Orthogonal Combinatorial Chemical Libraries

Benoît Déprez, Xavier Williard, Line Bourel, Hervé Coste, François Hyafil, and André Tartar*

Department of Chemistry, Institut Pasteur de Lille
Rue Calmette, 59045 Lille, France
GLAXO Centre de recherches
Avenue du Québec, 91951 Les Ulis, France

Received November 2, 1994

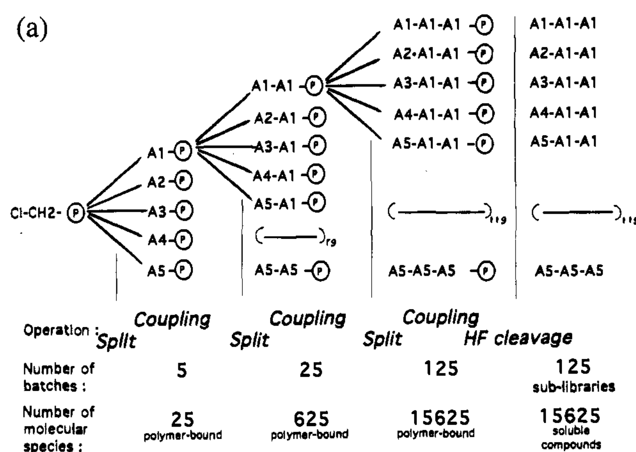
Although knowledge-based *de novo* design of high-affinity ligands for receptors or enzymes appears to be the most rational approach to the discovery of new pharmacologically active substances, high-throughput random screening of compounds having no known structural similarity with the natural ligand has proved to be a highly efficient method. To increase the input of new molecules, methods of combinatorial synthesis have been considerably developed in recent years.[1-4] Much work has been devoted to the development of various strategies designed for the identification of the active compounds in a mixture.[5] These strategies often reduce the scope of the chemical combinations and the biological assays. To circumvent these difficulties, we have developed a general concept which we name "orthogonal libraries". Herein we describe the design, synthesis, and first use of a pair of such orthogonal libraries,[6] leading to the discovery of a nanomolar ligand to the V2 vasopressin receptor.

Using a classical solid phase procedure, we prepared two libraries[7] of the same soluble 15 625 trimers resulting from the combination of 25 synthons at each position, linked by two amide bonds (Figure 1a). The synthons were 23 D-amino acids and 2 nonchiral amino acids, glycine and isonipecotic acid (Figure 1b). Each library was partitioned into 125 sublibraries of 125 trimers. For the first library (A), the 25 synthons were partitioned in five groups, A_1-A_5. Each sublibrary resulted from the incorporation of one group, A_n ($n = 1-5$), at each of the three variable positions of the synthesis ($5^3 = 125$). For the second library (B), the same 25 synthons were used, yielding the same 15 625 trimers. They were, however, partitioned *orthogonally* in five different groups, B_1-B_5 (Figure 1b).

The important property of the couple of libraries A and B is that, for any given n and n', A_n and $B_{n'}$ share only one synthon. *Thus, any sublibrary from A and any sublibrary from B share only one trimer.* In a given screen, an active trimer will confer activity to one single sublibrary of A and one single sublibrary of B. This active trimer can be unequivocally identified as the unique trimer shared between these sublibraries. The distribution of the 25 synthons in the groups A_n or $B_{n'}$ has been made in such a way that the structural diversity within a group of synthons is as large as possible. This avoids the synthesis of sublibraries containing structurally related compounds which could generate cumulative biological activities. Typically, a group consisted of one hydrophilic, one aromatic, one charged, and one small-chain α-amino acid in addition to a nonnatural side chain amino acid. Thus, each sublibrary of A or B explores a wide structural space, and if one trimer confers an activity to its sublibrary, the 124 other trimers of this sublibrary are less likely to interfere with the biological assays. This structural diversity within the mixtures submitted to the screening procedure, together with the limited number of chemical entities present in each sample (125), allows a low background and an easy detection of biological activity.

According to our strategy, library A was screened initially in a series of assays. When a positive result was obtained in

(a)



(b)

	A1	A2	A3	A4	A5
B1	D-Leucine	D-Proline	D-Serine	(4-nitro)-D-phenylalanine	D-Isoglutamic acid
B2	D-Arginine	D-Isoleucine	Glycine	D-Threonine	D-Isoglutamine
B3	D-Glutamine	D-Tyrosine	D-Valine	D-ε-Nicotinoyl-lysine	D-Histidine
B4	D-Tryptophane	D-Asparagine	D-Glutamic acid	D-Methionine sulfoxide	D-Alanine
B5	Isonipecotic acid	D-Lysine	D-Tetrahydro-isoquinoleic acid (D-Tic)	D-Aspartic acid	D-Phenylalanine

Figure 1. Synthesis of the two libraries. (a) The two libraries were synthesized by solid phase synthesis.[7] (b) The 25 amino acids were partitioned into five groups of five amino acids, each named A_1-A_5 for library A and B_1-B_5 for library B. In each group, amino acids were selected to maximize the structural diversity. For the synthesis of library A, each group (A_1-A_5) was coupled to a chloromethylated Merrifield resin (Cl-CH$_2$-Ⓟ), and the five mixtures of aminoacyl resins were split into five parts. To each part, one different group of amino acids were coupled. The resulting 25 dipeptidylresins were split again into five parts each and coupled with the mixtures of amino acids, leading, after deprotection and cleavage from the resin, to 125 sublibraries in a soluble form. Library B was prepared according to the same procedure, except that the orthogonal groups of amino acids, B_1-B_5, were used in this case.

(1) Houghten, R. A.; *et al. Nature* **1991**, *54*, 84–86.
(2) Lam, K. S.; *et al. Nature* **1991**, *354*, 82–84.
(3) Gallop, M. A.; *et al. J. Med. Chem.* **1994**, *37*, 1235–1251.
(4) Gordon, M. E.; *et al. J. Med. Chem.* **1994**, *37*, 1385–1401.
(5) Janda, K. D.; *et al. Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10779–10785.
(6) A related concept, named "indexed librairies" has recently been reported for the deconvolution of a mixture of 54 carbamates: Pirrung, M. C.; *et al. J. Am. Chem. Soc.* **1995**, *117*, 1240–1245. Smith, P. W.; *et al. BioMed Chem. Lett.* **1994**, *4*, 2821–2824. This group describes a similar strategy for the screening of 1600 esters or amides.
(7) Boc-protected amino acids were esterified to a chloromethylated resin. Benzyl-based protections were used for the side chains of trifunctional amino acids. 1-Hydroxybenzotriazolyltetramethylisouronium hexafluorophosphate (HBtu) was used for the couplings of Boc-protected amino acids mixtures. To avoid the preferential incorporation of amino acids having the most favorable coupling rates, a first coupling was performed with only 1.1 equiv of acylating species for 1 equiv of amino groups available on the resin. A second coupling step was then performed to force the reaction to completion. Amino acid analysis of 10 sublibraries showed that most amino acids were within 10% of the mean value, the most divergent ratios being within 25% of this value. At the end of the synthesis, the 250 polymer-bound sublibraries were treated for 1.5 h at 0 °C with anhydrous hydrogen fluoride in the presence of *p*-cresol and *p*-thiocresol. The resulting mixtures of peptides were precipitated with a mixture of diethyl ether and pentane, dissolved in trifluoroacetic acid, reprecipitated, and lyophilized from acetic acid. (a) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154. (b) Tam, J. P.; Merrifield, R. B. *The peptides*; Academic Press: New York, 1987; pp 185–249.

0002-7863/95/1517-5405$09.00/0 © 1995 American Chemical Society

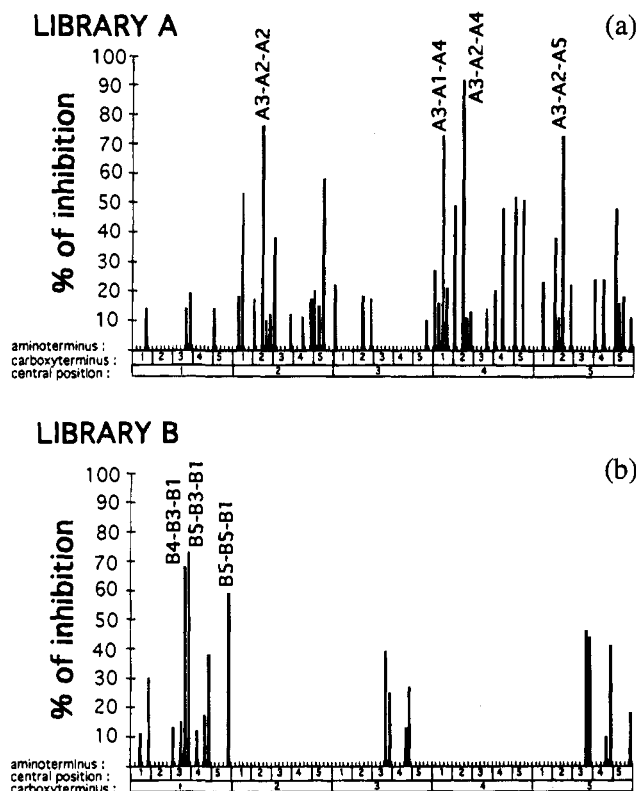


Figure 2. Biological activity of the two libraries in the V2 receptor binding assay. Cells were incubated in the presence of sublibraries from library A (a) or library B (b) at 25 μg/mL. Each bar represents the percentage of inhibition observed for one sublibrary.

one assay, library B was screened under the same conditions. In addition to allowing a direct identification of the active compound, this approach provides an internal control, as this compound is associated with a different set of 124 trimers in library A and in library B. As an example, we report the discovery of a new ligand for the V2 vasopressin receptor. All sublibraries of library A were tested at a concentration of 25 μg/mL, corresponding roughly to 0.2 μg/mL of each individual compound. Activities were determined by the inhibition of binding of radiolabeled vasopressin to renal LLCPK1 cells[8] (Figure 2a). The most active sublibrary was A_3-A_2-A_4, with 92% inhibition; three other sublibraries gave inhibitions higher than 70%: A_3-A_2-A_2, 76%; A_3-A_1-A_4, 73%; and A_3-A_2-A_5, 73%, indicating that some replacements could be tolerated in the middle and carboxy-terminus positions, while in the amino-terminus position, A3 contains a more stringent synthon. When the orthogonal library B was tested (Figure 2b), three sublibraries exhibited high activities, the most potent being B_4-B_3-B_1 with 73% inhibition, B_5-B_3-B_1 with 69%, and B_5-B_5-B_1 with 59%. This selection yielded 12 (4 × 3) possible inhibitors. To narrow the choice, we selected sublibraries on the basis of the dose-dependence of the inhibition and the potency at lower concentration. At 7.5 μg/mL, while only A_3-A_2-A_4 remained active in A (31% inhibition), in B, two sublibraries, B_5-B_3-B_1 (37%)

(8) Porcine kidney epithelial LLC-PK1 cells (American Type Culture Collection) were incubated with 3 nM [^{3}H]AVP in the presence of inhibitors. Following incubation at 4 °C for 180 min, the cells were filtered and washed. Radioactivity was determined by liquid scintillation counting. Lutz, W.; Londowski, J. M.; Sanders, M.; Salisbury, J.; Kumar, R. *J. Biol. Chem.* **1992**, *267*, 1109–1115.

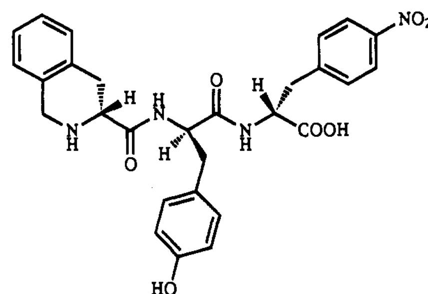


Figure 3. Structure of the trimer common to A_3-A_2-A_4 and B_5-B_3-B_1, H-D-Tic-D-Tyr-(4-nitro)-D-Phe-OH.

and B_4-B_3-B_1 (31%) remained active. The trimer common to A_3-A_2-A_4 and B_5-B_3-B_1, H-D-Tic-D-Tyr-(4-nitro)-D-Phe-OH (Figure 3), was synthetized and turned out to be a potent inhibitor (IC_{50} = 63 nM) of vasopressin binding to LLCPK1 cells. No inhibition of binding could be detected with the candidate common to A_3-A_2-A_4 and B_4-B_3-B_1, H-D-Glu-D-Tyr-(4-nitro)-D-Phe-OH. It is thus possible that activity in the sublibrary B_4-B_3-B_1 resides in an unwanted byproduct when the reactions are performed in pools. The unambiguous identification of D-Tic as a necessary pharmacophore for V2 antagonism is further substantiated by a classical approach[1] to the identification of the active species in A_3-A_2-A_4: among the five mixtures of 25 trimers where ambiguity at the first position has been resolved, only D-Tic-A_2-A_4 (and not D-Glu-A_2-A_4) was active.

In summary, the use of orthogonal self-deciphering libraries offers several advantages: it allows the screening of compounds in their soluble form, it is not limited to polypeptides or other sequencable polymers, and it does not introduce any restriction to the chemical strategies used to generate the diversity. It can even be used for the rapid screening of existing libraries of individual compounds which can be organized as orthogonal mixtures. The interest of this approach increases with the number of screens to which the libraries are submitted, as once the libraries are prepared, no further step of deconvolution or analysis is required.[9] Moreover, orthogonal libraries also offer an internal validation, as any positive result detected in one library must be confirmed by a positive result in the orthogonal library, thus allowing the detection of false positive results.

Acknowledgment. We thank M. G. Néliat (CEREP, Poitiers, France) for performing the binding assays and Ms. Florence Loriolle for technical assistance. B.D. and X.W. are supported by a Region Nord-Pas-de-Calais grant.

JA943561X

(9) Although also based on partitioning the library into sublibraries, orthogonal libraries differ from the positional scanning described by R. Houghten (Pinilla, C.; *et al. Biotechniques* **1992**, *13*, 901–905) in several aspects. A positional scan in this case would have required the synthesis of 75 sublibraries (OXX, XOX, and XXO) instead of 250. However, each sublibrary consists of 25 × 25 = 625 compounds instead of 5 × 5 × 5 = 125, introducing a 5-fold dilution factor (although in each case, the libraries contain 15 625 different compounds, in the positional scanning, each compound is synthesized three times). Moreover, the 625 compounds share the same defined position, introducing the risk that among the different possibilities, the determination of the most active sublibrary will favor the defined residue (O) which is compatible with the greatest number of combinations in the two variable positions (X) and not the residue belonging to the most active molecular entity. This risk is greatly diminished with orthogonal libraries, which are organized to maximize the diversity within each sublibrary. Moreover, each compound is tested in the presence of two different sets of 124 compounds in libraries A and B.