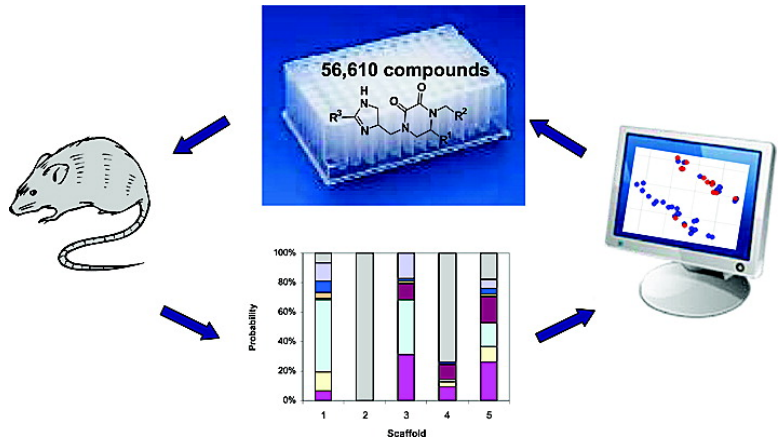
JOURNAL OF combinatorial CHEMISTRY

Accounts

Strategies for the Use of Mixture-Based Synthetic Combinatorial Libraries: Scaffold Ranking, Direct Testing In Vivo, and Enhanced Deconvolution by Computational Methods

Richard A. Houghten, Clemencia Pinilla, Marc A. Giulianotti, Jon R. Appel, Colette T. Dooley, Adel Nefzi, John M. Ostresh, Yongping Yu, Gerald M. Maggiora, Jose L. Medina-Franco, Daniela Brunner, and Jeff Schneider

J. Comb. Chem., **2008**, 10 (1), 3-19• DOI: 10.1021/cc7001205 • Publication Date (Web): 08 December 2007



More About This Article

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

- Supporting Information
- Links to the 3 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

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

Accounts

Strategies for the Use of Mixture-Based Synthetic Combinatorial Libraries: Scaffold Ranking, Direct Testing In Vivo, and Enhanced Deconvolution by Computational Methods

Richard A. Houghten,*,† Clemencia Pinilla,† Marc A. Giulianotti,† Jon R. Appel,† Colette T. Dooley,† Adel Nefzi,† John M. Ostresh,† Yongping Yu,⊥† Gerald M. Maggiora,‡ Jose L. Medina-Franco,‡ Daniela Brunner,§ and Jeff Schneider§,||

Torrey Pines Institute for Molecular Studies, 3550 General Atomics Court, San Diego, California 92121, BIO5 Institute and College of Pharmacy, University of Arizona, Tucson, Arizona 85721, PsychoGenics, Inc., 765 Old Saw Mill River Road, Tarrytown, New York 10591, and School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

Received July 19, 2007

Since its inception more than 20 years ago with high-throughput parallel synthesis for oligonucleotides and peptides,[1–3] synthetic combinatorial methods have fundamentally advanced the ability to synthesize and screen large numbers of compounds because of improvements made in technology, instrumentation, and library design strategies. This discipline was readily accepted initially and is now an embedded component of the drug discovery process worldwide. While there are a range of combinatorial approaches, the use of mixture-based libraries made up of tens of thousands to billions of compounds is the approach that enables the most rapid and economical acquisition of chemical and biological information. Mixture-based libraries represent powerful tools that can be used for the identification of active individual compounds for a wide range of important targets, as reviewed.[4] In the past decade, such approaches have been expanded to include the synthesis of low molecular weight acyclic and heterocyclic compounds.[5–8] As with most innovations, synthetic combinatorial methods developed for the synthesis and screening of mixture-based libraries were slow to gain acceptance because of the conceptual distance between these approaches and the traditional methods previously used in the pharmaceutical industry. This was, and is, especially true for mixture-based libraries composed of tens of thousands to billions of different compounds, but such methods are now being used by an increasing number of groups for the identification of highly active, novel compounds in research and drug discovery programs.[9,10]

Mixture-based libraries are systematically arranged mixtures of synthetic compounds having both defined and mixture positions of diversity. This permits information to be gathered regarding both the activity and importance of every functionality at each position of the library.[11,12] Post synthetic chemical modification of such existing mixture-based libraries using the "libraries from libraries" approach now enables the ever-increasing generation of low molecular weight compounds.[13] Thus, for the last 16 years, we have successfully used this approach for the design and the generation of a range of peptidomimetic and small molecule libraries from resin-bound polyamides.[7] We have also used this approach combining solid- and solution-phase synthesis methods for the synthesis of a nitrosamine library[14] and a platinum tetraamine coordination complex library.[15]

The power of synthetic mixture-based combinatorial libraries lies in their ability to accelerate the acquisition of information regarding specific functionalities at each variable position in the library that determines the activity of a specific chemical scaffold or pharmacophore. Another advantage of mixture-based libraries resides in the very high densities of compounds that can be synthesized in narrow areas of chemical space. When compared to existing high-throughput screening (HTS) programs, in which tens of thousands of individual compounds are screened against therapeutically important targets, millions of compounds formatted as mixtures can be examined using substantially less material and at much lower time/labor economics than if these same mixture-based diversities were made and screened as individual compounds. This unique combinatorial library approach can be applied to virtually any existing bioassay for the identification of novel ligands. For example, a novel, highly active tetrapeptide agonist for the κ-opioid receptor was identified from a positional scanning library of 6.25

* To whom correspondence should be addressed. Phone: 858-455-3805. Fax: 858-455-3804. E-mail: rhoughten@tpims.org.
† Torrey Pines Institute for Molecular Studies.
‡ University of Arizona.
§ PsychoGenics, Inc.
|| Carnegie Mellon University.
⊥ Current address: College of Pharmaceutical Science, Zijin Campus, Zhejiang University, Hangzhou 310058, P. R. China.

10.1021/cc7001205 CCC: $40.75
Published on Web 12/08/2007

Table 1. Selected Examples from Recent Studies Using Mixture-Based Libraries

receptors/other targets

compound class	target	ref
polyphenyl urea	X-linked inhibitor apoptosis protein (XIAP)	36
pyrrolidine bis-cyclic guandine	methicillin-resistant *S. aureus*	42
isatin-B-thiosemicarbazones	vaccinia virus-infected cells	63
polyamines	μ-opioid receptor	64
N-alkyl glycine (peptoid)	MDR P-gp	20, 21
N-alkyl glycine (peptoid)	lipopolysaccharide	22
N-alkyl glycine (peptoid)	*S. areus* MR10, *P. aeruginosa*	23
N-alkyl glycine (peptoid)	NMDA receptor vanilloid receptor subunit 1	24–26
dimeric iminodiacetic acid diamides	EPO receptor	19
rhodamine tetrapeptide	κ-opioid receptor	35
hexapeptide	DNA recombination	65
nonapeptide	*C. albicans*	66
nonapeptide	MHC, CD4 T cells, CD8 T cells	67–70
decapeptide	HIV CD4 T cells, Her-2/neu CD8 T, autoimmune CD4 T cells	71–74

enzymes

compound class	target	ref
indinavir	HIV protease	75
peptidomimetic	AICAR Tfase	18
tetrapeptide	cysteine proteases and proteases of parasitic origin	76
tetrapeptide	human prostatin, serine, cysteine, and threonine protease, thrombin, caspase 3	27, 28
tetrapeptide	human hepsin trypsin-like serine protease	29
tetrapeptide	granzyme M	30
tetrapeptide	protease KLK4 human kallikrein	31
tetra- and octapeptide	NSZB/NS3 Dengue virus	32
hexapeptide	angiotensin I converting enzyme, human cathepsin B	33, 34
hexapeptide	fibrinogen, vitronectin	77
hexapeptide	prohormone convertase 5/7, human furin	78, 79
hexapeptide	topoisomerase	80
undecapeptide	protein kinase C	81

million tetrapeptides.[16] This highly selective, all D-amino acid tetrapeptide (H-Phe-Phe-Nle-Arg-NH_2, $K_i = 1.2$ nM), which is structurally unrelated to any endogenous κ-opioid receptor-selective peptides or small molecule could not have been predicted to have activity nor have been identified using traditional drug discovery or computational methods. Minor modifications of this tetrapeptide has led to the development of an all D-amino acid analog (H-Phe-Phe-Nle-Arg-4-picolyl amide) having improved affinity and selectivity to the kappa receptor, and a 100-fold increase in peripheral selectivity compared to asimadoline[17] (manuscript in preparation). These compounds are now in phase II human clinical trials for neuropathic chronic pain.

Mixture-based libraries continue to find favor with researchers who have assays that are not suitable to HTS methods, not currently operable in HTS mode, or in which target reagents are limited by availability or cost. We have also determined that mixtures composed of very large numbers of compounds can be tested directly using in vivo assays yielding clear and reproducible results. The emphasis of this Account will be on recent advances in the use of mixture-based combinatorial libraries. These include scaffold-ranking strategies (which enable the rapid identification of active library scaffolds), the examination of computational analysis for the enhanced deconvolution of heterocyclic positional scanning libraries, and the direct in vivo screening of mixture-based libraries. We begin with a review of recent studies using mixture-based (primarily positional scanning) libraries.

Mixture-based libraries composed of peptides, peptidomimetics, and small molecules have been used in a wide range of in vitro bioassays for the identification of potential useful hits, lead compounds, or both. This was surveyed in earlier reviews.[4,9] Table 1 summarizes a number of studies published since 2002 using mixture-based libraries against receptors, enzymes, and other targets for the successful identification of novel compounds. Boger and co-workers recently published two studies using mixture-based libraries synthesized by solution-phase methods for the identification of erythropoietin mimetics and novel ligands that modulate protein–protein and protein−DNA interactions.[18,19] A number of studies by Perez-Paya and co-workers have used positional scanning libraries of peptoids for the identification of novel chemosensitizers and antibacterial compounds.[20–26] Several different groups have developed and used fluorescently labeled positional scanning libraries to probe enzyme specificity for a number of key proteases.[27–34] Recent work by Houghten and co-workers has involved the synthesis of a tetrapeptide positional scanning library, in which each peptide of the library was fluorescently labeled with rhodamine. The library was screened in a κ-receptor binding assay and directly yielded high affinity κ-specific ligands with intrinsic fluorescent properties.[35] Small molecule antagonists of XIAP-caspase inhibition (apoptosis suppression) that exhibited broad antitumor activity have also been identified from a positional scanning library composed of more than 85 000 polyphenyl ureas.[36] The mechanism of action[37] and efficacy of these compounds have been confirmed in an

increasing number of solid tumor [38,39] and hematologic [40,41] malignancies.

Scaffold Ranking Strategy for the Rapid Identification of Active Libraries

Mixture-based libraries, when arranged in a positional scanning format,[12] have been shown to provide extensive structure–activity information at each variable position in a given central scaffold in a wide range of assays. Such results are inherent with this approach because positional scanning libraries are composed of systematically arranged, very dense mixtures of compounds having both defined and mixture positions for a given scaffold. Thus, information is rapidly obtained regarding the activity of the functionalities for each varied position for a given scaffold. A simple conceptual description of a positional scanning synthetic combinatorial library (PS-SCL) is helpful. Thus, using a simple tripeptide PS-SCL composed, in total, of 27 tripeptides (3 amino acids at each position, $3^3 = 27$), this tripeptide PS-SCL is composed of three sublibraries (OXX, XOX, and XXO) and a total of nine separate mixtures. The defined positions (O) contain one of the three amino acids, and the mixture positions (X) contain a mixture of all three amino acids. It is important to note that each of the three sublibraries varies only in the location of the defined amino acid, but each one contains the same 27 peptides with only their defined positions varied. In this example, if the tripeptide RAT was the only active peptide, then this individual peptide (RAT) would be responsible for all the activity found for the active sublibrary mixtures RXX, XAX, and XXT. Thus, the active compound, in this case RAT, is identified by synthesis of the combination of the defined amino acids identified as the most active mixtures at each position. The positional scanning concept can be used to prepare libraries of any length of peptide or compound class. Thus, while the illustration above involving 3 amino acids at each of three position (resulting in 27 tripeptides) would require only 9 mixtures to be screened, a hexapeptide PS-SCL using 20 amino acids at each position (resulting in a library composed of 64 million hexapeptides) would require only 120 mixtures to be screened, and a decapeptide library using 20 amino acids at each position (resulting in over 10 trillion decapeptides) would require only 200 mixtures to be screened.

Over the past 16 years, we have synthesized numerous positional scanning libraries, covering a collection of approximately 7 500 000 small molecules and well over 10 trillion peptides.[7,9] Extensive optimization of reaction conditions are a necessary component of mixture library generation. This occurs before a library is synthesized to ensure a broad range of functional group incorporation and ensures that these functionalities withstand the synthesis conditions used. This step is used to determine the breadth of the reaction conditions for the functionalities tested by synthesizing individual controls using as wide a range of building blocks as is practical. These controls are prepared systematically, using a method in which one position was varied with each of the available building blocks, while the other reactive positions of a center scaffold molecule are fixed with a single building block. Each compound is assessed by LC-MS for the identity and purity, and those compounds that give yields and purities greater than 80% are included in the synthesis of the library. The reproducibility of the synthetic chemistry is the key to the successful generation and use of mixture based libraries.[4] The advantage of positional scanning synthetic libraries, each containing thousands to millions to even billions of compounds is that, when assembled as mixtures, each library can be screened using a much smaller set of samples, in most cases 100–200, relative to the total number of compounds making up the library. This has enabled us and others to use mixture-based libraries in a wide variety of assays, ranging from simple receptor-binding assays and cell-based assays to relatively low-throughput gel electrophoresis and now, as described below, directly in vivo animal models. This allows biologists to concentrate on designing assays that will provide the most relevant data rather than being forced to format virtually all assays into a classic high-throughput screening format. In addition, target-based assays, by their very nature, preclude that which is not known both in terms of multiple receptor interactions and or phenotypic activity.

Thus, of the more than 50 low molecular weight positional scanning libraries that have been prepared, which represent approximately 7 500 000 acyclic and heterocyclic compounds, the use of the mixture-based positional scanning format only requires one to screen ~6000 mixtures. A number of these small molecule libraries have been prepared using the "libraries from libraries approach",[7,13] in which different libraries are prepared using the same building blocks synthesized around differing chemical scaffolds. This is now often termed "diversity-oriented synthesis". The question inevitably arises as to which libraries are more likely to lead to the identification of active compounds or which libraries should be screened first. This is especially pertinent when there is no known molecule for a target of interest. One option is to screen all of the mixtures making up the positional scanning libraries (for example 6000 samples) and, on the basis of these screening results, decide which libraries will be deconvoluted to identify the active individual compounds. While 6000 samples even as mixtures is a small number compared to the hundreds of thousands to millions of compounds in typical high-throughput screening platforms, it can still be severely limiting, especially in the many assays available in academic, nonprofit research organizations or smaller biotechnology firms.

In an effort to further increase efficiency and utility as our collection of libraries expands, we have developed a new strategy termed "scaffold ranking" for the rapid identification and ranking of active library scaffolds. To illustrate this concept, 19 different mixture-based libraries, representing a total of approximately 4 million low molecular weight acyclic and heterocyclic compounds, can be rank ordered in terms of the activity of their scaffold following the screening of only 19 mixtures. Each of these scaffold mixtures are made up of differing small molecule libraries, in which each mixture is simply the entire library from each of the separate sublibraries of these respective 19 differing positional scanning libraries. Unlike positional scanning libraries that contain at least one defined position and several mixture

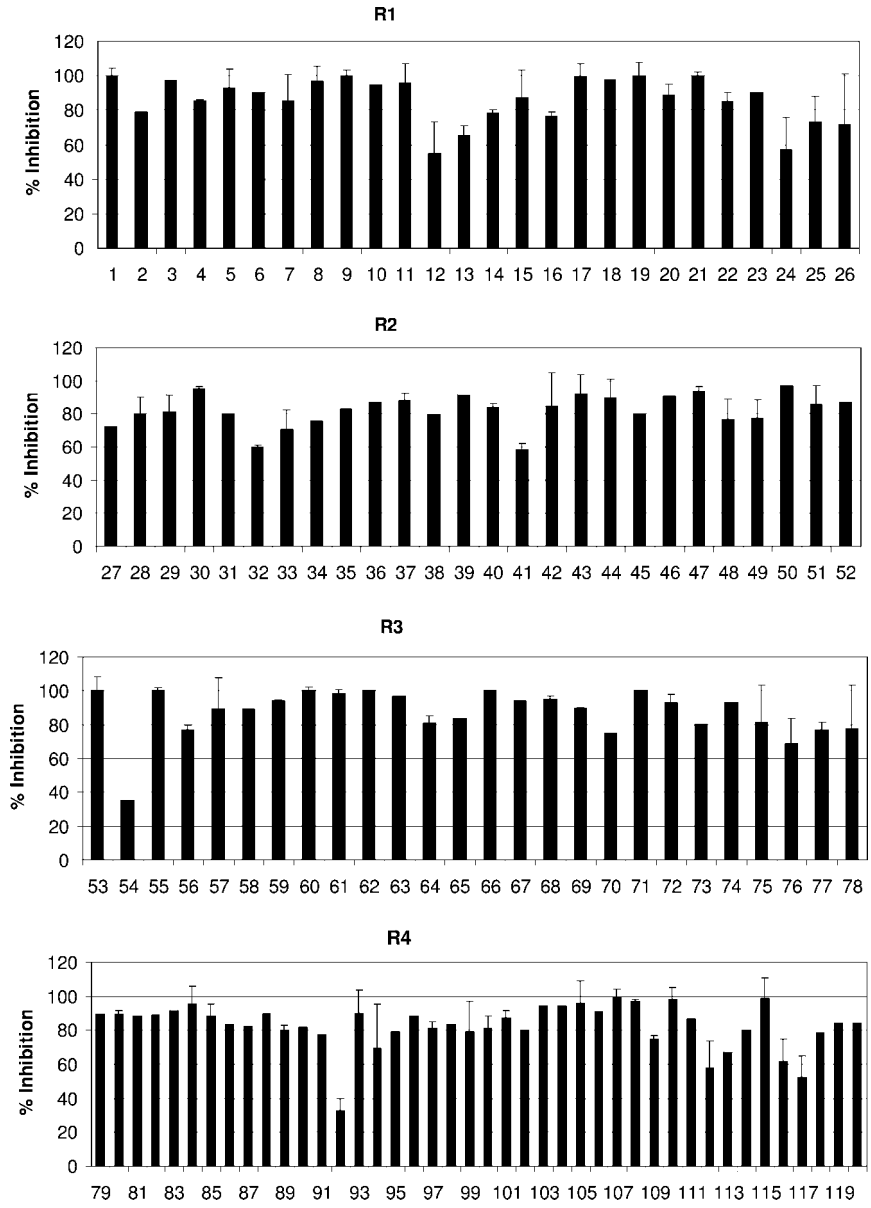


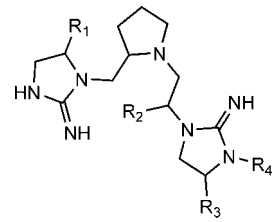
Figure 1. Screening profile of 1346 pyrrolidine bis-cyclic guanidine positional scanning library in a μ-opioid radioreceptor binding assay. This library has four positions of diversity (R1, R2, R3, R4) and is represented by four positional sublibraries, each differing in the location of the defined position with the other three positions as mixtures. See Table 2 for more details. Each bar represents the percent inhibition of a mixture displacing radiolabeled DAMGO binding to the μ-opioid receptor. The mixtures making up the library were screened in duplicate at 0.1 mg/mL.

positions, the scaffold mixtures have no defined positions. The results obtained following the testing of the scaffold mixtures allows one to rank order the libraries based solely on the overall activity of the various scaffolds. The most active scaffolds can then be screened as complete positional scanning libraries from the most active to the least active as time and resources permit for rapid lead identification of individual compounds.

To illustrate and validate the scaffold ranking strategy, we compared the activities obtained upon testing the individual mixtures of a complete positional scanning library relative to the activities of this same library as a single mixture having no defined positions. Figure 1 shows a positional scanning library (library 1346) with a pyrrolidine bis-cyclic guanidine scaffold with four variable positions of diversity and composed of four sublibraries each having one defined position and three mixture positions (OXXX, XOXX, XXOX, XXXO). In this example, sublibrary 1 of 1346 (R1 defined, represented as O) is made up of 26 separate mixtures each containing 28 392 compounds. Sublibraries 2 (R2 defined) and 3 (R3 defined) are similar to sublibrary 1 in that they are made up of 26 mixtures each containing 28 392 compounds. Sublibrary 4 (R4 defined) contains 42 mixtures each containing 17 576 compounds. Table 2 shows the building blocks used to synthesize this library and the resulting functionalities at each of the four positions using the synthetic methodology previously described.[42] It should be noted that each of the four sublibraries in total varies solely in the location of the defined functionality but contains exactly the same number and type of compounds (each having a total of 738 192 compounds).

Each of the 120 mixtures making up this positional scanning library were tested for their ability to inhibit radiolabeled [^{3}H-D-Ala^2,MePhe^4,Gly ol] enkephalin (DAMGO)

Table 2. Building Blocks Used to Synthesize 1346 Pyrrolidine Bis-cyclic Guanidine PS-SCL and Resulting Functionalities at the Four Positions of Diversity[a]



number				building block	functionality
1	**27**	**53**		Boc-L-Ala	*S*-methyl
2	**28**	**54**		Boc-L-Phe	*S*-benzyl
3	**29**	**55**		Boc-Gly	hydrogen
4	**30**	**56**		Boc-L-Ile	*S*-2-butyl
5	**31**	**57**		Boc-L-Leu	*S*-isobutyl
6	**32**	**58**		Boc-L-Ser(Bzl)	*R*-hydroxymethyl
7	**33**	**59**		Boc-L-Thr(Bzl)	(*R*,*R*)-1-hydroxyethyl
8	**34**	**60**		Boc-L-Val	*S*-isopropyl
9	**35**	**61**		Boc-L-Tyr(BrZ)	*S*-4-hydroxybenzyl
10	**36**	**62**		Boc-D-Ala	*R*-methyl
11	**37**	**63**		Boc-D-Phe	*R*-benzyl
12	**38**	**64**		Boc-D-Ile	*R*-2-butyl
13	**39**	**65**		Boc-D-Leu	*R*-isobutyl
14	**40**	**66**		Boc-D-Ser(Bzl)	*S*-hydroxymethyl
15	**41**	**67**		Boc-D-Thr(Bzl)	(*S*,*S*)-1-hydroxyethyl
16	**42**	**68**		Boc-D-Val	*R*-isopropyl
17	**43**	**69**		Boc-D-Tyr(BrZ)	*R*-4-hydroxybenzyl
18	**44**	**70**		Boc-L-phenylglycine	*S*-phenyl
19	**45**	**71**		Boc-L-norvaline	*S*-propyl
20	**46**	**72**		Boc-D-norvaline	*R*-propyl
21	**47**	**73**		Boc-L-norleucine	*S*-butyl
22	**48**	**74**		Boc-D-norleucine	*R*-butyl
23	**49**	**75**		Boc-L-naphthylalanine	*S*-2-naphthylmethyl
24	**50**	**76**		Boc-D-naphthylalanine	*R*-2-naphthylmethyl
25	**51**	**77**		Boc-L-cyclohexylalanine	*S*-cyclohexyl
26	**52**	**78**		Boc-D-cyclohexylalanine	*R*-cyclohexyl
			79	1-phenyl-1-cyclopropanecarboxylic acid	(1-phenyl-cyclopropyl)-methyl
			80	2-phenylbutyric acid	2-phenylbutyl
			81	3-phenylbutyric acid	3-phenylbutyl
			82	*m*-tolylacetic acid	*m*-tolylethyl
			83	3-fluorophenylacetic acid	2-(3-fluoro-phenyl)-ethyl
			84	3-bromophenylacetic acid	2-(3-bromo-phenyl)-ethyl
			85	(α-α-α-trifluoro-*m*-tolyl) acetic acid	2-(3-trifluoromethyl-phenyl)-ethyl
			86	*p*-tolylacetic acid	*p*-tolylethyl
			87	4-fluorophenylacetic acid	2-(4-fluoro-phenyl)-ethyl
			88	3-methoxyphenylacetic acid	2-(3-methoxy-phenyl)-ethyl
			89	4-bromophenylacetic Acid	2-(4-bromo-phenyl)-ethyl
			90	4-methoxyphenylacetic acid	2-(4-methoxy-phenyl)-ethyl
			91	4-ethoxyphenylacetic acid	2-(4-ethoxy-phenyl)-ethyl
			92	4-isobutyl-α-methylphenylacetic acid	2-(4-isobutyl-phenyl)-propyl
			93	3,4-dichlorophenylacetic acid	3,4-dichlorophenethyl
			94	3,5-bis(trifluoromethyl)-phenylacetic acid	2-(3,5-bis-trifluoromethyl-phenyl)-ethyl
			95	3-(3,4-dimethoxyphenyl)-propionic acid	3-(3,4-dimethoxy-phenyl)-propyl
			96	phenylacetic acid	phenethyl
			97	3,4,5-trimethoxybenzoic acid	3,4,5-trimethoxy-benzyl
			98	butyric acid	butyl
			99	heptanoic acid	heptyl
			100	isobutyric acid	isobutyl
			101	2-methylbutyric acid	2-methylbutyl
			102	isovaleric acid	3-methylbutyl
			103	3-methylvaleric acid	3-methylpentyl
			104	4-methylvaleric acid	4-methylpentyl
			105	*p*-toluic acid	4-methyl-benzyl
			106	cyclopentanecarboxylic acid	cyclopently methyl
			107	cyclohexanecarboxylic acid	cyclohexyl-methyl
			108	cyclohexylacetic acid	cyclohexyl-ethyl
			109	cyclohexanebutyric acid	cyclohexyl-butyl
			110	cycloheptanecarboxylic acid	cycloheptyl-methyl
			111	2-methylcyclopropanecarboxylic acid	(2-methyl-cyclopropyl)-methyl
			112	cyclobutanecarboxylic acid	cyclobutyl-methyl
			113	3-cyclopentylpropionic acid	3-cyclopentyl-propyl
			114	cyclohexanepropionic acid	cyclohexyl-propyl
			115	4-methyl-1-cyclohexanecarboxylic acid	4-methyl-1-cyclohexyl-methyl
			116	4-*tert*-butyl-cyclohexanecarboxylic acid	4-*tert*-butyl-cyclohexyl-methyl
			117	4-biphenylacetic acid	2-biphenyl-4-yl-ethyl
			118	1-adamantanecarboxylic acid	adamantan-1-yl-methyl
			119	1-adamantaneacetic acid	2-adamantan-1-yl-ethyl
			120	2-norbornaneacetic acid	2-bicyclo[2.2.1]hept-2-yl-ethyl

[a] Sublibrary 1: R1 defined for samples **1–26** (28 392 compounds each). Sublibrary 2: R2 defined for samples **27–52** (28 392 compounds each). Sublibrary 3: R3 defined for samples **53–78** (28 392 compounds each). Sublibrary 4: R4 defined for samples **79–120** (17 576 compounds each).

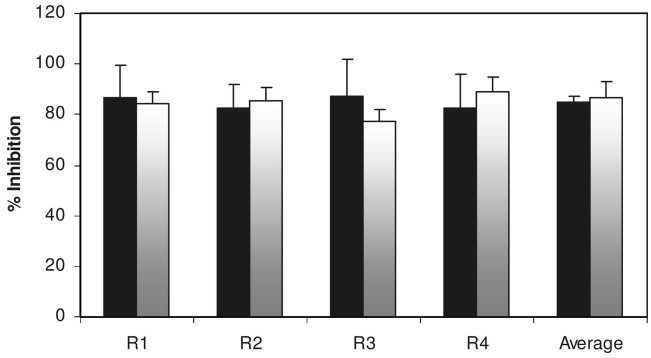


Figure 2. Comparison of library and scaffold activities for the pyrroldine bis-cyclic guanidine library in μ-opioid receptor. Library (black bars) activity is average percent inhibition for each diversity position (R1, R2, R3, and R4) from Figure 1. Scaffold (gray bars) activity is percent inhibition of each diversity position mixture that was prepared by physically combining the mixtures making up each diversity position and then testing them in the μ-opioid radioreceptor binding assay.

in a μ-opioid receptor-binding assay (Figure 1) as previously described.[16] The profile clearly shows the range of inhibition found for the set of mixtures (17 576–28 392 compounds/mixture) at each of the four diversity positions. When the resulting data is averaged for each position, one obtains a single "library" average activity for each diversity position (R1−R4) making up the library (738 192 compounds), which is shown as black bars in Figure 2. In addition, four mixtures with no defined positions were prepared by physical combination of the individual mixtures making up each of the four positions, namely, the 26 mixtures at positions 1, 2, and 3 and the 42 mixtures at position 4. These four "scaffold" mixtures, which theoretically are identical in number and type of compounds, were also tested in the μ-opioid receptor binding assay (shown as gray bars in Figure 2). The results obtained were in good agreement with the activity obtained by calculating the library average activity when screened as a positional scanning library. These results clearly indicate that activity can be obtained with more than a 10-fold increase in mixture complexity, and that a library of 738 192 compounds can yield activity that coincides well with that expected when compared to the average of the single position defined mixtures of this positional scanning library.

The scaffold ranking strategy is demonstrated here using some of our collection of libraries from libraries.[7,13] In this example, the scaffold mixtures were prepared using 19 different small molecule positional scanning libraries representing four different diversity families and more than 4 million low molecular weight compounds in total (Figure 3). Each diversity family included libraries containing the same number of diversity positions, identical side chain functionalities at each diversity position, and the same number of compounds within the library. Libraries within each diversity family differed only in the chemical nature of the central scaffold. As described conceptually above, each scaffold mixture is prepared by simply combining the mixtures from one of the defined positions from the positional scanning library to create a single "all X" mixture. The 19 scaffold mixtures were screened in the μ-opioid receptor binding assay, and these results are compared with the overall average percent inhibition of each library in the same assay (Figure 4). It is clear that the data for each library are in good agreement obtained by either the library or scaffold method. The scaffolds within each diversity family are ranked from most active to least active. In three of the four different diversity families, one can clearly find scaffolds that are more active than others within the same family. Thus, scaffold ranking enables one to rank order the scaffolds examined and then in turn enable prioritization based on the differing activities seen between the scaffolds. The most-active scaffolds from each diversity family for this assay, namely, 1169, 1421, and 1346, would logically thus be the first to be screened as complete positional scanning libraries. Clearly, the remaining libraries could be screened at a later date as time and resources allow. This approach would not preclude one of the lesser active mixtures from having active compounds; it simply allows one to rank the activities of the scaffolds ("triage") to facilitate future screening and effectively use available time and resources.

Table 3 illustrates the reduction of testing effort for the diversity found in the mixture-based scaffold ranking strategy. In this example, the μ-opioid receptor activity information was obtained from testing 19 different scaffold mixtures containing over 4 million small molecules, leading one to pursue several active libraries in complete positional scanning format, compared to testing over 2000 mixtures. The time and cost savings in screening efforts of mixtures versus individual compounds is clearly enormous, and the fact that the activities can be determined from mixtures of hundreds of thousands of compounds demonstrate the validity of testing mixtures. A demonstration of using this approach for low-throughput assays is presented later in this Account.

To complete the process for the identification of individual compounds from a positional scanning library, the pyrrolidine bis-cyclic guanidine library 1346 was chosen from the scaffold ranking strategy described above. The deconvolution of positional scanning libraries consists of the preparation of individual compounds derived from the combinations of the defined functionalities of the most active mixtures at each position of the library. In this example, upon screening the 1346 library at a fixed concentration in the μ-opioid receptor assay (Figure 1), 41 mixtures were found to inhibit >90% of the bound radioligand and were tested again in a dose–response manner to select the most-active mixtures based on IC_{50} values. For the 41 mixtures tested, IC_{50} values ranged from 200 to 3000 nM using an average molecular weight (MW) of 645.

For the deconvolution of the pyrrolidine bis-cyclic guanidine library, selection of the most-active functionalities was based on activity and differences in chemical character while also minimizing the number of individual compounds to be synthesized. In the R1 position, three of the four functionalities with IC_{50} values of less than 500 nM were chosen, namely, *S*-methyl (mixture **1**), *S*-4-hydroxybenzyl (**9**), and *R*-methyl (**10**). Hydrogen (**3**) was not chosen because of its similarity with the two methyl functionalities. In the R2 position, the three functionalities with IC_{50} values of less than 500 nM were selected, namely, *R*-2-naphthylalanine (**50**), *R*-propyl (**46**), and *R*-4-hydroxybenzyl (**43**) In the R3 position, the three most active of the eight functionalities

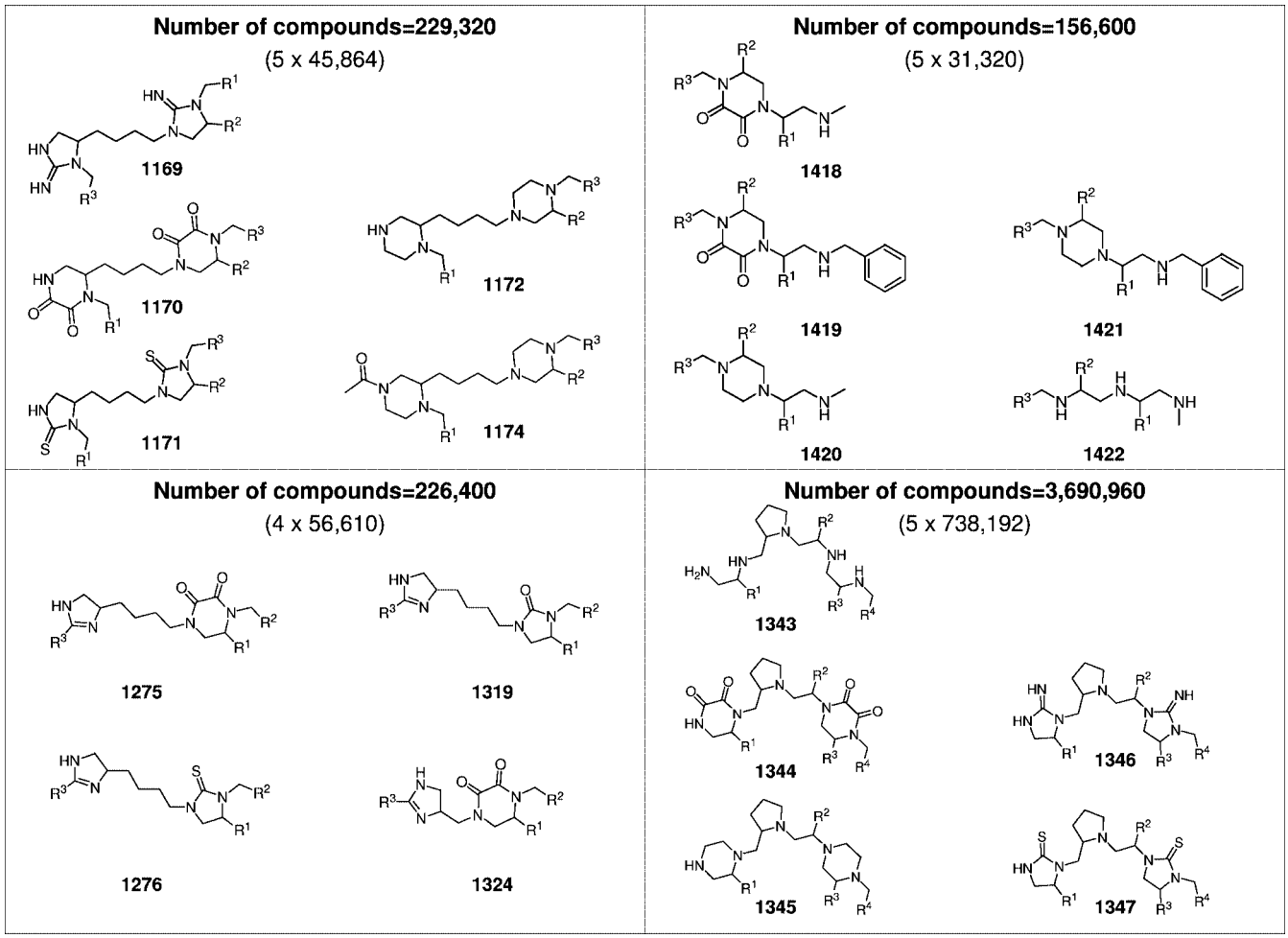


Figure 3. Positional scanning libraries used for scaffold ranking. Libraries are grouped into four different diversity families. The number of compounds in each library for a given family is the same, and the total number of compounds for each family is shown. Top left panel: Each library is made up of 45 864 compounds (R1 = 42 mixtures, R2 = 26 mixtures, and R3 = 42 mixtures). The libraries are 1169 bis-cyclic guanidine, 1170 bis-cyclic diketopiperazine, 1171 bis-cyclic thiourea, 1172 bis-cyclic piperazine, and 1174 N-acylated bis-piperazine. Top right panel: Each library is made up of 31 320 compounds (R1 = 29 mixtures, R2 = 27 mixtures, and R3=40 mixtures). The libraries are 1418 *N*-methyl-1,4,5-trisubstituted-2,3-diketopiperazine, 1419 *N*-benzyl-1,4,5-trisubstituted-2,3-diketopiperazine, 1420 *N*-benzyl-1,4,5-trisubstituted piperazine, 1421 *N*-methyl-1,4,5-trisubstituted piperazine, and 1422 *N*-methyl triamine. Bottom left panel: Each library is made up of 56 610 compounds (R1 = 43 mixtures, R2 = 37 mixtures, R3 = 45 mixtures). The libraries are 1275 dihydroimidazolyl-butyl-diketopiperazine, 1276 dihydroimidazolyl-butyl-cyclic thiourea, 1319 dihydroimidazolyl-butyl-cyclic urea, and 1324 dihydroimidazolyl-methyl-diketopiperazine. Bottom right panel: Each library is made up of 738 192 compounds (R1, R2, and R3 = 26 mixtures and R4 = 42 mixtures). The libraries are 1343 pyrrolidine pentamine, 1344 pyrrolidine bis-diketopiperazine, 1345 pyrrolidine bis-piperazine, 1346 pyrrolidine bis-cyclic guanidine, 1347 pyrrolidine bis-cyclic thiourea.

with IC_{50} values of less than 500 nM were selected, namely, *R*-methyl (**62**), *S*-methyl (**53**), and *S*-hydroxymethyl (**66**). The remaining five functionalities at this position were chemically similar to the three selected, for example, *R*-butyl (**74**) and *R*-propyl (**72**), and were not considered for deconvolution. In the final R4 positions, two functionalities, namely, 4-methyl-1-cyclohexyl-methyl (**115**) and cyclohexyl-ethyl (**108**), were selected from five chemically similar functionalities with IC_{50} values of less than 500 nM. The combination of the selected functionalities led to 72 ($3 \times 3 \times 4 \times 2 = 72$) individual pyrrolidine bis-cyclic guanidines that were synthesized and tested. The activities of the 72 individual pyrrolidine bis-cyclic guanidines range from active compounds (<100 nM) to less-active compounds (>15 μM).

The nature of positional scanning library deconvolution leads to a group of individual compounds that are inherently analogs of each other, arranged in groups having common functionalities at any of the four diversity positions. This inherent feature of positional scanning library deconvolution also provides preliminary structure–activity relationship (SAR) information. Here, the two most-active compounds differ only in the stereochemistry in the methyl group at the R1 position (1561-2 and 1561-50, K_i = 80–90 nM, Table 4). In addition, the single -substitution analogs of 1561-2 and another compound (1561-38), which were present in the set of 72 compounds identified from the library, are shown in Table 4. For compound 1561-2 one can conclude that the *S*-methyl group at R1 can be replaced with its stereoisomer equivalent while retaining activity, but its activity is lost upon substitution by *S*-4-hydroxybenzyl. Little difference in activity can be found upon substitution of the R4 position (2-(3-bromo-phenyl) ethyl to 4-methyl-1-cyclohexyl-methyl); however, the same substitution at the R4 position of 1561-38 (K_i = 132nM) results in a complete loss of activity. These are just a few illustrative examples of SAR information one

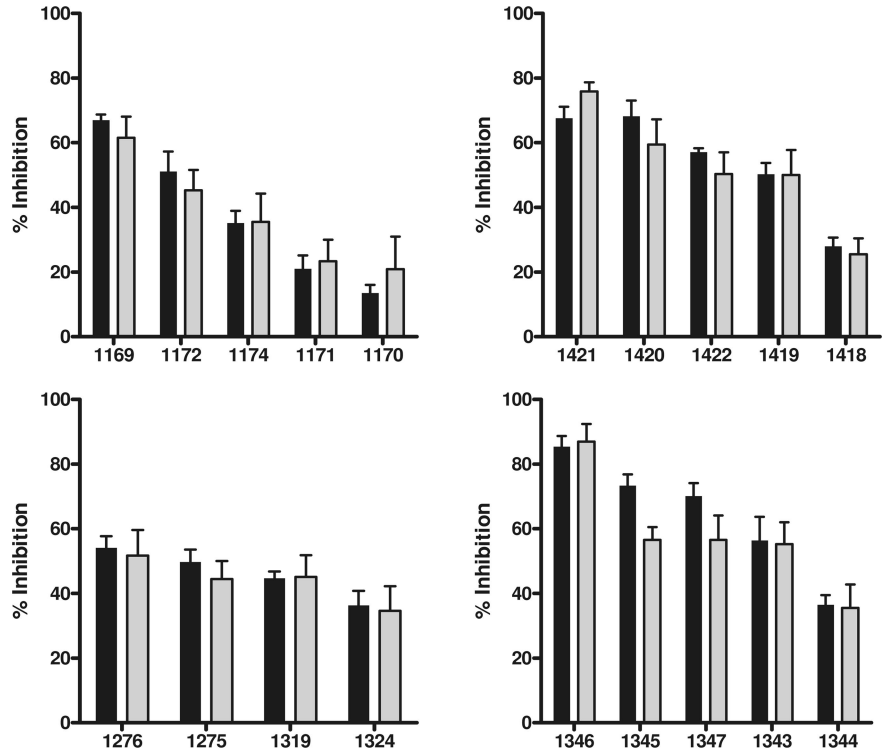


Figure 4. Comparison of library and scaffold μ-opioid activities for the 19 small molecule mixture-based libraries. Libraries are grouped into four diversity families, in which each library of a given family shares the same number of diversity positions, identical building block side chain functionalities, and the same number of compounds within each library. Library (black bars) activity is overall average percent inhibition for each library. Scaffold (gray bars) activity is overall average percent inhibition of each library that was prepared by physically combining mixtures making up each diversity position and then tested in the μ-opioid radioreceptor binding assay. Each bar represents the percent inhibition of a mixture displacing radiolabeled DAMGO binding to the μ-opioid receptor. Library was screened in duplicate at 0.1 mg/mL.

Table 3. Mixture-Based Libraries in Various Formats Reduce the Number of Samples

format	samples to test	no. of 96-well plates	reduction
total diversity	4 303 320	48 900	
PS−SCL	2094	24	2000-fold
R1 samples	621	8	7000-fold
scaffold ranking	19	1	70 000-fold

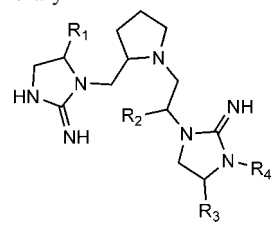
can readily obtain from the deconvolution of positional scanning libraries.

In Vivo Testing of Mixtures for Drug Discovery

In vivo testing of mixture-based chemically diverse libraries is now being explored in this laboratory as a more direct and value-added approach to drug discovery. Philosophically, this approach to drug discovery is analogous to the Chinese leader Deng Xiaoping's quote regarding a country's economic success: "It doesn't matter whether a cat is black or white, so long as it catches mice".[43] While the direct use of mixtures may seem counterintuitive to some, if this approach can identify therapeutic candidates, then its value in drug discovery is clear. If successful, as anticipated by our preliminary studies, this will result in significant savings of both time and resources compared to the traditional target-based drug discovery process. We expect that the in vivo discovery of active compounds will inherently yield more advanced therapeutic candidates. This is especially true when one notes that a large number of highly active compounds identified while using in vitro assays fail upon initial in vivo testing. The approach can be likened to the original identification of active compounds, the serendipitous discovery of active natural products by the ingestion of plants by humans. Natural products are also typically made up of hundreds to thousands of compounds with both varying structures and concentrations.

Previous studies of pharmokinetic in vivo screening of mixtures has proven to be beneficial for lead identification and optimization.[44,45] Preliminary experiments have been conducted that serve as proof of concept studies to validate the use of mixture-based libraries in direct in vivo models of pain modulation. The mouse tail flick assay was used to assess the in vivo antinociceptive effects of mixtures known to contain active compounds. This assay was used because it is well established, yields clear and reproducible end points, is highly representative of μ-opioid receptor activity and translates well to human studies. A positively biased mixture of 125 000 tetrapeptides (50 different amino acids, X, at three positions, 50^3) was synthesized in which 2,6-dimethyltyrosine (Dmt) was fixed at the N-terminal position (Dmt-XXX). This mixture contains the known active dermorphin-derived tetrapeptide analog H-Dmt-D-Arg-Phe-Lys-NH_2.[46] This tetrapeptide has high μ-opioid receptor affinity[47] with proven

Table 4. μ-Opioid Receptor Binding Activities for Single-Substitution Analogs of Two Active Compounds (1561-2 and 1561-38) Identified from the Pyrrolidine Bis-cyclic Guanidine Library



1561-	R1	R2	R3	R4	K_i (nM)	SEM
2	*S*-methyl	*R*-2-napthylmethyl	*R*-methyl	2-(3-bromo-phenyl)-ethyl	79.3	20.2
26	*S*-4-hdroxybenzyl	*R*-2-napthylmethyl	*R*-methyl	2-(3-bromo-phenyl)-ethyl	>1000	
50	*R*-methyl	*R*-2-napthylmethyl	*R*-methyl	2-(3-bromo-phenyl)-ethyl	89.9	5.2
10	*S*-methyl	*R*-butyl	*R*-methyl	2-(3-bromo-phenyl)-ethyl	677.9	69.4
18	*S*-methyl	*R*-4-hydroxybenzyl	*R*-methyl	2-(3-bromo-phenyl)-ethyl	120.0	15.2
4	*S*-methyl	*R*-2-napthylmethyl	*S*-methyl	2-(3-bromo-phenyl)-ethyl	233.1	32.7
6	*S*-methyl	*R*-2-napthylmethyl	*R*-hydroxymethyl	2-(3-bromo-phenyl)-ethyl	138.8	39.3
8	*S*-methyl	*R*-2-napthylmethyl	*R*-4-hydroxybenzyl	2-(3-bromo-phenyl)-ethyl	ND	
1	*S*-methyl	*R*-2-napthylmethyl	*R*-methyl	4-methyl-1-cyclohexyl-methyl	144.3	32.3

1561-	R1	R2	R3	R4	K_i (nM)	SEM
38	*S*-4-hdroxybenzyl	*R*-butyl	*R*-hydroxymethyl	2-(3-bromo-phenyl)-ethyl	132.2	27.6
14	*S*-methyl	*R*-butyl	*R*-hydroxymethyl	2-(3-bromo-phenyl)-ethyl	465.1	66.1
62	*R*-methyl	*R*-butyl	*R*-hydroxymethyl	2-(3-bromo-phenyl)-ethyl	362.6	97.8
30	*S*-4-hdroxybenzyl	*R*-2-napthylmethyl	*R*-hydroxymethyl	2-(3-bromo-phenyl)-ethyl	>1000	
46	*S*-4-hdroxybenzyl	*R*-4-hydroxybenzyl	*R*-hydroxymethyl	2-(3-bromo-phenyl)-ethyl	204.1	13.4
34	*S*-4-hdroxybenzyl	*R*-butyl	*R*-methyl	2-(3-bromo-phenyl)-ethyl	381.7	46.6
36	*S*-4-hdroxybenzyl	*R*-butyl	*S*-methyl	2-(3-bromo-phenyl)-ethyl	ND	
40	*S*-4-hdroxybenzyl	*R*-butyl	*R*-4-hydroxybenzyl	2-(3-bromo-phenyl)-ethyl	415.0	25.1
37	*S*-4-hdroxybenzyl	*R*-butyl	*R*-hydroxymethyl	4-methyl-1-cyclohexyl-methyl	>1000	

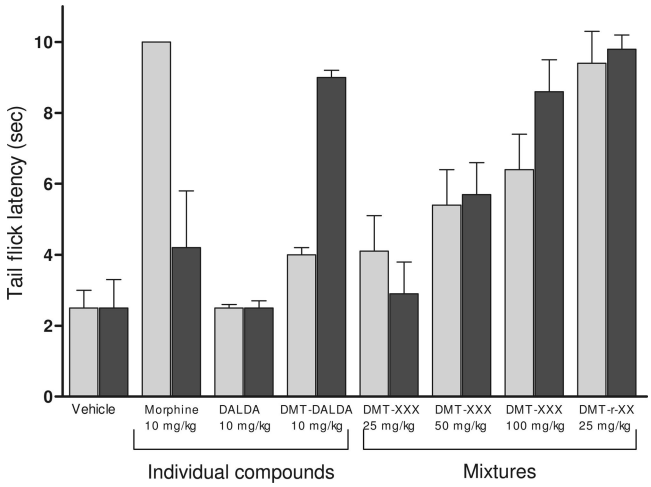


Figure 5. Comparison of in vivo activity of tetrapeptide mixtures and individual compounds using tail flick assay. Tail flick latency is the average of 2 readings per animal, 10 animals per compound or mixture. Vehicle is 10% trappsol. Gray bars = 30 min post injection; black bars = 5 h. post injection.

in vivo antinociceptive properties in mice.[48] The Dmt group has been shown to be an essential component of this compound, as demonstrated by both in vitro and in vivo analysis. Here the Dmt-XXX tetrapeptide mixture was examined for its activity in vivo and compared to the antinociceptive effect of the dermorphin-derived tetrapeptide analog and two mixtures lacking the peptide, namely, F-XXX and k-XXX. Dmt-XXX, at a dose of 100 mg/kg, exhibited activity and had duration of action comparable to dermorphin-derived tetrapeptide analog at 10 mg/kg (Figure 5). The effect was shown to be specific for Dmt-XXX because F-XXX and k-XXX were inactive at comparable doses (and 800 and 3580 times less active in vitro, respectively, see Table 5). It was notable that the mixture Dmt-XXX exhibited antinociceptive effects substantially longer than morphine, a property already reported for dermorphin-derived tetrapeptide analog.[48] As anticipated, the mixture required greater absolute doses than dermorphin-derived tetrapeptide analog.

Following these results, we carried out a single test case iteration that would be used to identify the next position of the active mixture. This was done by synthesizing a mixture likely to be the most active, namely, Dmt-r-XX, with D-arginine defined in the second position, as found in dermorphin-derived tetrapeptide analog. This iteration decreased the complexity of the mixtures from 125 000 compounds to 2500 compounds. The antinociceptive effect of this mixture was tested in the tail-flick assay as before. One would expect an increase in activity when tested at the same concentration because the tetrapeptide mixture is 50 times less complex with regards to the number of compounds. Indeed, increased activity was observed for this mixture. While the mixture size from Dmt-XXX to Dmt-r-XX decreased by 50-fold, the increase in activity was modest (4-fold in vitro). This can be explained by the possibility that arginine may not be the most active substitution at the second position of this mixture; D-arginine was defined at the second position simply as a test study to expand this proof of concept for the use of mixture-based combinatorial libraries directly in vivo. Once again, 5 h post injection at a dose of 25 mg/kg, this peptide mixture of 2500 tetrapeptides was clearly more active than morphine at this time point (Figure 5).

The use of mixture-based libraries for direct in vivo testing for the identification of inherently more advanced "hits", while clearly both exciting and promising, remains in the proof-of-concept phase at this time. If this approach is found

Table 5. μ-Opioid Receptor Binding (K_i) and cAMP (IC_{50}) Data for Individual and Tetrapeptide Mixtures

				binding K_i (nM)	cAMP IC_{50} (μM)	compounds
L-(Dmt)-Tyr	X	X	X	16.2	6.8	125 000
L-Phe	X	X	X	13 000	1027	125 000
D-Lys	X	X	X	58 030	ND	125 000
L-(Dmt)-Tyr	D-Arg	X	X	4.5	0.4	2500
Dmt-DALDA: L-(Dmt)-Tyr	D-Arg	L-Phe	L-Lys	0.2	0.002	1
DALDA: L-Tyr	D-Arg	L-Phe	L-Lys	5.5	0.8	1

to be as generally applicable as we expect, in vivo testing of mixtures will significantly accelerate the drug discovery process over traditional target-based in vitro assays. We are continually exploring the enormous potential inherent in the direct in vivo testing of chemically diverse libraries in an effort to reduce the time and resources required in the traditional drug discovery process. It is anticipated that this new approach has the potential to identify novel pain-modulating agents and the identification of ligands for orphan receptors that play a role in pain regulation.

By its very nature, in vivo assay screening is inherently low throughput, and to test a large diversity in vivo, the number of samples required can be further reduced using the scaffold ranking strategy, as previously described. Here, we present preliminary results using the scaffold ranking approach combined with a in vivo screening assay for psychiatric indications developed by PsychoGenics, Inc.[49] While this assay is a high-throughput assay relative to other in vivo assays, its focus is on the direct use of the behavior of animals as the assay. It uses specialized hardware, computer vision, and machine learning algorithms to produce predictions of therapeutic efficacy for several psychiatric indications. This study demonstrates the potential value of the scaffold ranking strategy as a method for triaging the screening of any collection of compounds having identical functionalities and differing central scaffolds.

To triage positional scanning libraries for the identification of psychiatrically relevant compounds, we prepared five different scaffold mixtures representing five different positional scanning libraries and screened them using PsychoGenics in vivo assay (Figure 6). This assay simultaneously screened and collected data for seven different CNS indications, and the predictions were expressed in the form of Bayesian posterior probabilities for each class. Scaffold ranking allows one to determine which scaffolds show the most pharmacological activity. In Figure 6, scaffolds 1, 3, and 5 were clearly active, scaffold 4 showed marginal activity, and scaffold 2 was completely inactive. Scaffolds 1, 3, and 5 were retested at a lower dose and scaffold 1 showed the highest activity of the three. The complete positional scanning library of scaffold 1 was subsequently screened, and the data for the mixtures with one position defined is shown in Figure 7a. The data revealed that only a few mixtures were active. Thus, specific functionalities at this position must be contributing to the activity of this scaffold, with the majority (20 of 27) of the mixtures having no activity. It is important to note that mixtures 3 and 4, as well as 9 and 10, differ solely on the stereochemistry in the moiety at the defined position. The activities found for mixtures 3 and 4 differ in their CNS activity profile, while mixture 10 is clearly active and mixture 9 is not. The

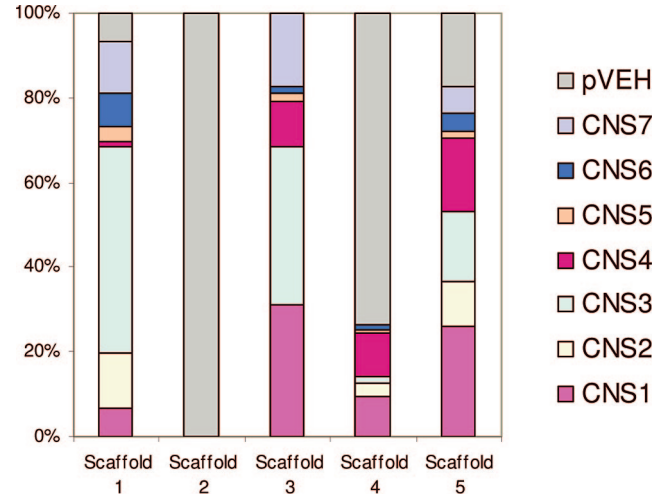


Figure 6. Screening results of 5 different scaffolds in Psychogenics in vivo screening assay for psychiatric indications. The five libraries (scaffolds 1−5) are from the same diversity family of libraries and therefore only differ at the core scaffold. All of them were screened at a dose of 30 mg/kg in mice, $n = 10$. The assay simultaneously screens for seven different CNS indications. "pVEH" indicates the vehicle, or control group, and thus indicates compounds with no observed behavioral effects. The *y*-axis represents predictions that are expressed in the form of Bayesian posterior probabilities from a single classifier trained on all eight classes. On the basis of the pVEH values, scaffold activity could be ranked 1, 3, 5 > 2, 4, leading one to test the complete libraries of scaffolds 1, 3, and 5 for further evaluation. However, a full measure of potency requires testing the scaffolds across a range of doses.

activities of 6 compounds from the initial positional scanning deconvolution are shown in Figure 7b. Five of the six compounds were active and showed a range of different CNS activities. Furthermore, the type of CNS activity (CNS 1, 3, and 4) found for the individual compounds correspond to those found in the mixtures. While this approach has clear value when used in relatively low-throughput assays, it should be noted that this approach is equally valuable for triaging libraries even when using mid- to high-throughput assays. Scaffold ranking is an approach that will direct the researcher toward pursuing the most-active libraries in a large collection of chemical diversity. It is possible that with this approach, as well as any approach using complex mixtures, that it may be difficult to distinguish mixtures having a few very high-affinity compounds from other mixtures containing many low-affinity compounds. Differentiation between these two types of mixtures becomes apparent upon the screening of the complete positional scanning library and deconvolution to individual compounds.

Computational Deconvolution of Positional Scanning Libraries

The technology of synthetic combinatorial chemistry and high-throughput screening has enabled access to the synthesis

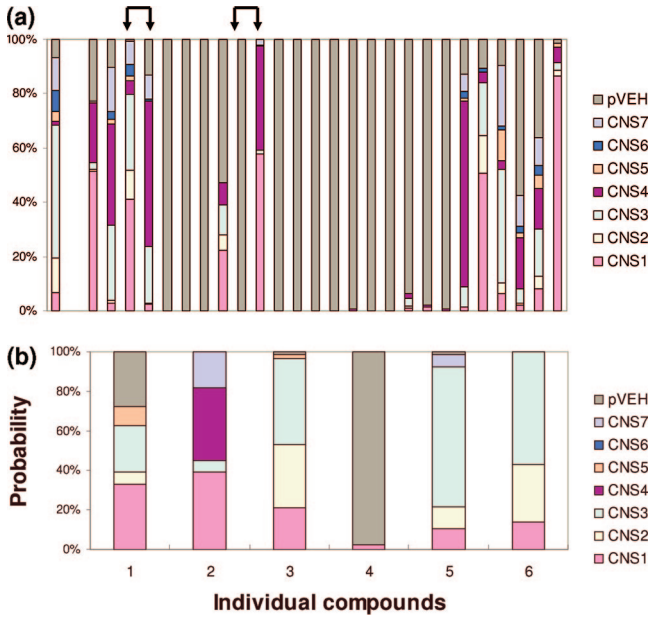


Figure 7. (a) Screening results from fixing one position of scaffold 1 in Psychogenics in vivo screening assay for psychiatric indications. As in Figure 6, the assay simultaneously screens for seven different CNS indications, and the predictions are expressed in the form of Bayesian posterior probabilities for each indication. In this figure, each vertical bar represents the different mixtures screened. The mixtures differ only by the substituent that was fixed at the given position. The arrows indicate mixtures that differ solely on the stereochemistry in the moiety at the defined position. (b) Screening results for six individual compounds initially identified from scaffold 1 in Psychogenics in vivo screening assay for psychiatric indications. All of them were screened at a dose of 30 mg/kg in mice, $n = 10$. See Figures 6 and 7a for more details.

and testing of thousands to millions of compounds in an extremely short period of time relative to testing this number of compounds individually. Practical considerations limiting the number of compounds to be synthesized and purified require that identification of a smaller number of structures having distinctly different chemical functionalities would clearly be beneficial. The visual inspection of each structure to determine unique functionality, however, can be limiting as well as being only qualitative in nature. Our laboratory has successfully employed the use of mixture-based libraries to circumvent a number of these limitations. Computational chemistry can now be used to apply more quantitative measurements to distinguish compounds in collections of compounds, including mixture-based libraries.

Each of the small molecule positional scanning libraries is designed around a core scaffold. Traditionally, such core scaffolds are chosen based on the following criteria: the core scaffold should be of biological importance (privileged structures), can be easily modified using readily available building blocks, and the core structure scaffold should be accessible whenever possible using straightforward synthetic conditions. Each positional scanning sublibrary contains positions that permit structural variations around the central core. Screening data from a library provides extensive and clear SAR information and enables identification of active individual compounds. Thus, the individual structural components, and their representative contributions to total biological activity within the positional scanning library, are revealed.

The use of quantitative measurements to distinguish compounds within a library of compounds is a highly useful and comprehensive approach.[50–52] These measurements involve molecular properties or descriptors.[53,54] It has been shown in many cases that a computational approach for the calculation of the descriptors, and the selection of a subset of the full library of compounds, can lead to a more enhanced representation of the library than if one selects compounds at random.[55] While mixture-based libraries eliminate a number of the limitations inherent with screening large collections of individual compounds, many of the quantitative aspects of chemoinformatics appeal to the methods developed by our laboratory. Chemoinformatic approaches are potentially an attractive means to enhance the information generated following the screening of our mixture-based libraries.

To assess the usefulness of combining chemoinformatic applications with our mixture-based libraries, we have conducted several pilot studies. One of the studies was based on the screening of a trisubstituted bicyclic guanidine PS-SCL against a κ-opioid receptor. Using positional scanning deconvolution methods, the screening of this library led to the selection of 48 individual compounds for synthesis. These individual compounds had binding affinities ranging from 37 to $>10\,000$ nM for the κ-opioid receptor.[4] The data obtained from this series of compounds is valuable in that active compounds were identified and SAR information was obtained from the range of activities. To enhance positional scanning deconvolution methods and obtain qualitative SAR data from our small-molecule libraries, we combined chemoinformatics approaches to develop a combined computational deconvolution method.

The trisubstitued bicyclic guanidine library (BCG) contains 102 459 individual compounds. To further reduce the number of compounds to be studied, as well as incorporate some of the information that was obtained from the original screening of the library, a technique similar to peptide-based biometrical analysis was applied. The peptide-based biometrical analysis is a method that scores each peptide within a sequence database using the data obtained from the screening of peptide positional scanning libraries.[9,56] Using the percent inhibition data obtained from screening of the BCG library in the κ-opioid receptor, a predicted relative activity score is calculated for each of the 102 459 compounds in the library. The predicted scores were calculated by addition of the normalized percent inhibition of the sample for each of the functionalities in the compound. Figure 8 represents the relative score distribution for all 102 459 compounds in the BCG library for the κ-opioid assay. In this example, the top 3% (2762 compounds with score ≥ 255) were selected as potentially the most-active compounds. This set of 2762 compounds will be referred to as the metrical analysis (MA) set. It should be noted that with sufficient computational power and time all 102 459 compounds could and should be studied. While 2762 compounds is significantly less than the original 102 459 compounds making up the entire BCG positional scanning library, the synthesis of even 2762

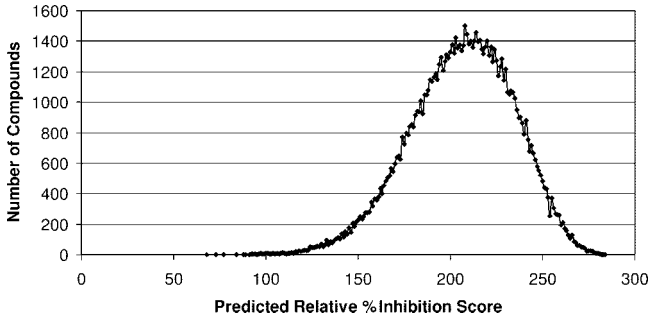


Figure 8. Relative score distribution for the BCG library. The predicted relative % inhibition score was calculated for each of the 102 459 compounds in the library. This was calculated simply by adding the score that each substituent in each compound received from the screening of the positional scanning library in the κ-opioid radioreceptor assay. Theoretically the higher the total score of the individual compound in this distribution the more likely it will be active when screened.

individual compounds is typically time-consuming and economically prohibitive for most groups in secondary-screening efforts.

To reduce the number of individual compounds to be synthesized, a computational subsetting technique was employed. A total of 199 molecular properties were calculated for the 2762 compounds in the MA set. However, 117 descriptors had no variance in the results. The 82 remaining descriptors were used for the principal component analysis (PCA).[57] Following the PCA, the next step was to select a diverse set of compounds from the original MA library of 2762. There are multiple options for selecting diverse compounds from a library including distance-based methods, cell-based methods,[58] and coverage-based algorithms.[59] For this project, we used the distance-based algorithm, MaxMin, as well as a novel coverage-based method, both available in Accelrys' Cerius2 package.

With the MaxMin function, nine subsets of differing sizes were created, namely, 30, 40, 48, 50, 60, 70, 80, 90, and 100 compounds in each subset (subset 48 was included for the 48 compounds identified through the use of positional scanning deconvolution). To select a set of compounds for synthesis from all the subsets, a weighted sum of percent differences was calculated for each of the different subsets (data not shown). The subset with 30 compounds had one of the lowest weighted sum values, and this subset was selected for synthesis and screening.

These 30 compounds were synthesized and then screened in the κ-opioid receptor. The results identified a range of activities similar to the activities found with the compounds selected using the positional scanning deconvolution methods in which the most active compound had a K_i value of 99 nM. It is important to note that in this type of study stereoisomers are not distinguished from one another and will occupy the same area of chemical space. The data obtained from both deconvolution approaches are shown in Figure 9A (positional scanning) and B (computational). There were several stereoisomers in the positional scanning set 9A, and these stereoisomers had differing activities. In some cases, one stereoisomer had an activity greater than 500 nM and the other less than 500 nM. A number of observations can be made when viewing the two data sets, 9A and 9B,

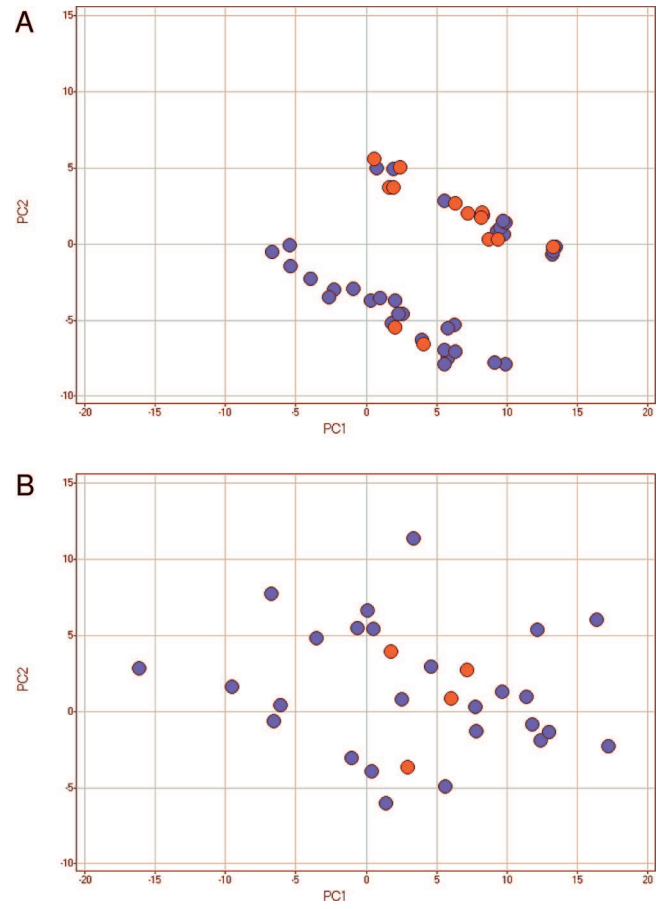
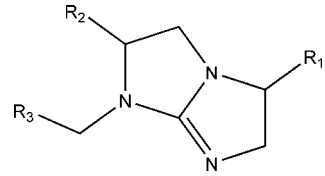


Figure 9. Comparison of the chemical space relationship for positional scanning and computational deconvolution approaches. Two principle components graphed on the x−*y* axis, respectively (approximately 91% of the total variance is represented in these components). Red circles represent compounds with K_i values lower than 500nM, and blue circles represent compounds with K_i values higher than 500nM. Figure 9A shows the data of the 48 compounds made strictly using positional scanning deconvolution. This set contains several stereoisomers. Figure 9B shows the data of the 30 compounds synthesized using the computational deconvolution approach.

together. First the compounds from the positional scanning approach, 9A, represent a less diverse area of chemical space than the compounds designed using the computational approach, 9B. This result is completely consistent with what was expected on the basis of how each set was devised. Second there appears to be a significant correlation between activity and spatial position. It appears that a compound that is in the area between PC1 = 0 and PC1 = 10 and PC2 = 0 and PC2 = 5 has a significantly greater chance of being "active" than a compound residing in any area less than PC2 = 0.

To validate the observed correlation between activity and spatial position and to demonstrate whether computational deconvolution could lead to identification of additional leads, as well as an enhanced SAR understanding, a set of 49 compounds were made, and the activities are shown in Table 6. Figure 10A graphically depicts the chemical space position of these compounds. The compounds depicted by red circles were synthesized to test the "positive predictive value" of this deconvolution method, while the blue circle compounds were chosen to test the "negative predictive value." In other

Table 6. κ-Opioid Receptor Binding (K_i) Data for Individual Bicyclic Guanidines Predicted Using Computational Deconvolution



no.	R1	R2	R3	K_i (nM)	prediction
47	*S*-methyl	*S*-4-methoxybenzyl	3-cyclohexylpropyl	39	+
46	*S*-methyl	*R*-4-ethoxybenzyl	1-adamantylethyl	151	+
15	*S*-methyl	*R*-4-ethoxybenzyl	2-norbornylethyl	185	+
11	*S*-ethyl	*R*-4-ethoxybenzyl	2-norbornylethyl	321	−
25	*R*-butyl	*R*-4-methoxybenzyl	2-norbornylethyl	430	+
30	*R*-cyclohexylmethyl	*R*-4-methoxybenzyl	1-adamantylethyl	457	+
9	*S*-ethyl	R-4-methoxybenzyl	2-norbornylethyl	598	+
19	*S*-propyl	*S*-4-methoxybenzyl	2-norbornylethyl	679	+
23	*S*-butyl	*S*-4-methoxybenzyl	2-norbornylethyl	811	+
38	*R*-cyclohexylmethyl	*R*-4-methoxybenzyl	cyclopentylmethyl	819	+
5	*R*-isopropyl	*R*-4-methoxybenzyl	cyclopentylmethyl	913	−
17	*S*-propyl	*R*-4-methoxybenzyl	2-norbornylethyl	921	+
37	*R*-cyclohexylmethyl	*R*-4-methoxybenzyl	cyclobutylmethyl	940	+
50	*S*-methyl	*R*-4-methoxybenzyl	2-norbornylethyl	1033	+
34	*S*-cyclohexylmethyl	*R*-butyl	isobutyl	1173	−
14	*S*-ethyl	*S*-4-methoxybenzyl	2-norbornylethyl	1368	+
45	*S*-isopropyl	*S*-4-methoxybenzyl	2-norbornylethyl	1426	+
39	*R*-cyclohexylmethyl	*R*-butyl	isobutyl	1442	−
10	*S*-ethyl	*S*-ethyl	1-adamantylethyl	1457	+
22	*S*-butyl	*S*-4-methoxybenzyl	cyclopentylmethyl	1643	+
12	*S*-ethyl	*S*-4-methoxybenzyl	phenylpropyl	1700	+
49	*S*-methyl	*S*-4-methoxybenzyl	2-norbornylethyl	1730	+
42	*S*-isopropyl	*R*-4-methoxybenzyl	cyclopentylmethyl	1840	+
8	*S*-ethyl	*R*-4-methoxybenzyl	cyclopentylmethyl	1929	+
43	*S*-isopropyl	*R*-4-methoxybenzyl	2-norbornylethyl	1933	+
28	*S*-cyclohexylmethyl	*R*-4-methoxybenzyl	1-adamantylethyl	2373	+
21	*S*-butyl	*R*-4-methoxybenzyl	2-norbornylethyl	2430	+
33	*S*-cyclohexylmethyl	*R*-4-methoxybenzyl	cyclopentylmethyl	3089	−
20	*S*-butyl	*R*-4-methoxybenzyl	cyclopentylmethyl	3539	+
48	*S*-methyl	*S*-4-methoxybenzyl	4-methyl-cyclohexylmethyl	4216	+
32	*S*-cyclohexylmethyl	*R*-4-methoxybenzyl	cyclobutylmethyl	4442	+
24	*R*-butyl	*R*-4-methoxybenzyl	cyclopentylmethyl	4698	+
7	*R*-isopropyl	*S*-4-methoxybenzyl	cyclopentylmethyl	4981	−
26	*R*-butyl	*S*-4-methoxybenzyl	cyclopentylmethyl	5114	+
35	*S*-cyclohexylmethyl	*S*-4-methoxybenzyl	cyclobutylmethyl	5121	−
31	*R*-cyclohexylmethyl	*S*-4-methoxybenzyl	1-adamantylethyl	5393	+
4	*S*-methyl	*S*-isopropyl	cyclohexylbutyl	5925	−
3	*S*-methyl	*S*-isobutyl	1-adamantylethyl	6054	−
29	*S*-cyclohexylmethyl	*S*-4-methoxybenzyl	1-adamantylethyl	6605	+
27	*R*-butyl	*S*-4-methoxybenzyl	2-norbornylethyl	7949	+
36	*S*-cyclohexylmethyl	*S*-4-methoxybenzyl	cyclopentylmethyl	7988	−
40	*R*-cyclohexylmethyl	*S*-4-methoxybenzyl	cyclobutylmethyl	8683	+
44	*S*-isopropyl	*S*-4-methoxybenzyl	cyclopentylmethyl	>10 000	−
41	*R*-cyclohexylmethyl	*S*-4-methoxybenzyl	cyclopentylmethyl	>10 000	+
2	*S*-methyl	*S*-isobutyl	3-cyclohexylpropyl	>10 000	−
6	*S*-methyl	*R*-isopropyl	cyclohexylbutyl	>10 000	+
1	*S*-methyl	*S*-methyl	1-adamantylethyl	>10 000	+
16	*S*-propyl	*R*-4-methoxybenzyl	cyclobutylmethyl	>10 000	+
18	*S*-propyl	*S*-4-methoxybenzyl	cyclobutylmethyl	>10 000	+

words, if these two sets of compounds are screened against the same κ-opioid receptor target as in the previous studies, there would be a high probability that the red circle set would have a larger percentage of active compounds than the blue circle set. The results from the screening effort are presented in Figure 10B. The data shows that for the negative predictive value set, 11 out of the 12 compounds were inactive as predicted. For the positive predictive value set, 5 out of the 37 compounds were active (K_i < 500 nM). Out of these five active compounds, three had K_i values of less than 200 nM. In addition, 11 out of the 37 positive predicted value set showed K_i values of less than 1 μM.

In addition, it should also be noted that all of the compounds represented in these sets that are deconvoluted from the library appear to be relatively similar structurally. Moreover, the distribution of similar compounds considered here is quite different from that typically found in a diverse compound collection (Figure 11). The curve located on the left-hand side of the figure is the cumulative distribution function (cdf) of the Tanimoto similarities,[60] based on MACCS fingerprints, as implemented in the MOE program available from Chemical Computing Group (http://www.chemcomp.com), of a random sample of 1000 compounds obtained from the NIH AIDS antiviral screening

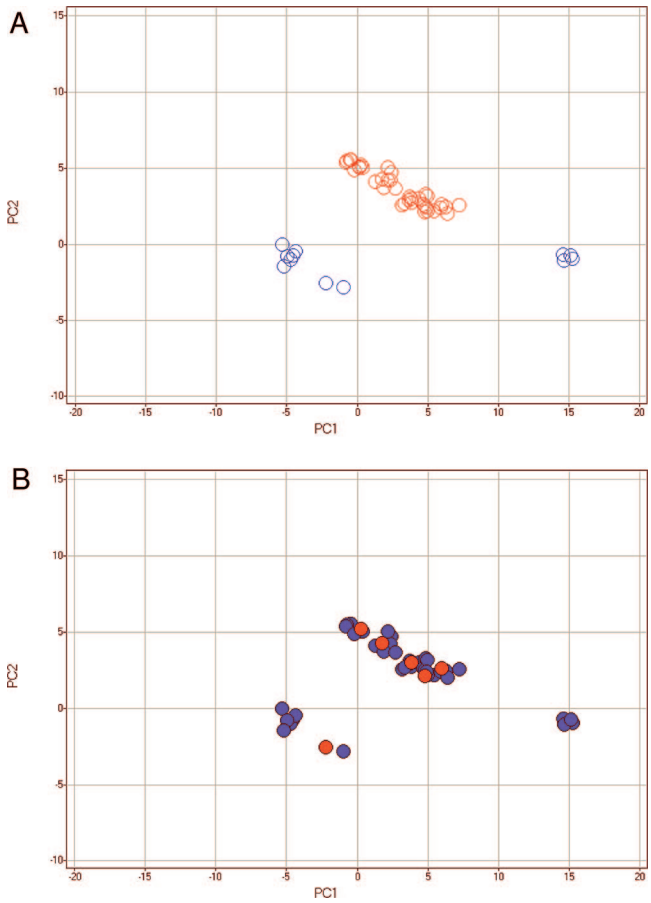


Figure 10. Testing the hypothesis of activity chemical space relationship. The 49 compounds are presented using the same principle components as in Figure 9. The red circles are compounds that were chosen to test the theoretically "active" area of chemical space; the blue circles were compounds chosen to test the theoretically "inactive" area of chemical space. Figure 10B shows the actual results after screening. The red circles now represent compounds with K_i values lower than 500 nM, and the blue circles represent compounds with K_i values greater than 500 nM.

database (http://dtp.nci.nih.gov/docs/aids/searches/active_compounds.html). It is clear from the figure that the median is at a value of approximately 0.2. Thus, the Tanimoto similarities of half the compounds in the NIH collection are less than 0.2, and half are greater than 0.2. From the shape of the curve, it is also clear that the cdf is monomodal and close to that of a normal distribution, another indication of structural diversity.[61] The curve located on the right-hand side of the figure is the corresponding cdf for the set of 48 compounds considered in Figure 9A, which have a median Tanimoto similarity of approximately 0.8. This value is quite distant from that of the NIH data set and indicates that the current data set is clustered *relative* to it. This means that even in a relatively dense area of chemical space we are still able to distinguish between active and inactive areas.

There is another significant benefit derived when using our type of mixture-based libraries, namely, they provide a high-density of compounds in local regions of chemical space. At first, this may not appear to be a very efficient procedure because it runs counter to the current emphasis on screening widely diverse sets of compounds, a strategy that is quite acceptable under conditions where changes in structure have a relatively small effect on activity. However,

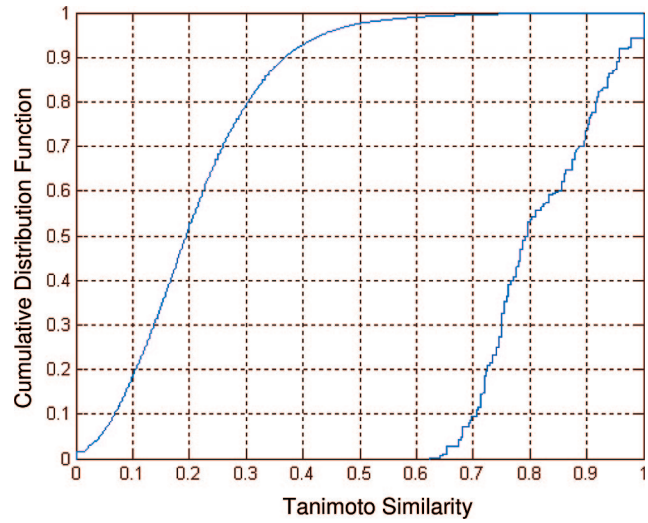


Figure 11. Distribution of a diverse set of compounds and a set of compounds from positional scanning deconvolution. The curve located on the left-hand-side of the figure is the *cumulative distribution function* (cdf) of the Tanimoto similarities, based on MACCS fingerprints, of a random sample of 1000 compounds obtained from the NIH HIV AIDS data set (http://dtp.nci.nih.gov/docs/aids/searches/active_compounds.html). The curve located on the right-hand side of the figure is the corresponding cdf for the set of 48 compounds considered in Figure 9A. Also note that the roughness of this curve in contrast to the first curve is caused by the considerably smaller number of compounds in the latter set.

in a growing number of cases, this is seen not to be true. In such cases, small changes in structure can lead to precipitous changes in activity giving rise to what have been called "activity cliffs" [62] and the very well known differences that single -position chirality variations cause in therapeutic activity. Such differences in chirality have been readily distinguishable in a variety of assays upon screening mixture-based positional scanning libraries.[4] If the density of compounds screened in such regions of chemical space is insufficient there is a significant likelihood that active compounds would be missed.

Figure 12 depicts the approximate 3-D chemical space of the set of the 48 compounds discussed above (Figure 9A). The chemical space was constructed using Tanimoto similarity based on MACCS key fingerprints, as noted earlier. The similarity matrix was then subjected to principal component analysis,[57] and the first three principal components, which explained 95% of the total variance of the sample, were plotted in the figure. Since MACCS fingerprints do not specifically account for stereochemical differences, stereoisomers in the figure are technically superimposed. In addition, there are inherent limitations in MACCS fingerprints for treatment of complicated multicyclic rings systems. However, the points are "jittered" to separate them and, hence, make them more visible. Active compounds ($<$500 nM) are red, and inactive compounds ($>$500 nM) blue. It is clear from the data that a small change in chemical structure or a change in stereochemistry can cause a significant change in activity in this assay.

Because chemical space is highly dependent on the nature of the representation used, the presence of activity cliffs may be obscured or may not be observed at all if different representations are employed, an unfortunate but unavoidable

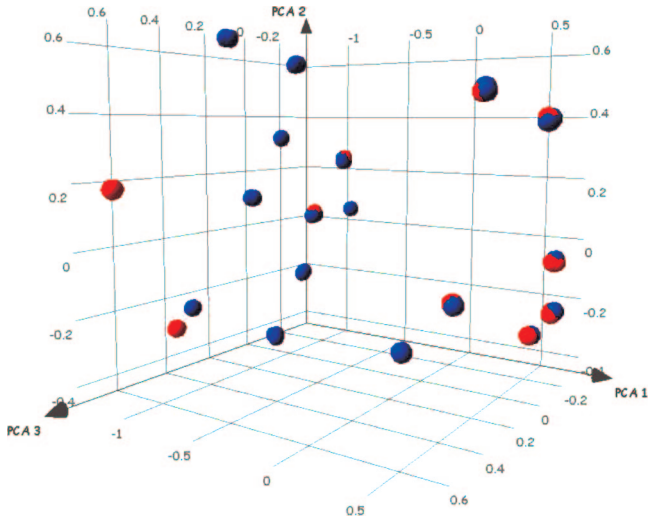


Figure 12. 3-D chemical space of the set of the 48 compounds synthesized using positional scanning deconvolution. The chemical space was constructed using 2-D Tanimoto similarity based on MACCS fingerprints. The similarity matrix was then subjected to principal component analysis, and the first three principal components, which explained 95% of the total variance of the sample, were plotted in the figure.

feature of chemical space.[62] Nevertheless, the concepts of chemical space and activity cliffs are useful because they can provide an intuitive, albeit imperfect, picture of the complex structure–activity relationships (SARs) present in the data set. This is the case here as exemplified by Figure 12, which shows numerous activity cliffs depicted as overlapping red/blue circles or red and blue circles in close proximity to each other. It is important to note that Figure 12 does not represent the true density of compounds investigated in the study because only a relatively small subset of compounds was characterized in detail (see above). Moreover, it graphically shows that unless a reasonable high density of compounds (and very importantly these should include stereoisomers) are screened in regions of activity cliffs there is a very significant likelihood that active compounds will be missed.

Three significant findings resulted from our initial studies aimed at combining computational approaches with the screening of our mixture-based libraries. First, the positional scanning deconvolution method developed by our group is as effective for the identification of active compounds from a large library of compounds as the computational deconvolution methods described here illustrated. Second, the combination of positional scanning deconvolution with computational deconvolution described here leads to information that is not available with positional scanning deconvolution alone. This data can be used to specifically identify groups of compounds within our current inventory of libraries, or other outside sources of compounds, having a high probability of similar activities to the most active compounds identified, but with differing chemical properties. And finally the studies confirm our belief that mixture based libraries enable a highly effective means to screen dense, localized regions of chemical space. Screening in dense regions of chemical space decreases the possibility that active compounds will be missed, a situation that is particularly critical when activity cliffs are prevalent in activity landscapes. Moreover, such dense screening allows for a more complete analysis of regions of chemical space, an analysis that would be far less if the regions were "cherry picked" for individual compounds.

Conclusions and Future Developments

The efforts of this laboratory over the past 16 years have been focused on systematically modernizing what has been known for the past 5000 years, namely, that compounds having useful therapeutic value can be identified from mixtures. Drug discovery has primarily been a process in which the activity of plants, plant extracts, or other materials were found to be active in humans. The systematic synthesis and screening of individual compounds began 50–75 years ago, and the concept of de novo drug design was conceived and continues to be developed over the past 20 years. The discovery of morphine for pain relief is a celebrated example. It has been known for over 5000 years that the secretion of opium poppies is a potent analgesic. Isolation of pure morphine from the hundreds to thousands of compounds in opium was accomplished only 200 years ago, it is structure deciphered only 80 years ago, discovery of its target, the μ-opiate receptor, was only accomplished in 1973, and endorphins, the natural ligands for this receptor, were identified in 1975. This is but one of a wealth of examples of the mixture of compounds making up plant extracts being used to identify palliatives or cures for many of mankind's physical and psychological afflictions.

A critical element of modern drug discovery has been "target-based" screening. While of immense importance, it is noteworthy that there remains a number of highly successful therapeutics that do not have a clearly delineated mechanism of action. We believe that the screening of systematically arranged synthetic mixtures directly in relevant in vivo assays will lead to the discovery of compounds that would not have been identified in any other manner. With this approach, it matters little what one's expectations or knowledge are prior to screening because it is the assay that determines the outcome, a principle that has always been and is likely to remain the mainstay of the drug discovery process.

To conclude, mixture-based synthetic combinatorial libraries that can be used directly in solution, as first presented by this laboratory in 1991 and used by many others today, allow for the rapid and economical identification of novel leads. We continue to explore and develop these methods and expand their utility; traditional positional scanning approaches have now been combined with a new scaffold ranking strategy, and computational analyses have been applied to minimize the synthesis of individual compounds associated with positional scanning deconvolution. These strategies facilitate the direct use of mixture-based libraries in animal models of disease. Along with the continuing advances in synthetic methodologies such as volatilizable solid- and solution-phase synthetic supports,[82] we believe that the tools that we have developed can be used directly as stand-alone procedures or in conjunction with existing drug discovery approaches. All benefit the drug discovery

process and therefore accrue to the benefit of the health and well being of mankind.

Acknowledgment. The authors' work was supported in part by National Institute of Drug Abuse (DA09410 and DA019620), National Institute of Mental Health (MH071041), National Cancer Institute (1U19CA113318), National Science Foundation (CHE0455072), Foundation of NCET-05-0523 of China, Arthritis and Chronic Pain Research Institute, Multiple Sclerosis National Research Institute, Alzheimer's and Aging Research Center, and Diabetes National Research Group.

References and Notes

(1) Frank, R. *Nucl. Acid. Res.* **1983**, *11*, 4365–4377.
(2) Geysen, H. M.; Meloen, R. H.; Barteling, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 3998–4002.
(3) Houghten, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5131–5135.
(4) Houghten, R. A.; Pinilla, C.; Appel, J. R.; Blondelle, S. E.; Dooley, C. T.; Eichler, J.; Nefzi, A.; Ostresh, J. M. *J. Med. Chem.* **1999**, *42*, 3743–3778.
(5) Thompson, L. A.; Ellman, J. A. *Chem. Rev.* **1996**, *96*, 555–600.
(6) Nefzi, A.; Ostresh, J. M.; Houghten, R. A. *Chem. Rev.* **1997**, *97*, 449–472.
(7) Nefzi, A.; Ostresh, J. M.; Yu, Y.; Houghten, R. A. *J. Org. Chem.* **2004**, *69*, 3603–3609.
(8) Dolle, R. E. *J. Comb. Chem.* **2004**, *6*, 623–679.
(9) Pinilla, C.; Appel, J. R.; Borras, E.; Houghten, R. A. *Nat. Med.* **2003**, *9*, 118–122.
(10) Boger, D. L.; Desharnais, J.; Capps, K. *Angew. Chem., Int. Ed.* **2003**, *42*, 4138–4176.
(11) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature* **1991**, *354*, 84–86.
(12) Pinilla, C.; Appel, J. R.; Blanc, P.; Houghten, R. A. *Biotechniques* **1992**, *13*, 901–905.
(13) Ostresh, J. M.; Husar, G. M.; Blondelle, S. E.; Dörner, B.; Weber, P. A.; Houghten, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11138–11142.
(14) Yu, Y.; Ostresh, J. M.; Houghten, R. A. *J. Org. Chem.* **2003**, *68*, 183–186.
(15) Nefzi, A.; Hoesl, C. E.; Pinilla, C.; Kauffman, G. B.; Maggiora, G. M.; Pasquale, E.; Houghten, R. A. *J. Comb. Chem.* **2006**, *8*, 780–783.
(16) Dooley, C. T.; Ny, P.; Bidlack, J. M.; Houghten, R. A. *J. Biol. Chem.* **1998**, *273*, 18848–18856.
(17) Vanderah, T. W.; Schteingart, C. D.; Trojnar, J.; Junien, J. L.; Lai, J.; Riviere, P. J. *J Pharmacol. Exp. Ther.* **2004**, *310*, 326–333.
(18) Capps, K. J.; Humiston, J.; Dominique, R.; Hwang, I.; Boger, D. L. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2840–2844.
(19) Goldberg, J.; Jin, Q.; Ambroise, Y.; Satoh, S.; Desharnais, J.; Capps, K.; Boger, D. L. *J. Am. Chem. Soc.* **2002**, *124*, 544–555.
(20) Masip, I.; Cortes, N.; Abad, M. J.; Guardiola, M.; Perez-Paya, E.; Ferragut, J.; Ferrer-Montiel, A.; Messeguer, A. *Bioorg. Med. Chem.* **2005**, *13*, 1923–1929.
(21) Abad-Merin, M. J.; Cortes, N.; Masip, I.; Perez-Paya, E.; Ferragut, J. A.; Messeguer, A.; Ferrer-Montiel, A. *J. Pharmacol. Exp. Ther.* **2005**, *313*, 112–120.
(22) Mora, P.; Masip, I.; Cortes, N.; Marquina, R.; Merino, R.; Merino, J.; Carbonell, T.; Mingarro, I.; Messeguer, A.; Perez-Paya, E. *J. Med. Chem.* **2005**, *48*, 1265–1268.
(23) Humet, M.; Carbonell, T.; Masip, I.; Sanchez-Baeza, F.; Mora, P.; Canton, E.; Gobernado, M.; Abad, C.; Perez-Paya, E.; Messeguer, A. *J. Comb. Chem.* **2003**, *5*, 597–605.
(24) Planells-Cases, R.; Montoliu, C.; Humet, M.; Fernandez, A. M.; Garcia-Martinez, C.; Valera, E.; Merino, J. M.; Perez-Paya, E.; Messeguer, A.; Felipo, V.; Ferrer-Montiel, A. *J. Pharmacol. Exp. Ther.* **2002**, *302*, 163–173.
(25) Montoliu, C.; Humet, M.; Canales, J. J.; Burda, J.; Planells-Cases, R.; Sanchez-Baeza, F.; Carbonell, T.; Perez-Paya, E.; Messeguer, A.; Ferrer-Montiel, A.; Felipo, V. *J. Pharmacol. Exp. Ther.* **2002**, *301*, 29–36.
(26) Garcia-Martinez, C.; Humet, M.; Planells-Cases, R.; Gomis, A.; Caprini, M.; Viana, F.; De La, P. E.; Sanchez-Baeza, F.; Carbonell, T.; De Felipe, C.; Perez-Paya, E.; Belmonte, C.; Messeguer, A.; Ferrer-Montiel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2374–2379.
(27) Shipway, A.; Danahay, H.; Williams, J. A.; Tully, D. C.; Backes, B. J.; Harris, J. L. *Biochem. Biophys. Res. Commun.* **2004**, *324*, 953–963.
(28) Petrassi, H. M.; Williams, J. A.; Li, J.; Tumanut, C.; Ek, J.; Nakai, T.; Masick, B.; Backes, B. J.; Harris, J. L. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3162–3166.
(29) Herter, S.; Piper, D. E.; Aaron, W.; Gabriele, T.; Cutler, G.; Cao, P.; Bhatt, A. S.; Choe, Y.; Craik, C. S.; Walker, N.; Meininger, D.; Hoey, T.; Austin, R. J. *Biochem. J.* **2005**, *390*, 125–136.
(30) Mahrus, S.; Kisiel, W.; Craik, C. S. *J. Biol. Chem.* **2004**, *279*, 54275–54282.
(31) Matsumura, M.; Bhatt, A. S.; Andress, D.; Clegg, N.; Takayama, T. K.; Craik, C. S.; Nelson, P. S. *Prostate* **2005**, *62*, 1–13.
(32) Li, J.; Lim, S. P.; Beer, D.; Patel, V.; Wen, D.; Tumanut, C.; Tully, D. C.; Williams, J. A.; Jiricek, J.; Priestle, J. P.; Harris, J. L.; Vasudevan, S. G. *J. Biol. Chem.* **2005**, *280*, 28766–28774.
(33) Bersanetti, P. A.; Andrade, M. C.; Casarini, D. E.; Juliano, M. A.; Nchinda, A. T.; Sturrock, E. D.; Juliano, L.; Carmona, A. K. *Biochemistry* **2004**, *43*, 15729–15736.
(34) Cotrin, S. S.; Puzer, L.; Souza Judice, W. A.; Juliano, L.; Carmona, A. K.; Juliano, M. A. *Anal. Biochem.* **2004**, *335*, 244–252.
(35) Houghten, R. A.; Dooley, C. T.; Appel, J. R. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 1947–1951.
(36) Schimmer, A. D.; Welsh, K.; Pinilla, C.; Wang, Z.; Krajewska, M.; Bonneau, M.-J.; Pedersen, I. M.; Kitada, S.; Scott, F. L.; Bailly-Maitre, B.; Glinsky, G.; Scudiero, D.; Sausville, E.; Salvesen, G.; Nefzi, A.; Ostresh, J. M.; Houghten, R. A.; Reed, J. C. *Cancer Cell* **2004**, *5*, 25–35.
(37) Wang, Z.; Cuddy, M.; Samuel, T.; Welsh, K.; Schimmer, A.; Hanaii, F.; Houghten, R.; Pinilla, C.; Reed, J. C. *J. Biol. Chem.* **2004**, *279*, 48168–48176.
(38) Berezovskaya, O.; Schimmer, A. D.; Glinskii, A. B.; Pinilla, C.; Hoffman, R. M.; Reed, J. C.; Glinsky, G. V. *Cancer Res.* **2005**, *65*, 2378–2386.
(39) Karikari, C. A.; Roy, I.; Tryggestad, E.; Feldmann, G.; Pinilla, C.; Welsh, K.; Reed, J. C.; Armour, E. P.; Wong, J.; Herman, J.; Rakheja, D.; Maitra, A. *Mol. Cancer Ther.* **2007**, *6*, 957–966.
(40) Carter, B. Z.; Gronda, M.; Wang, Z.; Welsh, K.; Pinilla, C.; Andreeff, M.; Schober, W. D.; Nefzi, A.; Pond, G. R.; Mawji, I. A.; Houghten, R. A.; Ostresh, J.; Brandwein, J.; Minden, M. D.; Schuh, A. C.; Wells, R. A.; Messner, H.; Chun, K.; Reed, J. C.; Schimmer, A. D. *Blood* **2005**, *105*, 4043–4050.
(41) Kater, A. P.; Dicker, F.; Mangiola, M.; Welsh, K.; Houghten, R.; Ostresh, J.; Nefzi, A.; Reed, J. C.; Pinilla, C.; Kipps, T. J. *Blood* **2005**, *105*, 1742–1748.
(42) Hensler, M. E.; Bernstein, G.; Nizet, V.; Nefzi, A. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5073–5079.
(43) Goodman, D. *Deng Xiaoping and the Chinese Revolution: A Political Biography*, 1st ed; Routledge: Oxford, U.K., 1994; p. 3.
(44) Berman, J.; Halm, K.; Adkison, K.; Shaffer, J. *J. Med. Chem.* **1997**, *40*, 827–829.

(45) Cheng, Y.; Rano, T. A.; Huening, T. T.; Zhang, F.; Lu, Z.; Schleif, W. A.; Gabryelski, L.; Olsen, D. B.; Stahlhut, M.; Kuo, L. C.; Lin, J. H.; Xu, X.; Jin, L.; Olah, T. V.; McLoughlin, D. A.; King, R. C.; Chapman, K. T.; Tata, J. R. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 529–532.
(46) Schiller, P. W.; Nguyen, T. M.; Berezowska, I.; Dupuis, S.; Weltrowska, G.; Chung, N. N.; Lemieux, C. *Eur. J. Med. Chem.* **2000**, *35*, 895–901.
(47) Zhao, G. M.; Qian, X.; Schiller, P. W.; Szeto, H. H. *J. Pharmacol. Exp. Ther.* **2003**, *307*, 947–954.
(48) Shimoyama, M.; Shimoyama, N.; Zhao, G. M.; Schiller, P. W.; Szeto, H. H. *J. Pharmacol. Exp. Ther.* **2001**, *297*, 364–371.
(49) Brunner, D.; Ghondhalekar, V.; Leahy, E.; La Rosa, D.; Ross, W. P. U.S. Patent 7,269,516, 2007.
(50) Martin, Y. C. *J. Comb. Chem.* **2001**, *3*, 231–250.
(51) Gorse, D.; Lahana, R. *Curr. Opin. Chem. Biol.* **2000**, *4*, 287–294.
(52) Willett, P. *Curr. Opin. Biotechnol.* **2000**, *11*, 85–88.
(53) Xue, L.; Bajorath, J. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 801–809.
(54) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Clinometric Series; Research Studies Press/John Wiley: New York, 1985.
(55) Martin, E. J.; Blaney, J. M.; Siani, M. A.; Spellmeyer, D. C.; Wong, A. K.; Moos, W. H. *J. Med. Chem.* **1995**, *38*, 1431–1436.
(56) Zhao, Y.; Gran, B.; Pinilla, C.; Markovic-Plese, S.; Hemmer, B.; Tzou, A.; Whitney, L. W.; Biddison, W. E.; Martin, R.; Simon, R. *J. Immunol.* **2001**, *167*, 2130–2141.
(57) Jolliffe, I. T. *Principal Component Analysis*; Springer Series in Statistics 487; Springer: New York, 2002.
(58) Buslepp, J.; Zhao, R.; Donnini, D.; Loftus, D.; Saad, M.; Appella, E.; Collins, E. J. *J. Biol. Chem.* **2001**, *276*, 47320–47328.
(59) Clark, R. D. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1181–1188.
(60) Willett, P.; Barnard, J. M.; Downs, G. M. *J. Chem. Inf. Model.* **1998**, *38*, 983–996.
(61) Agrafiotis, D. K. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 159–167.
(62) Maggiora, G. M. *J. Chem. Inf. Model.* **2006**, *46*, 1535.
(63) Pirrung, M. C.; Pansare, S. V.; Sarma, K. D.; Keith, K. A.; Kern, E. R. *J. Med. Chem.* **2005**, *48*, 3045–3050.
(64) Nefzi, A.; Ostresh, J. M.; Appel, J. R.; Bidlack, J.; Dooley, C. T.; Houghten, R. A. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4331–4338.
(65) Boldt, J. L.; Pinilla, C.; Segall, A. M. *J. Biol. Chem.* **2004**, *279*, 3472–3483.
(66) Kundu, B.; Srinivasan, T.; Kesarwani, A. P.; Kavishwar, A.; Raghuwanshi, S. K.; Batra, S.; Shukla, P. K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1473–1476.
(67) Duyar, H.; Dengjel, J.; de Graaf, K. L.; Wiesmuller, K. H.; Stevanovic, S.; Weissert, R. *Immunogenetics* **2005**, *57*, 69–76.
(68) Lieberman, S. M.; Takaki, T.; Han, B.; Santamaria, P.; Serreze, D. V.; DiLorenzo, T. P. *J. Immunol.* **2004**, *173*, 6727–6734.
(69) Takaki, T.; Lieberman, S. M.; Holl, T. M.; Han, B.; Santamaria, P.; Serreze, D. V.; DiLorenzo, T. P. *J. Immunol.* **2004**, *173*, 2530–2541.
(70) Rubio-Godoy, V.; Ayyoub, M.; Dutoit, V.; Servis, C.; Schink, A.; Rimoldi, D.; Romero, P.; Cerottini, J.-C.; Simon, R.; Zhao, Y.; Houghten, R. A.; Pinilla, C.; Valmori, D. *Eur. J. Immunol.* **2002**, *32*, 2292–2299.
(71) Venturini, S.; Allicotti, G.; Zhao, Y.; Simon, R.; Burton, D. R.; Pinilla, C.; Poignard, P. *Eur. J. Immunol.* **2006**, *36*, 27–36.
(72) Lustgarten, J.; Dominguez, A. L.; Pinilla, C. *J. Immunol.* **2006**, *176*, 1796–1805.
(73) Sospedra, M.; Muraro, P. A.; Stefanova, I.; Zhao, Y.; Chung, K.; Li, Y.; Giulianotti, M.; Simon, R.; Mariuzza, R. A.; Pinilla, C.; Martin, R. *J. Immunol.* **2006**, *176*, 1951–1961.
(74) Sospedra, M.; Zhao, Y.; zur Hausen, H.; Muraro, P. A.; Hamashin, C.; de Villiers, E. M.; Pinilla, C.; Martin, R. *PLoS. Pathog.* **2005**, *1*, e41.
(75) Raghavan, S.; Yang, Z.; Mosley, R. T.; Schleif, W. A.; Gabryelski, L.; Olsen, D. B.; Stahlhut, M.; Kuo, L. C.; Emini, E. A.; Chapman, K. T.; Tata, J. R. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2855–2858.
(76) Choe, Y.; Leonetti, F.; Greenbaum, D. C.; Lecaille, F.; Bogyo, M.; Bromme, D.; Ellman, J. A.; Craik, C. S. *J. Biol. Chem.* **2006**, *281*, 12824–12832.
(77) Lee, Y.; Kang, D. K.; Chang, S. I.; Han, M. H.; Kang, I. C. *J Biomol. Screen.* **2004**, *9*, 687–694.
(78) Fugere, M.; Appel, J.; Houghten, R. A.; Lindberg, I.; Day, R. *Mol. Pharmacol.* **2007**, *71*, 323–332.
(79) Kacprzak, M. M.; Peinado, J. R.; Than, M. E.; Appel, J.; Henrich, S.; Lipkind, G.; Houghten, R. A.; Bode, W.; Lindberg, I. *J. Biol. Chem.* **2004**, *279*, 36788–36794.
(80) Fujimoto, D. F.; Pinilla, C.; Segall, A. M. *J. Mol. Biol.* **2006**, *363*, 891–907.
(81) Fujii, K.; Zhu, G.; Liu, Y.; Hallam, J.; Chen, L.; Herrero, J.; Shaw, S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13744–13749.
(82) Houghten, R. A.; Yu, Y. *J. Am. Chem. Soc.* **2005**, *127*, 8582–8583.

CC7001205