Synthetic Receptor Binding Elucidated with an Encoded Combinatorial Library

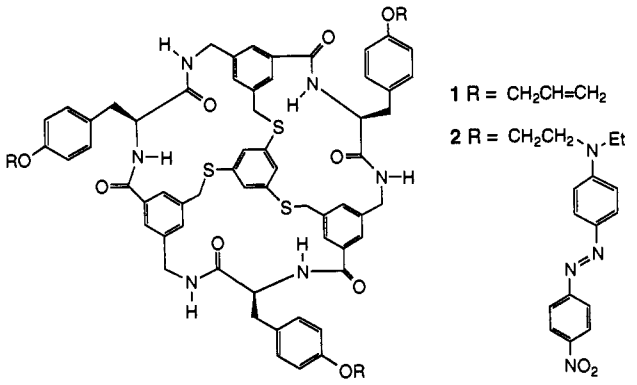
Allen Borchardt and W. Clark Still*

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

Received September 30, 1993

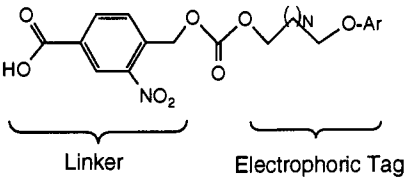
For more than a decade, researchers have studied the binding of synthetic receptors and substrates to learn about molecular complexation. But because binding studies are labor intensive, a small number of binding measurements are usually carried out and often uncover only those binding phenomena explicitly sought. In this communication, we describe an alternative method for evaluating binding using libraries of substrates.[1] We show that the method not only allows the screening of >50 000-member substrate libraries for binding but also can uncover selectivities that would be very difficult to find using traditional means.

Our assay involves linking a receptor (e.g., **1**) to a dye and then adding a library of substrates attached to Merrifield beads. Beads bearing substrates which tightly bind the receptor will become colored. The only difficulty is in determining the structure of the 10–100 pmol of substrate on each colored bead.[2] To solve this problem, we use a binary-encoded combinatorial substrate synthesis which tags each bead with a unique molecular record of its synthesis.[3,4]



Beginning with **1**,[5b] we removed the allyl protecting groups ($(Ph_3P)_4Pd$/dimedone) and alkylated (Bu_4NF, THF) the remaining part with the mesylate of the azo dye, disperse red, to yield the red-colored **2**. Next we prepared an encoded substrate library of N-acylated tripeptides. The split synthesis method[6] was used and resulted in a set of Merrifield beads having only one type of tripeptide per bead and having ~50 000 different kinds of beads incorporating all possible combinations of the 15 acylating agents (R′COX) and 15 amino acids (AA1–AA3) shown in Table 1.

Library synthesis was preceded by preparation of 16 different tag molecules having the structure shown below and functionalized with electrophoric polyhalobenzenes (Ar) as described previously.[3]



We designate these compounds as T1–T16, where their designation number corresponds to the gas chromatography (GC) retention order of the tags that are released by photolysis.

Next we divided ~3 g of 50–80-μm (aminomethyl)polystyrene beads into 15 equal portions and placed each into a different synthesis vessel. We then coupled different combinations of tags T1–T4 such that the beads in each vessel were labeled by a unique combination of tags as indicated in Table 1. We used diisopropylcarbodiimide to attach each different tag to the aminomethyl resin at a level corresponding to ~0.5% of the free amino groups. Next we coupled[7] the Fmoc amino acids (AA1) given in Table 1 in the vessels indicated. We then combined and mixed the 15 portions of beads.

Next the library was deprotected (piperidine) and once more portioned into 15 vessels. The beads in each vessel were again tagged (using T5–T8), and the corresponding amino acid (AA2) was coupled as in Table 1. The procedure was repeated once more using tags T9–T12 for the AA3 coupling and then using T13–T16 to encode the final acylation step. Because 15 different reactants were used in each of the four stages of the synthesis, the final substrate library had as many as 15^4 (50 625) different members.

We then screened the library for binding by mixing a 0.1-g sample (~10^6 beads) with 50 μM **2** in $CHCl_3$. After 24 h, we found that ~10% of the beads had developed pink-red colorations and that ~0.5% of the beads were stained deep red. We picked 50 of these deep red beads and photolyzed (350 nm) each separately in 1–2 μL of DMF to release the tag alcohols. After silylation ($CH_3C(OTMS)NTMS$), electron-capture GC was used to analyze the tag complement of each bead and thus define the structures of each tight-binding substrate.[8] The residues found at each position of these 50 substrates are summarized in Table 2 along with the number of instances each residue was found (see also supplementary material).

These data indicate that **2** discriminates between substrates most effectively when structural differences occur near the free end of the substrate chain. Thus the number of accepted residues is minimal in the case of the terminal acylating group (R′) and increases with distance from the terminus. Such high selectivity for the N-terminal group is consistent with our previous studies of **1**.[5a,c] Our current results, however, indicate that the binding properties of **2** (and **1**) are more remarkable than we previously realized. For example, in 49 of the 50 deep red beads we decoded, the terminal R′ was composed of exactly three non-hydrogen atoms. We also observed extraordinary selectivity among the three-atom R′ groups in the library in the form of complete selection of R′ = cyclopropyl over R′ = isopropyl. Other notable trends include exclusive binding of L-amino acids at the N-terminal AA3 site. Remarkably, AA3 is most commonly Gln, but the

(1) Previous multiple substrate binding assays: (a) Smith, P. W.; Chang, G.; Still, W. C. *J. Org. Chem.* **1988**, *53*, 1587. (b) Schwabacher, A. W.; Lei, H. *J. Org. Chem.* **1990**, *55*, 6080. (c) Chu, Y.-H.; Whitesides, G. M. *J. Org. Chem.* **1992**, *57*, 3524.

(2) This scheme is similar to that used to establish the binding properties of antibodies: Geysen, H. M.; Mason, T. J. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 397 and references therein.

(3) 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.

(4) (a) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5381. (b) Kerr, J. M.; Banville, S. C.; Zuckermann, R. N. *J. Am. Chem. Soc.* **1993**, *115*, 2529.

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

(6) (a) Furka, A.; Sebestyen, M.; Asgedom, M.; Dibo, G. *Abstr. 14th Int. Congr. Biochem.* (Prague, Czechoslovakia) **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).

(7) Steward, J. M.; Young, J. D. Solid Phase Peptide Synthesis, 2nd ed.; Pierce Chemical Co.: Rockford, IL, 1984.

(8) More precisely, the tags record of the sequence of steps to which a given bead has been subjected.

0002-7863/94/1516-0373$04.50/0

Table 1. Components of Encoded Combinatorial Library of Substrates

R′(C=O)-AA3-AA2-AA1-NH$(CH_2)_5$CONH-Support

vessel	N-terminal R′	T13–T16[a]	AA3[b]	T9–T12[a]	AA2[b]	T5–T8[a]	AA1[b]	T1–T4[a]
1	methyl	1 0 0 0	Gly	1 0 0 0	Gly	1 0 0 0	Gly	1 0 0 0
2	ethyl	0 1 0 0	D-Ala	0 1 0 0	D-Ala	0 1 0 0	D-Ala	0 1 0 0
3	isopropyl	0 0 1 0	L-Ala	0 0 1 0	L-Ala	0 0 1 0	L-Ala	0 0 1 0
4	*tert*-butyl	0 0 0 1	D-Ser	0 0 0 1	D-Ser	0 0 0 1	D-Ser	0 0 0 1
5	*tert*-amyl	1 1 0 0	L-Ser	1 1 0 0	L-Ser	1 1 0 0	L-Ser	1 1 0 0
6	trifluoromethyl	0 1 1 0	D-Val	0 1 1 0	D-Val	0 1 1 0	D-Val	0 1 1 0
7	isobutyl	0 0 1 1	L-Val	0 0 1 1	L-Val	0 0 1 1	L-Val	0 0 1 1
8	$MeOCH_2$	1 0 0 1	D-Pro	1 0 0 1	D-Pro	1 0 0 1	D-Pro	1 0 0 1
9	cyclopropyl	1 0 1 0	L-Pro	1 0 1 0	L-Pro	1 0 1 0	L-Pro	1 0 1 0
10	cyclobutyl	0 1 0 1	D-Asn	0 1 0 1	D-Asn	0 1 0 1	D-Asn	0 1 0 1
11	cyclopentyl	1 1 1 0	L-Asn	1 1 1 0	L-Asn	1 1 1 0	L-Asn	1 1 1 0
12	$AcOCH_2$	0 1 1 1	D-Gln	0 1 1 1	D-Gln	0 1 1 1	D-Gln	0 1 1 1
13	phenyl	1 0 1 1	L-Gln	1 0 1 1	L-Gln	1 0 1 1	L-Gln	1 0 1 1
14	Me_2N	1 1 0 1	D-Lys	1 1 0 1	D-Lys	1 1 0 1	D-Lys	1 1 0 1
15	morpholino	1 1 1 1	L-Lys	1 1 1 1	L-Lys	1 1 1 1	L-Lys	1 1 1 1

[a] Tagging molecules used for encoding (0 and 1 denote absence and presence of tag, respectively). [b] Main-chain protection, Fmoc; side-chain protection, Ser (tBu), Asn (β-trityl), Gln (γ-trityl), Lys (ε-Boc) (see supplementary material).

Table 2. Residues Found in Substrates Bound by Receptor 2 and Frequencies of Occurrence of Each Residue (in brackets)

R′	AA3	AA2	AA1
cyclopropyl [23]	L-Gln [21]	L-Pro [21]	L-Ala [8]
Me_2N [15]	L-Ala [15]	L-Ala [13]	L-Ser [8]
$MeOCH_2$ [11]	Gly [6]	Gly [6]	D-Ala [8]
ethyl [1]	L-Lys [4]	L-Val [5]	Gly [6]
	L-Ser [4]	L-Asn [1]	L-Gln [4]
		D-Gln [1]	D-Pro [3]
		L-Lys [1]	D-Ser [2]
		D-Lys [1]	D-Asn [2]
		D-Ala [1]	L-Lys [2]
			L-Val [2]
			D-Val [2]
			L-Pro [1]
			D-Gln [1]
			D-Lys [1]

closely related Asn is not found at all. We also found high selectivity for L-Pro and L-Ala at the AA2 site, a substrate position where previous studies of **1** had detected no selectivity.

To verify that our assay reflected simple receptor–substrate complexation, we resynthesized several library peptides as their *n*-hexylamides and measured their association with **1** in $CDCl_3$ by NMR titration. The results in Table 3 confirm that the color assay mimics solution-phase binding. These studies also suggest that binding energy differences as small as 1 kcal/mol can be distinguished by our solid-phase assay. Interestingly, the preference of **1** for R′ = cyclopropyl extends to cyclopropanoyl derivatives of L-Ala itself which are bound more enantioselectively ($\Delta\Delta G$ = 3.4 kcal/mol) than any other Ala derivative yet studied.[5a,c]

By conducting the assay at a lower receptor concentration, beads bearing only the most tightly binding substrates could be found. Thus with **2** at 10 μM in $CHCl_3$, ~0.02% of the substrate library was deeply stained. Upon decoding these beads, we found a subset of the binding sequences detected by the previous assay including cyclopropanoyl-L-Ala-L-Pro-L-Ala. We synthesized this sequence as the dodecylamide and measured its binding to

Table 3. Binding of **1** and Peptides in Free $CDCl_3$ Solution

peptide substrate	$-\Delta G_{association}$ (kcal/mol)
cyclopropylCO-L-Gln(γ-trityl)-L-Pro-Gly-NHC_6H_{13}	3.8
isopropylCO-L-Gln(γ-trityl)-L-Pro-Gly-NHC_6H_{13}	~1.5
Me_2NCO-L-Gln(γ-trityl)-L-Pro-Gly-NHC_6H_{13}	3.8
cyclopropylCO-L-Asn(β-trityl)-L-Pro-Gly-NHC_6H_{13}	2.9
cyclopropylCO-D-Gln(γ-trityl)-L-Pro-Gly-NHC_6H_{13}	~1
cyclopropylCO-L-Ala-L-Pro-L-Ala-$NHC_{12}H_{25}$	5.7
cyclopropylCO-D-Ala-L-Pro-L-Ala-$NHC_{12}H_{25}$	no binding observed
cyclopropylCO-L-Ala-OtBu	6.6
cyclopropylCO-D-Ala-OtBu	3.2

1 in $CHCl_3$ by NMR titration. As shown in Table 3, this substrate is very tightly bound by **1** in comparison with the other peptides listed and with its diastereomer, cyclopropanoyl-D-Ala-L-Pro-L-Ala-$NHC_{12}H_{25}$, which showed no evidence of binding under the same conditions. This stereoselectivity, which measures >5 kcal/mol, stands among the highest yet observed with a synthetic receptor.

These results demonstrate that combinatorial library methods can dramatically enhance our knowledge of receptor binding. With **1**, they have led to a considerably altered view of its binding despite extensive previous study. The method's main limitation is that not all substrates are easy to prepare by combinatorial methods. Nevertheless, as a method for establishing the scope and general trends of binding with substrates such as ours, the combinatorial library approach is hard to beat.

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

Supplementary Material Available: Complete sequences of 50 substrates 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.