Virtual combinatorial libraries: Dynamic generation of molecular and supramolecular diversity by self-assembly

IVAN HUC AND JEAN-MARIE LEHN

Université Louis Pasteur, Institut Le Bel, 4 rue Blaise Pascal, 67000 Strasbourg, France

Contributed by Jean-Marie Lehn, December 26, 1996

ABSTRACT Molecular and supramolecular diversity may be generated, respectively, by reversible, covalent or noncovalent self-assembly of basic components whose various potential combinations in number and nature represent a virtual combinatorial library. This concept is applied to the induction of inhibitors of carbonic anhydrase (CA) by reversible recombination of aldehyde and amine components. It is found that the presence of CA favors the formation of those condensation compounds that may be expected to present the strongest binding to the CA active site. The virtual combinatorial library approach may represent a powerful methodology for the discovery of substrates, inhibitors, receptors, catalysts, and carriers for a variety of processes.

Combinatorial chemistry has experienced an explosive growth in recent years (1–4). It provides a powerful methodology for exploring the molecular geometrical and interactional spaces through molecular diversity generation in particular for the discovery of new biologically active substances and medical drugs. It rests on the constitution of vast combinatorial libraries (CLs), extensive collections of molecules derived from a set of units connected by successive and repetitive application of specific chemical reactions. It is thus based on large populations of different molecules that are present as discrete entities. Virtual combinatorial chemistry is a conceptually different approach that rests on supramolecular chemistry (5). It relies on a reversible connection process for the spontaneous and continuous generation of all possible combinations of a set of basic components, thus making virtually available all structural and interactional features that these combinations may present. Such multicomponent self-assembly amounts to the presentation of a virtual combinatorial library (VCL; i.e., a potential library made up of all possible combinations in number and nature of the available components) and the selection from it of that entity, among all those possible, that possesses the features most suitable for formation of the optimal supramolecular entity with the target site, by recruiting the correct partners from the set of those available (Fig. 1). The degree of completeness of the set of components/subunits depends on the extent to which the possible combinations cover the geometrical and interactional spaces of the target site.

Self-assembly in a multicomponent system is a combinatorial process with a search procedure directed by the kinetic and thermodynamic parameters imposed by the nature of the components and their interactions.

Whereas CLs are molecular, real and based on preformation of their constituents through covalent nonreversible connections, VCLs are by essence of supramolecular nature since the entity expressed is that giving the strongest binding—i.e., the thermodynamically most stable supramolecular species. They require a reversible assembly process and their constituents may be either molecular or supramolecular. The formation of the virtual constituents of a VCL rests on the combination of its components either by covalent assembly through a reversible chemical reaction for a molecular VCL, or by self-assembly through reversible noncovalent binding interactions for a supramolecular VCL. VCLs thus rely on the spontaneous generation of molecular or supramolecular diversity for the efficient and economical exploration of the structure–energy hypersurface through reversible combination of a set of components.

In a CL there is a one-to-one correspondence between a member of the library and the target, whereas a VCL is a multiplexing system that establishes an n-to-one adaptive correspondence between a set of components and the target site.

Reversibility is an essential feature of VCLs; it operates either at two levels, in the course of the formation of the constituents of a VCL as well as in their binding to the target, or at a single level when the constituents do not exist to a significant amount in the free state outside the supramolecular substrate–receptor entity.

In addition to their difference in constitution, CLs and VCLs also differ in the requirements for screening of the library and retrieval of a given entity. Whereas the complex mixture represented by a CL requires encoding procedures (such as nucleotide sequences, chemical “barcodes,” or radiofrequency microchips) (1–4), a VCL may in principle be reduced to a few constituents or even one constituent, thus greatly facilitating detection and identification of the “active” substance produced.

The generation of CLs or of VCLs may be applied either to the discovery of a substrate for a given receptor or to the construction of a receptor for a given substrate. In the spirit of Emil Fischer’s “Lock and Key” metaphor, the constitution of a CL of substrates amounts to the fabrication of a large collection of keys, with the goal (hope?) that one of them will fit the target lock/receptor and be retrievable from the mixture. On the other hand, a substrate VCL consists in a set of parts that may spontaneously and reversibly assemble to generate a potentially large set of different keys one of which would fit the lock/receptor, the degree of complementarity depending on the way in which the features of the available parts are able to cover the space of all possible shapes. The same image can be applied to CLs and VCLs of receptors for a given substrate.

Thus, in the framework of VCLs, two processes may be considered, depending on whether a receptor or a substrate acts as template for the assembly of the other partner: casting consists in the receptor-induced assembly of a substrate that fits the receptor; conversely, molding consists in the substrate-induced assembly of a receptor that optimally binds/fits the substrate in the substrate (Fig. 1). Both processes involve (i) a
still imperfect and requiring improvement, the present results
nevertheless illustrate the feasibility of the basic concept and
point out paths for future developments.

MATERIALS AND METHODS

Materials. CAII from bovine erythrocytes was purchased
from Sigma and used without further purification. The con-
centration of CA stock solutions was estimated spectrophoto-
metrically with \( \varepsilon_{280} = 5.7 \times 10^4 \text{ M}^{-1}\text{cm}^{-1} \) (24). \( \text{para} \) and
\( \text{meta} \)-sulfamoylbenzaldehydes were prepared from the corre-
sponding nitriles according to a literature procedure (25). The
starting 4-sulfamoylbenzonitrile was purchased from May-
bridge Chemical (Tintagel, UK) and 3-sulfamoylbenzonitrile
was synthesized from 3-carboxybenzenesulfonic acid (26).
Hexyl 4-sulfamoylbenzoate was prepared by heating the com-
mercial acid in a HCl-saturated solution of 1-hexanol. Crystals
were obtained upon addition of hexane to the cooled reaction
mixture.

Compounds 3a, 3b, 3c, and 3d (Fig. 2) were synthesized
according to the following general procedure. To a mixture of
4-sulfamoylbenzaldehyde (1.0 mmol), the hydrochloride salt of
a, b, c, or d (2 equivalents), and Et\(_3\)N (1 equivalent) in MeOH
(5 ml) at 0°C Na\(_2\)BH\(_3\)CN (3 equivalents) was added. After 24 h
at 25°C, the mixture was evaporated to dryness. For 3a, 3c, and
3d, the residue was chromatographed on silica gel eluting with
MeOH/CH\(_2\)Cl\(_2\) (1 to 102 of MeOH in CH\(_2\)Cl\(_2\) depending on
compound). The final product was converted to its hydrochlo-
ride salt. For compound 3b, the residue was chromatographed
on C\(_18\) reversed-phase silica gel, eluting with a MeOH/H\(_2\)O
(from 0 to 70% MeOH) gradient. The purity of these compo-
unds was controlled by reversed-phase HPLC analysis.

Choice of the Library Generating Reaction. For the purpose
of the present experiments, the reaction on which the library
is based has to be reversible and to take place in or near
physiological conditions. This is the case for the fast and
reversible condensation of amines and aldehydes to imines that
allows the system to equilibrate in presence of the receptor.
In addition, the imine(s) formed may subsequently be irreversibly
reduced by NaBH\(_3\)CN (27), thus fixing the compound(s)
generated in the form of amines, which also facilitates the
isolation and analysis of the final equilibrated mixture ob-
tained (9). In addition, in view of the small steric difference

![Virtual combinatorial libraries.](fig1.png)

**FIG. 1.** Virtual combinatorial libraries. (Upper) Schematic representation of the casting process: receptor-induced self-assembly of the complementary substrate from a collection of components/fragments. This process amounts to the selection of the optimal substrate from a virtual substrate library. (Lower) Schematic representation of the molding process: substrate-induced self-assembly of the complementary receptor from a collection of structural components/fragments. This process amounts to the selection of the optimal receptor from a virtual receptor library. The diverse potential constituents of the libraries (center) are either covalently linked or noncovalently bound reversibly generated species that may or may not exist in significant amount(s) in the free state, in absence of the partner.

set of components, (ii) their reversible combination for spon-
taneous diversity generation, (iii) their recognition-directed
selection of one partner by the other one (in fact, both partners
could in principle be self-assembled supramolecular species).
In addition one may wish to lock-in the self-assembled struc-
tures by performing a chemical reaction that irreversibly links
together the components of the entity generated in the optimal
combination.

Both processes also amount to the generation of the fittest and
present adaptation and evolution by spontaneous recom-
bination under changes in the partner(s) or in the environ-
mental conditions. They thus embody a sort of (supra)mo-
elar Darwinism!

The VCL concept described herein bears relation to pro-
cesses and approaches such as the formation of clathrates by
assembly of a solid-state host structure around included sub-
strates (6–8), the template-directed oligonucleotide synthesis
(9), the generation of RNA aptamers (10), the computational
progressive build-up of a host in a binding site (11), the
self-assembly of an inorganic catalyst (12) or of ionophores
(13), the template-induced formation of a metallomacrocyle
(14) or of macrocyclic lactams (15), the construction of a
binding site via metal ion coordination of the components (16),
the equilibrated condensation of multiple components to yield
macrocyclic cryptands (17) or macrocyclic oligocholates
(18), or the formation of a binding site by noncovalent self-
assemble in a molecular monolayer (19) (see also section
9.4.1 in ref. 5). It is also implicit in the previously described
self-recognition operating in helicate self-assembly, which
involves the spontaneous selection of the correct partners in
a mixture (ref. 20, see also section 9.5 in ref. 5). It has in
particular been already exemplified in the self-assembly of a
circular double helicate of specific size directed by the strong
binding of a chloride ion in the central cavity, a process that
amounts to a molding event (21).

We present herein an implementation of the VCL concept
involving a casting process based on the recognition directed
assembly of inhibitors of carbonic anhydrase II (CAII, EC
4.2.1.1). CAII is a well-characterized Zn(II) metalloenzyme
\( M_r = 30,000 \) whose inhibition especially by para-substituted
sulfonamides (22, 23) has been extensively studied. Although

![Precursor amines a-d and aldehydes I-3 and resulting components of the combinatorial library.](fig2.png)

**FIG. 2.** Precursor amines a–d and aldehydes I–3 and resulting components of the combinatorial library. The products are given as the amines resulting from the hydride reduction of the initially formed imines. The alcohols I’–3’ are obtained as side products.
between imines and amines the interference with the cast–
mold interaction is minimized.

**Library Precursor Components.** To generate the library, a
set of aldehydes 1–3 and amines a–d was selected (see Fig. 2)
by the following several criteria to present: (i) comparable
reactivities to avoid bias in the competition (i.e., aromatic
aldehydes and equally nucleophilic aliphatic amines); (ii)
structural variability involving different combinations of polar,
nonpolar, and charged subunits; (iii) similar UV–visible ab-
 sorption spectra to allow an equally sensitive detection of the
different products at a given wavelength; the conjugation
between the aromatic moiety and the electron withdrawing
 group results in absorption maxima between 220 and 240 nm
between the aromatic moiety and the electron withdrawing
 group results in absorption maxima between 220 and 240 nm
for the reductive amination products. The aldehydes and
amines react reversibly to form imines whose proportions
should vary according to their affinity for the casting site on
CA. As final products one may expect to obtain the corre-
sponding reductive amination products as well as alcohols
CA. As final products one may expect to obtain the corre-
sponding reductive amination products as well as alcohols
CA.

**Reaction Conditions and Analysis.** The reaction was per-
formed in the presence of the CA enzyme in water at pH 6 (20
mM phosphate buffer). A 15-fold excess of the starting pri-
mary amines was used to limit side reactions between the
aldehydes and the product secondary amines or amino residues
at the periphery of the enzyme, since besides the terminal
amine, CA has 18 lys e-NH$_2$ groups (23). This excess, which
establishes pseudo-first-order conditions for each aldehyde
with respect to any amine, is also necessary to avoid in the
course of the reaction a shortage of one amine that could
introduce a bias in the library. Because the reaction should
selectively the best ligand(s) for the casting site, no ef-
nicient turnover was expected (for examples of product
inhibition in bisubstrate template systems see refs. 28 and 29).
Thus, a stoichiometric quantity of CA with respect to each
aldehyde was used.

**Reaction Work Up.** Reactions were performed with
and without enzyme: NaBH$_3$CN is added to a solution of amines,
aldehydes, and enzyme in a phosphate buffer and the mixture
is incubated at 25°C for 14 days. The reactions without enzyme
are completed within 24 h. But some reactions were slowed
down by the enzyme when one of the reagents was bound to
it.

Prior to analysis of the library, thermal denaturation of the
enzyme (2 min at 80°C) ensured the release from the casting
site of some possibly tightly bound ligands. The reaction
products were subsequently separated from CA by microcen-
trifuge filtration through a cellulose membrane. Although
slower and less convenient, a simple dialysis should lead to
the same result and allow the recovery of the enzyme. In the
filtration step, control experiments showed that the enzyme
retains more than 80% of one equivalent of a known inhibitor
(hexyl 4-sulfamoylbenzoate) (30) when it was not denatured
and 0% when it was denatured.

**HPLC Analysis of the Reaction Products.** The final mixtures
containing the reductive amination products, the excess of
starting amines and the alcohols produced by direct reduction
of the aldehydes were analyzed by reversed-phase HPLC
with detection at 230 nm. The chromatography was performed on
an Hewlett-Packard HP 1100 apparatus equipped with a diode-
array detector. A C$_8$-reverse-phase column was used (Merck
RP-Select B, 5 μm, 25 × 4 mm) under an optimized binary
gradient (CH$_3$CN/aqueous 50 mM phosphate, pH 6). To
validate the analytical method, some products were synthe-
sized and characterized separately (see above). The other
peaks were all assigned to the corresponding compounds by
comparison to the retention time and UV profile of the peaks
obtained in experiments where one aldehyde, or one aldehyde
and one amine were mixed with the hydride. In general, the
HPLC traces of the mixtures did not contain significant
amounts of impurities from the enzyme, and peaks that could
not be attributed to a product were of small or negligible
intensity.

**RESULTS AND DISCUSSION**

In a first set of experiments, the library was prepared in
absence of CA. The corresponding HPLC trace is shown in Fig.
3A. With the gradient of elution solvent used, the 15 reduction
products and the four starting amines of Fig. 2 are seen as
separate signals. The differences in the signal intensities arise
especially from differences in the extinction coefficients. For
instance, the amines a, c, and d absorb much less at 230 nm
(peaks at 2.2, 12.2, and 20.2 min, respectively) than the reaction
products. None of the starting aldehydes can be detected,
which indicates completion of the reaction, and peaks from
impurities or side reactions are small. When the library is
prepared in the presence of CA, the distribution of products
yields the HPLC trace shown in Fig. 3B. The two chromatograms
look similar, but a closer comparison reveals several novel
features.

(i) Using the signal of amine b (at 11.8-min retention time,
rt) as an internal standard, one observes a lower overall yield
of reductive amination in the presence of CA. For example, the
yield of 1a–1d and 2a–2d is 20–30% lower in the presence of
the enzyme. On the other hand, the relative abundance of
aldehydes 1’ (4 min rt) and 2’ (8.5 min rt) is unchanged in the
presence of CA. This loss of material likely originates from
reactions of the aldehydes with amine groups at the periphery
of the enzyme.

(ii) The proportions of amine products obtained from the
sulfonate aldehyde 1 are unchanged by the enzyme. The same
holds for the carboxylate aldehyde 2.

(iii) The most striking difference concerns the products of
reaction of the aldehyde sulfonamide 3 (peaks marked with a
star on Fig. 2A). The presence of CA causes a 80% decrease
in the abundance of 3a (11 min rt) and 3d (35.5 min rt), it more
than doubles the abundance of 3b (22.5 min rt), and it does not
change significantly the abundance of 3c. Thus, the enzyme has
induced some selectivity, and the reaction of 3 with ben-
zylamine c has been favored compared with the reactions with
a, d, and to a lesser extent, b.

While an absence of effect of the enzyme, as for 1 and 2, does
not allow to eliminate the possibility of an association, the
selectivity observed for aldehyde sulfonamide 3 suggests that
binding does occur for this compound, as could be expected
from the similarities between its structure and that of known
inhibitors (23). The selectivity observed between 3a–3d re-
lects the different binding affinities of the intermediate
imines for CA, the best ligands being more abundant in the
end.

Further support for this interpretation was provided by
experiments where only one aldehyde and two amines were
allowed to compete. With the aldehyde 1 or 2, the presence of
CA did not result in a change of the proportions of the products
but simply resulted in a slight loss of material. When the
aldehyde was 3, competitions between a + c, b + e, or d + c
were influenced by CA. As shown in Table 1, the reaction of
3 with benzylamine c was favored by CA at the expense of the
other amines with the same trend than that observed for the
complete library.

The results of two of these competition experiments are
presented in Fig. 4. After benzylamine c, glycyld-D-phenylala-
nine b is the second best competitor for aldehyde 3 and the
ratio 3c/3b is only increased by a factor of 4.5 in the presence
of CA (see chromatograms in Fig. 4A and B). Amine d is the
poorest reagent in the condensation with 3 for binding to CA,
and 3d is a very minor product in the chromatogram in Fig. 4D.
the ratio 3c/3d is multiplied by 20 in the presence of CA. Also, the effect of CA is drastically diminished in the presence of 1 equivalent of the inhibitor hexyl 4-sulfamoylbenzoate, which competes for the binding site of CA and, thus, partially secludes the aldehyde to the solution where it reacts with a particular amine as determined by the reactivities in the free state.

The emergence of 3e as a major competitor in the library is consistent with previous studies of inhibitors of CA. The Zn(II) ion is located at the bottom of a conical cleft where para-substituted aromatics such as aldehyde 3 are bound with dissociation constants in the submicromolar range. In addition, two secondary hydrophobic binding sites have been located in the vicinity of this cleft. One of them is very near the sulfonamide binding site and is responsible for the high affinity for CA of 4-sulfamoylbenzoic acid benzylamide ($K_d = 1.1 \text{ nM}$) (31), a compound very similar to the imine precursor of 3c. By contrast, glycinamide substituents on an aromatic sulfonamide as in 3a and 3b, do not enhance the affinity for CA (32). The benzyl groups of 3b and 3d are too far from the arenesulfonamide moiety to fit in the nearest hydrophobic secondary binding site, and too close to reach the more distant one (33). The binding of amines b and d in the hydrophobic sites may impose a disposition that does not favor imine formation with the aldehyde group of 3 protruding out of the zinc pocket.

In a last set of experiments, the different amines were allowed to compete for 3-sulfamoylbenzaldehyde, the meta-substituted analog of 3. In the presence of CA (1 equivalent), the reaction was slowed down even more than for 3 and yielded 3-sulfamoylbenzyl alcohol as a very major product. In this case, the aldehyde may be located too deeply in the zinc pocket so that imine formation is hindered but reaction with the hydride is still possible. This is again consistent with the lower affinity of meta-substituted arenensulfonamides for CA (30).

**CONCLUSION**

The results described herein illustrate the feasibility of the dynamic virtual combinatorial approach to combinatorial chemistry and positions it within the framework of supramolecular chemistry. A great variety of extensions may be envisaged. The present case of a virtual library of nonnatural substrates for a biological receptor may be complemented by the design of systems producing virtual libraries of artificial receptors for biological substrate molecules. If the basic components of a virtual receptor library bear functional groups capable of performing a reaction on the substrate to be bound, application of VCLs to the discovery of novel substrate-
selective reagents and reactions may be considered (e.g.,
virtual "artificial enzyme" libraries). The same holds for
self-assembled carrier species susceptible to performing selectiv-
membrane transport processes.

The VCL approach also bears more or less close conceptual
analogies to a number of processes belonging to other areas of
science. To consider just one such case, one may note that in
the assembly-forming corrections scheme of brain function
(34), individual brain cells rapidly change the partners with
which they synchronize their responses, so that the same cells
are used in different constellations, in a sort of VCL of
neurons.

Further exploration of the specific applications and of the
broad implications of the VCL concept is being pursued.

Sci. 19, 57–64.
3. Terrett, N. K., Gardner, M., Gordon, D. W., Kobylecki, R. J. &
Special Issue 29, 112–170.
9197–9198.
31, 979–988.
5545–5571.
mun., 1697–1698.
118, 1545–1546.
17. Jazwinski, J., Lehn, J.-M., Lilienbaum, D., Ziessel, R., Guilhem,
1694.
Soc. 117, 11833–11838.
Acad. Sci. USA 90, 5394–5398.
21. Hasenkopf, B., Lehn, J.-M., Kneisel, B. O., Baum, G. & Fenske,
(1991) The Carbonic Anhydrases: Cellular Physiology and Molec-
ular Genetics (Plenum, New York).
5823.
12, 954–966.
116, 10296–10297.
B 193, 107–125.
6, 559–564.
Soc. 116, 5057–5062.