $m = a$, which implies addition of the same pair of elements twice. If $b - n$ is an integer, then the left-hand side of eq 2 is a multiple of p, but the right-hand side of eq 2 cannot be a multiple of p, since $1 \leq b \leq p$ and $1 \leq n \leq p$, and hence $1-p \leq (b - n) \leq p - 1$, that is, the absolute value of $b - n$ is less than p. Hence $S(m,n)$ cannot be the same as $S(a,b)$, proving that all pairwise additions of numbers from set 1 and set 2 give unique sums.

Synthesis of 50-Component Self-Coded Library

The practical application of the methodology is illustrated by the synthesis of a self-coded bead library of bisamides, according to Scheme 1, using a split-and-mix strategy to generate a single compound on each bead. The *N*-protected benzyl bromide **7** was attached to polystyrene resin via a traceless phosphonium salt linker group.[17] Following deprotection of the product, **8**, the amino group was acylated with five *N*–Fmoc amino acids (R_{1a} to R_{1e}, Table 12), and the Fmoc group was removed with piperidine. The free amino group was then acylated with 10 acid chlorides (R_{2a} to R_{2j}, Table 12) to give the bead-bound library **10**.

The selection of R-groups was made according to rules 1–3 using a periodicity of 12. The masses of groups for R_1 were selected so as to have **different** remainders when divided by the periodicity, 12 (i.e., they would belong to **different** columns if arranged in a chart format—rule 3), and groups R_2 were chosen to have masses giving the **same** remainder on division by 12 (and hence would appear in the **same** column in a periodicity 12 chart—rule 2).

Scheme 1

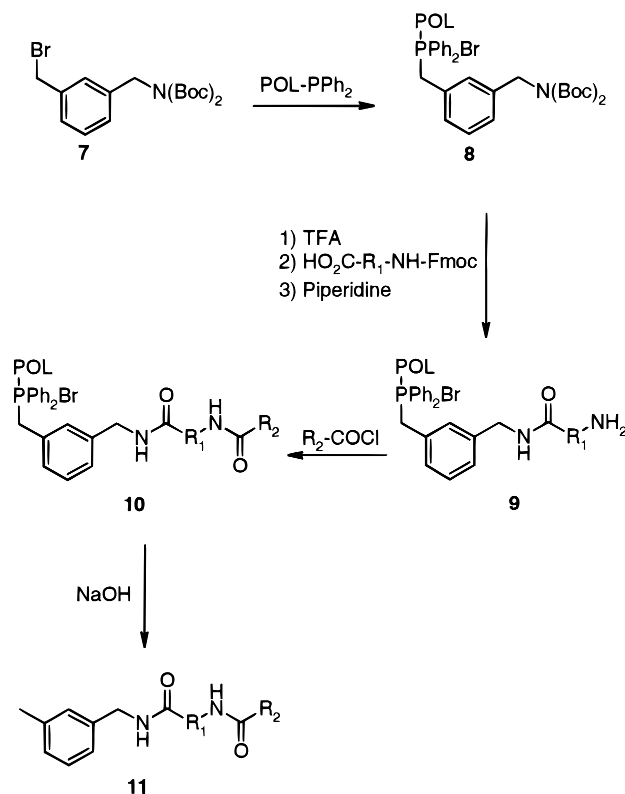


Table 12. Reagents Used in Synthesis of Self-Coded Library **10**

R-group	reagent	R-group mass	mass/12
R_{1a}	*N*-Fmoc-glycine	14	1 rem 2
R_{1b}	*N*-Fmoc-β-alanine	28	2 rem 4
R_{1c}	*N*-Fmoc-6-aminohexanoic acid	70	5 rem 10
R_{1d}	*N*-Fmoc-4-aminomethyl-benzoic acid	90	7 rem 6
R_{1e}	*N*-Fmoc-4-aminomethylcyclo-hexanecarboxylic acid	96	8 rem 0
R_{2a}	butyryl chloride	43	3 rem 7
R_{2b}	2-furoyl chloride	67	5 rem 7
R_{2c}	phenylacetyl chloride	91	7 rem 7
R_{2d}	cinnamoyl chloride	103	8 rem 7
R_{2e}	methyl adipyl chloride	115	9 rem 7
R_{2f}	1-naphthoyl chloride	127	10 rem 7
R_{2g}	10-undecenoyl chloride	139	11 rem 7
R_{2h}	2,5-dimethoxyphenylacetyl chloride	151	12 rem 7
R_{2i}	4-fluoro-3-(trifluoromethyl)-benzoyl chloride	163	13 rem 7
R_{2j}	4-*n*-heptylbenzoyl chloride	175	14 rem 7

Fifteen randomly selected beads from **10** were individually cleaved by hydrolysis of the phosphonium linker to give the toluene derivatives **11**. Analysis of the cleavage products by APCI-MS gave very clear M+H signals (Figure 1 shows a typical result) that could be unambiguously assigned to components of the library (Table 13, beads 1–15). In one case (bead 14) a very weak response was obtained, suggesting a poorly functionalized bead or inefficient cleavage.

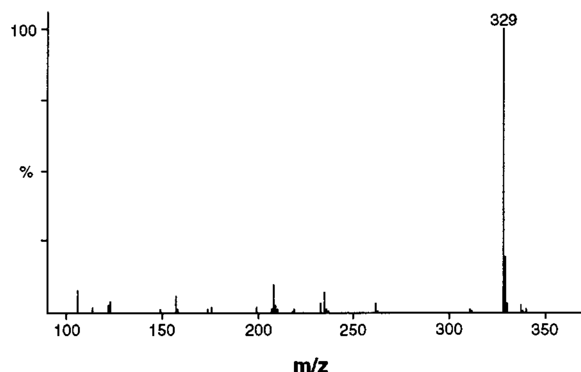


Figure 1. Mass spectrum for bead 2.

Table 13. Assignment of Cleavage Products **11** from Library **10**

bead no.	M+H (found)	assignment	mass (calc)
1	365	$R_{1c}-R_{2d}$	364
2	329	$R_{1c}-R_{2b}$	328
3	379	$R_{1e}-R_{2c}$	378
4	439	$R_{1e}-R_{2h}$	438
5	391	$R_{1e}-R_{2d}$	390
6	379	$R_{1e}-R_{2c}$	378
7	409	$R_{1d}-R_{2f}$	408
8	329	$R_{1c}-R_{2b}$	328
9	457	$R_{1d}-R_{2j}$	456
10	325	$R_{1d}-R_{2a}$	324
11	357	$R_{1a}-R_{2h}$	356
12	403	$R_{1e}-R_{2e}$	402
13	397	$R_{1d}-R_{2e}$	396
14	not found		
15	433	$R_{1d}-R_{2h}$	432

Conclusion

Using a combination of nominal mass and isotope pattern to distinguish between members of a self-coded library provides an additional, powerful means of encoding combinatorial libraries that is very simple to implement. The method relies critically on the ability to measure the molecular ion of a ligand selected from the library, a process which is rarely problematic with modern mass spectrometric techniques. While the method has great flexibility, there will be certain situations, such as peptide libraries, where restrictions on the available reagent pools limit its usefulness. In such circumstances or where very large combinatorial libraries are required, a tagging strategy would be more appropriate. However, for libraries within the scope described above, the method offers a number of significant advantages over existing encoding and deconvolution protocols: no additional steps of resynthesis or introduction of tags are required; no limitations are imposed by the need for orthogonal synthetic and release chemistries as is the case for tagging methods; the method is equally applicable to solution and solid-phase synthesis methods, since all the information to decode the structures is stored in the ligands themselves and can thus be applied to screening strategies, such as affinity selection, where association of ligands and beads is not possible.

Experimental Section

Mass spectra were recorded with a Micromass Platform 1 spectrometer. 4-Bromopolystyrene was a copolymer of 4-bromostyrene and divinylbenzene purchased from Polymer Laboratories with a loading of 2 mmol g^{-1}.

Polymer-Bound 3-[*N,N*-Di(*tert*-butyloxycarbonyl)aminomethyl]benzyl Triphenylphosphonium Bromide (8). A suspension of polymer-bound triphenylphosphine (2.96 g) prepared from 150−300 μm 4-bromopolystyrene[18] in DMF (25 mL) was treated with 3-bromomethyl-*N,N*-di(*tert*-butyloxycarbonyl)benzylamine (4 g, 10 mmol). The mixture was heated at 70 °C for 48 h and cooled, and the polymer was filtered and washed alternately with toluene (50 mL) and CH_2Cl_2 (50 mL) (3 times), then with CH_2Cl_2 (50 mL) and Et_2O (50 mL) (3 times), and finally with 1:1 CH_2Cl_2/Et_2O (50 mL) and Et_2O (2 × 50 mL) and dried to give **8** (4.21 g). Anal. Br (ionic): found, 7.3. Loading = 0.91 mmol g^{-1}.

Mixture of Polymer-Bound 3-[*N*-(Aminoacyl)aminomethyl]benzyl Triphenylphosphonium Salts (9). In five separate experiments, **8** (385 mg, 0.35 meq) was washed with CH_2Cl_2 (2 × 25 mL) and then treated with 30% CF_3CO_2H and 2% anisole in CH_2Cl_2 (2 × 25 mL) for 1 and 30 min. The resin was washed with CH_2Cl_2 (3 × 25 mL), 10% Et_3N in CH_2Cl_2 (2 × 25 mL), and CH_2Cl_2 (2 × 25 mL). The resin was suspended overnight in a solution of 1-hydroxy-7-azabenzotriazole (109 mg, 0.8 mmol), 1,3-diisopropylcarbodiimide (0.13 mL, 0.8 mmol), and one of the carboxylic acids (R_{1a} to R_{1e}, Table 12) (0.75 mmol) in DMF (10 mL) and CH_2Cl_2 (3 mL). After the resin was washed with DMF (2 × 25 mL) and CH_2Cl_2 (3 × 25 mL), the five products were combined as a slurry in CH_2Cl_2, filtered, and treated with 20% piperidine in DMF (2 × 30 mL) for 1 and 30 min. The resin was washed with DMF (2 × 30 mL) and CH_2Cl_2 (3 × 30 mL) and then dried to give **9** (1.83 g, approximately 1.75 mmol) as a mixture of five polymer-bound components.

Mixture of Polymer-Bound 3-[*N*-(Acylaminoacyl)aminomethyl]benzyl Triphenylphosphonium Salts (10). In 10 separate experiments, **9** (180 mg, approximately 0.17 mmol) was suspended in CH_2Cl_2 (10 mL) and treated with Et_3N (0.1 mL, 0.7 mmol) and one of the acid chlorides (R_{2a} to R_{2j}, Table 12) (0.35 mmol). After 90 min the resin was filtered and washed with CH_2Cl_2 (2 × 20 mL). The 10 products were combined as a slurry in CH_2Cl_2 and washed with DMF (3 × 30 mL) and CH_2Cl_2 (2 × 30 mL). The resin was finally washed successively with 3:1, 1:1, and 1:3 mixtures of CH_2Cl_2 and Et_2O, then washed with Et_2O, and dried to give **10** as a mixture of 50 polymer-bound components.

Analysis of Library 10 by Mass Spectrometry. Individual beads from **10** were treated in glass inserts with a 10 mM solution of sodium hydroxide in 10% aqueous dioxane (50 μL) overnight to effect cleavage to individual members of library **11**. Representative results of APCI-MS analysis along with the assignment of the molecular ions detected are given in Table 13.

Acknowledgment. I thank Ian Macpherson and Michael Smith of the Statistical Sciences Department, SB, for useful discussions regarding the mathematical basis of this method, and Catriona Thom for recording the mass spectra.

References

(1) (a) Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. General Method for Rapid Synthesis of Multicomponent Peptide Mixtures. *Int. J. Pept. Protein Res.* **1991**, *37*, 487−493. (b) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. A New Type of Synthetic Peptide Library for Identifying Ligand-Binding Activity. *Nature* **1991**, *354*, 82−84. (c) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. Generation and Use of Synthetic Peptide Combinatorial Libraries for Basic Research and Drug Discovery. *Nature* **1991**, *354*, 84−86.

(2) Lam, K. S.; Lebl, M.; Krchnak, V. The "One-Bead-One-Compound" Combinatorial Library *Method. Chem. Rev.* **1997**, *97*, 411−448.

(3) (a) Pinilla, C.; Appel, J. R.; Blanc, P.; Houghten, R. A. Rapid Identification of High Affinity Peptide Ligands Using Positional Scanning Peptide Combinatorial Libraries. *BioTechniques* **1992**, *13*, 901−905. (b) Pinilla, C.; Appel, J. R.; Blondelle, S. E.; Dooley, C. T.; Eichler, J.; Ostresch, J. M.; Houghten, R. A. Versatiliity of Positional Scanning Synthetic Combinatorial Libraries for the Identification of Individual Compounds. *Drug Dev. Res.* **1994**, *33*, 133−145.

(4) (a) Freier, S. M.; Konings, D. A. M.; Wyatt, J. R.; Ecker, D. J. Deconvolution of Combinatorial Libraries for Drug Discovery: A Model System. *J. Med. Chem.* **1995**, *38*, 344−352. (b) Erb, E.; Janda, K. D.; Brenner, S. Recursive Deconvolution of Combinatorial Chemical Libraries. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11422−11426.

(5) (a) Chu, Y.-H.; Kirby, D. P.; Karger, B. L. Free Solution Identification of Candidate Peptides from Combinatorial Libraries by Affinity Capillary Electrophoresis/Mass Spectrometry. *J. Am. Chem. Soc.* **1995**, *117*, 5419−5420. (b) Kelly, M. A.; Liang, H.; Sytwu, I.-I.; Vlattas, I.; Lyons, N. L.; Bowen, B. R.; Wennogle, L. P. Characterisation of SH2-Ligand Interactions via Library Affinity Selection with Mass Spectrometric Detection. *Biochemistry* **1996**, *35*, 11747−11755. (c) Chu, Y.-H.; Dunayevskiy, Y. M.; Kirby, D. P.; Vouros, P.; Karger, B. L. Affinity Capillary Electrophoresis-Mass Spectrometry for Screening Combinatorial Libraries. *J. Am. Chem. Soc.* **1996**, *118*, 7827−7835. (d) Blackburn, C.; Pingali, A.; Kehoe, T.; Herman, L. W.; Wang, H.; Kates, S. A. Libraries of Angiotensin Converting Enzyme Inhibitors: Solid-Phase Synthesis and Affinity Selection. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 823−826. (e) Weiboldt, R.; Zweigenbaum, J.; Henion, J. Immunoaffinity Ultrafiltration with Ion Spray HPLC/MS for Screening of Small-Molecule Libraries. *Anal. Chem.* **1997**, *69*, 1683−1691.

(6) (a) Ohlmeyer, M. H. J.; Swanson, R. N.; Dillard, L. W.; Reader, J. C.; Asouline, G.; Kobayashi, R.; Wigler, M.; Still, W. C. Complex Synthetic Chemical Libraries Indexed with Molecular Tags. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10922−10926. (b) Nestler, H. P.; Bartlett, P. A.; Still, W. C. A General Method for Molecular Tagging of Encoded Combinatorial Chemistry Libraries. *J. Org. Chem.* **1994**, *59*, 4724. (c) Burbaum, J. J.; Ohlmeyer, M. H. J.; Reader, J. C.; Henderson, I.; Dillard, L. W.; Li, G.; Randle, T. L.; Sigal, N. H.; Chelsky, D.; Baldwin, J. J. A Paradigm for Drug Discovery Employing Encoded Combinatorial Libraries. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6027−6031. (d) Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. Generation and Screening of an Oligonucleotide-Encoded Synthetic Peptide Library. *Proc. Nat. Acad. Sci. U.S.A.* **1993**, *90*, 10700−10704. (e) Ni, Z.-J.; Maclean, D.; Holmes, C. P.; Murphy, M. M.; Ruhland, B.; Jacobs, J. W.; Gordon, E. M.; Gallop, M. A. Versatile Approach to Encoding Combinatorial Organic Syntheses Using Chemically Robust Secondary Amine Tags. *J. Med. Chem.* **1996**, *39*, 1601−1608. (f) Nikolaiev, V.; Stierandova, A.; Krchnak, V.; Seligmann, B.; Lam, K. S.; Salmon, S. E.; Lebl, M. Peptide-Encoding for Structure Determination of Nonsequenceable Polymers Within Libraries Synthesised and Tested on Solid-Phase Supports. *Pept. Res.* **1993**, *6*, 161−170. (g) Vagner, J.; Barany, G.; Lam, K. S.; Krchnak, V.; Sepetov, N. F.; Ostrem, J. A.; Strop, P.; Lebl, M. Enzyme-Mediated Spatial Segregation on Individual Polymeric Support Beads: Application to Generation and Screening of Encoded Combinatorial Libraries. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8194−8199. (h) Edwards, P. N.; Main, B. G.; Shute, R. E. GB 2,297,551 A, 1996. (i) Geysen, H. M.; Wagner, C. D.; Bodnar, W. M.; Markworth, C. J.; Parke, G. J.; Schoenen, F. J.; Wagner, D. S.; Kinder, D. S. Isotope or Mass-Encoding of Combinatorial Libraries. *Chem. Biol.* **1996**, *3*, 679−688.

(7) (a) Brenner, S.; Lerner, R. A. Encoded Combinatorial Chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5381−5383. (b) Nielsen, J.; Brenner, S.; Janda, K. D. Synthetic Methods for the Implementation of Encoded Combinatorial Chemistry. *J. Am. Chem. Soc.* **1993**, *115*, 9812−9813. (c) Kerr, J. M.; Banville, S. C.; Zuckermann, R. N. Encoded Combinatorial Peptide Libraries Containing Non-Natural Amino Acids. *J. Am. Chem. Soc.* **1993**, *115*, 2529−2531.

(8) (a) Campian, E.; Sebestyen, F.; Major, F.; Furka, A. Synthesis of Support-Bound Peptides Carrying Color Labels. *Drug Dev. Res.* **1994**, *33*, 98−101. (b) Egner, B. J.; Rana, S.; Smith, H.; Bouloc, N.; Frey, J. G.; Brocklesby, W. S.; Bradley, M. Tagging in Combinatorial Chemistry: The Use of Coloured and Fluorescent Beads. *J. Chem. Soc., Chem. Commun.* **1997**, 735−736.

(9) (a) Nicolaou, K. C.; Xiao, X.-Y.; Parandoosh, Z.; Senyei, A.; Nova, M. P. Radiofrequency Encoded Combinatorial Chemistry. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2289−2291. (b) Moran, E. J.; Sarshar, S.; Cargill, J. F.; Shahbaz, M. M.; Lio, A.; Mjalli, A. M. M.; Armstrong, R. W. Radio Frequency Tag Encoded Library Method for the Discovery of Tripeptide-Substituted Cinnamic Acid Inhibitors of the Protein Tyrosine Phosphatase PTP1B. *J. Am. Chem. Soc.* **1995**, *117*, 10787−10788. (c) Armstrong, R. W.; Tempest, P. A.; Cargill, J. F. Microchip Encoded Combinatorial Libraries: Generation of a Spatially Encoded Library from a Pool Synthesis. *Chimia* **1996**, *50*, 258−260.

(10) (a) Brummel, C. L.; Lee, I. N. W.; Zhou, Y.; Benkovic, S. J.; Winograd, N. A Mass Spectrometric Solution to the Address Problem of Combinatorial Libraries. *Science* **1994**, *264*, 399−402. (b) Egner, B. J.; Langley, G. J.; Bradley, M. Solid Phase Chemistry: Direct Monitoring by Matrix-Assisted Laser Desorption/Ionisation Time of Flight Mass Spectrometry. A Tool for Combinatorial Chemistry. *J. Org. Chem.* **1995**, *60*, 2652−2653. (c) Haskins, N. J.; Hunter, D. J.; Organ, A. J.; Rahman, S. S.; Thom, C. Combinatorial Chemistry: Direct Analysis of Bead Surface Associated Materials. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1437−1440. (d) Brummel, C. L.; Vickerman, J. C.; Carr, S. A.; Hemling, M. E.; Roberts, G. D.; Johnson, W.; Weinstock, J.; Gaitanopoulos, D.; Benkovic, S. J.; Winograd, N. Evaluation of Mass Spectrometric Methods Applicable to the Direct Analysis of Non-Peptide Bead-Bound Combinatorial Libraries. *Anal. Chem.* **1996**, *68*, 237−242.

(11) (a) Winger, B. E.; Campana, J. E. Characterisation of Combinatorial Peptide Libraries by Electrospray Ionisation Fourier Transform Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1813. (b) Nawrocki, J. P.; Wigger, M.; Watson, C. H.; Hayes, T. W.; Senko, M. W.; Benner, S. A.; Eyler, J. R. Analysis of Combinatorial Libraries Using Electrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1860−1864. (c) Dunayevskiy, Y. M.; Vouros, P.; Wintner, E. A.; Shipps, G. W.; Carell, T.; Rebek, J., Jr. Application of Capillary Electrophoresis-Electrospray Ionisation Mass Spectrometry in the Determination of Molecular Diversity. *Proc. Nat. Acad. Sci. U.S.A.* **1996**, *93*, 6157.

(12) Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. Crystal Structure of HIV-1 Protease in Complex with VX-478, as a Potent and Orally Bioavailable Inhibitor of the Enzyme. *J. Am. Chem. Soc.* **1995**, *117*, 1181−1182.

(13) Kurth, M. J.; Ahlberg Randall, L. A.; Chen, C.; Melander, C.; Miller, R. B.; McAlister, K.; Reitz, G.; Kang, R.; Nakatsu, T.; Green, C. Library-Based Lead Compound Discovery: Antioxidants by Analogous Synthesis/Deconvolutive Assay Strategy. *J. Org. Chem.* **1994**, *59*, 5862−5864. In a library of 27 antioxidants derived from 3 carboxylic acids and 9 aldehydes, only 19 have unique nominal masses. The remainder fall into isobaric pairs (3a, 9b; 4a, 6b; 7a, 9c; 8b, 3c).

(14) Brown, R. D.; Martin, Y. C. Designing Combinatorial Library Mixtures Using a Genetic Algorithm. *J. Med. Chem.* **1997**, *40*, 2304−2313.

(15) (a) Hughes, I. Self-Coded Combinatorial Libraries. Poster presented at CHI 4th Annual "Exploiting Molecular Diversity", Coronado, CA 1997. (b) Hughes, I. WO 97/08190, 1997.

(16) Zuckermann, R. N.; Kerr, J. M.; Kent, S. B. H.; Moos, W. H. Efficient Method for the Preparation of Peptoids [Oligo(N-Substituted Glycines)] by Submonomer Solid-Phase Synthesis. *J. Am. Chem. Soc.* **1992**, *114*, 10646−10647.

(17) Hughes, I. Application of Polymer-Bound Phosphonium Salts as Traceless Supports for Solid-Phase Synthesis. *Tetrahedron Lett.* **1996**, *37*, 7595−7598.

(18) Bernard, M.; Ford, W. T. Wittig Reagents Bound to Cross-Linked Polystyrenes. *J. Org. Chem.* **1983**, *48*, 326−332.

JM9800954