Pergamon

Tetrahedron Vol. 51, No. 2, pp. 567-578, 1995

0040-4020/95 $9.50+0.00

0040-4020(94)00916-3

Sequence-Selective Peptide Binding with a Synthetic Receptor

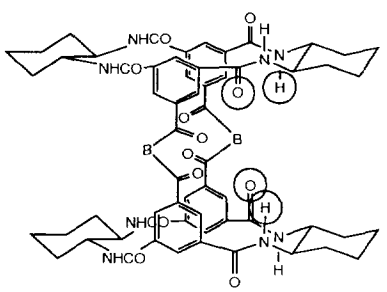
Seung Soo Yoon and W. Clark Still
Department of Chemistry, Columbia University, New York, NY 10027

Abstract: *A dye-labeled variant of a highly substrate-selective synthetic receptor for peptides is described and its binding selectivity for certain di- and tripeptide sequences is elucidated using an encoded combinatorial library.*

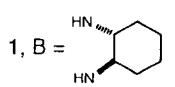
Perhaps the most intriguing challenge in molecular recognition is the creation of synthetic molecules that have properties similar to those of such remarkable biological substances as antibodies or enzymes. These substances manage to bind or process certain substrates with exquisite selectivity in the presence of a biological medium swimming with a vast array of different, and often closely related, molecules. One ubiquitous class of substrates are the oligomers of α-amino acids. While biological receptors manage to distinguish them by differences in their amino acid sequences, analogous selectivities are difficult to attain with synthetic receptors. While significant strides have been made in selectively binding derivatives of single amino acids,[1,4] synthetic receptors which bind particular multiple amino acid sequences are unknown.

One reason that sequence-selective binding of peptide substrates is underdeveloped in host-guest chemistry is that peptides are physically large molecules and the receptors which would bind them would need binding sites of comparable dimensions. Unfortunately, most contemporary host molecules have rather small substrate-binding regions and thus cannot easily distinguish substrates having differentiating structural features which are distributed over a large area. The problem is ultimately one of molecular design - in particular, it is not easy to formulate molecular structures that are both readily synthesized and have large, well-structured binding sites that are complementary to a given substrate.[2]

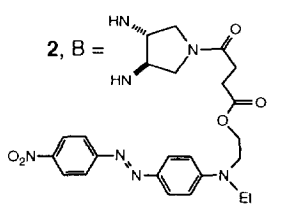
While the general problem of receptor design remains unsolved, several receptors are known which bind certain biological molecules with high selectivity.[3] Among these is receptor **1**.[4]



1, B =



2, B =



This molecule is a cyclooligomer of trimesic acid and (1*R*,2*R*)-diaminocyclohexane, and can be synthesized in one step from commercially available materials. Its most interesting property, however, is that it binds certain α-amino acid derivatives with high selectivity. In particular, **1** was found to bind L-amino acids enantioselectively (70-99% ee) and to select for amino acid sidechains having a particular size (*e.g.* phenyl>>benzyl, ethyl>>methyl). These properties were rationalized by a three-strand, β-sheet-like binding mode involving peptide association with the circled hydrogen bond donors/acceptors and having an L-peptide sidechain buried within the center of the receptor. We also observed very strong binding of **1** to a certain tripeptide (N-Boc-Gly-(L)Val-Gly-NHBn) which presumably involved additional hydrogen bonding to the outlying amides which are part of the B fragments of **1**. While these findings suggested the possibility of sequence-selective peptide binding, studies using **1** and various alanine- and valine-containing tripeptides failed to detect it. In conjunction with other receptor work, Borchardt and Still recently described a new method for studying the scope of receptor binding using an encoded combinatorial library of substrates.[5] Because such a library-based assay enables many thousands of binding experiments to be performed simultaneously, it is ideal for discovering binding phenomena about which little is known. In the following paragraphs, we use the method to search for sequence-selective peptide binding by **2**, a dye-labeled analog of **1**.

The Method: *A Solid Phase Binding Assay and Combinatorial Chemistry.*

Our approach to evaluating the binding properties of a synthetic receptor is closely related to methods developed for finding good ligands to biological receptors such as antibodies.[6] The general scheme involves labeling a receptor (*e.g.* with a fluorescent dye or radioisotope) so that it may be sensitively detected and then treating the labeled receptor with a large collection of potential substrates. If these substrates are spatially separated (e.g. on different solid particles or in different locations on a plate), then those areas occupied by substrates which bind the receptor will themselves become labeled. The appeal of the approach is its simplicity and the fact that very large numbers of different substrates can be screened for binding simultaneously. Even such simple procedures as examining substrate-bearing particles (e.g. Merrifield synthesis beads) through a low power microscope for the presence of label allows millions of substrates to be screened for binding by one person in a few hours.

By varying the concentration of the labeled receptor used, it is possible to control the minimum association constant ($K_{a\text{-}min}$) which can be detected by the binding assay. Assuming simple bimolecular receptor/substrate binding, K_a is traditionally defined:

$$K_a = \frac{[\text{Complex}]}{[\text{Receptor}]_{free}[\text{Substrate}]_{free}}$$

Obviously, when $[\text{Complex}] = [\text{Substrate}]_{free}$, K_a simplifies to $1/[\text{Receptor}]_{free}$. In the context of a solid phase assay using labeled receptor in solution and substrates on solid particles, this situation

obtains when the substrate on a particular particle is half bound by labeled receptor. Hence, one may estimate the minimum K_a (here termed $K_{a\text{-}min}$) of receptor for substrates on any fully labeled particles by measuring the equilibrium concentration of the labeled receptor in free solution over the particles and taking its reciprocal.

The particular solid phase assay we use here generally involves labeling the receptor with a colored dye and mixing it in dilute solution with a library of peptide-like substrates attached to Merrifield synthesis beads. After 48 hours of agitation to equilibrate a dilute solution of the colored receptor and the initially colorless substrate bead library, we find that a small percentage of the beads take on deep colorations. We then measure the concentration of free receptor remaining in solution to determine $K_{a\text{-}min}$ and then pick those beads having the deepest coloration. By determining the structures of the substrates on those beads, we know which substrates in the library bind the receptor with association constants of at least $K_{a\text{-}min}$ assuming the substrates on the deeply colored beads are at least 50% saturated by labeled receptor.

While the receptor labeling and assay procedures are straightforward, creation of an appropriate substrate library is less obvious. In recent years, a variety of methods for creating large collections of diverse molecules have been developed and could be used;[7] but one method, combinatorial *split synthesis*,[8] is particularly relevant to substrate library preparation. Given that organic synthesis proceeds via a series of steps and that many of these steps can often be carried out using a range of alternative reagents (or synthons, residues, etc.) to yield different products, split synthesis prescribes a simple protocol for preparing the library of products resulting from all possible combinations of all alternative reagents used. Split synthesis is carried out on small solid support particles (e.g. Merrifield beads) and yields a particle-supported library in which any particular particle carries the product from one particular set of reagents.

The split synthesis method was developed originally for oligopeptide synthesis and can yield very large libraries. For example, if a pentapeptide library is prepared by split synthesis in which 20 different amino acids are used at each of the five residue-coupling steps, the final library will contain 3,200,000 (20^5) different pentapeptides. Furthermore, any particular synthesis particle will bear only one type of pentapeptide (or at least have been submitted to only one particular, well-defined series of chemical steps).

Until recently, split synthesis has been effectively limited to libraries of sequencable biopolymers such as peptides and nucleotides. The limitation arose because individual members of the library are produced on *single* solid support particles and these particles carry only picomolar quantities of product. Such quantities are far too small for most classical structure elucidation techniques. While mass spectroscopy offers a partial solution to the product analysis problem, it is limited by its inability to distinguish isomers or to deal with the mixtures of products and byproducts which commonly result from solid phase synthesis.

The best general solution to the structure analysis problem involves a recently developed technique known as *encoding*.[9] Encoding entails attaching arrays of molecular tags to the solid support particles during each synthetic step to create unique, tag-encoded records of the

particular reagents used in the synthesis of each library member. By analyzing the tag complement of any particular solid support particle after the synthesis, one can determine the particular synthetic steps which were used in the synthesis of the library member on that particle.

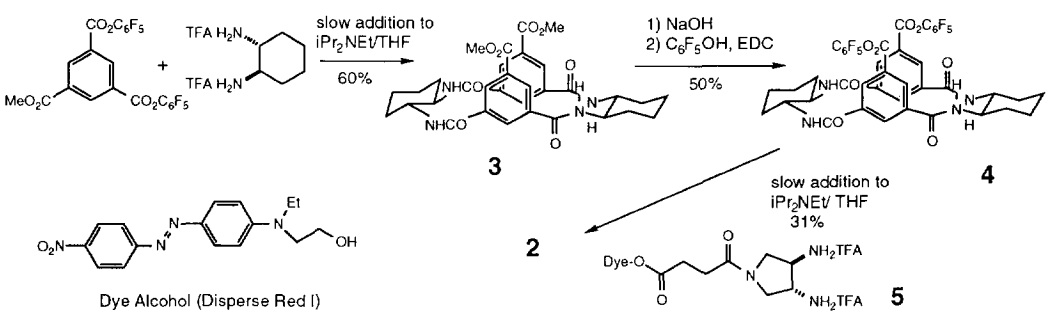
In the study of our labeled receptor **2**, we use a substrate library of 50,625 (15^4) terminally acylated tripeptides prepared by split synthesis on 50-80μ polystyrene (Merrifield) beads as previously described.[5] Our library is encoded using a set of sixteen highly electrophoric tagging molecules which can be detached from single synthesis beads and analyzed using electron capture capillary gas chromatography (ECGC).[9e]

The Means: *A Labeled Receptor and Encoded Substrate Library.*

Our first task was to prepare a labeled variant of receptor **1** that could be visually detected by simple inspection. Since **1** itself has no appropriate label attachment site, we prepared a relative of **1** in which the two spanning *trans*-diaminocyclohexanes (B in **1**) were replaced by stereochemically similar *trans*-3,4-diaminopyrrolidines. The pyrrolidine ring nitrogen could then serve as the label attachment point. For the label, we chose an intense red dye, Disperse Red 1, and the final structure of the labeled receptor thus became **2**.

The synthesis of red-labeled receptor **2** is straightforward and is outlined in the Scheme. Because **2** is not significantly strained and has few rotatable bonds, we were able to construct it by a few simple reactions which both couple fragments and macrocyclize in a single step. Thus, the macrocyclic tetramide **3** was prepared in >50% yield by a single reaction which linked and cyclized two molecules of *trans*-1,2-diaminocyclohexane and two molecules of diactivated trimesic ester. The synthesis was completed via diactivation of **3** as a *bis*-pentafluorophenyl ester (**4**) followed by another single step coupling/macrocyclization (~30% yield) using labeled diamine **5** to give the desired dye-labeled receptor **2**.

Scheme. Synthesis of Labeled Receptor **2.**



Our polymer-supported, encoded substrate library has been described previously[5] and has the general structure:

Polystyrene-$CH_2NHCO(CH_2)_5NH$ - AA1 - AA2 - AA3 - C(=O)R

where AA*n* represents any one of the following fifteen sidechain-protected[10a] amino acids (standard single letter codes for amino acids in parentheses):

Gly (G), D-Ala (a), L-Ala (A), D-Ser (s), L-Ser (S), D-Val (v), L-Val (V), D-Pro (p), L-Pro (P), D-Asn (n), L-Asn (N), D-Gln (q), L-Gln (Q), D-Lys (k), L-Lys (K)

and where R represents any one of the following fifteen groups:

methyl (Me), ethyl (Et), i-propyl (iPr), t-butyl (tBu), i-butyl (iBu), neopentyl (neoPe), trifluoromethyl (TFM), methoxymethyl (MOM), cyclopropyl (cPr), cyclobutyl (cBu), cyclopentyl (cPe), acetoxymethyl (AcOM), phenyl (Ph), dimethylamino (Me_2N), morpholino (Mor)

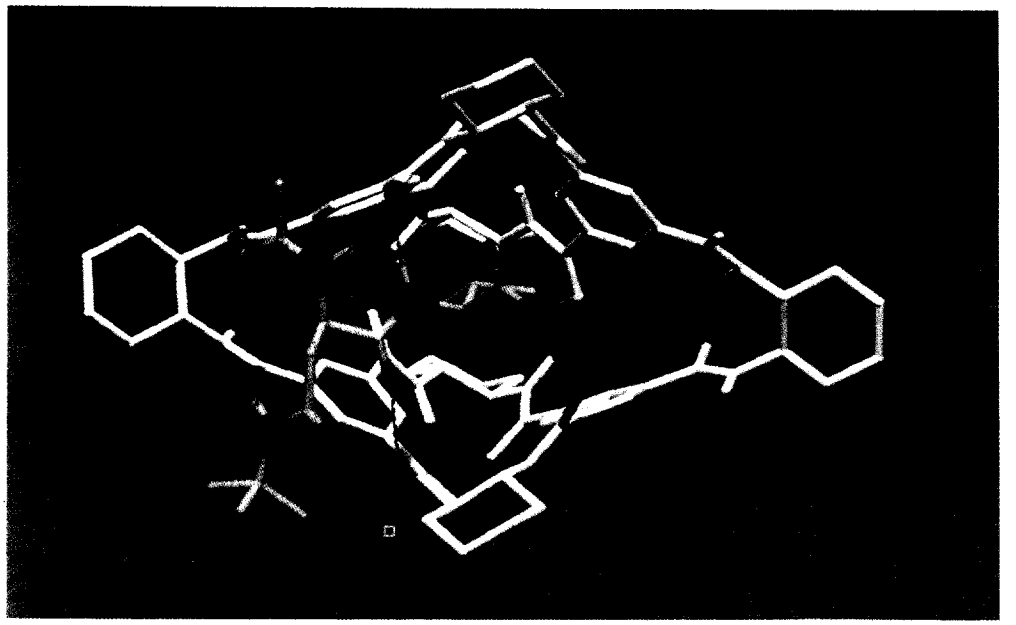
Because we prepared our substrate library by split synthesis using fifteen different amino acids at each of the three AA sites and terminated the tripeptide chain with fifteen different acylating agents, the total number of different substrates in the library is 15^4 or 50,625. To encode this library, we used sixteen GC-distinct electrophoric tagging molecules.

To carry out the binding assay, we suspended a 10 mg sample (~10^5 beads) of the substrate library in an Eppendorf tube containing ~0.3 mL $CHCl_3$ and then added ~25 μL of a 600 μM $CHCl_3$ solution of red **2**. Upon mixing, the bead library immediately extracted most of the colored receptor from solution, but examination of the mixture through a low power microscope showed that all beads looked essentially the same. However, after 30 min of agitation on a wrist-action shaker, ~5% of the beads had turned light orange. After 48 hours of agitation, ~1% of the beads were stained deep red-orange along with more having various lighter orange colorations as shown in Figure 1. At this point, a UV/VIS measurement of receptor remaining in solution revealed [**2**] = 21 μM and thus $K_{a\text{-}min}$ would be ~48000 for receptor-saturated beads. To estimate the degree of saturation for our most deeply stained beads, we compared their colors with those of standardized beads having 1%, 10% and 100% dye loading. This comparison suggested our most deeply stained beads to be ~5% saturated. We found it difficult to estimate the extent saturation to better than a factor of 2-4. Assuming 5% saturation, the actual $K_{a\text{-}min}$ would be ~2400 which corresponds to a minimum binding energy for our red beads of ~4.5 kcal/mol. Fifty-five of these most deeply stained beads were picked manually under a 4X microscope for decoding.

After washing the selected beads extensively with DMF, they were placed individually in separate melting point capillary tubes containing 1-2 μL of pure DMF for tag detachment. Because the tags were attached by photolabile *ortho*-nitrobenzylic carbonate linkages, they could be released for ECGC analysis by long wavelength ultraviolet (350 nm) irradiation. After silylation (*bis*-trimethylsilylacetamide) to increase tag volatility, the solution over each bead was injected into an electron capture capillary gas chromatograph for tag analysis.

Figure 1. Photomicrograph of substrate library on Merrifield beads after treatment with red receptor **2**. The most intensely colored beads bear substrates that bind **2** with association energies of at least -4 kcal/mol.

Figure 2. Possible structure of the complex between receptor **1** (white) and Ac-(D)Asn(tBu)-(L)Val-NHMe (light blue) as found by molecular mechanics.



The resulting chromatograms showed which tags were present and which were absent and thus revealed[10b] the structure of each substrate which had been selected by the solid phase binding assay.

The Result: *Sequence-Selective Binding of Peptides.*

Using the bead-supported peptide library and dye-labeled receptor **2**, the above-described solid phase assay and ECGC decoding of 55 of the most intensely stained beads revealed the substrates whose sequences are listed in Table 1. Among these sequences, 53 unique sequences were found indicating that the 55 beads we picked included only a small fraction of all tightly binding sequences. Nevertheless, the data in Table 1 shows a number of clear trends.

Table 1. Peptide substrate sequences found to bind **2** with $K_{a\text{-}min} \sim 2400$.

	R	AA3	AA2	AA1		R	AA3	AA2	AA1		R	AA3	AA2	AA1
1.	Me	V	n	G	19.	Mor	n	V	Q	37.	TFM	A	n	V
2.	Ph	V	n	G	20.	Ph	n	V	n	38.	AcOM	a	n	V
3.	cPe	V	n	K	21.	AcOM	L	V	n	39.	Me_2N	S	n	V
4.	AcOM	V	n	Q	22.	cPe	L	V	n	40.	AcOM	Q	n	V
5.	Et	V	n	A	23.	iPr	q	V	S	41.	Mor	k	n	V
6.	MOM	V	P	Q	24.	iBu	q	V	S	42.	cPr	A	q	V
7.	AcOM	n	V	S	25.	TFM	q	V	S	43.	Et	G	q	V
8.	Et	n	V	S	26.	MOM	q	V	S	44.	Me	S	q	V
9.	cPr	n	V	S	27.	MOM	q	V	G	45.	MOM	S	q	V
10.	iBu	n	V	S	28.	AcOM	q	V	G	46.	cPe	p	q	V
11.	Me	n	V	S	29.	Me	q	V	G	47.	AcOM	s	G	N
12.	Ph	n	V	S	30.	Et	q	V	q	48.	MOM	N	k	G
13.	Me_2N	n	V	S	31.	TFM	q	V	Q	49.	AcOM	G	K	S
14.	iPr	n	V	S	32.	Me_2N	Q	V	N	50.	tBu	q	G	G
15.	iBu	n	V	G	33.	Me_2N	G	V	S	51.	iBu	q	G	S
16.	Ph	n	V	G	34.	Ph	A	n	V	52.	Ph	q	S	G
17.	MOM	n	V	G	35.	cPr	A	n	V					
18.	Mor	n	V	K	36.	Et	A	n	V					

The most obvious trend is a strong preference of **2** for L-valine (V). Thus, 46 of the 52 sequences found (entries 1-46) contain V somewhere in the sequence and most commonly at the central AA2 site. This finding is in line with our previous NMR studies of **1** which revealed the tightest binding for peptides containing V (among those amino acids used in our library).[4a] Curiously, all non-valine-containing substrates (entries 47-52) incorporate at least one glycine (G). By comparing the observed frequencies of the various amino acids at the three different sites with their statistical expectation frequencies and standard deviations (σ), we can discern which residue-site selectivities are statistically significant. Summarized below are the significant selectivities (>90% confidence level) at all substrate sites along with their deviations (units of σ) and the corresponding confidence levels (%) as derived from counting statistics.[11]

R	AA3	AA2	AA1
none	D-Asn, 3.3σ, 99+%	L-Val, 4.6σ, 99+%	L-Ser, 3.0σ, 99+%
	D-Gln, 1.9σ, 94%	D-Asn, 2.9σ, 99+%	L-Val, 2.9σ, 99+%
			Gly, 2.3σ, 98%

While these results indicate that **2** has significant selectivity for certain residues at each amino acid position, **2** has even more selectivity for particular amino acids relative to the L-valine (V) position. In particular, in 43 (93%) of the 46 V-containing sequences, the amino acid either immediately preceding or immediately following the V is D-(trityl)asparagine (n, 65% of sequences) or D-(N-trityl)glutamine (q, 32% of sequences). These two residues (n,q) are closely related in that they are the only D amino acids in the library having (N-trityl)carboxamide-bearing sidechains. In 36 (84%) of these 43 n/q-V sequences, n or q precedes V. In such n/q-V sequences, the site following V (i.e. AA1) also shows significant selectivity and is most commonly L-(O-tBu)serine (S) or glycine (G). There is essentially no selectivity among the various N-terminal R groups.

These results indicate that **2** preferentially binds the tripeptide consensus sequence n-V-S and its relatives q-V-S, n-V-G and q-V-G. The most selectivity is seen at the dipeptide level where **2** binds n-V and q-V in the both frame-shifted sites AA3-AA2 and AA2-AA1. Considering that q and n are similar D-amino acids with carboxamide sidechains, **2**'s dipeptide selectivity for n/q-V is remarkable.

To confirm our findings and to estimate the energetic extents of some of the selectivities observed, we synthesized a series of diastereomeric acetoxymethyl-(N-trityl)Asp-Val-(O-tBu)Ser-linker-bead substrates. By treating known quantities of these supported substrates with **2** in $CHCl_3$, we could use UV/VIS spectrophotometry (λ = 476 nm) to measure [**2**] both before and after treatment with supported substrate and thus evaluate binding energies. The results of these solid phase binding measurements are summarized below:

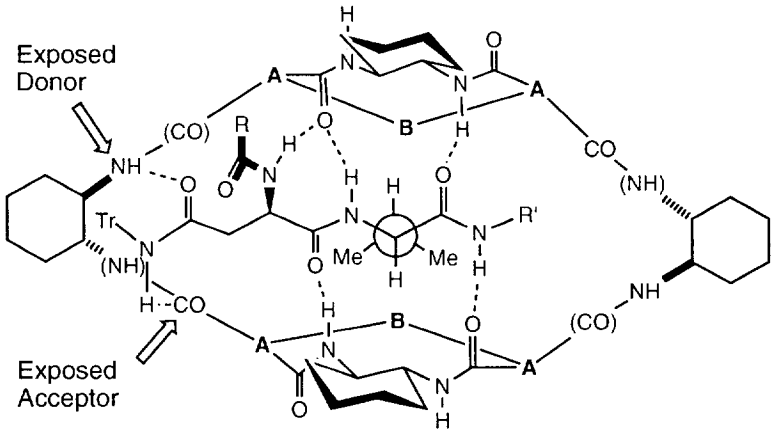
Peptide Substrate	Binding energy (kcal/mol)	Found in Assay?
AcOM - n - V - S - polymer	-4.4	Yes
AcOM - n - v - S - polymer	-0.8	No
AcOM - N - V - S - polymer	-3.4	No
AcOM - n - V - s - polymer	-2.8	No
AcOM - N - v - s - polymer	-3.0	No
AcOM - N - v - S - polymer	-3.1	No

These measurements show that **2** binds the consensus sequence n-V-S (found by the solid phase, color assay) more tightly than any of the diastereomers studied (not found by assay). Stereochemical inversion at V is most damaging to binding which is thereby reduced by 3.6 kcal/mol. Inversion of n or S also weakens binding but only by 1.1 and 1.6 kcal/mol respectively.

Interestingly, the residue binding energies are not simply additive as the fully inverted (enantiomeric) sequence N-v-s binds less tightly by only 1.4 kcal/mol.

The binding energies measured by solid phase experiments are not directly comparable to those from solution phase experiments because the environment inside the bead is not quite the same as free solution. In particular, the concentration of supported peptide *in the bead* is ~0.1M and this relatively high concentration favors peptide substrate aggregation which would diminish binding to receptor. In fact, comparison measurements of binding of several (L)Ala- and (L)Val-containing peptides in free solution and on polystyrene supports using receptor **1** showed that peptide binding is 2-3 kcal/mol stronger when binding is measured in dilute free solution.

The origin of **2**'s n/q-V selectivity is a matter of conjecture at this time, but it is likely that the outer B amides are involved as suggested in the following cartoon:



This proposal is based on molecular mechanics studies and ^{1}H nmr measurements on the N-acetyl-(L)Val-NHtBu complex with **1** which fix the valine sidechain within the center of the receptor.[5] A conformational search on Ac-(D)Asn(tBu)-(L)Val-NHMe bound to **1** found a structure closely related to the cartoon above as the global minimum and this structure is shown in Figure 2.[12] The other low energy structures found in the search had similar geometries which differed only in the details of the hydrogen bonding between the substrate sidechain carboxamide and receptor amides marked as "Exposed Donor" and "Exposed Acceptor" in the cartoon above.

Whatever the origin of the observed discrimination in binding, it is clear that receptor **2** binds peptide substrates having particular amino acid sequences with remarkable selectivity. This selectivity includes selection based on sidechain stereochemistry (favoring D-Asn/Gln, L-Val), size (Val>>Ala) and functionality (favoring the carboxamides of Asn and Gln). We believe that these results define the first example of a synthetic receptor which binds peptidic substrates with significant sequence-selectivity.

Conclusion.

The foregoing studies establish that receptor **2** binds certain sidechain-protected di- and tripeptides with remarkable selectivity and shows a large preference for n/q-V-containing substrates. Based on the number of replicate substrate sequences found in the binding assay, we estimate the total number of different substrate sequences which are bound by **2** at -4 kcal/mol or better to be 500-1000 out of the entire library of 50,625 sequences.

Highly selective complexation was also found by related binding studies using the analogous sidechain-deprotected library. In that case, **2** showed even higher substrate selectivity with a replicate analysis indicating that only 40-70 sequences out of 50,625 are bound at the -4 kcal/mol level of binding. The most commonly found sidechain-deprotected sequences differed from the ubiquitous n/q-V sequences found with the protected library and had AA2 = glycine (86%) and AA1 = L-serine (49%) or glycine (37%).

The results we describe here not only show that **2** is a sequence-selective receptor for peptides but also demonstrate the power of directed screening of large chemical libraries as a method to find novel molecules having sought-after properties.[14]

Experimental Section.

A_2B_2 dimethyl ester (3). A solution of 5.2 g of 1,3,5-benzenetricarboxyl acid *bis*-pentafluorophenyl monomethyl ester (9.34 mmol) and 1.1 g of 1R,2R-diaminocyclohexane *bis*-trifluoroacetic acid (TFA_2) salt (9.34 mmol) in 50 mL of dimethylacetamide (DMA) was added to a solution of 7.2 mL of iPr_2NEt (41.1 mmol, 4.4 equiv.) in 200 mL THF at 25 °C over 20 hrs by syringe pump. After stirring for an additional 8 hrs, all volatiles were removed at reduced presure. The residue was purified by flash chromatography on silica gel using 5% methanol in CH_2Cl_2 to give **3** as an amorphous white solid (1.52 g, 54%): R_f = 0.65 (silica gel, 5% MeOH in CH_2Cl_2); IR (neat) 3440, 2987,1720, 1672 cm^{-1}; ^1H NMR ($CDCl_3$, ppm) δ 8.54 (s, 1H), 8.47 (s, 2H), 7.11 (m, 2H), 3.99 (m, 2H), 3.94 (s, 3H), 2.28 (m, 2H), 1.87 (m, 2H), 1.44 (m, 4H); ^{13}C NMR ($CDCl_3$, ppm) δ 24.78, 31.88, 52.60, 55.59, 129.58, 130.53, 130.91, 135.39, 165.27, 168.16; MS (FAB) *m/z* 605 (M^++1); HRMS Calc for $C_{32}H_{37}N_4O_8$ 605.2612, found 605.2607.

A_2B_2 bis(pentafluorophenyl)ester (4). To a solution of 1 g of dimethyl ester (1.65 mmol) in 30 mL of THF and 15 mL of MeOH was added 3.3 mL of 1N NaOH solution. After stirring for 5 hrs at 25 °C, the reaction mixture was acidified with 1N HCl solution and extracted with EtOAc (3x50 mL). After solvent removal, the crude dicarboxylic acid was dissolved in 30 mL of THF and 30 mL of CH_2Cl_2, and 0.61 g of pentafluorophenol (3.3 mmol) and 0.63 g of EDC (3.3 mmol) were added. After stirring for 5 hrs at 25 °C, all volatiles were removed at reduced pressure and the residue was purified by flash chromatography on silica gel using 20% acetone in CH_2Cl_2 yielding **4** as an amorphous white solid (0.77 g, 51%): R_f = 0.87 (silica gel, 10% acetone in CH_2Cl_2); ^1H NMR ($CDCl_3$, ppm) δ 9.14(s, 2H), 8.87 (s, 1H), 7.12 (m, 2H), 4.01 (m, 2H), 2.29 (m, 2H), 1.90 (m, 2H), 1.50 (m, 4H); HRMS Calc for $C_{42}H_{31}N_4O_8F_{10}$ 909.6772, found 909.6767.

N-Succinyl dye-3R,4R-pyrrolidine diamine diTFA salt (5). To a solution of 0.8 g of Disperse Red 1 (Aldrich, 2.09 mmol), 0.35 mL of Et_3N (2.75 mmol) and 0.035 g of DMAP (0.25 mmol) in 30 ml of CH_2Cl_2 was added 0.25 g of succinic anhydride (2.50 mmol). After stirring

overnight at 25 °C, all volatiles were removed at reduced pressure. The residue was purified by flash chromatography on silica gel using 5% MeOH in CH_2Cl_2 to give the mono-dye ester of succinic acid as an amorphous red solid (0 .95 g, 96%): R_f = 0.31 (silica gel, 5% MeOH in CH_2Cl_2); ^1H NMR ($CDCl_3$, ppm) δ 8.34 (d, J = 8.9 Hz, 2H), 7.97 (dd, J = 12.0, 9.0 Hz, 4H), 6.87 (d, J = 8.9 Hz, 2H), 4.52 (t, J = 6.3 Hz, 2H), 3.61 (t, J = 6.4 Hz, 2H), 3.44 (q, J = 7.2 Hz, 2H), 2.64 (t, J = 5.9 Hz, 2H), 2.49 (t, J = 6.0 Hz, 2H), 1.16 (t, J = 7.2 Hz, 3H)

To a solution of 0.095 g of above mono-dye ester (0.201 mmol), 0.030 g of HOBT (0.221 mmol) and 0.087 g of 3R,4R-di(Boc-amino)pyrrolidine (0.221 mmol, prepared by Boc protection (Boc_2O, NEt_3) and debenzylation (ammonium formate, Pd/C) from the known N-benzyl 3R,4R-diaminopyrrolidine[13]) in 30 mL of CH_2Cl_2 was added 0.042 g of EDC (0.221 mmol) at 0 °C. After stirring for 2 hr at 0 °C and additional 8 hr at 25 °C, all volatiles were removed at reduced pressure. The residue was purified by flash chromatography on silica gel using 5 % MeOH in CH_2Cl_2 to give Boc_2-**5** as an amorphous red solid (0 .12 g, 79%): R_f = 0.62 (silica gel, 5% MeOH in CH_2Cl_2); ^1H NMR ($CDCl_3$, ppm) δ 8.33 (d, J = 8.9 Hz, 2H), 7.92 (dd, J = 12.0, 9.0 Hz, 4H), 6.80 (d, J = 8.9 Hz, 2H), 5.29 (m, 1H), 4.75 (m, 1H), 4.32 (t, J = 6.3 Hz, 2H), 4.02-3.91 (m, 4H), 3.71 (t, J = 6.4 Hz, 2H), 3.54 (q, J = 7.2 Hz, 2H), 3.23-3.10 (m, 2H), 2.65-2.49 (m, 4H), 1.44 (s, 18H), 1.26 (t, J = 7.2 Hz, 3H); ^{13}C NMR ($CDCl_3$, ppm) δ 172.83, 169.98, 156.73, 151.25 126.66, 126.22, 125.37, 124.64, 122.61, 119.58, 111.42, 61.43, 54.86, 50.34, 48.70, 45.62, 28.76, 28.52, 28.28, 28.22, 12.24; MS (FAB) *m/z* 698 (M^++1); HRMS Calc for $C_{34}H_{48}N_7O_9$ 698.3514, found 698.3512.

To a solution of 0.12 g of Boc_2-**5** (0.159 mmol) and 0.15 ml of anisole in 15 ml of CH_2Cl_2 was added 5 ml of TFA at 25 °C. After stirring for 2 hr, all volatiles were removed at reduced pressure and the residue (crude **5**·TFA_2) was washed with ether and used for next step without further purification.

Dyed receptor (2). A solution of 0.12 g of *bis*(pentafluorophenyl) ester **4** (0.131 mmol) and 0.11 g of **5**·TFA_2 (0.146 mmol, 1.1 eq.) in 10 mL of DMA was added via syringe pump over 20 h to a solution of 0.23 mL of iPr_2NEt (1.31 mmol) in 200 ml of THF at 25 °C. After stirring for additional 8 hr, all volatiles were removed at reduced presure. The residue was then purified by flash chromatography on silica gel using 3% MeOH in CH_2Cl_2 to give receptor **2** as an amorphous red solid (31 mg, 20.4%): R_f = 0.41 (silica gel, 5% MeOH in CH_2Cl_2); IR (neat) 3391, 2893, 1721, 1673, 1530 cm^{-1}; ^1H NMR (1:1=$CDCl_3$:CD_3OD, ppm) δ 8.49 (s,4H), 8.27 (s, 4H), 8.01(s, 4H), 7.98 (d, 4H, J = 9.0 Hz), 7.64 (dd, 8H, J = 8.8, 1.4 Hz), 6.55 (d, 4H, J = 8.9 Hz), 4.56 (m, 4H), 4.08 (m, 8H), 3.76 (m, 8H), 3.48 (m, 8H), 3.11 (q, 4H, J = 7.0 Hz) 2.40 (m, 4H), 2.31 (m, 4H), 1.95 (m, 8H), 1.61 (m, 8H), 1.21 (m, 16H), 1.01 (t, 6H, J = 7.1Hz); ^{13}C NMR ($CDCl_3$:CD_3OD=1:1, ppm) δ 172.83, 170.45, 167.97, 165.32, 156.55, 151.20, 146.99, 134.17, 134.05, 133.67, 129.85, 128.81, 128.68, 128.51, 128.32, 125.97, 124.32, 122.22, 111.20, 61.34, 57.07, 54.49, 52.00, 45.28, 32.18, 31.47, 30.79, 28.41, 28.07, 24.81, 24.42, 24.29, 11.73; MS(FAB) *m/z* 2077 (M^++1); HRMS Calc for $C_{108}H_{120}N_{22}O_{22}$ 2076.8850, found 2076.8776.

Solid Phase Binding Assay. The solid phase substrate library was prepared by the encoded split synthesis as described previously and included 50,625 different acylated tripeptide sequences corresponding to all possible combinations of the 15 acylating agents and 15 amino acids (used three times) listed in the text.[5,9e]

To screen the substrate library for binding, a 10 mg sample of the library (~10^5 beads) was mixed in a 1.5 mL Eppendorf tube with 0.3 mL of ~50 μM **2** in $CHCl_3$. After agitation on a wrist-action shaker for 48 h, ~1% of the beads were found to be stained deep red. Fifty-five of these deep red beads were picked by hand under a 4X wide-field microscope and photolyzed (350 nm, 4 hrs) in 1-2 μL of DMF to release the tag molecules. After silylation ($CH_3C(OTMS)NTMS$, ~0.1 μL), electron capture GC was used to analyze the tag complement of each picked bead.

Notes and References.

1. S.C. Peacock, L.A. Domeier, F.C.A. Gaeta, R. C. Helgeson, J.M. Timko and D.J. Cram, *J. Am. Chem. Soc.*, **100**, 8190 (1978); J. Rebek, B. Askew, P. Ballester and M. Doa, *J. Am. Chem. Soc.*, **109**, 4119 (1987); J.-I. Hong, S.K. Namgoong, A. Bernardi and W.C. Still, *J. Am. Chem. Soc.*, **113**, 5111 (1991); Y. Murakami, T. Ohno, O. Hayashida and Y. Hisaeda, *J. Chem. Soc., Chem. Commun.*, 950 (1991): A. Galan, D. Andreu, A.M. Echavarren, P. Prados and J. de Mendoza, *J. Am. Chem. Soc.*, **114**, 1511 (1992); R. Liu and W.C. Still, *Tetrahedron Lett.*, **34**, 2573 (1993).

2. Notable successes include: W.L. Mock and N.-Y. Shih, *J. Am. Chem. Soc.*, **110**, 4706 (1988); M.A. Petti, T.J. Shepodd, R.E. Barrans and D. A. Dougherty, *J. Am. Chem. Soc.*, **110**, 6825 (1988); J.K. Judice and D.J. Cram, *J. Am. Chem. Soc.*, **113**, 2790 (1991); J.-I. Hong, S.K. Namgoong, A. Bernardi and W.C. Still, *J. Am. Chem. Soc.*, **113**, 5111 (1991); T.H. Webb, H. Suh and C.S. Wilcox, *J. Am. Chem. Soc.*, **113**, 8554 (1991); M.E. Tanner, C.B. Knobler and D.J. Cram, *J. Org. Chem.*, **57**, 40 (1992); reference 4.

3. *E.g.* K.-S. Jeong, A.V. Muehldorf and J.E. Rebek, *J. Am. Chem. Soc.*, **112**, 6144 (1990); R.P. Dixon, S.J. Geib, A.D. Hamilton, *J. Am. Chem. Soc.*, **114**, 365 (1992); reference 1.

4. a. S.S. Yoon and W.C. Still, *J. Am. Chem. Soc.*, **115**, 832 (1993); b. S.S. Yoon and W.C. Still, *Tetrahedron Lett.*, in press (1994).

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

6. H.M. Geysen and T.J. Mason, *Bioorg. Med. Chem. Lett.*, **3**, 397 (1993) and references therein.

7. Reviews: G. Jung and A.G. Beck-Sickinger, *Angew. Chem. Int. Ed. Engl.*, **31**, 367 (1992); M.R. Pavia, T.K. Sawyer and W.H. Moos, *Bioorg. Med. Chem. Lett.*, **3**, 387 (1993).

8. a. A. Furka, M. Sebestyen, M. Asgedom and G. Dibo, *Abstr. 14th Int. Congr. Biochem.*, Prague, Czechchoslovakia, **5**, 47 (1988); b. A. Furka, M. Sebestyen, M. Asgedom and G. Dibo, *Abstr 10th Int. Symp. Med. Chem.*, Budapest, Hungary, p 288 (1988); c. A. Furka, M. Sebestyen, M. Asgedom and G. Dibo, *Int. J. Pept. Protein Res.*, **37**, 487 (1991); d. K.S. Lam, S.E. Salmon, E.M. Hersh, V.J. Hruby, W.M. Kazmierski and R.J. Knapp, *Nature*, **354**, 82 (1991).

9. a. S. Brenner and R.A. Lerner, *Proc. Natl. Acad. Sci. USA*, **89**, 5381 (1992); b. J.M. Kerr, S.C. Banville and R.N. Zuckermann, *J. Am. Chem. Soc.*, **115**, 2529 (1993); c. J. Nielson, S. Brenner and K.D. Janda, *J. Am. Chem. Soc.*, **115**, 9812 (1993); d. M.C. Needels, D.G. Jones, E.H. Tate, G.L. Heinkel, L.M. Kochersperger, W. J. Dower, R. W. Barrett and M.A. Gallop, *Proc. Natl. Acad. Sci. USA*, **90**, 10700 (1993); e. M.H.J. Ohlmeyer, R.N. Swanson, L.W. Dillard, J.C. Reader, G. Asouline, R. Kobayashi, M. Wigler and W.C. Still, *Proc. Natl. Acad. Sci. USA*, **90**, 10922 (1993).

10. a. Sidechain protection: Asn (trityl), Gln (trityl), Lys (Boc), Ser (tBu); b. Binary tag code given in reference 5.

11. This analysis was conducted with the SACCS (Statistical Analysis of Combinatorial Chemical Sequences) computer program developed by Dr. Peter Shenkin in this laboratory.

12. AMBER* force field, GB/SA CHCl3 solvation treatment, Monte Carlo conformational search, MacroModel/BatchMin V4.5. We thank Dr. Quentin McDonald for carrying out this search.

13. D.R. Reddy and E.R. Thornton, *J. Chem. Soc., Chem. Commun.*, 172 (1992).

14. This work was supported by NSF Grant CHE92 08254.

(*Received* 29 *March* 1994)