# Directed combinatorial chemistry

*Joseph C. Hogan Jr*

**Combinatorial chemistry has given chemists access to vast numbers of molecules, but selecting the right one has proved more difficult. As chemists have gained experience in the technique, however, it has become possible to use solid- or solution-phase syntheses with different chemistries and scaffolds to produce libraries tailor-made for finding or optimizing a lead directed at almost any class of target.**

COMBINATORIAL chemistry is based on two premises: that the probability of finding a molecule by random screening is proportional to the number of places one looks for it; and that the simultaneous generation of numerous molecules provides numerous places to look. It relies on producing all possible combinations of a basic set of modular components. As illustrated in Table 1, the number of products attainable from a given set of components increases exponentially, while the number of components required increases only arithmetically. The result is a fundamental shift away from traditional stepwise organic synthesis to reaction and process design strategies that allow for the simultaneous production of large sets of related molecules.

Such sets were initially viewed as a source of new molecules for high-throughput screening ('random' screening, but in a sense quite different from the 'random' screening of natural molecules, which have been preselected for biological utility by nature and are therefore not random). The present challenge for the combinatorial chemist is to extend this technology from being simply a highly productive way of generating random 'hits' in high-throughput screens to a powerful general method for accelerating all aspects of lead generation, lead optimization and drug development.

TABLE 1 Creating chemical diversity from sets of building blocks

| No. of blocks linked | Basic set of 10 | Basic set of 100 |
|---|---|---|
| 3 | $10^3$ (1,000) | $100^3$ (1 million) |
| 4 | $10^4$ (10,000) | $100^4$ (100 million) |
| 5 | $10^5$ (100,000) | $100^5$ (10 billion) |

## Diversity versus information

A number of descriptors have been developed to characterize the diversity of a molecular set mathematically (reviewed in ref. 1). In its simplest terms, diversity is usually equated with the number of different molecules. But large numbers of molecules do not in themselves necessarily provide a drug candidate (or any other desired product). In fact, the ability to generate molecules has inherent upper limits. A complete library of 60-residue (60-mer) permutations of the 20 coded amino acids, for instance, simply cannot be produced, as it would consist of $20^{60}$ (= $1.15 \times 10^{78}$) different peptides, and there is not enough mass in the Universe to provide even one copy of each such peptide. Preselection is clearly necessary.

Large numbers of molecules also present a major challenge for information management, especially if the structural information obtained from biological screening is to be maximized. It is as if a 'combinatorial genie' has been conjured up that can take the chemist wherever he or she wishes to go. The chemist can imagine an unlimited number of molecular destinations, each with a characteristic shape, size and physical properties. The problem is to direct the genie. The variety of molecules that must be produced to generate a hit for a given target is highly dependent on the information with which the chemist starts. In fact, molecular diversity and information are inversely proportional:

$$\text{required diversity} \propto \text{library size} \propto 1/\text{available information}$$

Thus, with no information about a target, one needs 'infinite' diversity, or as many different molecular shapes and sizes as possible (a problem sometimes referred to as the 'diversity issue'). The challenge is to find selectors with which to pick discrete structural themes to produce a manageable number of possibilities. The diversity requirements of libraries designed to generate hits differ from those designed to optimize lead compounds to produce a drug candidate. Any information about the structure–activity relationship abruptly changes the requirement for diversity, directing the process towards producing an envelope of analogues around the active structure.

## Generating large numbers of molecules

**Split-and-recombine approach.** This method, first introduced by Furka[2], uses a sequence of solid-phase reactions, carried out in parallel on aliquots of any of a variety of solid-phase supports, typically porous resin beads. Each aliquot is functionalized with one of a set of modular building blocks using a common connection chemistry. The functionalized aliquots are then combined to produce a uniform pool that is subsequently divided into further aliquots, each containing examples of all of the functional blocks from the first step. A second synthetic step is then carried out as before, and so on. The final product is a mixture of individual molecules attached to the support, with each individual support unit having multiple copies of an individual product attached to its surface. The multiplication provided by repetitive splitting and recombination can generate very large numbers of molecules.

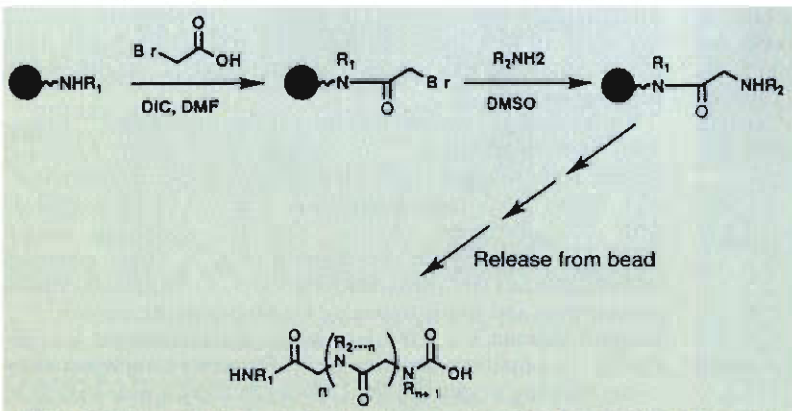


FIG. 1 Synthesis of oligomeric *N*-alkyl glycines. DIC, 5-(3,3-Dimethyl-1-diazenyl)-1H-imidazole-4-carboxamide (dacarbazine); DMF, dimethylformamide; DMSO, dimethylsulphoxide.

Originally, this method was used to produce huge libraries of peptides (using traditional connection chemistries[3–5]). Active peptides were isolated and identified using various deconvolution, sorting and microsequencing strategies. These libraries were quite successful in generating hits in several assays, but their use was severely limited by the unsuitabil-

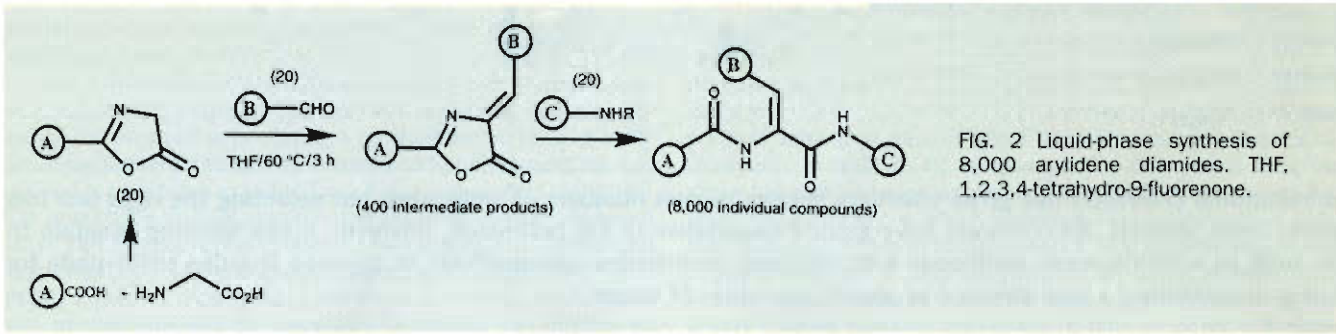


FIG. 2 Liquid-phase synthesis of 8,000 arylidene diamides. THF, 1,2,3,4-tetrahydro-9-fluorenone.

ity of peptides as drugs and the extreme difficulty of translating them into non-peptidic drug candidates. Large libraries of oligonucleotides have also been produced in this way, but these molecules have generally not produced useful leads.

The approach using organic building blocks was therefore modified to produce large numbers of small organic molecules, which are inherently better potential drug candidates. Because the individual molecules could not be sequenced like peptides and oligonucleotides, methods to tag individual beads for post-assay identification were necessary. Among the approaches developed were: (1) deconvolution by multiple iterations of the synthesis, each with only one substituent at one position, guided by iterative bioassays to arrive at the bioactive molecule; (2) deconvolution through archiving at each step of the split-and-recombine process, obviating the need for multiple resynthesis; and (3) direct analysis of the active molecule by mass spectrometry. The active molecule may be identified either by screening the functionalized support elements or by releasing the synthesized product molecules from the support and screening them in solution. The latter procedure has proved preferable, particularly for cell-based assays. Although the product molecules may be tested individually or as mixtures in either case, mixtures frequently produce equivocal or false results (partly due to synergistic interactions between different molecules and the target).

**Parallel unit synthesis.** This method involves performing modular chemical reactions in parallel, using discrete reaction chambers laid out in a spatially addressable format (such as a 96-well microtitre plate). Typical libraries contain between 1 and 10,000 individual molecules. The design and construction of such single-compound combinatorial arrays requires at least two sets of suitable building blocks, and three or four sets are more common.

Either solution- or solid-phase reactions may be used; for a given set of building blocks and reactions, the choice between them is based on a combination of process considerations (such as the importance of mass action effects, the need for filtration, and so on) and reactivity considerations (such as the incorporation of bireactive blocks, which produces cleaner results using a solid-phase approach).

For solution-phase combinatorial syntheses, attention must be paid to purity and yield requirements for the building blocks and reactions chosen, to the number of reactions involved, and to the nature and amounts of by-products, spent reagents and catalysts left in the product. Methods must also be developed for analysis and scaling up of the production. The spatially addressible format allows direct read-out of structure–activity data for both positive and negative results, as well as rapid access to larger quantities of materials.

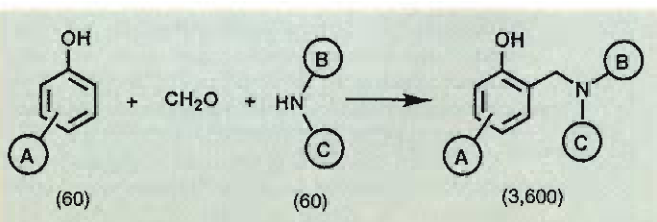


FIG. 3 Solution-phase synthesis of a 3,600-compound Mannich library using the combinatorial reaction of formaldehyde with 60 substituted phenols and 60 secondary amines.

**Comparison of the two methods.** The split-and-recombine method can produce very large numbers of molecules, and is useful in random screening, particularly for targets of unknown structure. On the other hand, only picomole amounts of compounds are produced, follow-up quantities are difficult and costly to synthesize, deconvolution or decoding techniques must be used to extract structural information about active compounds, and solid-phase syntheses suffer from inherent limitations, restricting the method's utility for lead optimization and drug development.

The parallel synthesis method generates fewer molecules, but the spatially addressable format provides structure–activity data immediately, simplifies follow-up production of larger amounts of material, and can in principle access the entire gamut of synthetic reactions. This method can be useful for lead generation, particularly for targets of known structure, and is highly useful for lead optimization and drug development.

## Assembly strategies

In principle, chemical building blocks may be assembled by either of two strategies. The first approach, which is inherently more amenable to solid-phase synthesis, produces oligomers by repeated application of a single coupling chemistry. A classical example is the peptide library referred to above; the synthesis of oligomeric *N*-alkyl glycines[1] shown in Fig. 1 is an excellent non-peptidic example.Using this approach, libraries of oligocarbamates[6], peptide phosphonates[6], vinylogous polypeptides[8] and other oligomeric scaffolds have been produced. These molecules generally possess flexible backbones, which can weaken target binding and can also hinder the application of structure-guided techniques. The approach also tends to add considerable unit weight per variable group.

The second approach produces monomeric molecules in which multiple variable groups are arranged about a central scaffold or core. These molecules can be constructed using either stepwise or concerted reaction schemes. An example of the former is the solution-phase synthesis of 8,000 arylidene diamides with an average purity of 85% using two successive reactions between three blocks connected to a common scaffold (J. C. H., manuscript in preparation) (Fig. 2).

Other examples include libraries of benzodiazepines[9,10], hydantoins[10] and thiazolidines[11]. An example of a concerted reaction scheme is the solution-phase synthesis of a 3,600-compound Mannich library using the combinatorial reaction of formaldehyde with 60 substituted phenols and 60 secondary amines (J. C. H., manuscript in preparation) (Fig. 3). Other examples include libraries of hydroxyaminimides (J. C. H., manuscript in preparation) and multicomponent Ugi/Passerini chemistries.

**General chemistries.** The basic chemical requirements for producing combinatorial libraries are: (1) reactive complementarity of the building blocks; (2) the capacity to carry a wide variety of functional groups; and (3) non-equivocal reaction pathways. The number and variety of building blocks, scaffolds and reactive groups with these properties is systematically being expanded, together with the range of synthetic reactions amenable to both

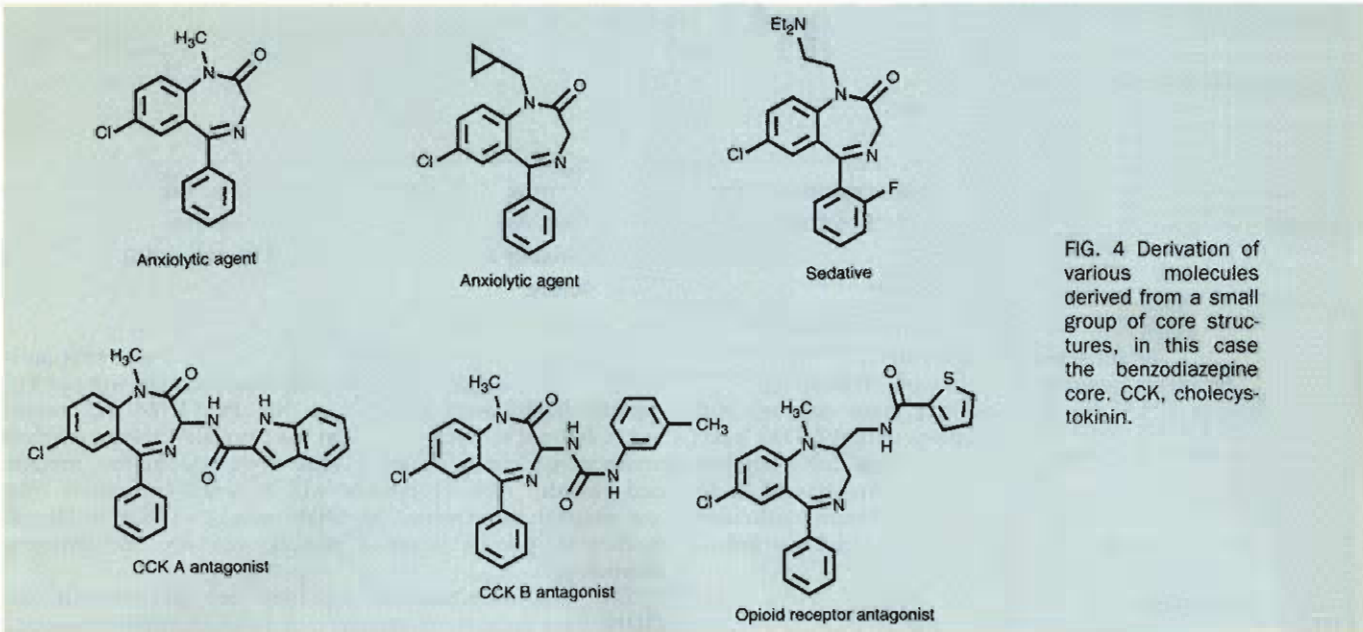


FIG. 4 Derivation of various molecules derived from a small group of core structures, in this case the benzodiazepine core. CCK, cholecystokinin.

solid- and solution-phase processes. There is no reason why most synthetic reactions in organic chemistry cannot be made compatible with the requirements of combinatorial chemistry (reviewed in ref. 12).

**Scaffolds versus substituents.** Many natural molecules are derived from a small group of core structures, including peptides, nucleosides, carbohydrates, steroids and alkaloids. Each of these core structures is generally associated with biological activity, and systematic structural variation of their substituents has been invaluable in guiding the development of new therapeutic agents. The same approach has been used with synthetic core structures such as the benzodiazepine core, which provides potent and selective ligands for a wide range of receptors[13] (Fig. 4).

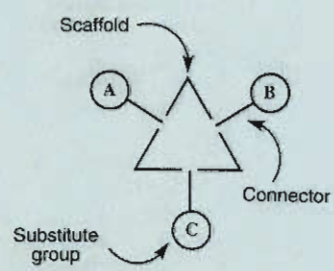


FIG. 5 A simplistic diagram of the relationship between scaffold and substituents.

Nature's preference for certain structural motifs is clearly associated with the fundamental modularity of biological architecture and processes. Any small organic molecule may also be regarded as being constructed from virtual modular components. In the simplest case, these consist of a scaffold and a set of substituents (Fig. 5). The scaffold can have two roles. It may confer biological activity in itself; in this case, varying the substituent groups can produce analogues with improved stability, potency, selectivity, toxicity, bioavailability or effectiveness against resistant organisms. In other cases, the scaffold has no inherent biological activity, and serves as a spatial framework for the display of attached functional groups. In such instances, the scaffold itself must be varied so as to sample as many different shapes and configurations as possible. Any active motifs can then become scaffolds for subsequent optimization through traditional medicinal chemistry, aided by directed combinatorial assays.

But modularity also allows the use of structure-guided design, making combinatorial chemistry into a potent process for the accelerated design of synthetic molecular recognition agents. By selecting an appropriate set of building blocks based on information about a target, a focused combinatorial library can be designed and produced for screening and subsequent iterative optimization, leading to a drug candidate. The required information can be obtained from the structure of the target or its ligands, its biochemical mechanism of action, the structures of active compounds found by screening, or a computational model. This process may eventually improve understanding of the structural rules underlying evolutionary design processes.

## Future developments

As we have seen, combinatorial chemistry is rapidly evolving from a tool for generating large numbers of molecules to a powerful general method for the accelerated creation of molecules with desirable functional properties. The advent of this process in itself is likely to shorten the integrated pharmaceutical development process by about two years, and the availability of specialized cell-based assays to screen for pharmacological and toxicological properties together with the rapidly accelerating discovery of targets using genomics should lead to even more dramatic gains. Similar principles may also have an equally dramatic effect on other areas of chemistry. □

*Joseph C. Hogan is at ArQule, 200 Boston Avenue, Medford, Massachusetts 02155, USA.*

1. Martin, E. *J. Med. Chem.* **38**, 1431 (1995).
2. Furka, A. *et al. Int. J. Pept. Prot. Res.*
3. Houghton, R. A. *Proc. Natl Acad. Sci. USA* **82**, 5131 (1985).
4. Houghton, R. A., Bray, M. K., Degraw, D. T. & Kirby, C. J. *Int. J. Pept. Prot. Res.* **27**, 673 (1986).
5. Houghton, R. A. *et al. Nature* **354**, 64 (1991).
6. Cho, C. Y. *et al. Science* **261**, 1303 (1993).
7. Campbell, D. A. & Bermack, J. C. *J. Org. Chem.* **59**, 658 (1994); **57**, 6331 (1992).
8. Hagihara, M., Anthony, N. J., Stout, T. J., Clardy, J. & Schreiber, S. L. *J. Am. Chem. Soc.* **114**, 6568 (1992).
9. Bunin, B. A. & Ellman, J. A. *J. Am. Chem. Soc.* **114**, 10997 (1992).
10. Hobbd De Witt, S. *et al. Proc. Natl Acad. Sci. USA* **90**, 6909 (1993).
11. Patek, M., Drake, B. & Lebl, M. *Tetrahedr. Lett.* **36**, 2227 (1995).
12. Hermkens, P. H. H., Ottenheijm, H. C. J. & Rees, D. *Tetrahedron* **52**, 4527 (1996).
13. Evans, B. E. *et al. J. Med Chem.* **31**, 2235 (1988).