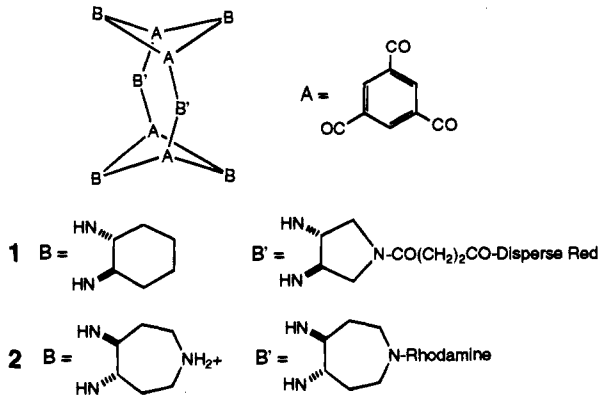
Sequence-Selective Binding of Peptides in Water by a Synthetic Receptor Molecule

Mercedes Torneiro and W. Clark Still*

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

Received March 6, 1995

Substantial progress has been made in recent years in the design and construction of synthetic receptors for small biological molecules including derivatives of simple amino acids and sugars.[1] However, for synthetic receptors to be used in biological systems, it is necessary to design structures that bind biooligomers and biopolymers on the basis of their residue sequence in water. In this area, less progress has been made. Several hostlike structures have been found to bind peptides sequence selectively in organic solvents,[2] but few examples of such binding in the more relevant aqueous environment have been reported.[3] One readily available family of synthetic receptors that bind peptides sequence selectively in chloroform is based on polycyclic oligomers of trimesic acid ($A(OH)_3$) and 1,2-diamines (*e.g.*, BH_2).[2a-d] One of these receptors, the D_2-symmetric **1**, incorporates a large, hydrophobic binding cavity and in chloroform selectively binds certain amino acid sequences containing (L)Val.[2c] In this report, we describe the tetraammonium ion **2**, a water-soluble analog of **1**, and show that it binds related amino acid sequences with high selectivity in water.



As implied by the structural diagrams, the design for **2** was based on insertion of hydrophilic ammonium groups into the cyclohexanes of **1** and use of a water soluble dye (rhodamine B) to label the receptor. The synthesis of **2** is detailed in the supplementary material but generally involved the scheme outlined in Figure 1. The route started with *trans*-3-hexene-1,6-diol[4] and used chiral osmylation[5] to set the chirality of diol derivative **3**. Azide displacement of the diol dimesylate and reduction led the inverted diamine **4** that served as the precursor of both **5** ($B'H_2$) and **6** (cf. ref 2c). These compounds cyclooligomerized in 0.5 mM THF containing iPr_2NEt to provide the tetra-Boc derivative of receptor **2** in 18% yield after chromatography on silica gel and an LH20 sizing gel. For technical reasons, it was more convenient to prepare **2** in the enantiomeric series opposite from **1**.

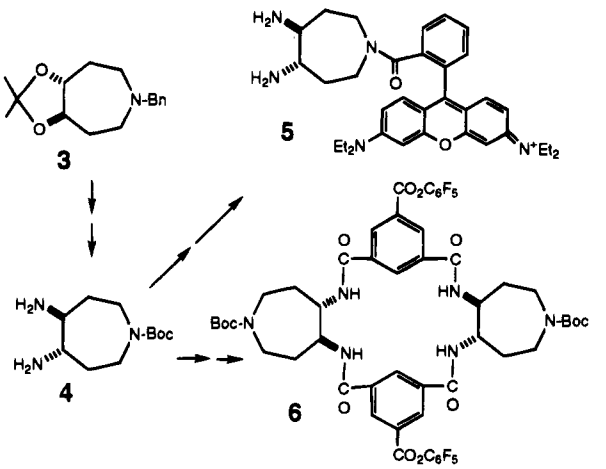


Figure 1. Outline of the synthesis of **2**.

To assay the peptide binding properties of **2** in water, we prepared tripeptide libraries on hydrophilic poly(ethylene glycol)polystyrene (TentaGel)[6] having the general structure AA_3-AA_2-AA_1-$NH(CH_2)_2$-TentaGel. Encoded split synthesis[7] was used to prepare the libraries that consisted of all possible combinations of 29 amino acids[8] at all three positions (AA_n) and produced maximally 29^3 (24 389) tripeptides on TentaGel beads. One such library (the "protected library") was side-chain-protected and acetylated at the N-terminus. Another library (the "deprotected library") had side-chain protection removed and the N-terminus free.

The receptor/library binding assay[2c-f] was carried out by equilibrating a ~10 μM solution of **2** in pH 4[9] water with the TentaGel-supported tripeptide libraries for 24 h. Under these conditions, only one bead per ~1000 picked up the scarlet color of the rhodamine-dyed receptor with both the protected and the deprotected libraries. The color contrast between beads that bound the receptor and those that did not was very large, thus, except for the few beads that were deep scarlet, the libraries were essentially colorless. Such a finding implies selectivities between bound and unbound beads on the order of 3 kcal/mol or more.

The scarlet beads from each library were picked and their chemical tags decoded by electron capture gas chromatography to yield the sequences of the most tightly bound tripeptides. The results in terms of the most frequently found sequences

(1) E.g.: (a) Peacock, S. C.; Domeier, L. A.; Gaeta, F. C. A.; Helgeson, R. C.; Timko, J. M.; Cram, D. J. *J. Am. Chem. Soc.* **1978**, *100*, 8190. (b) Rebek, J.; Askew, B.; Ballester, P.; Doa, M. *J. Am. Chem. Soc.* **1987**, *109*, 4119. (c) Hong, J.-I.; Namgoong, S. K.; Bernardi, A.; Still, W. C. *J. Am. Chem. Soc.* **1991**, *113*, 5111. (d) Murakami, Y.; Ohno, T.; Hayashida, O.; Hisaeda, Y. *J. Chem. Soc., Chem. Commun.* **1991**, 950. (e) Galan, A.; Andreu, D.; Echavarren, A. M.; Prados, P.; de Mendoza, J. *J. Am. Chem. Soc.* **1992**, *114*, 1511. (f) Liu, R.; Still, W. C. *Tetrahedron Lett.* **1993**, *34*, 2573. (g) Konishi, K.; Yahara, K.; Toshishige, H.; Aida, T.; Inoue, S. *J. Am. Chem. Soc.* **1994**, *116*, 1337. (h) Mizutani, T.; Ema, T.; Tomita, T.; Kuroda, Y.; Ogoshi, H. *J. Am. Chem. Soc.* **1994**, *116*, 4240. (i) Cristofaro, M. F. Chamberlin, A. R. *J. Am. Chem. Soc.* **1994**, *116*, 5089 (ref 2).

(2) (a) Yoon, S. S.; Still, W. C. *J. Am. Chem. Soc.* **1993**, *115*, 823. (b) Yoon, S. S.; Still, W. C. *Tetrahedron Lett.* **1994**, *35*, 2117. (c) Yoon, S. S.; Still, W. C. *Tetrahedron* **1995**, *51*, 567. (d) Yoon, S. S.; Still, W. C. *Tetrahedron Lett.* **1994**, *35*, 8557. (e) Yoon, S. S.; Still, W. C. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 2458. (f) Borchardt, A.; Still, W. C. *J. Am. Chem. Soc.* **1994**, *116*, 7467.

(3) Tabet, M.; Labroo, V.; Sheppard, P.; Sasaki, T. *J. Am. Chem. Soc.* **1993**, *115*, 3866. Albert, J. S.; Goodman, M. S.; Hamilton, A. D. *J. Am. Chem. Soc.* **1995**, *117*, 1143. LaBrenz, S. R.; Kelly, J. W. *J. Am. Chem. Soc.* **1995**, *117*, 1655.

(4) Gassman, P. G.; Bonser, S. M.; Mlinaric-Majerski, K. *J. Am. Chem. Soc.* **1989**, *111*, 2652.

(5) Sharpless, K. B.; Amberg, W.; Benanni, Y. L.; Crispino, G. A.; Hartung, J.; Jeong, K-S.; Kwong, H.-L.; Morikawa, K.; Wang, Z.-M.; Xu, D.; Zhang, X.-L. *J. Org. Chem.* **1992**, *57*, 2768.

(6) TentaGel S NH_2 from Rapp Polymere. Bayer, E. *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 113.

(7) A total of 15 tag molecules (five tags per AA_n) were used to encode the library synthesis according to the method reported: 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.

(8) Protected library: AA_n = D-Ala, L-Ala, D-Phe, L-Phe, D-Val, L-Val, D-Leu, L-Leu, D-Pro, L-Pro, D-Asp(*O*-tBu), L-Asp(*O*-tBu), D-Glu(*O*-tBu), L-Glu(*O*-tBu), D-Asn(*N*-Tr), L-Asn(*N*-Tr), D-Gln(*N*-Tr), L-Gln(*N*-Tr), D-His(*N*-Tr), L-His(*N*-Tr), D-Lys(*N*-Boc), L-Lys(*N*-Boc), D-Arg(*N*-Pmc), L-Arg(*N*-Pmc), D-Ser(*O*-tBu), L-Ser(*O*-tBu), D-Thr(*O*-tBu), L-Thr(*O*-tBu), Gly.

(9) At higher pH, selective binding of the rhodamine dye moiety to the deprotected library was observed. See: Wennemers, H.; Still, W. C. *Tetrahedron Lett.* **1994**, *35*, 6413.

0002-7863/95/1517-5887$09.00/0

Table 1. Peptide Sequences Selectively Bound by Receptors **1** and **2**[a]

entry	AA_3	AA_2	AA_1	frequency found[b]	frequency expt[c]
Receptor **2** (pH 4 Water, Protected Tripeptide Library on PEG-PS[d])					
1	(D)Leu,(D)Val	(L)Gln	X	25	0.2
2	(L)Gln	(D)Leu,(D)Val	X	17	0.2
3	X	(L)Gln,Gly	(D)Leu	28	0.2
Receptor **2** (pH 4 Water, Deprotected Tripeptide Library on PEG-PS[e])					
4	(D)Gln,(D)Glu,(D)Asp	(D)Leu	(L)Asp,(L)Glu	34	0.02
5	(D,L)Asp	(D,L)Asp	(D,L)Asp	27	0.03
Receptor **1** ($CHCl_3$, Protected Tripeptide Library on PS[f])[2c]					
6	(L)Val	(D)Asn	X	10	0.4
7	(D)Asn,(D)Gln	(L)Val	(L)Ser,Gly	34	0.1
8	X	(D)Asn,(D)Gln	(L)Val	25	0.9
Receptor **1** ($CHCl_3$, Deprotected Tripeptide Library on PS[f])					
9	X	Gly	(L)Ser	51	0.4
10	(L)Asn,(L)Ser	(L)Pro	Gly	17	0.06

[a] X indicates all possible residues (*i.e.*, little or no selectivity). [b] Percentage of beads selected by receptor binding assay with indicated tripeptide sequence. [c] Percentage of beads having indicated peptide sequence in library (expectation frequency for random bead picking). [d] 24 389 member library on poly(ethylene glycol)polystyrene with standard side-chain protecting groups[8] and N-terminal acetylation. [e] 24 389 member library on poly(ethylene glycol)polystyrene. [f] 50 625 member library[2c] with N-terminal acylation on polystyrene.

for both protected and unprotected libraries in water are given in Table 1. For comparison, tripeptide binding results for receptor **1** in chloroform are also listed.

As anticipated from the small number of beads that bound receptor **2**, a relatively small number of tripeptide sequences were found among the scarlet beads. Indeed, replicate analysis of these sequences provided an estimate that **2** in water binds a total of ~75 and ~30 members of the protected and deprotonated libraries, respectively, each of which contained ~24 000 tripeptides.[10]

Among the tripeptide sequences actually found, there was a great deal of redundancy, especially at the dipeptide level. With the protected tripeptide library in water, receptor **2** had high selectivity for several dipeptide sequences that were remarkably similar to those preferentially bound by receptor **1** in chloroform. In particular, 87% of the beads that bound **2** in water contained (D)Leu or (D)Val adjacent to (L)Gln or Gly (Table 1, entries 1–3). Among these beads, (D)Leu and (L)Gln occurred ~4 times as frequently as (D)Val and Gly. The (D)Leu/(D)Val affinity was particularly large, as every bead decoded contained one of these residues. In comparison with **2**'s binding properties in water, receptor **1** (entries 6–8) in chloroform displays similar peptide selectivity in the form of preferential binding of (L)Val adjacent to (D)Asn or (D)Gln. Thus both receptors **1** and **2** have a strong preference for binding hydrophobic amino acids of similar size (Leu, Val) adjacent to an amide–substituted amino acid having the opposite chirality (Gln, Asn). The different binding enantioselection of **1** and **2** follow simply from their differing chiralities.

With a deprotected tripeptide library, the sequences bound by **2** in water (Table 1, entries 4, 5) differed substantially from those bound by **1** in chloroform (entries 8, 10). Thus, while **1** preferentially bound very different sequences of protected and deprotected peptides, **2** showed a strong preference for binding a hydrophobic D-amino acid (Leu) in both solvents. With **2**, selectivity for binding the deprotected tripeptides was higher in that the hydrophobic amino acid occurred at the AA_2 site in ~90% of the scarlet beads. This position is flanked on both sides by other amino acids, which as entry 4 shows are primarily (81% of the beads) the carboxylate-bearing residues Asp or Glu, residues that likely associate with the ammonium ions at the periphery of **2**'s hydrophobic binding site. There is significant selectivity for AA_3 (especially), being a D, and AA_1, being an L-amino acid.

As shown in Table 1, entry 5, we find also a second type of deprotected tripeptide that binds **2** strongly in water. These sequences have a run of three aspartates with no particular stereoisomer being preferred. In some ways this binding is trivial and reflects simple ionic association between the anionic peptide and the cationic **2**. On the other hand, it is noteworthy that of the 12 scarlet beads we decoded that bore no hydrophobic amino acid, 11 carried diastereomers of Asp-Asp-Asp. The sole remaining bead carried Asp-Asp-Glu. Thus the shorter carboxylate-bearing side chain of Asp is significantly preferred by **2** and may reflect geometrical requirements of direct interactions between peptidic side-chain carboxylate and receptor ammonium ions.

Interestingly, the selectivity of **2** for binding Asp-Asp-Asp disappears at lower pH: only deprotected peptides resembling that in Table 1, entry 4, were found at pH 2, and all these contained a hydrophobic D-amino acid at the AA_2 site. Indeed, all pH 2 scarlet beads had AA_3 = Gln or Glu (83% D), AA_2 = Leu, Phe, or Val (100% D), and AA_1 = Asp or Glu (57% L). Presumably this pH effect follows from reduction of charge–charge receptor–substrate interactions due to increased protonation of Asp or Glu side-chain carboxylates. Replicate analysis of the 30 scarlet beads that we picked at pH 2 indicated that 10 μM **2** bound only 12 sequences out of the entire ~25 000 member deprotected tripeptide library.

These findings demonstrate that highly sequence-selective synthetic receptors for peptides in water can be constructed from certain designs developed for use in organic solvents. In this work, the designs for **1** and **2** are typical of many hostlike molecules (and biological receptors) in that they incorporate a nonpolar, conformationally rigidified binding cavity surrounded by polar functionality for association with substrate functionality. While the driving forces for binding peptides by **1** in chloroform and **2** in water are certainly different, it is interesting that the two receptors have similar peptide-binding selectivities—in particular, the remarkable preference for particular chiralities of the hydrophobic amino acids Val and Leu. Given such binding similarities, it is likely that **1** and **2** bind Leu/Val-containing peptides in similar ways, via insertion of peptidic aliphatic side chains into the receptors' large, lipophilic binding cavities.[2a] With **1** in chloroform, the driving force is formation of amide–amide hydrogen bonds, with van der Waals attractions between the peptide side chain and the binding cavity playing a minor role. With **2** in water, the driving force is likely hydrophobic (*e.g.*, removal of the aliphatic peptide side chain from water), with amide–amide hydrogen bonding acting more gratuitously to orient the peptide within the binding site. In any case, **2** may be considered to be a prototypical host for peptides in water, and its key structural features (rigidified three-dimensional binding cavity, surrounding polar functionality, hydrophilic surface) should be readily incorporated into other receptor designs having different sequence selectivities.

Acknowledgment. This work was supported by NSF Grant CHE92 08254 (W.C.S.) and an F.P.I. Postdoctoral Fellowship from the Spanish Ministerio de Educación y Ciencia (M.T.).

Supplementary Material Available: Protected and deprotected tripeptide sequences that are bound by **2** and a full synthesis scheme for **2** (2 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, can be ordered from the ACS, and can be downloaded from the Internet; see any current masthead page for ordering information and Internet access instructions.

JA9507269

(10) A referee has suggested that the peptide-binding selectivity of **2** may be much less than these numbers indicate because many library peptides are hydrophobic and might be buried in locally hydrophobic regions of the PEG-PS resin, thus being inaccessible to hexacationic **2**. While we cannot completely rule out the operation of such an effect, available evidence suggests that it is not a major determinant of selectivity, as **2** in pH 4 water has no trouble binding highly hydrophobic protected peptide sequences that are related to entry 1 of Table 1 including (D)Leu-(L)Gln(*N*-trityl)-(D)Phe and (D)Leu-(L)Gln(*N*-trityl)-(D)Leu (see supplementary material for other hydrophobic sequences bound by **2**).