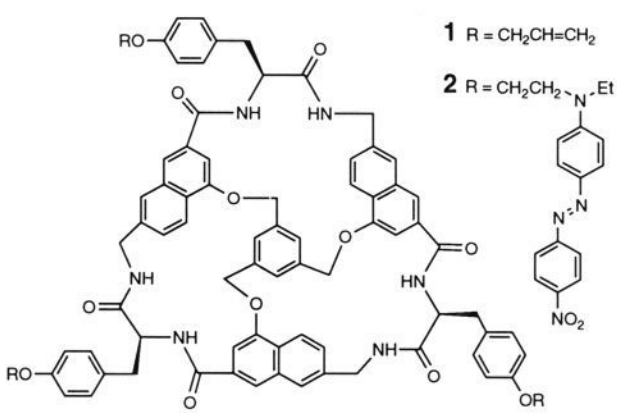
Synthetic Receptor for Internal Residues of a Peptide Chain. Highly Selective Binding of (L)X-(L)Pro-(L)X Tripeptides

Allen Borchardt and W. Clark Still*

Department of Chemistry, Columbia University New York New York 10027

Received May 2, 1994

Highly stereoselective binding of a neutral substrate by a receptor in an organic solvent often depends on the substrate's ability to participate in an array of favorable intermolecular electrostatic interactions such as hydrogen bonds.[1,2] Biological receptors use similar means but also distinguish substrates or fragments by size and seem to make such distinctions based in part upon a substrate's ability to precisely fill a binding cavity. Here we describe a new synthetic receptor (**1**) that shows an analogous capability to distinguish closely related peptidic substrates in organic solvents based on stereochemistry and, in some cases, on subtle differences in residue size. We will show that **1** not only exhibits high selectivity for binding tripeptides containing an internal L-proline (>99% diastereomeric excess (de) for L-Pro vs D-Pro) but also binds L-Pro more tightly than cyclic analogs which are either smaller or larger than Pro itself. Furthermore, **1** stereoselectively binds substrates having L-amino acids adjacent to L-Pro (90–99% de for L-Ala) and with binding constants ($K_a = 2.5 \times 10^5$ for iPrCO-(L)Ala-(L)Pro-(L)Ala) that are among the largest reported for binding of a neutral guest by a synthetic host.



Our design for **1** was taken from a previously described peptide-binding receptor having essentially the same structure as **1** but with the naphthalenes replaced by benzenes.[1c] The previous receptor had a cuplike shape with an ~6 Å diameter cavity surrounded by six unassociated hydrogen bond donors (D) and acceptors (A). According to molecular mechanics, the new design (**1**) had a similar conformation, except that the three naphthalenes widened the binding cavity (~8 Å diameter).[3] Whereas the original receptor interacted primarily with small terminal substituents of peptides (cartoon **O**), it appeared that the new, enlarged

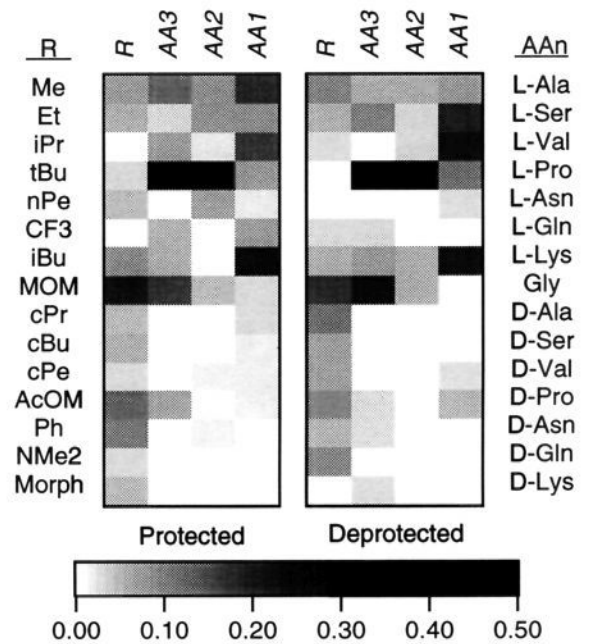
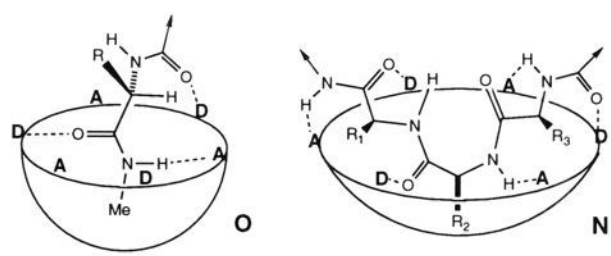
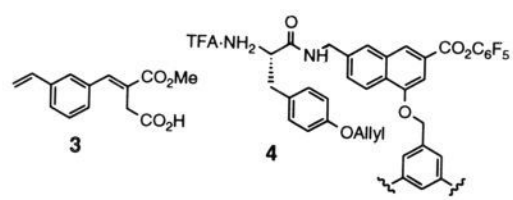


Figure 1. Gray-scale histograms of substrate sequences binding **2** (see text).

design might bind internal residues (R_2) with formation of as many as six hydrogen bonds (cartoon **N**).



Synthesis of **1** (see supplementary material) began with naphthol construction via Friedel–Crafts cyclization of the Stobbe-derived half-ester **3**.[4] That material was elaborated to **4**, which



underwent a triple macrolactamization with iPr_2NEt to provide **1** in 50% yield. Phenol deprotection ($(Ph_3P)_4Pd$/dimedone) and alkylation (Bu_4NF, THF) with the mesylate of the azo dye Disperse Red 1 gave the intensely red receptor derivative **2**.

To survey **2**'s binding properties, we used a solid-phase color assay employing an encoded combinatorial library of ~50 000 acylated tripeptide substrates.[2] This library had the general structure R(C=O)-AA3-AA2-AA1-NH$(CH_2)_5$CONH-polystyrene, where R represents 15 different acyl substituents and AA1–AA3 each represents 15 different D and L amino acids.[5] The library was supported on 50–80 μm polystyrene beads and was prepared both with and without side chain protection. Each bead carried only one type of tripeptide substrate.

(1) *E.g.*: (a) Rebek, J.; Askew, B.; Ballester, P.; Doa, M. *J. Am. Chem. Soc.* **1987**, *109*, 4119. (b) Jeong, K.-S.; Muehldorf, A. V.; Rebek, J. *J. Am. Chem. Soc.* **1990**, *112*, 6144. (c) Hong, J.-I.; Namgoong, S. K.; Bernardi, A.; Still, W. C. *J. Am. Chem. Soc.* **1991**, *113*, 5111. (d) Dixon, R. P.; Geib, S. J.; Hamilton, A. D. *J. Am. Chem. Soc.* **1992**, *114*, 365. (e) Yoon, S. S.; Still, W. C. *J. Am. Chem. Soc.* **1993**, *115*, 832.

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

(3) A 5000-step Monte Carlo conformational search (Goodman, J. M.; Still, W. C. *J. Comput. Chem.* **1991**, *12*, 1110) for **1** (L-Tyr modeled by L-Ala) using the MacroModel/AMBER* force field (McDonald, D. Q.; Still, W. C. *Tetrahedron Lett.* **1992**, *33*, 7743 and references therein) found a single conformation within the lowest 2 kcal/mol.

(4) *Cf.*: Johnson, W. S.; Graber, R. R. *J. Am. Chem. Soc.* **1950**, *72*, 925.

0002-7863/94/1516-7467$04.50/0

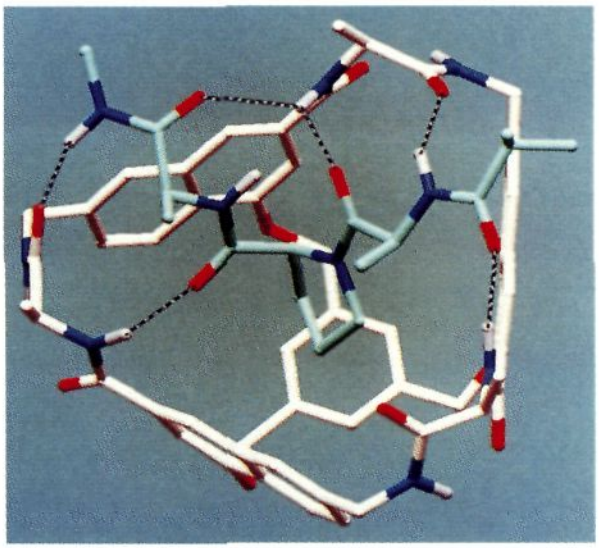


Figure 2. Global minimum of **1**/iPrCO-(L)Ala-(L)Pro-(L)Ala-NHMe.

Both side chain protected and deprotected libraries were screened for binding by treatment with ~50 μM solutions of receptor **2** in $CHCl_3$. After 24 h of equilibration with each library, ~10% of the beads had become colored, with ~1% being deep crimson. The most deeply colored beads were picked and decoded using gas chromatography to yield the sequences of the most tightly binding substrates. The results were extraordinary. With the protected library, 63 out of 68 tight-binding peptide sequences contained L-Pro, and 71% of these placed L-Pro at the AA2 site. Furthermore, in all AA2 = L-Pro sequences, the AA1 and AA3 residues were either L-amino acids or Gly. With the deprotected library, we picked 38 beads and found that all contained L-Pro. As before, L-Pro appeared most commonly at AA2 (79% of the sequences), and AA1 and AA3 were nearly always occupied by L-amino acids.

Our solid-phase binding data for both libraries are represented in Figure 1 as gray-scale histograms which give the fractional occupancy of each substrate site (R, AA1–AA3) by each possible residue (R = Me, Et, ...; AA*n* = L-Ala, L-Ser, ...) as a gray-scale running between white (0% occupancy) and black (≥50% occupancy). Thus the darkest areas under AA1 and AA2 for L-Pro indicate that most peptides bound by **2** had L-Pro at these sites. The selectivity for AA1–AA3 = L-amino acids can be seen as a generally dark region in the upper half of the histograms just below AA1-AA3. The large number of gray areas below R indicates little selectivity for the N-terminal substituent. The general similarity of the protected and deprotected histograms indicates that **2** binds both libraries with similar selectivity.

The preference of **2** for L-Pro-containing substrates and the absence of significant selectivity for the N-terminal substituent suggested that our receptor could be binding substrate approximately as shown in cartoon **N** (R_2 = Pro ring). Indeed, ^1H NMR of the complex of **1** with iPrCO-(L)Ala-(L)Pro-(L)Ala-$NHC_{12}H_{25}$ in $CDCl_3$ supported such a binding mode: thus the L-Pro ring CH's shifted upfield to the region between –0.4 and –2.4 ppm upon binding, a result consistent with the Pro ring being buried deep within **1**'s aromatic-lined binding cavity. Furthermore, a 40 000-step Monte Carlo conformational search of the complex using AMBER* in GB/SA $CHCl_3$ provided additional support by locating just such a structure as the global minimum for the complex (Figure 2).

Table 1. Binding Energies (kcal/mol) of **1** Plus Peptides in $CHCl_3$

entry	peptide substrate	$-\Delta G$
1	iPrCO-(L)Ala-(L)Pro-(L)-Ala-$NHC_{12}H_{25}$	7.4
2	iPrCO-(D)Ala-(L)Pro-(L)Ala-$NHC_{12}H_{25}$	5.5
3	iPrCO-(L)Ala-(D)Pro-(L)Ala-$NHC_{12}H_{25}$	4.3
4	iPrCO-(L)Ala-(L)Pro-(D)Ala-$NHC_{12}H_{25}$	4.4
5	iPrCO-(L)Pro-(L)Pro-(L)Ala-$NHC_{12}H_{25}$	6.0
6	iPrCO-(L)Pro-(L)Pro-(D)Ala-$NHC_{12}H_{25}$	5.4
7	iPrCO-(L)Ala-(L)azetidinyl-(L)Ala-$NHC_{12}H_{25}$	5.3
8	iPrCO-(L)Ala-(L)pipecolinyl-(L)Ala-$NHC_{12}H_{25}$	6.3

To quantify **1**'s selectivity for an internal L-proline flanked by L-amino acids, we synthesized *N*-dodecylamides of several tight-binding substrates and certain diastereomers. We then measured their binding energies in $CHCl_3$ solution ([**1**] = 100 nM) by CD-monitored[6] titration (Table 1). Notably, we found that **1** bound L-Pro at AA2 3.1 kcal/mol more strongly than D-Pro (entry 1 vs 3), L-Ala at AA1 3.0 kcal/mol more tightly than D-Ala (entry 1 vs 4), and L-Ala at AA3 1.9 kcal/mol more tightly than D-Ala (entry 1 vs 2). Taken together, these findings suggest a strong, highly ordered interaction between **1**'s binding site and the entire tripeptide chain (as suggested by **N** and Figure 2). Interestingly, binding decreased significantly when we replaced L-Pro by either its four- or its six-membered ring analog (entry 1 vs 7 and 8). These results underscore the importance of filling a receptor's binding cavity precisely—binding is disfavored not only when a substrate is too large but also when it is too small.[7]

In conclusion, we have prepared a new receptor (**1**) having an enlarged binding cavity but that is otherwise similar to a previously described peptide receptor.[1c] Though the structural similarities of these two receptors far outweigh the differences, their binding properties are very different. Whereas the previous receptor selects for small terminal substituents (as in **O**), **1** shows little selectivity for the terminal substituent but instead selects for internal L-Pro with high stereochemical and steric selectivity (as in **N**). **1** also shows sequence-selective tripeptide binding in the form of a strong preference for the L-configuration of flanking amino acids. More generally, these results illustrate how known principles of receptor design and the massive data gathering capacity of combinatorial library screening can be combined to create a powerful approach to problems in molecular recognition. Here, we have used the approach to create a functioning tripeptide receptor and firmly established its binding selectivity by screening >50 000 different tripeptidic substrates for complexation. The results of these experiments have led us to a well-defined model that suggests how **1** binds its favored substrates and that can be used to design receptors for other tripeptide sequences.

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

Supplementary Material Available: Synthesis of **1** and complete sequences of tripeptides bound by **2** (2 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, and can be ordered from the ACS; see any current masthead page for ordering information.

(5) R = methyl (Me), ethyl (Et), isopropyl (iPr), *tert*-butyl (tBu), neopentyl (neoPe), trifluoromethyl (CF_3), isobutyl (iBu), methoxymethyl (MOM), acetoxymethyl (AcOM), cyclopropyl (cPr), cyclobutyl (cBu), cyclopentyl (cPe), phenyl (Ph), morpholino (Morph), dimethylamino (Me_2N). AA1–AA3 = Gly (G), D-Ala (a), L-Ala (A), D-Ser(OtBu) (s), L-Ser(OtBu) (S), D-Val (v), L-Val (V), D-Pro (p), L-Pro (P), D-Asn(*N*-trityl) (n), L-Asn(*N*-trityl) (N), D-Gln(*N*-trityl) (q), L-Gln(*N*-trityl) (Q), D-Lys(*N*-Boc) (k), L-Lys(*N*-Boc) (K).

(6) Kearney, P. C.; Mizoue, L. S.; Kumpf, R. A.; Forman, J. E.; McCurdy, A.; Dougherty, D. A. *J. Am. Chem. Soc.* **1993**, *115*, 9907.

(7) *Cf.*: Yoon, S. S.; Still, W. C. *Tetrahedron Lett.* **1994**, *35*, 2117.