Combinatorial chemistry

Combinatorial chemistry is a new subfield with the goal of synthesizing very large numbers of chemical entities by condensing a small number of reagents together in all combinations defined by a small set of reactions. This article provides a general overview of the subfield, describing the chemical library, and then examines how the necessary screening and assaying is carried out.

Chemical Libraries

Combinatorial chemistry is sometimes referred to as matrix chemistry. If a chemical synthesis consists of three 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 $10 \times 10 \times 10 = 1000$ products as the result of as few as $10 + 10 + 10 = 30$ total reactions; 100 types of each reagent will yield 1,000,000 products as the result of as few as 300 total reactions. While the concept is 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 earliest work was motivated by a desire to discover novel ligands (that is, compounds that associate without the formation of covalent bonds) 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 may be useful as a drug. For this reason, pharmaceutical applications provided early and strong motivation for the development of combinatorial chemistry.

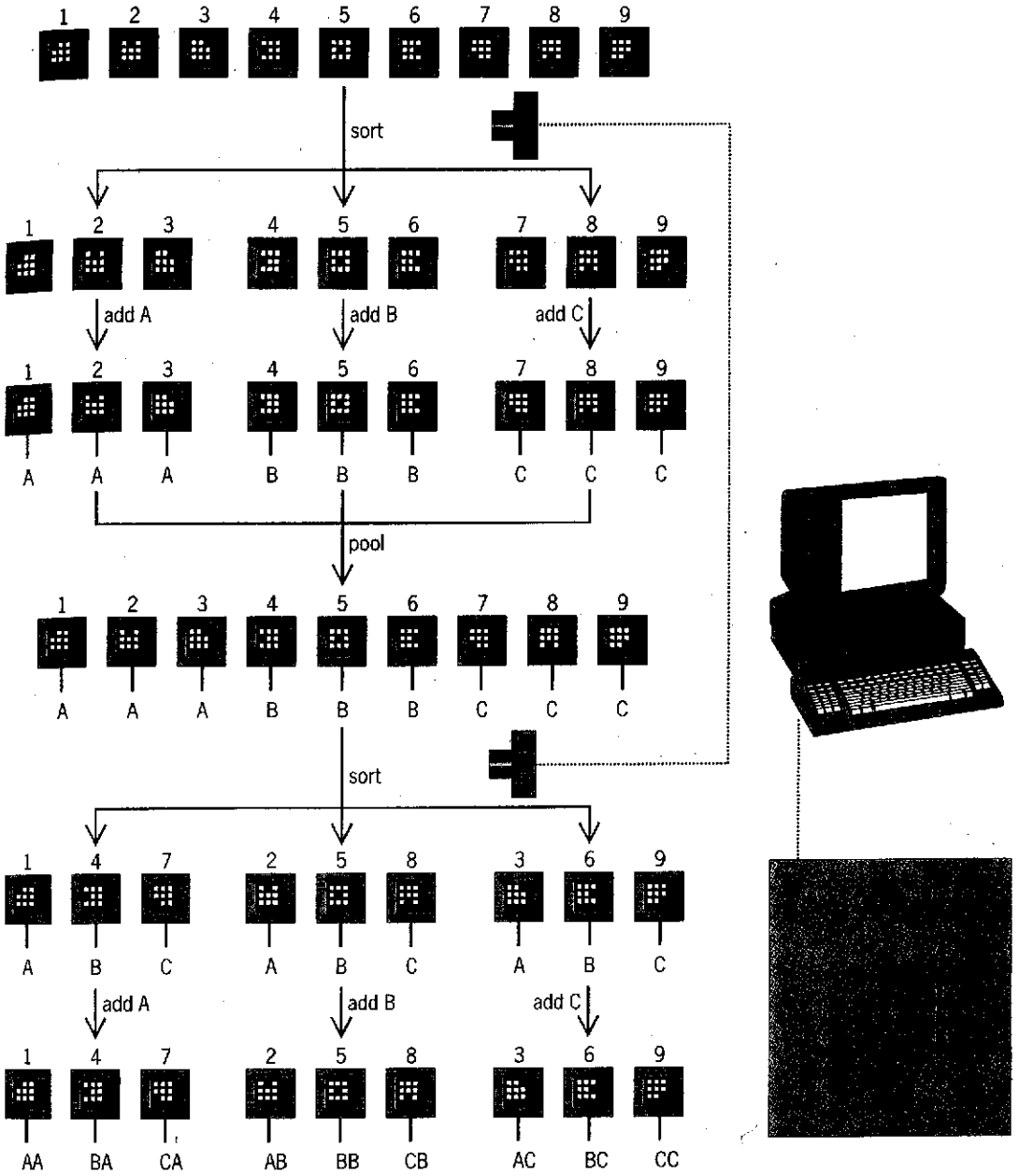
Peptide libraries. In 1984, H. M. Geysen reported on the synthesis of a library of peptides and the screening of that library to probe how changes in a single amino acid in a peptide would change its association strength with an antibody. Biochemists assume that the detailed chemical structure of any ligand will affect that ligand's binding strength to a macromolecular receptor, but even with powerful modern computer modeling methods it is normally not possible to predict either the magnitude or even the direction of the effect. By synthesizing a large number of peptides, each varying from another by only one amino acid, it was possible to determine empirically which amino acid substitutions made binding stronger and which made it weaker. The synthesis of this library benefited from the fact that chemical methods for linking amino acids together to make peptides has been optimized such that little optimization of the reaction conditions was required. Instead, the challenge was how to conveniently make thousands of peptides in a format that facilitated their use in the subsequent binding studies. The solution was to synthesize the peptides on a rack of plastic pins and to test their binding ability while the peptides remained chemically attached to the pin. While solving technical challenges, perhaps more importantly this approach highlighted a philosophy in studying ligand-receptor binding problems: 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.

Because there are 20 naturally occurring amino acids, the synthesis of a linear peptide that is n amino acids long can be done in 20^n different ways. Thus, there are 64,000,000 possible hexapeptides ($n = 6$). It would not be convenient, perhaps not even possible, to synthesize that many peptides on individual pins. In 1985, a synthesis method was reported in which the plastic used for synthesis was encased in an inert mesh resembling a tea bag. Collections of such tea-bag reactors were subjected to the chemical addition of an amino acid at the same time. After the addition, the bags were washed thoroughly and the bags (not the bag contents) mixed. The bags were redistributed to new beakers, and another amino acid was chemically added. In this method, each bag contains only one peptide. By chemically cleaving the peptide from the polymeric support, the peptide itself can be obtained. In a variant of this procedure, a mixture of reagents can be used in any given beaker such that each bag contains a mixture of peptides bound to the polymer. While the production of such mixtures complicates the subsequent assay step, it does provide for a much greater number of assays to be accomplished. Still unresolved is the question of whether the time economy afforded by using mixtures compensates for the pitfalls inherent in their testing for binding activity.

To make very large numbers of individual peptides would require very small bags for purely practical reasons. In 1988, work involving the use of polymer resin beads was reported; approximately one-tenth the width of a pinhead, these beads served as a kind of bag. By utilizing the "split-pool" approach, it was possible to synthesize extremely large peptide libraries in which each bead possessed a single peptide. The amount of peptide on one bead is only around 200 picomoles; however, 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.

Organic libraries. Because much of the practical application in discovering tightly binding ligands derives from the pharmaceutical industry, the combinatorial synthesis of druglike compound libraries is of great interest. Two practical considerations make this a greater experimental challenge than the synthesis of peptide libraries.

The synthetic methods required to make druglike molecules (that is, low-molecular-weight, organic molecules) on a polymer support have not been optimized. While solid-supported peptide synthesis



Scheme for using optical-pattern encoding for the synthesis of a combinatorial library. ***(Courtesy of X. Xiao)***

originated in the early 1960s and has been extensively developed since, initial experiments with organic solid-phase synthesis in the early 1970s were not followed up widely. In addition, while there are only 20 naturally occurring amino acids and therefore a finite number of reactions required to use them efficiently, an almost infinite number of organic chemical reagents exists and a very large number of reaction types. However, by the early 1990s several groups had reported the synthesis of moderately sized organic libraries by the solid-phase synthesis method. Each approach utilized a strategy like that of the Geysen pin method and thus was amenable to the parallel synthesis of hundreds to thousands of compounds.

The synthesis of much larger-membered libraries using the polymer-resin-bead approach was inhibited by the second practical issue: while 200 picomoles of an organic compound is enough for the ligand–receptor study, it is not enough to identify the structure of the ligand. The reason is that analytical methods for structure determination are both easier and more advanced for use with biological macromolecules such as polypeptides and polynucleotides. Beginning in the mid-1990s, the solution of bead tagging, or encoding, solved this issue. The strategy is simple: if the result of a chemical synthesis step cannot be easily read at low concentration, something should be added to the bead that conveniently encodes the reaction history of that bead for

later analysis. The first reported methods of bead encoding involved the use of biological macromolecules themselves. After each step in the organic synthesis, either an amino acid or a nucleotide was added to a growing oligopeptide or oligonucleotide on the same bead so that the specific sequence could be read later, and from that sequence the reaction history could be elucidated. However, neither oligopeptides nor oligonucleotides are chemically inert enough to survive the conditions required for organic synthesis. Later, tags were introduced that were much more inert chemically.

More recent methods of encoding include the use of radio-frequency memory microchips and optical bar-coding strategies. The **illustration** shows a scheme that uses optical-pattern encoding for the synthesis of a combinatorial library. Each of the nine chips has a different pattern written on it with a laser (a 3×3 matrix pattern as shown could encode for 2^9, or 512, unique chips). With the use of pattern-recognition software, the chips are sorted into appropriate reaction vessels so that chemical groups A, B, and C can be covalently added to the core organic molecule on them. After pooling the chips, optical scanning again permits sorting into appropriate reaction vessels so that a new set of three chemical groups can be added. At the end of this combinatorial synthesis involving two steps with three reagents at each, a total of 3^2, or 9, different products is obtained. Larger libraries are obtained by employing additional reaction steps, additional reagents per step, and an appropriately larger matrix pattern to enable the unique encoding of each product in the library.

Major challenges in combinatorial chemistry focus on both the characterization of compounds and the screening of very large compound libraries. In addition, the use of combinatorial chemistry methods is under active study for the discovery of new catalysts, chemosensors, and other chemical substances in which prior binding of the cognate ligand (the molecule that fits into another, usually larger molecule) is necessary. Anthony W. Czarnik

Assaying Problems

In combinatorial chemistry attention is now being focused on the problem of how to identify the set of molecules that possess the desired combination of properties. In a drug-discovery effort, the library members that strongly bind to a particular biological receptor are of interest. In a search for new materials that behave as superconductors at relatively high temperatures, the special combination of elements yielding the best electrical properties is a goal. In each case, the library might consist of up to a million members, while the subset of target molecules might consist of several thousand contenders or just a single highly selective binder. This subset could then be studied in more detail by conventional means.

To take advantage of the parallel nature of the combinatorial chemistry strategy, rapid screening and assaying protocols are necessary. For example, gas chromatography/mass spectrometry of bead-encoded libraries is currently limited to several analyses per hour, a difficult proposition when the researcher is faced with analyzing several thousand samples. The problem is to find ways of keeping track of each library member so that parallel screening can be coupled with parallel assaying.

Spatial array. Several emerging strategies promise to address this problem. In the first case, a library is constructed in a spatial array such that the chemical composition of each location in the array is noted during the construction. The binding molecules, usually labeled with a fluorescent tag, are exposed to the entire assay. The locations that light up can then be immediately identified from their spatial location. This approach is being actively developed for libraries of proteins and nucleotides. A problem is that the chemistry required to attach various molecules to the solid surface, usually silicon, is quite tricky and difficult to generalize. The assaying strategy is intertwined with the available procedures for synthesizing the libraries themselves.

Polystyrene-bead support. A conceptually straightforward approach is to first synthesize the library by using polystyrene beads as the solid support. The product molecules are then stripped from the support and pooled together into a master solution. This complex mixture consisting of a potentially large selection of ligand molecules could then be exposed to an excess of a target receptor. The next step is to devise a method for identifying the ligand-receptor pairs that point to molecularly specific binding. One approach is to examine a part of the mixture en masse by using affinity capillary electrophoresis. With this technique, the migration times of the ligand-receptor pair are significantly longer than the unreactive ligands, and can be interrogated by electrospray mass spectrometry.

The mass spectrometric method often provides a direct structural identification of the ligand, either by determination of its molecular weight or by collision-induced dissociation experiments. In the latter case, the molecular ion is selected by a primary mass spectrometer and is driven into a region of high-pressure inert gas for fragmentation. The fragment ions are then used to reconstruct the original molecular structure. This direct approach to screening and assaying has the advantage that the screening is carried out in solution rather than on a solid support, and it avoids steric problems associated with resin-bound molecules. At present the approach seems limited to libraries of about 1000 compounds because of interference from unbound ligands and by sensitivity issues. New strategies using mass spectrometry may eliminate this limit.

Mass spectrometry. A different tack involves assaying the polystyrene beads one by one after the resin-bound molecules are exposed to a receptor. With this approach, active beads may be identified by color or by fluorescence associated with the receptor, and are subsequently indexed in standard

96-well titre plates. Identification is then possible by using a variety of spectroscopic techniques; at present, the most popular methods are electrospray mass spectrometry and matrix-assisted laser desorption ionization mass spectrometry.

Successful structure determination using mass spectrometry is often routine, particularly for protein libraries where sequencing strategies have been worked out; however, ambiguities often arise for nonpeptide libraries. Encoding strategies promise to resolve this issue. For example, it is now possible to construct the library such that the first molecule attached to the resin can be identified by its ratio of carbon-13 (^{13}C) to carbon-12 (^{12}C). By using isotope enrichment protocols, a particular isotope ratio can be associated with a specific molecule. Then, for a library synthesized with three steps, the molecule in the first step can be identified from its isotope ratio, the molecule in the third step is known from the synthesis, and the molecular weight is measured by mass spectrometry. Hence, the molecular weight of the molecule used in the second step can be determined by the difference. This approach has been successfully employed to assay a 1000-member library. There are many ways that this approach could be extended to larger systems.

Another advantage of the isotope-encoding strategy is that the code is actually carried with the molecule after it is cleaved from the bead. This property may have important implications when performing bioassays. This scheme appears to be functional in a practical sense. The mass-spectrometry methods are generally sensitive enough (≈100 femtomole) to provide a reliable assay, and robotic techniques can be developed to increase the number of assays to perhaps a few dozen per hour.

Another level of sophistication involves the assay of a single polystyrene bead by using a method such as infrared spectroscopy. The advantages of optical techniques are that they are nondestructive and can be employed without removing the target molecules from the linker. There are, however, several obvious drawbacks, including lack of specificity and sensitivity for single bead studies. Recently, however, experiments using infrared microspectroscopy have partially overcome the sensitivity limitations. The selectivity issue has been addressed by using a carbon-deuterium stretching frequency to determine the deuterium content of a given compound. This signature can lead to identification under suitable circumstances.

Improving assaying power. It would be desirable to speed up the whole process. To fully capitalize on the elegant concept implicit in combinatorial chemistry, parallel screening methods capable of identifying large numbers of library members need to be developed. When the libraries are synthesized on beads, it would also be preferable to assay the compound without cleaving it from the polymeric resin support. This process inherently destroys the library and opens the possibility of incomplete cleaving reactions. One approach is to take advantage of a special type of mass spectrometry where it is possible to record spectra from very small areas of a solid surface. With this technique, the sample is bombarded by a focused energetic ion beam having a kinetic energy of several thousand electronvolts. The energetic particle loses some of its momentum in the top layers of the solid, causing desorption of molecular ions near the point of impact. If the energetic ion beam is formed in a short (nanosecond) pulse, the secondary ions may be measured by using time-of-flight detection. The resulting mass spectra can be recorded from an area that is much less than 1 square micrometer. Hence, it is feasible to spatially resolve the chemical components on a single resin particle that is typically in the 20–300-μm size range.

Experiments using this idea have been attempted in the last few years. The first attempts were successful only if the covalent bonds attaching the molecules to the resin were first clipped. This was accomplished by exposing the bead to a vapor of trifluoracetic acid, a standard release agent for acid-sensitive linking moieties. After clipping, the molecules were found to remain on their respective beads, even if prior to their treatment they were essentially touching. For example, in this method, two distinct types of coated beads can be used—one with phenylalanine and one with leucine. It might be possible to extend this method to a collection of thousands of beads arrayed onto a plate.

None of the above approaches provides assaying power that satisfies the need to characterize massive combinatorial libraries. Many schemes appear to have enough sensitivity or selectivity to perform the job, but whether any will be truly practical remains to be seen. It is likely, given the high level of activity in this field, that one of the above methods (or perhaps a completely new one) will become practical.

For background information *see* ANALYTICAL CHEMISTRY; ELECTROPHORESIS; LIGAND; MASS SPECTROMETRY; MOLECULAR RECOGNITION; OPTICAL INFORMATION SYSTEMS; ORGANIC SYNTHESIS in the McGraw-Hill Encyclopedia of Science & Technology. Nicholas Winograd

Bibliography. C. L. Brummel et al., A mass spectrometric solution to the address problem of combinatorial libraries, *Science*, 264:399–402, 1994; Y. Chu et al., Affinity capillary electrophoresis: Mass spectrometry for screening combinatorial libraries, *J. Amer. Chem. Soc.*, 118:7827–7835, 1996; S. H. DeWitt and A. W. Czarnik (eds.), *A Practical Guide to Combinatorial Chemistry*, 1997; P. A. Fodor et al., Light directed spatially addressed parallel chemical synthesis, *Science*, 251:767–773, 1991; H. M. Geysen et al., Isotopes or mass in coating of combinatorial libraries, *Chem. Biol.*, 3:679–688, 1996; K. Russell et al., Analytical techniques for combinatorial chemistry: Quantitative infrared spectroscopic measurements of deuterium-labeled protecting groups, *J. Amer. Chem. Soc.*, 118:7941–7945, 1996; S. H. Wilson and A. W. Czarnik (eds.), *Combinatorial Chemistry: Synthesis and Application*, 1997.