Discovery of Sequence-Selective Peptide Binding by Synthetic Receptors Using Encoded Combinatorial Libraries

W. CLARK STILL

Department of Chemistry, Columbia University, New York, New York 10027

Received September 21, 1995

Introduction

One of the most characteristic properties of biological macromolecules is that most bind their preferred substrates with extraordinarily high selectivity. While the antibodies of the immune system leap to mind in this context, virtually all biological systems rely on selective binding events; and, more specifically, many rely on the sequence-selective binding of oligopeptides or peptidic substructures of proteins. In recent decades, chemists have worked to create synthetic small molecules with similar binding properties. While such molecules would have applications in the pharmaceutical, diagnostic, and separations industries, the main driving force has been a quest for the basic principles that govern binding. Through many host/guest and molecular recognition studies, much has been learned about the nature of binding and the characteristic features of effective receptors or host molecules. Indeed, we can now design and prepare impressively selective receptors for many small organic molecules. The types of selectivities that have been designed into host molecules include functional group selectivity,[1] enantioselectivity,[2] and biooligomer residue selectivity.[3] However, the rational, structure-based design methods that have proven so useful in making receptors for small organics are not yet up to the task of designing more challenging receptor molecules that function like the binding sites of antibodies. So how are we to approach the next step of making synthetic molecules that behave like real biological receptors?

One of the main distinctions between chemists' host molecules and biological receptors (*e.g.*, antibodies) is the size and complexity of the substrates these molecules bind. Whereas typical host molecules bind one functional group or a part of a single biooligomer residue, most biological receptors bind large, multiple-residue arrays in a sequence- and/or conformation-dependent way. Thus one direction that might be taken is a move toward larger receptors: to bind a large substrate selectively, a receptor must interact with a significant proportion of its surface. Making significantly larger receptors can be a major challenge because the best receptors are expected also to incorporate conformationally restricted binding cavities: structural features that can be nontrivial to design and prepare in real molecules. But the problem exceeds size alone. Many biological substrates (*e.g.*, peptides) are also conformationally flexible. Thus even the three-dimensional geometry of the substrate one wishes to bind is often unknown. Stated this way, the prospect of creating synthetic molecules that behave like biological receptors in the near term sounds rather bleak. Actually, I think the prognosis for creating such molecules is quite good thanks to recent advances in the field of combinatorial synthesis.[4] These advances are providing techniques for dealing with problems like this one where too little is known for a purely deterministic, structure-based solution.

In this Account, I summarize my research group's work on a key question that must be answered before the creation of synthetic receptor molecules having antibody-like binding properties can become reality. That question is, How large and complex must a synthetic molecule be to bind oligopeptides sequence-selectively? To answer this question, my research group has been preparing a variety of host-like receptor molecules and studying their binding to large collections of oligopeptides that are prepared by encoded combinatorial synthesis. These experiments not only have established the size and nature of

Clark Still was born in Augusta, GA, in 1946. He received a B.S. in 1969 and a Ph.D. in 1972 from Emory University with David Goldsmith. After postdoctoral studies at Princeton University and Columbia University, he began his academic career at Vanderbilt University in 1975. He returned to Columbia in 1977, where he now works in the areas of molecular recognition, combinatorial chemistry, and computational chemistry.

(1) Examples of functional group selectivity. Ammonium ions: Garel, L.; Lozach, B.; Dutasta, J.-P.; Collet, A. *J. Am. Chem. Soc.* **1993**, *115*, 11652. Phosphates: Schiessl, P.; Schmidtchen, F. P. *J. Org. Chem.* **1994**, *59*, 509. Vance, D. H.; Czarnik, A. W. *J. Am. Chem. Soc.* **1994**, *116*, 9397. Ethers or ketones: Maruoka, K.; Nagahara, S.; Yamamoto, H. *J. Am. Chem. Soc.* **1990**, *112*, 6115. Bell, D. A.; Anslyn, E. V. *J. Org. Chem.* **1994**, *59*, 512. Acids: Fan, E.; Van Arman, S. A.; Kincaid, S.; Hamilton, A. D. *J. Am. Chem. Soc.* **1993**, *115*, 369. Hamann, B. C.; Branda, N. R.; Rebek, J. *Tetrahedron Lett.* **1993**, *34*, 6837. Albert, J. S.; Hamilton, A. D. *Tetrahedron Lett.* **1993**, *34*, 7363. Alcohols: Webb, T. H.; Suh, H.; Wilcox, C. S. *J. Am. Chem. Soc.* **1991**, *113*, 8554. Mendez, L.; Singleton, R.; Slawin, A. M. Z.; Stoddart, J. F.; Williams, D. J.; Williams, M. K. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 478. Huang, C.-Y.; Cabell, L. A.; Anslyn, E. V. *J. Am. Chem. Soc.* **1994**, *116*, 2778. Bell, D. A.; Dfaz, S. G.; Lynch, V. M.; Ansyln, E. V. *Tetrahedron Lett.* **1995**, *36*, 4155.

(2) Examples of enantioselection. Ammonium ions: Peacock, S. C.; Domeier, L. A.; Gaeta, F. C. A.; Helgeson, R. C.; Timko, T. M.; Cram, D. J. *J. Am. Chem. Soc.* **1978**, *100*, 8190. Petti, M. A.; Shepodd, T. J.; Barrans, R. E.; Dougherty, D. A. *J. Am. Chem. Soc.* **1988**, *110*, 6825. Li, Y.; Echegoyen, L.; Martinez-Diaz, M. V.; de Mendoza, J.; Torres, T. *J. Org. Chem.* **1991**, *56*, 4193. Diketopiperazines: Jeong, K.-S.; Muehldorf, A. V.; Rebek, J. *J. Am. Chem. Soc.* **1990**, *112*, 6144. Cristofaro, M. F.; Chamberlin, A. R. *J. Am. Chem. Soc.* **1994**, *116*, 5089. α-Amino acids: Burger, M. T.; Armstrong, A. A.; Guarnieri, F.; McDonald, D. Q.; Still, W. C. *J. Am. Chem. Soc.* **1994**, *116*, 3593. Mizutani, T.; Ema, T.; Tomita, T.; Kuroda, Y.; Ogoshi, H. *J. Am. Chem. Soc.* **1994**, *116*, 4240. References 10 and 11. See also: Pirkle, W. H.; Pochapsky, T. C. *Chem. Rev.* **1989**, *89*, 347.

(3) Examples of biopolymer residue selectivity. Nucleic acid bases: Park, T. K.; Schroeder, J.; Rebek, J. *J. Am. Chem. Soc.* **1991**, *113*, 5125. Park, T. K.; Schroeder, J.; Rebek, J. *Tetrahedron* **1991**, *47*, 2507. Kobayashi, K.; Asakawa, Y.; Kato, Y.; Aoyama, Y. *J. Am. Chem. Soc.* **1992**, *114*, 10307. Sugars: Bonnar-Law, R. P.; Davis, A. P.; Murray, B. A. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 1407. Kikuchi, Y.; Tanaka, Y.; Sutarto, S.; Kobayashi, K.; Toi, H.; Aoyama, Y. *J. Am. Chem. Soc.* **1992**, *114*, 10302. Coteron, J. M.; Vicent, C.; Bosso, C.; Penades, S. *J. Am. Chem. Soc.* **1993**, *115*, 10066. Liu, R.; Still, W. C. *Tetrahedron Lett.* **1993**, *34*, 2573. α-Amino acids: Murakami, Y.; Ohno, T.; Hayashida, O.; Hisaeda, Y. *J. Chem. Soc., Chem. Commun.* **1991**, 950. Reference 11.

(4) Reviews: (a) Pavia, M. R.; Sawyer, T. K.; Moos, W. J. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 387. (b) Gallop, M. A.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gordon, E. M. *J. Med. Chem.* **1994**, *37*, 1233. (c) Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. *J. Med. Chem.* **1994**, *37*, 1385.

0001-4842/96/0129-0155$12.00/0

molecules that exhibit sequence-selective peptide binding but have provided a convincing example of the power of combinatorial approaches to problems in molecular recognition.

Synthetic Receptors for α-Amino Acid Derivatives

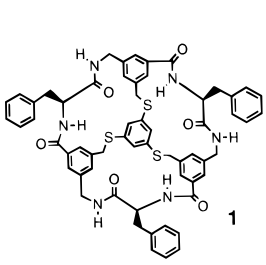
My research group has been concentrating on synthetic receptors or host molecules for simple peptides. Over the years, we and others have developed a working model that defines the structural features associated with effective receptor molecules. Generally, such molecules are preorganized[5] to fit the preferred substrate both sterically and electronically, and they have a thermodynamic driving force for binding that is appropriate to the relevant solvent medium. I summarize them below.

First of all, effective receptors are characterized by energetically accessible conformers having concave binding sites that are complementary in size and electrical charge distribution to the substrates they bind (Cram's preorganization[5]). The best receptors have a binding site that is just large enough (and appropriately shaped) to accommodate the substrate: too small a binding site obviously prevents binding, but too big is also bad for binding. Electrical complementarity can involve ion pairing (salt bridges) but more commonly is incorporated as receptor−substrate hydrogen bonding. Such intermolecular hydrogen bonding can be built into complexes by designing receptors having spatially separated donor and acceptor groups that disfavor intramolecular hydrogen bonding structurally. A key requirement for high selectivity is that the binding site be structurally well defined: a binding site that exists in many structurally distinct conformers is likely to bind many different substrates.[6]

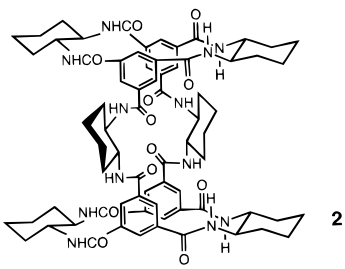
Equally important to receptor−substrate complementarity is an energetic driving force for binding, and that depends largely on the solvent medium. While the thermodynamics of binding depend upon a subtle balance of complex enthalpic and entropic effects, the following generalities are helpful in designing many molecular complexes. In water, the driving force for binding is usually the removal of hydrophobic surface area from exposure to the aqueous medium. In less polar organic solvents (*e.g.*, chloroform), the driving force is typically electrical and often derives from the formation of intermolecular hydrogen bonds. In either case, the energetic driving force must be quite large for binding to occur (to be favorable in *free energy*) because it must overcome the large loss of entropy (20−30 cal/(deg·mol) or 6−9 kcal/mol at room temperature) that accompanies bimolecular complex formation. Such considerations imply that significant binding will be found only if accompanied by the formation of at least two good hydrogen bonds (in a nonpolar organic solvent) or with burial of at least 200 Å^2 of hydrophobic, solvent accessible surface area (in water).

In the following paragraphs, I describe several synthetic receptor molecules for peptides that we have designed using the above ideas as working hypotheses. Other work in this general area has been summarized in several recent reviews.[7] Some of our receptors were made to bind substrates in organic solvents, and some were tailored for water. In all cases, the design procedure was the same. We used readily available organic molecules of limited conformational flexibility that could be assembled into a structure that incorporated a large, conformationally well defined cavity or cleft. Usually such molecules incorporated macrocycles or conformational locking mechanisms[8] to restrict conformational heterogeneity. Since conformational stability in such structures cannot be estimated reliably by simple inspection, we turned to computational methods involving conformational searching[9] to distinguish those structures that were highly flexible (and thus poor candidates for selective synthetic receptors) from those that were not. Structures without a low-energy binding cavity or cleft were discarded or modified in an attempt to induce such a binding site. We also sought structures having unassociated hydrogen bond donors and acceptors in or near the binding site that would orient the bound substrate and thus favor high substrate selectivity.

The structures of two of the synthetic receptors we thus designed and prepared are shown herein **1**[10] and **2**.[11] These molecules were very simple to prepare. The key step in the synthesis of **1** was a triple macrolactamization that proceeded in >60% yield on a gram scale. Compound **2** is even easier to make: it forms in one step (13% yield) on mixing commercially available (*R*,*R*)-cyclohexane-1,2-diamine and trimesic acid chloride. These synthetic receptors have little conformational flexibility according to molecular mechanics. Their significantly populated conformations include ones characterized by deep binding cavities with peripheral arrays of hydrogen bond donors and acceptors for interacting with donor/acceptor substrates such as peptides.

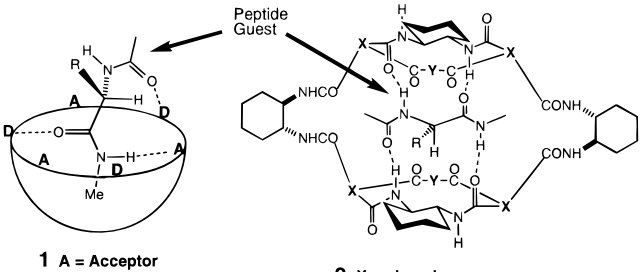


1



2

The binding of these molecules to simple α-amino acid derivatives appears to involve the formation of three or four intermolecular hydrogen bonds. ^{1}H NMR studies of the peptide complexes of **1** and **2** in chloroform suggest the binding modes indicated generally in the cartoons below:



1 A = Acceptor
D = Donor

2 X = phenyl
Y = diaminocyclohexane

The design of **1** and **2** was originally aimed at enantioselective receptors for amino acid derivatives.

(5) Cram, D. J. *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 1009.

(6) *Caveat*: Structural rigidity in a binding site favors highly selective binding only if a substrate precisely fits the binding site. Because designing host molecules with such a high-precision fit is often difficult, some flexibility is desirable. But extensive conformational heterogeneity is not.

Indeed, binding experiments with single D- and L-amino acid amides showed that both **1** and **2** bound such substrates with selectivity for the L configuration in the 90−99% ee range. These experiments also established that **1** and **2** had large preferences for amino acids having certain substituents or side chains. Receptor **1**, for example, preferentially bound terminal (L)-amino acids bearing a small main chain substituent which it pulled into the binding cavity as indicated in the cartoon above. Receptor **2**, on the other hand, preferentially bound amides of L-valine (Val) and L-phenylglycine (PGly), but not L-phenylalanine (Phe), in spite of the fact that Phe and PGly differ by only one methylene. In the case of **2** with Val or PGly, it was the side chain (R) that occupied the receptor binding cavity.

In this Account, however, we are more interested in the possibility that **1**, **2**, or related compounds might bind oligopeptides sequence-selectively. Receptor **2** in particular would seem to have the potential to bind an oligopeptide because it contains additional amide groups that do not appear to be involved in binding single amino acid substrates. Indeed, our early studies with **2** indicated that the tripeptide Ac-Gly-(L)Val-Gly-NHMe was bound much more tightly than Ac-(L)Val-NHMe.

While it is possible (though tedious) to determine the binding of a receptor for various single amino acid substrates one at a time, there are so many di- and tripeptides that individual analyses of even a simplified, representative set of peptides is not really feasible. The solution to this problem is to prepare many thousands of different oligopeptides by encoded combinatorial synthesis and then to screen them all for binding in a single experiment. I describe how this is done in the following section.

Encoded Combinatorial Substrate Libraries

One of the most powerful techniques available for preparing combinatorial libraries is solid phase *split synthesis*. It was developed to make large collections (*libraries*) of peptides.[12] The result of a split synthesis is a collection of synthesis beads, each of which bears a single library member, which in the case of a peptide library corresponds to one sequence of amino acids. Thus split synthesis has often been described as a "one bead, one peptide" method. The beauty of the method is that it is simple and effective. It leads naturally to a library having individual library members segregated on individual synthesis beads and (in the limit of a large number of beads) to the formation of library members consisting of every possible combination of every amino acid used in the synthesis. Thus split synthesis using 20 different amino acids at each site of a pentapeptide would yield as many as 20^5 or 3 200 000 different pentapeptides.

As originally described, split synthesis has one serious limitation: it is only applicable to the synthesis of sequenceable oligomers. The problem is that one synthesis bead carries only ~100 pmol (~10^{13} molecules) of each library member. While that quantity is adequate for modern Edman sequencing of a small peptide, it is generally too little for structure determination with more complex types of molecules. In simple systems, modern mass spectroscopic techniques have been used for structure elucidation of compounds from single synthesis beads.[13] Other approaches to structure determination using mixtures of compounds or spatial segregation of library members have also been described.[4b,c] However, when libraries contain many thousands or millions of different compounds, many of which may be impure or isomeric in some way with other library members, then only one method would seem to solve the structure elucidation problem. That method is *encoding*.

The idea behind encoding follows from the way many proteins are sequenced: not by sequencing the protein itself but by sequencing the gene that encodes for it.[14] Thus, while a library member itself may not be analyzable, it is possible to associate certain unique chemical markers or *tags* with that library member to specify its identity. The idea of encoding synthetic information with a chemical tag was first proposed in the literature by Brenner and Lerner.[15] In the earliest implementations, the tag was a sequenceable biopolymer (DNA or oligopeptide), and the encoded information was contained in the sequence of residues in the biopolymer tag.[16] Because of worries about the stability of such a tag to the often vigorous conditions of organic synthesis, my group developed a different tagging method based on multiple, chemically inert small molecule tags.[17] During the past two years, that method has been used to make many dozens of different libraries and has proven itself to be a robust and practical method of encoding. I describe it in the following paragraphs.

Split synthesis makes one and only one library member on a particular bead because any particular library member (obviously) results from treatment of some individual synthesis bead with one particular set of chemical reagents. Encoding involves attaching

(7) Reviews: Schneider, H.-J. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 848. Webb, T. H.; Wilcox, C. S. *Chem. Soc. Rev.* **1993**, 383.

(8) Iimori, T.; Erickson, S. D.; Rheingold, A. L.; Still, W. C. *Tetrahedron Lett.* **1989**, *30*, 6947. Alder, R. W.; Maunder, C. M.; Orpen, A. G. *Tetrahedron Lett.* **1990**, *31*, 6717. Hoffmann, R. W.; Sander, T.; Brumm, M. *Chem. Ber.* **1992**, *125*, 2319. Hoffmann, R. W. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 1124. Tang, S.; Still, W. C. *Tetrahedron Lett.* **1993**, *34*, 6701.

(9) Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379.

(10) (a) Hong, J.-I.; Namgoong, S. K.; Barnardi, A.; Still, W. C. *J. Am. Chem. Soc.* **1991**, *113*, 5111. (b) Liu, R.; Still, W. C. *Tetrahedron Lett.* **1993**, *34*, 2573. (c) Erickson, S. D.; Simon, J. A.; Still, W. C. *J. Org. Chem.* **1993**, *58*, 1305.

(11) Yoon, S. S.; Still, W. C. *J. Am. Chem. Soc.* **1993**, *115*, 823. Yoon, S. S.; Still, W. C. *Tetrahedron Lett.* **1994**, *35*, 2117.

(12) (a) Furka, A.; Sebestyen, M.; Asgedom, M.; Dibo, G. *Abstr. 14th Int. Congr. Biochem., Prague, Czechchoslovakia* **1988**, *5*, 47. (b) Furka, A.; Sebestyen, M.; Asgedom, M.; Dibo, G. *Abstr. 10th Int. Symp. Med. Chem., Budapest, Hungary,* **1988**, 288; (c) Furka, A.; Sebestyen, M.; Asgedom, M.; Dibo, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487. (d) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354*, 82 (see also: Nature **1991**, *358*, 434). (e) Zuckermann, R. N.; Kerr, J. M.; Siani, M. A.; Banville, S. C. *Int. J. Pept. Protein Res.* **1992**, *40*, 498. See also: Lebl, M.; Krchnak, V.; Sepetov, N.; Seligmann, B.; Strop, P.; Felder, S.; Lam, K. S. *Biopolymers (Pept. Sci.)* **1995**, *37*, 177.

(13) Brummel, C. L.; Lee, I. N. W.; Zhou, Y.; Benkovic, S. J.; Winograd, N. *Science* **1994**, *264*, 399. Zambias, R. A.; Boulton, D. A.; Griffin, P. R. *Tetrahedron Lett.* **1994**, *35*, 4283. Youngquist, R. S.; Fuentes, G. R.; Lacey, M. P.; Keough, T. *J. Am. Chem. Soc.* **1995**, *117*, 3900.

(14) See, for example: Oldenburg, K. R.; Loganathan, D.; Goldstein, I. J.; Schultz, P. G.; Gallop, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5393.

(15) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5381.

(16) Tagging with biopolymers: (a) Kerr, J. M.; Banville, S. C.; Zuckermann, R. N. *J. Am. Chem. Soc.* **1993**, *115*, 2529. (b) Nikolaiev, V.; Stierandova, A.; Krchnak, V.; Seligmann, B.; Lam, K. S.; Salmon, S. E.; Lebl, M. *Pept. Res.* **1993**, *6*, 161; (c) Nielsen, J.; Brenner, S.; Janda, K. D. *J. Am. Chem. Soc.* **1993**, *115*, 9812.

(17) Tagging with chemically inert organic small molecules: (a) Ohlmeyer, M. H. J.; Swanson, R. N.; Dillard, L. W.; Reader, J. C.; Asouline, G.; Kobayashi, R.; Wigler, M.; Still, W. C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10922. (b) Nestler, H. P.; Bartlett, P. A.; Still, W. C. *J. Org. Chem.* **1994**, *59*, 4723.

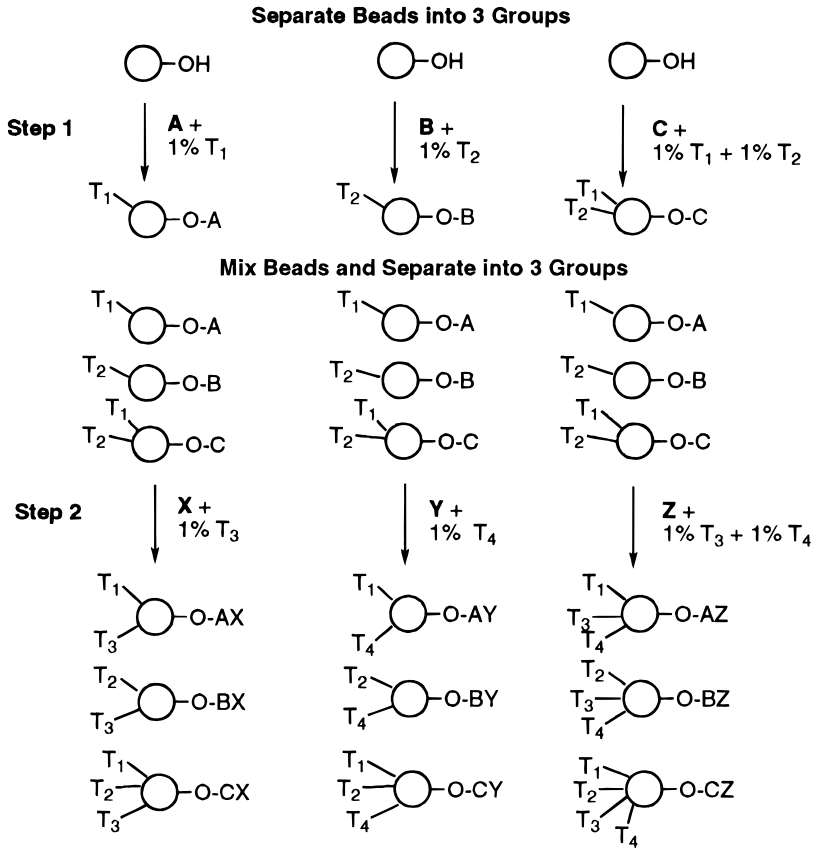


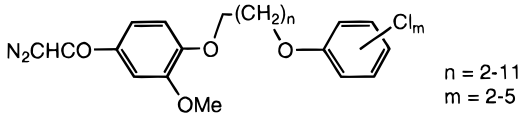
Figure 1. Split synthesis encoded with multiple tagging molecules (T_1−T_4).

unique arrays of readily analyzable molecular tags to each bead that designate the particular set of reagents used in the synthesis of that bead. Thus, groups of tags defining each reagent used are attached to beads at each combinatorial step in a split synthesis and create a tag-encoded record of the reagents used in the synthesis of that particular library member. Thus analyzing any bead for its tag content (*decoding*) yields the recipe for the synthesis of that type of bead.

The actual scheme we use[17] is best described with a simple split synthesis example that uses two steps with three alternative reagents for each step. Such a synthesis would yield $3 \times 3 = 9$ different library members. First imagine that you have four tagging molecules (T_1−T_4) that can be easily distinguished and analyzed, and that can somehow be attached to synthesis beads (described later). The scheme is diagramed in Figure 1; the synthesis beads are represented by circles, and their reactive functionality is represented by −OH. The first step in the synthesis is accomplished by first splitting the starting mass of beads into three fractions and placing each fraction into a different reaction vessel, one vessel for each of the three different reagents (here A, B, and C) to be used in the first synthetic step. After the reactions are carried out with the different reagents, the three flasks will contain three kinds of beads bearing −OA, −OB, and −OC, respectively. Next we do the encoding. As shown in the figure, we attach T_1 to beads in the first vessel, T_2 to beads in the second vessel, and a 1:1 mixture of T_1 and T_2 to beads in the third vessel. Thus each different kind of bead has a different array of tags, and if any bead were picked at random, the reagent used to make that bead could be determined by analyzing the tag content of that bead. While one could use a different tag for each reagent, it is much simpler to use mixtures of tags because mixtures of *N* different tags can represent 2^N different reagents. Thus only seven tags are needed to represent as many as 128 (2^7) different reagents.

The next step in the example is the *mix and split* procedure that characterizes split synthesis. Here, all the beads are combined, thoroughly mixed, and split into new fractions for the next step. In our example, there are three reagents (X, Y, and Z) in the second combinatorial step of the synthesis, so we split the mixed beads into three fractions. After placing each fraction in a different reaction vessel, we carry out the three reactions. In the first vessel, reagent X gives −OAX, −OBX, and −OCX. In the second vessel, where reagent Y is used, −OAY, −OBY, and −OCY are produced, and so on. Encoding is then performed using T_3 and T_4 as shown. This procedure thus yields nine different kinds of beads, each having a different library member and a unique array of synthesis-defining tags. Though this is a very simple example, the procedure is readily extendable. Using more combinatorial steps, more alternative reagents, and more tagging molecules, one can make libraries having as many as 10^9 members.

In principle, almost any type of molecule can be used as a tag. However, there are practical limitations because tags need to be chemically inert and reliably analyzed on femtomolar scales from single synthesis beads. They also need a mechanism for attachment to the synthesis beads. One class of compounds that fulfills these requirements are the diazoketones shown below:[17b]



Such reagents are readily preparable in 40 different forms (T_1-T_{40}) using 10 different 1,ω-diols, four different chlorophenols, and vanillic acid. The diazoketone serves to attach the tag to the polymer backbone of the synthesis beads via the corresponding reactive acylcarbene (generated using rhodium trifluoroacetate). The catechol diether allows release of the actual tag ($HO(CH_2)_{n+1}OArCl_m$) for the decoding analysis upon oxidation (ceric ammonium nitrate). We have also employed related tagging reagents that allow photochemical tag release.[17a] The tags themselves are analyzed by capillary gas chromatography using electron capture detection (ECGC), a detection method that is particularly sensitive for heavily chlorinated aromatics. The high detection sensitivity of ECGC for such molecules allows each tag to be used in a quantity that is quite small relative to that of the library member.

In the case of peptide libraries built on polystyrene or poly(ethylene glycol)/polystyrene synthesis beads, the acylcarbene tags add with little selectivity to both the polymer bead and its attached library member. But because the bulk of the bead material is the polymer matrix, it is the polymer that picks up most of the tag via an addition reaction involving attack on the aromatic rings of the polystyrene. Furthermore, tags are added at molar levels (~1 pmol/bead) corresponding to ~1% of the library members. Thus even in a worst case scenario with the acylcarbene tag avoiding the polymer matrix altogether and attacking only the library member, the bulk of the material on the bead after tagging would still be the unmodified library member.

Carrying out encoded split synthesis as described above is a simple business once an adequate stock of tagging molecules is synthesized. We have been using it routinely for library production since 1993.

Screening Receptors for Peptide Binding

To establish the binding properties of our synthetic receptors, we use encoded split synthesis to make substrate libraries of variously substituted tripeptides. These libraries have included D and L-amino acids, unnatural amino acids, and a variety of N-terminal substituents. All of these elements were encoded using the ECGC-readable tags described above.

The general structure of our tripeptide substrate libraries is given below:

$$R\text{-}AA_3\text{-}AA_2\text{-}AA_1\text{-polymer bead} = \mathbf{L}$$

Here R represents an N-terminal substituent (*e.g.*, an acyl group), AA_{1-3} represents three amino acids numbered in the order of addition to the bead, and the polymer bead is a standard solid phase synthesis bead composed of polystyrene (PS, used to study binding in organic solvents) or poly(ethylene glycol)/polystyrene (PEG-PS, used to study binding in water). Two of the libraries we prepared (for binding in chloroform and water, respectively) are defined by the general library formula **L** and the residues listed below.

library 1 (on PS)	libary 2 (on PEG-PS)
R = Ac, EtCO, iPrCO, tBuCO, iBuCO, neoPeCO, CF_3CO, $MeOCH_2CO$, $AcOCH_2CO$, cycloPrCO, cycloBuCO, cycloPeCO, PhCO, Me_2NCO, morpholinoCO	R = Ac
AA_{1-3} = Gly, (D)Ala, (L)Ala, (D)Ser, (L)Ser, (D)Val, (L)Val, (D)Pro, (L)Pro, (D)Asn, (L)Asn, (D)Gln, (L)Gln, (D)Lys, (L)Lys	AA_{1-3} = Gly, (D)Ala, (L)Ala, (D)Ser, (L)Ser, (D)Val, (L)Val, (D)Pro, (L)Pro, (D)Asn, (L)Asn, (D)Gln, (L)Gln, (D)Lys, (L)Lys, (D)Phe, (L)Phe, (D)Leu, (L)Leu, (D)Asp, (L)Asp, (D)Glu, (L)Glu, (D)His, (L)His, (D)Thr, (L)Thr, (D)Arg, (L)Arg

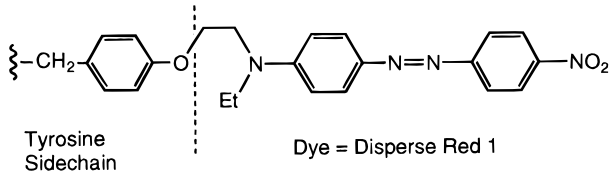
Because split synthesis yields all possible combinations of all reagents, library 1 (**L1**) had $15^4 = 50\,625$ members and library 2 (**L2**) had $29^3 = 24\,389$ members. These libraries were prepared using Fmoc chemistry with standard side-chain protection (Ser, Thr, Asp, Glu = *O*-tBu; Asn, Gln, His = *N*-Tr; Lys = *N*-Boc; Arg = *N*-Pmc) starting from commercial PS and PEG-PS synthesis beads carrying reactive primary amino groups. The libraries were screened for receptor binding both before and after side-chain deprotection. Libraries **L1** and **L2** were encoded with 16 and 15 different tagging molecules, respectively.

Given a library, the next step is to select those library members that have a desired property. This is a challenging area for future work because a large library essentially necessitates *selection by inspection*. The problem is that there can be so many different beads and compounds in a library that individual analyses are impractical. What is needed is the ability to find those beads that have the desired property simply by looking at them. In the case of binding, such a distinction can be easy.[18]

To find library beads having substrates that bind our receptors, we label a receptor in some easily identifiable way, equilibrate the receptor with the library, and then pick those library beads that accumulate the label. This is conveniently done with a colored dye.[19] Thus we chemically attach a commercial dye to each of our receptors and then treat solutions of those dye-labeled receptors with a substrate library on beads. When these solutions are dilute (1−100 μM), only those library members that bind the receptors most tightly will take on the color of the dye. Such colored beads can then be picked and their tags decoded to determine the structures of the preferred substrates.

Sequence-Selective Peptide Binding Discovered in Chloroform

We carried out our first combinatorial library binding assay using a labeled derivative of the C_3-symmetric receptor **1**.[20] This compound was prepared by resynthesis of an analog of **1** having the (L)Phe fragment replaced by *O*-allyl (L)Tyr.[10c] We then deprotected the three phenolic side chains and alkylated with a bright red azo dye to give the labeled receptor (red-**1**) having the dye appendages shown below:



The binding assay was carried out using the 50625-member, side-chain-protected library **L1** by agitating the library with 50 μM red-**1** in chloroform for 24 h. Under these conditions, only a small fraction of **L1** bound the receptor as indicated by the fact that only ~0.5% of the beads developed the deep red coloration that we have come to associate with tight binding. We

(18) This method is analogous to ones described for establishing the binding properties of antibodies: Geysen, H. M.; Mason, T. J. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 373 and references therein.

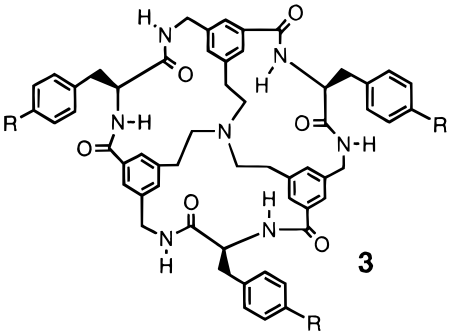
(19) Lam, K. S.; Zhao, Z.-G.; Wade, S.; Krchnak, V.; Lebl, M. *Drug. Dev. Res.* **1994**, *33*, 157. Dyes must be chosen with care because they can exhibit significant (and selective) binding themselves. Wennemers, H.; Still, W. C. *Tetrahedron Lett.* **1994**, *35*, 6413.

(20) Borchardt, A.; Still, W. C. *J. Am. Chem. Soc.* **1994**, *116*, 373.

picked 50 of these deep red beads and decoded their tags using ECGC.

In some ways, the tripeptides bound by red-**1** were anticipated from our previous, single amino acid binding experiments. Thus the N-terminal acylating agent was generally small and the N-terminal amino was always L (or Gly). On the other hand, the library screening showed much binding selectivity that previous binding experiments had not revealed. In particular, in 98% of the red beads, the terminal acylating group RC═O had exactly three non-hydrogen atoms in R (cyclopropyl, dimethylamino, and methoxymethyl). Thus red-**1**'s binding site is exceptionally selective for one particular size of R. The N-terminal amino acid (AA_3) was not only L but was (L)Ala or (L)-Gln(*N*-Tr) in 72% of the red beads. Interestingly, **L1** also contained closely related (L)Asn(*N*-Tr) at AA_3 but that residue was never found among the red beads. The penultimate substrate residue (AA_2) was either (L)Pro or (L)Ala in 68% of the beads picked, though many other residue types showed up there with low frequencies. There was virtually no selectivity for the C-terminal residue (AA_1). These selectivities were verified by resynthesis of certain tripeptides and solution phase binding experiments with **1**. These solution phase studies also established that beads acquiring deep red colorations in the solid phase binding assay carry peptides that bind **1** with significant binding energies ($-\Delta G_{binding} = 4-6$ kcal/mol in $CDCl_3$). Thus receptor **1** shows significant selectivity that spans three residues—the N-terminal acylating group and next two amino acids—a result we regard as remarkable considering the small dimensions of **1**'s binding site.

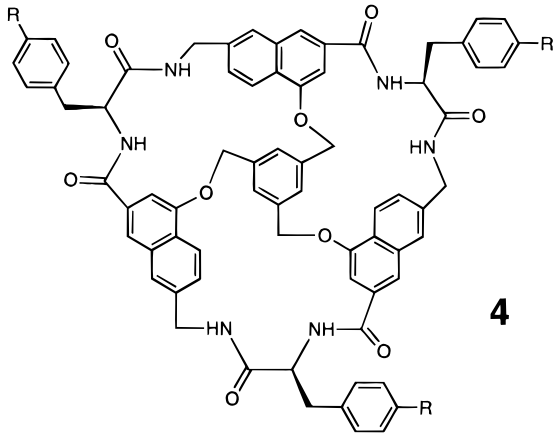
In contrast with **1**, whose binding cavity is lipophilic, receptor **3** (R = O−Disperse Red 1) carries a binding cavity with a tertiary amine that could act as a hydrogen bond acceptor and thus bind peptide side chains carrying donor functionality (*e.g.*, carboxamides from Asn or Gln).[21] When it was equilibrated with



3

side-chain-deprotected **L1** in chloroform ([**3**] ~100 μM), highly selective binding was observed in that ~0.2% of the beads accumulated the red color of **3**. ECGC decoding of those beads showed that **3** was most selective for the AA_2 and AA_1 residues at the C-terminal end of the tripeptidic substrates. Thus we found that AA_2 was (D)Pro or Gly, while AA_1 was (D)-Asn or (D)Ser. Even more interesting than the selectivity for two of the 15 possible residues for these sites was the fact that (D)Pro was always found with (D)-Asn, and Gly was always found with (D)Ser. Thus **3** selectively binds dipeptides (D)Pro-(D)Asn (~40% of the red beads) and Gly-(D)Ser (~20% of the red beads). ^{1}H NMR studies of the preferred (D)Pro-(D)Asn complex with **3** suggest that it binds via insertion of the asparagine side chain into **3**'s binding site to allow hydrogen bonding with the receptor's tertiary amine.

Receptor **4** is closely related to **1** but has a much larger binding cavity.[22] When **4** was equilibrated with **L1** in chloroform, we again found sequence-selective peptide binding in the form of a high preference for an internal (L)Pro (AA_2) flanked by two L-amino acids.

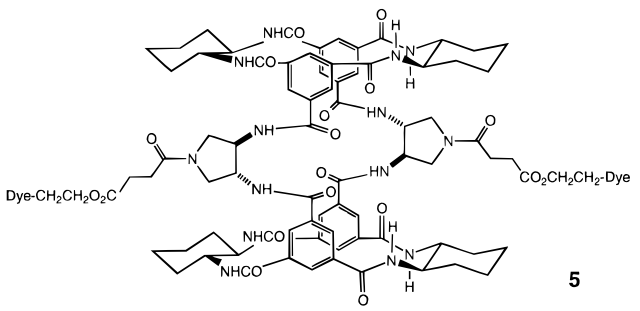


4

When such peptides were resynthesized and binding with **4** was studied in chloroform solution, we found strong evidence from ^{1}H NMR experiments that the preferred (L)Pro was binding within **4**'s lipophilic binding cavity. Reminiscent of **1**'s sensitivity to the size of the R that bound within its binding cavity, **4**'s binding was highly dependent on the size of the AA_2 residue, with four- or six-membered analogs binding 1−2 kcal/mol more weakly than the corresponding five-membered Pro-containing tripeptides.

Binding studies with the above C_3-symmetric receptors and others[23] indicate that relatively simple host-like receptors can bind short oligopeptides with sequence selectivity that extends over as many as three residues. Considering the small dimensions of the binding sites involved, the selectivity is remarkable. It appears to result from fixation of a small part of the peptide substrate by the receptor's concave binding site and association of proximate amino acids by hydrogen bonding to receptor functionality around the binding site periphery. These results also indicate the power of the encoded combinatorial library methodology: many of the novel binding selectivities we discovered would have been difficult if not impossible to find without it.

The sequence-selective binding described above has involved a series of C_3-symmetric receptors that are rather similar in structure. To show that other receptor topologies can also bind peptides sequence-selectively, I summarize related library binding studies with a D_2-symmetric receptor based on **2**. Its structure (**5**) is shown below:[24]



5

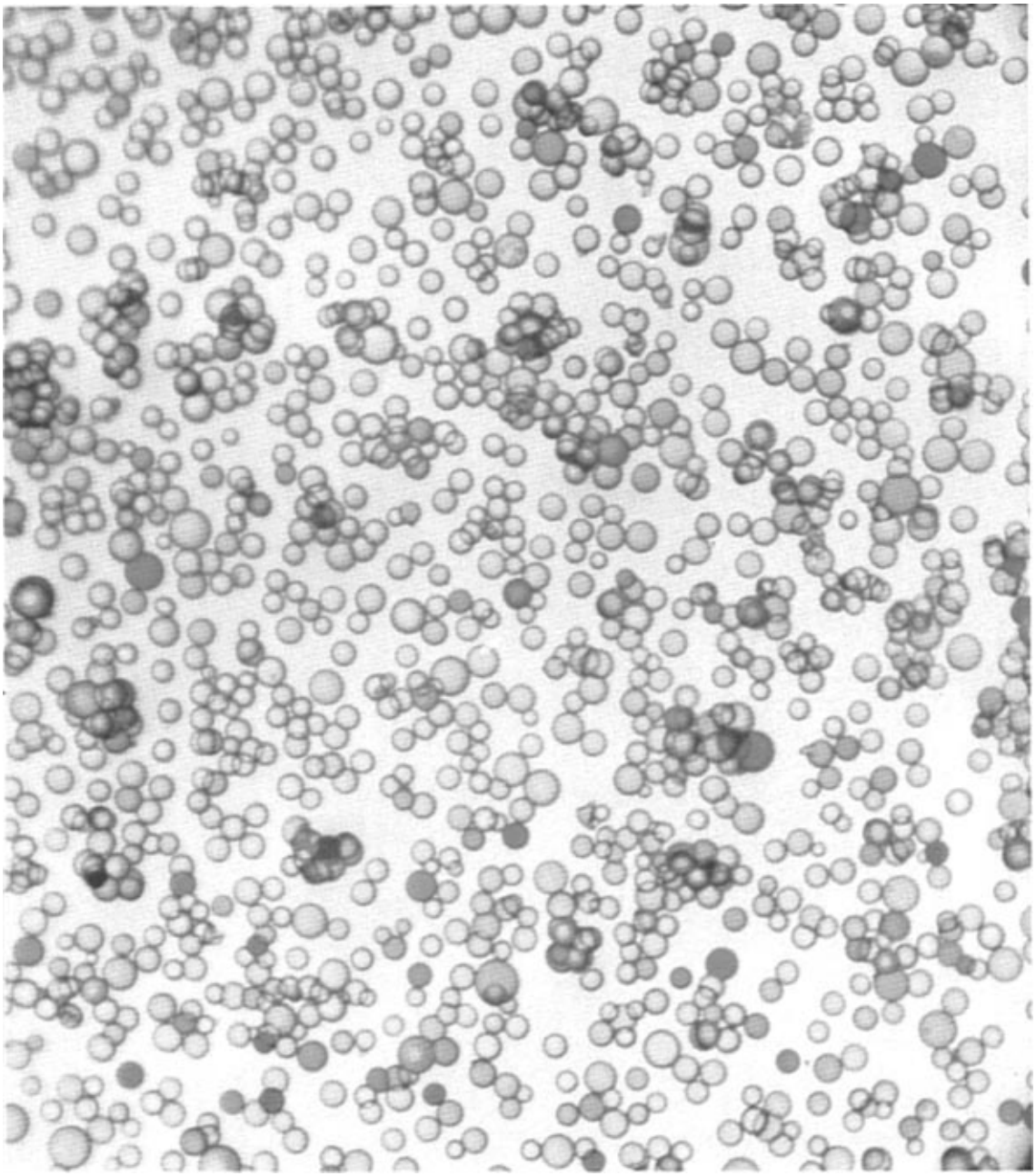


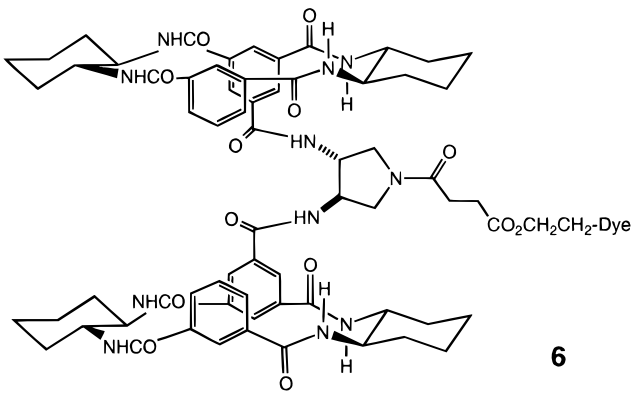
Figure 2. L1 beads upon equilibration with **5** in $CHCl_3$.

Previously described solution phase binding studies[11] had found that **2** bound particular L-amino acids such as (L)Val and L-phenylglycine with high side-chain selectivity and enantioselectivity. Solid phase binding studies with **5** showed similar selectivities for (L)Val, but they uncovered highly sequence-selective binding as well. In particular, **5** tightly bound ~1% of the **L1** tripeptides in chloroform; a photomicrograph of the binding assay is shown in Figure 2. Note that some beads are not colorless but are colored less intensely than others. Solution phase binding experiments indicate that these beads bear peptides that bind as little as 1 kcal/mol less tightly than peptides on the reddest beads. When the reddest beads were picked and decoded by ECGC, we found that 77% of the associated peptides had (L)Val somewhere in their sequence. Interestingly, 95% of those beads having AA_1 or AA_2 = (L)Val also had a (D)-carboxamide-bearing amino acid (*i.e.*, (D)Asn or (D)Gln) as the N-terminal substituent. Furthermore, for such AA_3-AA_2 = (D)Asn/Gln-(L)Val sequences, AA_1 was (L)Ser in 52% of such sequences. Thus **5** shows marked selectivity for binding certain oligopeptidic substructures as large as tripeptides.

Changing receptor topology yet again, we simplified the structure of **5** by eliminating one of the macrocyclically linking diamines to yield analog **6**.[25] When equilibrated with **L1** in chloroform, **6** showed exceptionally high peptide-binding selectivity. Figure 3 shows the result of such equilibration: not only are only a few beads deeply colored (<0.1% of all beads), but there is also high contrast between the deeply colored beads and all other beads. This high contrast implies that these nearly colorless beads carry pep-

(21) Carrasco, M. R.; Still, W. C. *Chem. Biol.* **1995**, *2*, 205.
(22) Borchardt, A.; Still, W. C. *J. Am. Chem. Soc.* **1994**, *116*, 7467.
(23) Yoon, S. S.; Still, W. C. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 2458.
(24) Yoon, S. S.; Still, W. C. *Tetrahedron* **1995**, *51*, 567.
(25) Wennemers, H.; Yoon, S. S.; Still, W. C. *J. Org. Chem.* **1995**, *60*, 1108.

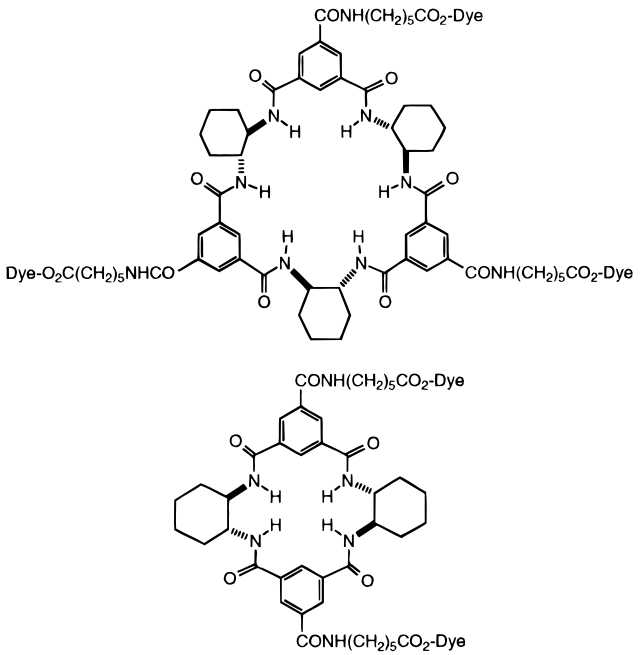
Figure 3. L1 beads upon equilibration with **6** in $CHCl_3$.



6

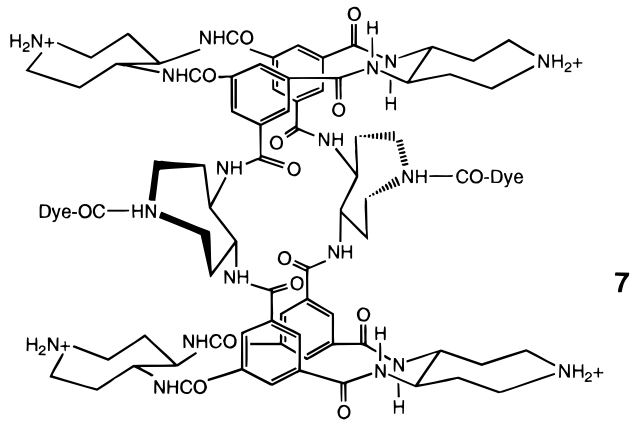
tides that bind **6** much less tightly than do the few red beads (compare Figures 2 and 3). When the deeply colored beads were picked and decoded, 78% of the beads were found to carry only two tripeptide sequences: AA_3-AA_1 = (D)Pro-(L)Val-(D)Gln and (L)Lys-(L)Val-(D)Pro.

The results above establish that a variety of rather simple host-like molecules can bind peptides with significant sequence-selectivity in chloroform. While these molecules are all characterized by macrocyclic structures or substructures and hydrogen-bonding amide linkages, not all such molecules bind peptides. The two related macrocyclic compounds below, for example, do not visibly bind any beads in **L1** at concentrations as high as 0.1 M in chloroform.



Sequence-Selective Peptide Binding in Water

As our ultimate objective is to bind peptidic substrates in the biologically more relevant solvent water, we have also started to develop water-soluble receptors and have studied their binding properties using the encoded tripeptide library **L2** on water-swellable PEGPS beads. The first such receptor (**7**) we prepared is a hydrophilic analog of **5**.[26] For practical reasons, **7**



7

was synthesized in the enantiomeric series opposite to that of **5**, and when it was equilibrated with **L2** in pH 4 water, highly selective binding was again observed. In particular, 0.1−0.2% of the beads in both side-chain-protected and -deprotected **L2** were stained scarlet by rhodamine-labeled **7** at a concentration of ~10 μM. With side-chain-protected **L2**, every stained bead contained a (D)Val or (D)Leu somewhere in the tripeptide sequence. In addition, ~80% of these beads had an (L)Asn adjacent to the (D)Val/Leu. In this regard, the peptide-binding selectivity of **7** in water is remarkably close to that of **5** in chloroform. With side-chain-deprotected **L2**, two types of sequences were bound. One was stereorandom Asp-Asp-Asp, and it reflects ionic association with the cationic **7**. The other sequence was (D)Asp/Glu-(D)Leu-(L)Asp/Glu. It too is favored by ionic interactions, but the high selectivity for (D)Leu at AA_2 and the selectivity for a D carboxylic acid at AA_3 and an L carboxylic acid at AA_1 is notable.

Though our work on receptors for peptides in water is just beginning, results with **7** and others to be reported in the near future suggest that highly sequence-selective peptide binding in water is feasible with small-molecule synthetic receptors.[27]

Conclusion: Synthetic Receptor Design and Combinatorial Chemistry

The results described above and others[28] clearly establish that simple, host-like molecules can bind peptides with significant sequence-selectivity. Thus the prognosis is good for creating synthetic small molecules with binding properties similar to those of natural antibodies. While selectivity appears highest for those parts of a peptidic substrate that bind within a receptor binding cavity, significant selectivity for residues proximate to the binding site is also observed when those residues can interact with external receptor functionality. Because such external interactions may involve flexible substructures and thus be difficult to design into complexes, combinatorial methods provide a powerful approach for their exploitation.

More generally, combinatorial synthesis and screening offers an effective approach to many chemical problems where too little is known for a deterministic solution. It also provides an important tool for testing specific hypotheses via batteries of measurements that can be applied in single experiments and thus to test hypotheses in many different contexts. For example, the (D)Asn selectivity of **3** that we sought would have been difficult to find directly because it is most effective when the (D)Asn is preceded by (D)Pro. In any case, it is clear that combinatorial synthesis and screening is a highly useful tool for molecular discovery. We have used it here to answer a question about the general structure of sequence-selective peptide-binding small molecules. We are now using it to find small-molecule receptors for particular peptide substrates via preparation of encoded combinatorial libraries of receptors. In these libraries, we use what we know about effective peptide receptors to design a relevant, basic receptor structure and then vary that structure combinatorially to make up for what we do not know about binding a particular substrate.

The work described above has been supported by Grant CHE92 08254 from the National Science Foundation. I also acknowledge Mike Ohlmeyer, Larry Dillard, and John Reader, who developed the ECGC encoding methodology that made the results described here possible. Finally, I wish to note the seminal contributions of Allen Borchardt, who developed the combinatorial binding methodology that revealed for the first time the sequence-selective peptide-binding properties of synthetic small-molecule receptors.

AR950166I

(26) Torneiro, M.; Still, W. C. *J. Am. Chem. Soc.* **1995**, *117*, 3887.
(27) See also: LaBrenz, S. R.; Kelly, J. W. *J. Am. Chem. Soc.* **1995**, *117*, 1655.
(28) Yoon, S. S.; Still, W. C. *Tetrahedron Lett.* **1994**, *35*, 8557. Boyce, R.; Li, G.; Nestler, H. P.; Suenaga, T.; Still, W. C. *J. Am. Chem. Soc.* **1994**, *116*, 7955. Gennari, C.; Nestler, H. P.; Salom, B.; Still, W. C. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1765.