Primer

Combinatorial chemistry

Anthony W. Czarnik* and Jack D. Keene†

Combinatorial chemistry, which did not even exist fifteen years ago, is a new subfield of chemistry that is starting to provide insights into molecular interactions, biochemical catalysis and drug development. Instead of finding and testing naturally existing compounds for useful activity, the basic aim of combinatorial chemistry studies is to create vast numbers of new, synthetic compounds and screen them for activity, to identify molecules on the basis of function (see green box).

Diverse collections of some biological macromolecules (for example, phage display peptide libraries, antibody libraries and oligonucleotide libraries) can be created via biosynthetic methods but such work is more commonly referred to as 'combinatorial biochemistry'. By contrast, combinatorial chemistry uses synthetic chemistry methods to generate organized collections, or libraries, of compounds including peptides and small organic molecules. This Primer will focus on combinatorial chemistry.

The earliest work in combinatorial chemistry was aimed at discovering ligands for biological macromolecules, such as proteins. Such ligands can be useful tools in understanding the structure and function of proteins and, if the ligand meets certain physiochemical constraints, it might be useful as a drug.

One strategy at the heart of combinatorial chemistry is the concept of combining chemical building blocks, like 'beads on a string', in all possible combinations. This is sometimes referred to as 'matrix chemistry'. If a chemical synthesis route consists of three discrete steps, each employing one class of reagent to accomplish the conversion, then employing one type of each reagent class will yield $1 \times 1 \times 1 = 1$ product as the result of $1 + 1 + 1 = 3$ total reactions. Combining 10 types of each reagent class will yield 1,000 products as the result of as few as 30 total reactions; 100 types of each reagent will yield 1,000,000 products as the result of as few as 300 total reactions. Although conceptually simple, considerable strategy is required to identify 1,000,000 products worth making, and to carry out their synthesis in a manner that minimizes labor and maximizes the value of the resulting organized collection, called a chemical library.

The central approach of combinatorial chemistry — and one which involves a paradigm shift for most organic chemists — is to start with a source of molecular diversity (that is, lots of compounds) organized in a way that makes their empirical testing straightforward, then test them all and analyze the results at the end. This approach is well illustrated by the development of peptide libraries to study ligand–receptor binding.

Peptide libraries

By synthesizing a large number of peptides, each varying from another by only one amino acid, it is possible to determine empirically which amino acid substitutions make binding to a macromolecular receptor stronger and which make it weaker. The problem is how to conveniently make thousands of peptides in a format that can be easily used in the subsequent binding studies. One solution is to synthesize the peptides on a rack of plastic pins (usually called Geysen pins), and to test their binding ability while the peptides remain chemically attached to the pin.

Because there are 20 naturally occurring amino acids, the synthesis of a linear peptide n amino acids long can be done in $20n$ different ways. Thus, there are 64,000,000 possible hexapeptides. It would not be convenient, perhaps not even possible, to synthesize that many peptides on individual pins. An alternative to synthesis on pins is to encase the plastic support used for synthesis in an inert mesh resembling a tea-bag. Collections of such 'tea-bag' reactors are subjected to the chemical addition of an amino acid at the same time. After the addition, the tea-bags are washed thoroughly and the bags (not the bag contents) mixed. The bags are redistributed to new beakers and another amino acid chemically added. In this method, each tea-bag contains only one peptide, which can be obtained by chemically cleaving it from the polymeric support. This type of divide–couple–recombine approach is known as split-pool synthesis.

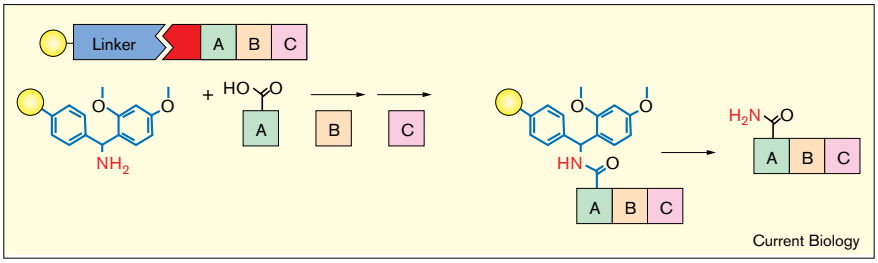
Synthesis and use of a combinatorial library

The goal of combinatorial chemistry is to synthesize very large numbers of chemical entities by condensing a small number of reagents together in all combinations defined by a given reaction sequence. A single starting material is subjected to a library synthesis route, using a range of reagents in each diversity step. The resulting collection of library members may contain from a few hundred to millions of samples. Automated sample distribution provides microtiter plates ready for the screen. After addition of all the 'constant' components for the screening reaction, followed by incubation, the results of each assay are read using absorbance, fluorescence, or radioactivity.

Because there is only one compound per aliquot of solid phase support, individual compounds are easily isolated. In the case of peptides, the structure of a promising compound can be determined from the individual 200 picomole sample. But organic compounds must be tagged during synthesis so that the reaction history of any promising compound is known, allowing its identity to be determined.

To make large numbers of individual peptides would require impossibly small tea-bags, for purely practical reasons. The solution is to use the same split-pool approach, but with polymer resin beads, each only microns in diameter, as a kind of tea-bag. After addition of an individual amino acid to separate reaction vessels, each containing an aliquot of the beads, all the aliquots are recombined, then split again into individual vessels for the next step of the synthesis. In this way, it is possible to synthesize large peptide libraries in which each bead possesses a single peptide that potentially can be recognized selectively by a biological target (antibody, enzyme, or receptor). Each bead carries only about 200 picomoles of peptide but this is enough for both a simple ligand–receptor binding assay and for the analytical techniques required to establish the exact chemical structure of that peptide.

Figure 1



An example of a combinatorial library design, in which chemical building blocks (reagents A,B,C), which make up the potential compounds of the library, are attached sequentially to a polystyrene bead (yellow) through a chemical linker (blue). Attachment of the reagents to the linker is via an amino group (red). A optical bar code or a stable isotope ratio code can be inserted in the linker region, and can be deciphered by mass spectroscopy. Varying ratios of stable isotopes allow thousands of separate codes to be inserted in the linker region. In this example, the red block represents an atom transferred into the ligand that is made up from a varying ratio of isotopes. This and other encoding methods allow tagging of the compounds in a combinatorial library and eventual decoding of their individual history of synthesis. (Adapted, with permission, from Geysen *et al.*, *Chem Biol* 1996, **3**:679-688.)

Organic libraries

Because much of the impetus for discovering tightly binding ligands derives from the pharmaceutical industry, the combinatorial synthesis of 'drug-like' (low molecular weight, organic molecules) compound libraries is of great interest. Two practical considerations make this a greater experimental challenge than the synthesis of peptide libraries.

First, the synthetic methods required to make drug-like molecules on a plastic support have not been optimized. Although solid-supported peptide synthesis saw its origins in the early 1960s and has been extensively developed since then, initial experiments with organic solid-phase synthesis in the early 1970s were not followed up widely. In addition, although there are only 20 naturally occurring amino acids and therefore a finite number of reactions required to use them efficiently, an enormous number of organic chemistry reagents and a very large number of reaction types exist. By the early 1990s, however, the synthesis of moderately-sized organic libraries using the solid-phase synthesis method had been reported. Each approach used a strategy like that of the pin method, and was therefore amenable to the parallel synthesis of hundreds to thousands of compounds.

The second consideration that inhibited the synthesis of much larger libraries using a one-bead-one-compound approach is that, whereas 200 picomoles of an organic compound is enough for a ligand–receptor study, it is not enough to identify the structure of the ligand; in other words, it's not easy to characterize any promising compounds that emerge from the screen. Recently, 'bead tagging', or encoding, solved this problem. The strategy is simple: if the result of a chemical synthesis step cannot be easily read at low concentration, one should add something to the bead that conveniently encodes the reaction history of that bead for later analysis. Indeed, the first reported methods of bead encoding involved the use of biological macromolecules as tags. After each organic synthesis step, either an amino acid or a nucleotide is added to a growing oligopeptide or oligonucleotide on the same bead, such that the sequence can be read later to reveal the identity of the tagged compound.

But neither oligopeptides nor oligonucleotides are chemically inert enough to survive the conditions required in organic synthesis, so more recent methods of encoding include the use of radiofrequency memory microchips and optical bar-coding strategies. A clever method of tagging organic compounds using stable (non-radioactive) isotopes incorporated either into a bead, into a chemical linker or into the chemical backbone of each compound allows encoding and decoding by mass spectroscopy (see Figure 1). For example, by adjusting the amounts of ^{13}C, ^{15}N and deuterium in a series of amino acids that link a bead to a compound, a ratio barcode can be embedded in the library members, with little or no effect on their chemical reactivity.

From flask to well

A library only brings value when screened. The way library members are screened for activity depends on the form in which they are synthesized. In most instances, the compounds are cleaved from the solid support on which they were made and eluted into microtiter plates, such that each well contains one compound.

For a typical high-throughput screen, an experimenter will start with a set of microtiter plates containing library members. All wells are then loaded with the constant components (for example, target protein, buffer, assay reagent), incubated for the appropriate time, and read. This method is most appropriate when screening for activities such as binding or catalysis because the compounds react with a single purified target molecule.

A variation on this method is the use of a bead-based library. In such libraries, the synthesis occurs on a polystyrene bead support; each bead contains one compound. The synthesis is done such that the synthetic library member is connected to the bead via a chemically or photochemically cleavable linker. In this scenario, one bead (or a small number of beads) is placed in a well. The plate is irradiated so that compounds are released from the bead. All of the constants for assay are added, and the assay carried out.

One powerful application of such 'releasable' libraries is that lawn assays on petri dishes can be performed and local effects on living cells observed. If an interesting biological effect is observed surrounding the bead containing the released compound, the bead can be decoded to reveal the identity of the chemical. The released compound can then be resynthesized for subsequent biological experiments. One can readily imagine the potential of this method for the discovery of novel antibiotics or anti-proliferative compounds using cultured cells.

An important paradigm of combinatorial screening is the ability to produce secondary 'focused' libraries using information gained from the primary screening libraries. Focused libraries represent substructures of the original primary library. They can allow one to refine the properties of a selected compound, such as by increasing its affinity of binding or catalytic rate.

Future innovation

Major challenges in combinatorial chemistry focus on both the compound characterisation and screening of very large compound libraries; some believe that it involves statistics and mathematics more than it does chemistry.

As combinatorial chemistry has so recently been applied to drug discovery, there are, as yet, no fully developed success stories in clinical trials to cite. It is claimed, however, that several proprietary examples exist which are in clinical development and soon to be available to the public. In theory, combinatorial chemistry has the potential to uncover numerous biological agonists and antagonists. It could also reveal a wide variety of synthetic receptors, including catalysts and chemosensors. Many combinatorially-derived reagents will prove useful in the research laboratory for understanding basic biological processes, and a few might end up as drugs or drug leads.

One can imagine a tremendous advantage of using combinatorial libraries and high-throughput screening to rapidly derive molecules that inhibit or activate gene products discovered through the international genome projects. For example, the genomes of many of the pathogenic microbes, or 'superbugs', that are becoming resistant to all known antibiotics have been sequenced. As microbes evolve antibiotic resistance at very high rates, humans will need robust technologies such as combinatorial chemistry, using genomic information, to out-evolve the bugs' rapid adaptation to each generation of new antibiotics. One can imagine similar approaches to the treatment of diseases, using genomic information about cancer-susceptibility genes and genes involved in other diseases with complex traits. In this way, the rapid application of combinatorial chemistry and high-throughput screening could conceivably allow therapies to be customized for individual patients.

As only a very small number of biologically-active compounds has ever been sampled from the universe of possible chemicals, the potential to discover novel biological modifiers and highly specific drugs using combinatorial technologies has opened a new frontier in biology and medicine.

Key references

Bunin BA, Ellman JA: **A general and expedient method for the solid-phase synthesis of 1,4-benzodiazepine derivatives.** *J Am Chem Soc* 1992, **114:**10997-10998.

Czarnik AW, DeWitt SH (eds): *A Practical Guide to Combinatorial Chemistry.* Washington: ACS Books; 1997.

DeWitt SH, Kiely JS, Stankovic CJ, Schroeder MC, Cody DMR, Pavia MR: **Diversomers: an approach to non-peptide, non-oligomeric chemical diversity.** *Proc Natl Acad Sci USA* 1993, **90:**6909-6913.

Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D: **Light-directed, spatially addressable parallel chemical synthesis.** *Science* 1991, **251:**767-773.

Geysen HM, Meloen RN, Barleing SJ: **Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid.** *Proc Natl Acad Sci USA* 1984, **81:**3998-4002.

Geysen HM, Wagner CD, Bodnar WM, Markworth CJ, Parke GJ, Schoenen FJ, Wagner DS, Kinder DS: **Isotope or mass encoding of combinatorial libraries.** *Chem Biol* 1996, **3:**679-688.

Houghten RA: **A general method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen–antibody interaction at the level of individual amino acids.** *Proc Natl Acad Sci USA* 1985, **82:**5131-5135.

Joyce GF: **Directed molecular evolution.** *Sci Amer* 1992, 90-97.

Kenan DJ, Tsai DS, Keene JD: **Exploring molecular diversity with combinatorial shape libraries.** *Trends Biochem Sci* 1994, **19:**57-64.

Lander ES: **The new genomics: global views of biology.** *Science* 1996, **274:**536-539.

Ni Z-J, Maclean D, Holmes CP, Murphy MM, Ruhland B, Jacobs JW, Gordon EM, Gallop MA: **Versatile approach to encoding combinatorial organic synthesis using chemically robust secondary amine tags.** *J Med Chem* 1996, **39:**1601-1608.

Ohlmeyer MHJ, Swanson RN, Dillard LW, Reader JC, Asouline G, Kobayashi R, Wigler M, Still WC: **Complex synthetic chemical libraries indexed with molecular tags.** *Proc Natl Acad Sci USA* 1993, **90:**10922-10926.

Wilson SH, Czarnik AW (Eds): *Combinatorial Chemistry: Synthesis and Application.* New York: Wiley & Sons; 1997.

Addresses: *Illumina, 9390 Towne Centre Drive, STE 200, San Diego, California 92121-3015, USA; †Department of Microbiology, Duke University Medical Center, Durham, North Carolina 27710, USA.