Sequence-Selective Receptors of Peptides. A Simple Molecular Design for Construction of Large Combinatorial Libraries of Receptors

Yuefei Shao and W. Clark Still*

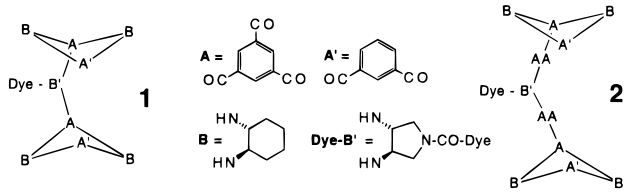
Department of Chemistry, Columbia University, New York, New York 10027

Received April 29, 1996

One of the greatest challenges in the area of molecular recognition is the creation of synthetic receptor molecules that bind biopolymers sequence-selectively. While sequence-selective binding of oligonucleotides by synthetic receptors is well established,[1] much less progress has been made with the other major classes of biopolymers. Recently, however, the situation has begun to change for oligopeptides. In particular, several relatively simple hostlike molecules have been described that bind di- and tripeptides with significant sequence-selectivity.[2] As part of a program directed toward large combinatorial libraries of such receptors,[3] we studied several designs for receptors that could be readily prepared on a solid support and also could be shown to bind diverse peptidic substrates sequence-selectively. One of the most promising of those designs is receptor **1**.[4] This molecule has a two-armed structural motif that we often find associated with significant binding properties.[3ac,5] In **1**, a conformationally restricted linker (**B′**) joins the two arms that each consist of $\mathbf{A_2B_2}$ cyclooligomers of 1,2-diamines (**B**) and isophthalic acids (**A**, **A′**). Preliminary work further indicated that by altering the structures of the **B′** linker and $\mathbf{A_2B_2}$ arms, it was possible to modify the preferred binding sequences (which often contained L-Val or L-Leu). We therefore envisioned making a combinatorial library of such receptors from an array of different linkers and arms as source of selective receptors for diverse peptidic sequences.

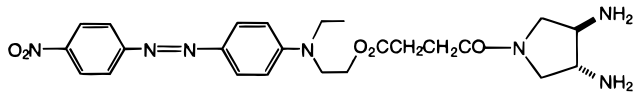
The main problem with a receptor library based on **1**, however, is that its members would be prepared by a combinatorial synthesis having only three combinatorial steps.[6] Thus, when 10 different linkers (**B′**) and 10 different arms ($\mathbf{A_2B_2}$, used twice) are used, a library of **1** would contain only 10^3 members. While more than 10 linkers and arms could be used, it is easy to see that the effort necessary to produce a much larger library (*e.g.*, having 10^6 members) in this way would be prohibitive.

Here we address this problem with an expanded two-armed receptor (**2**) whose synthesis involves not three but five combinatorial steps. Thus, starting with 10 linkers, 10 arms, and 20 different amino acids (**AA**), receptor libraries **2** having as many as 4×10^5 members



should be readily preparable. For such a library to be useful, however, the various library members must have differing binding selectivities. By screening selected examples of **2** with a combinatorial tripeptide library, we show that binding selectivities do indeed vary sensitively with the identity of the **AA** residue used.

Syntheses of receptors **2** are straightforward and employ the same enantiomerically pure, dye-labeled pyrrolidine that was used to prepare **1**:[4,7]



Using Fmoc-protected amino acids and standard peptide coupling chemistry, selected amino acids (**AA**) were first coupled to the above pyrrolidine core. After Fmoc deprotection, the A_2B_2 macrocycles were then added as pentafluorophenyl esters to give **2**. In these syntheses, we used **AA** = D- and L-phenylalanine (**f** and **F**), asparagine (**n** and **N**), *N*-tritylasparagine (**n(trityl)** and **N(trityl)**), proline (**p** and **P**), and *trans*-hydroxyproline (**hp** and **HP**) to prepare receptors **2f**, **2F**, **2n**, **2N**, **2n(tr)**, **2N(tr)**, **2p**, **2P**, **2hp**, and **2HP**, respectively.[8]

Each of these receptors was screened as described previously[2] for binding members of an encoded, acetylated tripeptide library having the following form:[9]

Ac-AA3-AA2-AA1-NH$(CH_2)_5$CONH-polystyrene

The binding assays were carried out by equilibrating a receptor variant (**2X**) with 10 copies of the tripeptide library in chloroform for 48 h on a wrist action shaker. The concentration of **2X** was made as low as possible so that only the strongest receptor-peptide complexes would be detected. Under these conditions, a relatively small number of library beads acquired the red color of **2X**. These red beads were picked under a low power microscope and decoded by electron capture gas chromatography to determine the sequences of the peptides bound by **2X**. The results are summarized in Table 1 for all receptors except the proline-linked receptors **2p** and **2P**, which turned out to have relatively poor affinity and selectivity.

The data in Table 1 indicate that receptors **2** are generally quite selective in binding certain di- and tripeptide sequences. In particular, every receptor studied was able to selectively bind a preferred subset of <1% of the 3375 peptide sequences in the library. In fact, most variants of **2** were able to distinguish subsets consisting of <0.5% of the peptides in the library. With receptors **2f**, **2N(tr)**, **2HP**, and **2hp**, the most tightly bound

(1) Review: Zimmer, C.; von Wahnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31.

(2) Review: Still, W. C. *Acc. Chem. Res.* **1996**, *29*, 155.

(3) For previously described combinatorial libraries of synthetic receptors, see: (a) Boyce, R.; Li, G.; Nestler, H. P.; Suenaga, T.; Still, W. C. *J. Am. Chem. Soc.* **1994**, *116*, 7955. (b) Burger, M. T.; Still, W. C. *J. Org. Chem.* **1995**, *60*, 7382. (c) Cheng, Y.; Suenaga, T.; Still, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 1813.

(4) This receptor selectively binds the two tripeptide sequences D-Pro-L-Val-D-Gln and L-Lys-L-Val-D-Pro: Wennemers, H., Yoon, S. S.; Still, W. C. *J. Org. Chem.* **1995**, *60*, 1108.

(5) See also: LaBrenz, S. R.; Kelly, J. W. *J. Am. Chem. Soc.* **1994**, *116*, 7955. Gennari, C.; Nestler, N. P.; Salom, B.; Still, W. C. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1765.

(6) The combinatorial steps would be addition of linker (**B′**) to a solid support and then sequential addition of the two different arms ($\mathbf{A_2B_2}$ or equivalent) to the linker.

(7) Yoon, S. S.; Still, W. C. *Tetrahedron* **1995**, *51*, 567.

(8) Designations taken from the standard single-letter code for amino acids; X indicates any amino acid.

(9) Fifteen different amino acids were used at each of the three AAn sites to yield a library having 3375 (15^3) different tripeptides segregated on Merrifield synthesis beads where AA1−3 were (D and L) Ala, Val, Pro, Ser(*O*-tBu), Asn(*N*-trityl), Gln(*N*-trityl), Lys(*N*-Boc), and Gly.

S0022-3263(96)00775-X CCC: $12.00

Table 1. Peptide-Binding Selectivities of Receptor 2 in Chloroform

receptor (concn, μM)[a]	total sequences bound[b]	preferred sequences bound			frequency[c] (%)
		AA3	AA2	AA1	
2F (5)	23 ± 6 (0.7%)	X	Gly	D-Pro	67
		D-Pro	L-Ala	D-X	15
2f (66)	2 ± 0 (0.06%)[d]	L-Gln	Gly	D-Val	56
		D-Asn	L-Pro	D-Val	44
2N (6)	7 ± 1 (0.2%)	D-Pro	L-X	D-Gln	42
		D-Pro	D-Ala, Lys	D-Pro	37
2n (3)	22 ± 6 (0.7%)	L-X	L-Asn	D-Pro	28
		Gly, D-X	Gly	Gly	24
2N(tr) (60)	2 ± 0 (0.06%)[d]	L-Pro	Gly	D-Pro	57
		D-Pro	Gly	D-Pro	43
2n(tr) (5)	13 ± 2 (0.4%)	L-Ala	L-Pro	X	64
		Gly, D-X	Gly	D-Pro	20
2HP (3)	2 ± 0 (0.06%)[d]	D-Pro	Gly	D-Pro	55
		D-Ala	Gly	D-Pro	45
2hp (25)	1 ± 0 (0.03%)[d]	L-Lys	D-Asn	L-Ala	100

[a] Equilibrium [**2**] during assay. [b] Total number of different tripeptide sequences tightly bound by **2** at the concentration indicated (percentage of library bound). [c] Percentage of all beads picked having the indicated sequence; sequences indicated were found at least four times. [d] Assays in which both dark red beads (strong binding) and light red beads (weaker binding) were found; results given here are for the dark red beads only, and peptide sequences from both dark and light red beads are given in the supporting information.

members of the library consisted of only one or two tripeptide sequences (0.03 and 0.06% of the library).

Interestingly, these exceptionally selective receptors did not always have the tightest binding as measured by the minimum concentration of **2** giving significant bead staining.

The binding selectivities of AA-linked receptors **2** can be seen to be very different from those of **1** and its derivatives, which showed a strong preference for L-Val-containing peptides. While each variant of **2** had its own unique peptide-binding selectivity, many of these receptors (*e.g.*, **2F**, **2N(tr)**, **2n(tr)**, and **2HP**) displayed a significant, consensus-like preference for dipeptide sequences having AA2 = Gly and AA1 = D-Pro. Receptors **2N(tr)** and **2HP** had the most similar binding properties in that both selectively bound the tripeptide sequence Ac-D-Pro-Gly-D-Pro. However, **2N(tr)** also bound Ac-L-Pro-Gly-D-Pro, while **2HP** also bound Ac-D-Ala-Gly-D-Pro. Other receptors had more distinct binding properties. Among the most highly selective receptors, **2f** preferentially bound Ac-L-Gln-Gly-D-Val and Ac-D-Asn-L-Pro-D-Val, while **2hp** selected the single sequence Ac-L-Lys-D-Asn-L-Ala.

While the ability of receptors **2** to distinguish peptides based on their sequence is significant, additional sequences can become tightly bound at higher concentrations of receptor. Evidence of such lower affinity (and less selective) binding could be seen in many of our binding assays. In such cases, two populations of receptor-binding beads could be visibly distinguished—one population consisting of dark red beads that was saturated with receptor and another, typically larger, consisting of light red beads that was significantly less saturated. Previous studies in this laboratory have shown that binding energies of peptides on such dark and light red beads can differ by as little as 1 kcal/mol as measured by solution phase binding experiments. In many cases, the sequences of the more weakly bound peptides (light red beads) were related to the sequences of the most strongly bound peptides (dark red beads). For example, receptor **2N(tr)** that most selectively bound Ac-D/L-Pro-Gly-D-Pro (dark red beads) was found to bind 10 additional related sequences (mostly Ac-L-X/Gly-Gly-D-Pro/Gln) on light red beads. In the case of **2hp** (most selective for Ac-L-Lys-D-Asn-L-Ala), light red beads carried 20 additional sequences that were structurally distinct (mostly Ac-D-X/Gly-Gly-D-X/Gly. In general, many of the more weakly binding sequences are related to the Gly-D-Pro consensus sequence mentioned above. Indeed, every receptor we studied except **2N** showed at least weak binding to peptides having Gly-D-X subsequences.

This work provides additional support for our general observation that molecules containing a cyclic linker and two (or more) functionalized, oligomeric arms make good receptors for peptides in organic solvents. Compared with more traditional, cagelike host molecules, two-armed receptors are relatively simple to prepare because the connections between linkers and arms are acyclic. In the work described here, we found that introduction of additional diversity elements (**AA**) between the linker and the arms of **1** gave new receptors **2** whose peptide-binding selectivities were quite different from those of **1** and varied significantly **AA**'s structure. Interestingly, most of the new receptors had binding similarities too—in particular, in 8 out of the 10 receptors studied bound tripeptides had AA2 = Gly and in most of these the AA2 Gly was followed by a D-amino acid (AA1) that was frequently D-Pro. These results indicate that **2** and **1** (with various linkers and macrocyclic arm substitutions) form distinct families not only structurally but also based on their binding properties: while **2** prefers Gly-D-X sequences, **1**-like receptors generally select L-Val/Leu-containing peptides. To make such receptors select for other types of peptides, it will be necessary make more substantial changes in the linkers, arms, or AA-like components. Just how large these changes must be in order to break away from the consensus binding behavior observed here is the subject of ongoing investigations.

Acknowledgment. This work was supported by NSF Grant No. CHE95 44253.

Supporting Information Available: Sequences for all tripeptides binding receptors **2F**, **2f**, **2N**, **2n**, **2N(tr)**, **2n(tr)**, **2HP**, **2hp**, **2P**, and **2p** and experimental details for the synthesis of **2F** (6 pages).

JO960775J