Combinatorial Technology: An Overview

A.K. GALANDE*
Bombay College of Pharmacy, Kalina, Mumbai - 400 098

The concept behind combinatorial chemistry burgeoned in the early 1980s when Mario Geysen, invented the method of simultaneous synthesis of diversified peptides. And now the approach has developed so much that combinatorial chemistry is being heralded as a future of pharmaceutical research. For the preliminary cognizance of this exciting technology, I have described techniques involved in combinatorial technology, which will briefly encompass the methodology of combinatorial technology as a tool for drug discovery. Emphasis is on the aspects of library design, syntheses involved, screening, and enumeration procedures which are an integral part of overall methodology. Article also tries to present pros and cons of combinatorial technology and emphasis is on the limitations and disadvantages for adopting combinatorial technology as a developing approach for drug discovery.

A recent advance in drug discovery has been the development of combinatorial technologies. Combinatorial technology involves an approach to chemical synthesis that enables the creation of a large number of compounds by putting chemical building blocks together in a given reaction sequence in every possible combination and screening them to get a "hit" - a compound which has the potential of becoming a future drug molecule. Thus, in combinatorial technology there is a generation of diverse collection of molecules, a molecular 'library', and then use of some process of selection to find the molecules with the desired properties.

The development of combinatorial chemistry has its roots in peptide chemistry. Bruce R. Merrifield published a seminal paper on solid-phase peptide synthesis and for his pioneering work in the solid-phase peptide synthesis he won the Nobel Prize in 1984. From 1972-1976, papers dealing with the development of solid-phase synthesis were published by Leznoff and Rapoport. There seemed to be a lull, till in 1982-1985 when Geysen and his associates presented a paper, which represents the first description of the preparation of an encoded library. The 1984 Geysen article on pin technology for parallel synthesis led to Richard Houghten's famous "tea-bag" technique for parallel synthesis. But it was not until an elegant work by Jon Ellman and his student Barry Bunin on the solid-phase synthesis of diazepines, published in 1992 that the sleeping giant was awakened to become a powerful tool for drug discovery in the pharmaceutical industry.

The practical application of combinatorial chemistry has required the development of a number of enabling technologies, many of which have emerged from small start-up companies. These enabling technologies include automated synthesis of chemical libraries, spatially arrayed libraries, computational chemistry software, solid-phase chemical synthesis, chemical encoding, software for managing and analyzing combinatorial library data, and microfluidic techniques for miniaturized chemical synthesis. This review provides an overview of some of the core technologies and their relationship to applications of combinatorial chemistry.

* For Correspondence
E-mail: amitkg@giabmc.vsnl.net.in
COMBINATORIAL TECHNOLOGY

Combinatorial technology broadly consists of four major steps, which can be given in the sequential order as follows:

1) Library Design
2) Synthesis
3) Screening
4) Enumeration

Detailed explanation of these components is given below:

Library design:

Consider a typical chemical reaction $A + B$ leading to the single compound $C$, and consider a typical combinatorial case of this equation for generation of library of 25000 compounds created by 50 A's and 500 B's. A molecular library of 25000 molecules created in such a manner is called a virtual library. It is very easy to create such virtual libraries on paper or on computer screen. However, it should be noted that not all the compounds in such libraries could actually be synthesized. This could be due to various reasons like incompatibility of certain reactants with other reagents or with the reaction conditions. Also some times it may not be possible for a chemist to test all the compounds synthesized for their activity. Under conditions like these, it is necessary to select optimum number of compounds so that the resulting library will have molecules, which can all be tested and can be practically synthesized. In addition, such a selected library should have maximum diversity, minimum compounds, and minimum cost of production. This requires excellent selection strategies.

The term 'library design' here implies acquiring the collection of compounds creating 'ideal' molecular diversity. Diversity is an important reason to do library design, but controlling other pharmaceutical relevant properties is an even more important reason. The basic principle behind the process of library design is to implement some degree of rationality in the process. Library design also helps in consolidating the feasibility of synthesis and increase the chances of getting desired results in biological testing of such libraries. Such combinatorial libraries could be either reactant-based or product-based. Fig. 1 shows the library design at the reactant level and at the product level.

**Fig. 1:** Subset selection can be performed at either the reactant level (1) or at the product level (2). Reactant pool has N1 number of R1 reagents and N2 number of R2 reagents. At the reactant level, n1 number of R1 reagents are selected, represented by r1. And same is the case for R2 reagents. Library synthesized in such a manner is represented by 'd'. At product level, 'D' represents library obtained by reacting all the reagents. Subset is then selected from the product library D, which is denoted by E.

In a reactant-based combinatorial library the selection is done at the level of reactants. Reactants are chosen in such a way that the resultant combinatorial libraries will have above mentioned attributes and will represent ideal molecular diversity ready for synthesis. In product-based combinatorial libraries there is no selection at the level of reactants. All the available building blocks are reacted. Selection is made from the combinatorial products so formed. Product-based combinatorial libraries assume the complete synthesis of all the available building blocks.

When the compounds associated with virtual library are astronomical, subset selection either at the reactant level or at the product level is indispensable and most of the times there is no bypass around this problem of the library design. The problem of 'how to select?' can be solved by using diversity tools, which will select a representative sample of the most diverse and optimum compounds within a theoretical or virtual library. Softwares are available that will select and focus extremely large corporate or commercial theoretical libraries. Diversity analysis introduces a rationale into the combinatorial chemistry process by using property data to design new libraries and reduce the final number of compounds that need to be synthesized. The ability to use shape and pharmacophoric activity descriptors facilitates the incorporation of SAR information into library design.
integration of SAR data allows the design of smaller focused libraries. There is an infinite number of descriptors that can be used to relate structure with activity, and most tools now provide comprehensive libraries of these. The search for new compounds, however, warrants the development of novel descriptors that capture the essential characteristics of the compounds. Tools to quickly develop such "customized" descriptors can provide a significant competitive advantage. With the development of various computational tools based on several algorithms, selecting combinatorial libraries to optimize diversity and related pharmaceutical properties, has now become a specialized task of a computational chemist.

Rational approaches if used for subset selection, entails use of computational softwares and this computer-assisted library design gives much better compound selection than randomly selecting reactants or products and building combinatorial libraries.

**Synthesis:**

Basically, two syntheses technologies allow the rapid creation of combinatorial libraries: 'Split and Pool' synthesis and 'Parallel' synthesis. Fig.2 is a diagrammatic representation of split and pool synthesis. In 'Split and Pool' method, microscopic resin beads are divided into N reaction vessels, one for each monomer in the monomer library. (N would be 20 for peptide synthesis using 20 different naturally occurring amino acids). The first monomer is coupled to the resin, and then the beads in each reaction vessel are repooled. The cycle of divide, couple and repool is repeated several times. (n-times to create an n-mer library), using a suitable deprotecting group each time a new monomer is added. Separation into the N reaction vessels guarantees an equimolar representation of each of the N-mers at each position in the combinatorial library. This solid-phase synthesis method is especially suitable for peptide and oligonucleotide libraries.

In case of small molecule 'Parallel' synthesis, compounds are synthesized in separate vessels in liquid-phase without remixing, typically by using a robotic arm to add reagents to different wells of a 96-well microtiter plate. Hits from the library are then identified by their position on the well plate. Multiple parallel synthesis restricts the production of compounds to smaller numbers than 'Split and Pool' synthesis. However, greater quantities of compound are created and there is no need to deconvolute or tag the individual compounds. Also, since

**Screening:**

Synthesized combinatorial libraries are assayed for their activity by the process of screening. In combinatorial synthesis, a series of steps in chemical synthetic pathways are interwoven resulting in a greatly expanded set of products. The 96-well plate grid positions provide a working identity for each resulting chemical sample, allowing the researcher to return to an identification key upon selection of a sample of interest. Thus for this
application, one of the primary roles of the multi-well plate is the assignment of identity to samples. Information management systems can track the identity, source, history, ownership, and activities of the contents of each individual well. Over the past decades, this has proven to be a godsend in the pharmaceutical industry, allowing drug discovery groups to examine escalating numbers of samples for the select few that will progress to the next round of screening, and eventually to market. This standardized format has allowed automation to be installed for at least some component of all major screening programs and rapid screening methodologies such as high throughput screening (HTS) and Ultra high throughput screening (UHTS) have emerged.

Over the last few years, there has been a growing demand for assay miniaturization in high throughput screening. The successful adaptation of the 96-well microplate serves as a practical example of the power of miniaturization. Assays that once required 2 to 3 ml volumes are now done in 1/10th of this volume. However, before the 96-well plate became a standard in HTS there were many questions about the suitability of this small format to screening assays and the availability of all the necessary peripheral equipment to support it. These problems were quickly solved and now improved versions of microtiter plates such as 256, 384 have developed which can now allow working with assay volumes as low as 50 μl. And in the future, assay miniaturization will move to 1536-well (2-5 μl) and 6144-well (0.2-0.9 μl) plates followed by a move to microchip arrays in the next 7-10 years.

HTS procedures are highly compatible with combinatorial synthesis mainly because it can cope with the large number of chemical entities synthesized and also because it considerably reduces the amount of molecular targets required for the assay procedure.

However, there are several critical problems regarding these advanced screening technologies which need to be answered. If screening is performed in a typical well-plate format, liquid handling speed should be made compatible with that of the screening. Increased speed provided by the advances in liquid handling are still hampered by the limitations in the signal read time. Because advances in signal analysis equipment such as spectrophotometers, scintillation counters, or fluorometers, have not kept pace with the advances in liquid handling. On the order hand, if screening is performed on pooled combinatorial mixtures, inherent synergistic, antagonistic or masking effects of multiple compounds within the pool might occur. Also, the installation of robotic workstations, the extent of automation of these robotic systems, and the data management, are the areas where improvement is needed to make HTS a great success.

Enumeration: 17, 21, 22

Enumeration refers to the identification of “hits” (i.e. active compounds obtained in the screening procedures) and a number of techniques can be used for this and most of them are based on monoclonal antibodies or soluble receptors. The most common one is based on classical ELISA test in which antibody or enzyme conjugated to a fluorescein conjugate is targeted against the binding monoclonal antibody, allowing its ready identification in light microscope. The stained beads are then selected and their attached molecules sequenced (if they are peptides) using an automated peptide sequencer.

In the case where the library is being made from the non-sequenceable compounds, the beads must be labeled or tagged. This is usually done with linear peptide or nucleotide sequence whose order maps to the order of synthesis of monomer in the library. Encoded combinatorial chemistry has opened a new dimension for the study of intermolecular interactions of small organic molecules with chemical and biological ligands. Although biological research has known the use of pools of proteins or genes for some years, the use of complex mixtures of small organic molecules has been hindered by the problems of purifying active compounds and of elucidating their structures. "Encoded combinatorial synthesis" solves this problem by creating "one-bead-one-compound" libraries that display the individual library members physically separated on beads of a solid support resin. Easily detectable molecular tags are attached to the resin beads as they proceed through the combinatorial synthesis, thereby recording the reaction history of each individual bead. The screening of the bead-supported library allows to identify individual beads which carry molecules interacting with the ligands in solution. The molecular tags can be then cleaved from each of the selected beads and can be analyzed to report the structures of the library members on these beads. There has been a considerable improvement in this 'tagging technology'. And chemical compounds can now be tagged by
photoreactive, radio-emitting, optically active and certain types of electronic tags with memory chips.

Encoded combinatorial libraries became the method of choice for 'Split and Pool' synthesis. However, in non-support bound combinatorial libraries any sort of tagging became difficult because of absence of solid support and hence the methodologies of iterative deconvolution and positional scanning were developed for the determination of active sequences. Using either method, the library is synthesized in subsets (or pools) such that the structural similarities of the compounds within each subset leads to the identification of active compounds.

There are certain other alternative methods like Markush structure-handling techniques which are used to generate structure "fingerprints" for the compounds in the library, without enumerating the compounds themselves, and thus offer substantial savings in processing time. Markush or "generic" structures are essentially structures involving R-groups, where a part of the molecule is defined by a series of alternatives. They occur frequently in chemical patent claims, where the inventor wants to protect an entire class of compounds.

**ADVANTAGES AND APPLICATIONS**

Combinatorial chemistry has evolved into a technology that is used for a widening range of applications. One of the biggest advantages of combinatorial chemistry over classical synthetic chemistry is that it can lead to compounds that otherwise might not be synthesized using the traditional method of medicinal chemistry. And one of the most important properties of combinatorial chemistry is that it provides a practical method for dealing with molecular design problems that are too complex or too underdeveloped for traditional 'deterministic' structure-based methods to work well. The term 'deterministic' here, refers to a classical approach of organic chemists who spend much of their time trying to synthesize molecules with very specific purposes in mind. Also, using traditional organic chemistry takes a chemist one month to produce four leads, at a cost per compound that is estimated to be about $7,500. In that same period, an organic chemist employing combinatorial approaches would have produced about 33,000 compounds ... at a cost of about $12 per compound!

The putative advantages of the combinatorial approach are 1) as a synthetic substitute for natural samples in random discovery screening or 2) as a targeted systematic exploration of the Structure-Activity Relationship (SAR) of a known lead compound. In the lead-optimization paradigm, a combinatorial library serves a similar role as the production of analogs in traditional medicinal chemistry, with the added advantages of parallelism in the time consuming synthetic and screening steps.

With the advent of benzodiazapine libraries, solid-phase organic synthesis has not looked back. Combinatorial approach is becoming favourite among scientists and yielding good results. Hydantoins, cyclic peptides, antibiotics, thrombin inhibitors, metalloprotease inhibitors are a few examples where combinatorial approach has been successfully used to give potentially useful lead molecules. Additionally, the chemistry is also applied to complex molecules like taxol for getting better analogues. Combinatorial chemistry is not only helping the medicinal chemist but the combinatorial approach is proving its potential even in several other areas of research involving catalysts, insulators, microwave devices and few superconducting materials. Combinatorial chemistry is proving to be a revolutionary concept with immense potential and with its diverse applicability it is advancing rapidly. Hence it is not surprising if one thinks of combinatorial chemistry as the solution for 'everything'. However, "It's not a panacea" as said by Michael Savage, chief executive officer of privately owned Molecular Simulations Inc. According to him combinatorial chemistry should work hand in hand with rational drug design and other established drug discovery paradigms for its optimum results. This skepticism will become more explicit by looking at the following disadvantages and limitations of combinatorial chemistry.

**DISADVANTAGES**

Problems associated with the library design:

a) If the subset selection in the library design is random, it attenuates the 'power' of diversity because one can never be sure of whether the selected subset is the one with the maximum diversity (Eric Martin, Chiron Corp. Emeryville CA, personal communication).

b) Rational subset selection, although used extensively, may attenuate the 'purpose' of combinatorial chemistry. Where combinatorial chemistry is appreciated as a tool for rapid 'lead' discovery, rational selection requires 2-3 days just to accomplish the subset selection.
Combinatorial Synthesis - a tough job ahead:

a) Chemistries suitable for combinatorial synthesis are limited, especially when it comes to solid phase synthesis. Hence the feasibility of 'directed synthesis' (where the next building block should compulsorily get incorporated into the partial lead molecule in order to accomplish a 'hit') cannot be taken for granted.

b) The automated synthesis system must be installed integrated with laboratory workflow. This is difficult because many of available instruments have poor user interfaces and must be programmed at very low level.

c) To make 10,000 compounds via a three-step solution-phase method requires 10,000 beakers, vials or wells, each used three times. This entails at least 30,000 liquid handling steps. There is no avoiding these physical requirements.

d) The principal drawback of parallel synthesis is the limited number of compounds that can be synthesized per unit time and per reaction. Automating parallel synthesis can help to eliminate the tedious of the process, but does little to address these time and throughput limitations.

e) The Split and Pool technique, conversely, addresses these limitations elegantly, but imposes other limitations of its own. Compounds produced using this process are no longer discrete, but are rather present as mixtures that must be deconvoluted using various tagging and screening techniques.

f) The 'Split and Pool' technique reduces problems associated with differential reaction rates but it is labor intensive.

g) Successful solid phase synthesis requires highly efficient assembly of a desired sequence and minimization of deleterious side reactions that occur during assembly, side chain deprotection and cleavage.

h) One more disadvantage of solid-phase synthesis is that a hydroxyl, amine, carboxyl, or other polar group must be present on a molecule to be able to attach it to a solid support. This is a potentially undesirable constraint on the structure of compounds synthesized on solid phase, because products retain the polar group even after they are cleaved from the support.

i) Resins associated with solid-phase synthesis have their own problems like poor homogeneity, swelling or shrinking, low loading capacity; this may hamper reaction rate and site accessibility.

Limitations in the process of screening:

a) It should be appreciated that one now works with the volume as small as 10 μl using 384 well plate for screening. But the difficulties in the modifications from 96 well plate to 384 well plate have shown that it is not going to be very easy to cope up with the further needs of miniaturization.

b) The more compounds that are made, the larger is the possible resulting information, particularly if the compounds are well selected. However, the larger the number of compounds, the less deeply they can be biologically tested, posing problems for screening procedures.

c) Although the chemical explorations prior to synthesis of the library, examine the scope and conditions of the reaction, there is no guarantee that all of the compounds intended to be in the library are in fact present. Only the most active are generally identified.

d) The biggest challenge to the interpretation of data from parallel synthesis is the lack of purity of the compounds tested. Although some purity checks are made, frequently one accepts samples of which 10-20% of the mass, is not the intended product. One must then be sure that the activity resides in the predominant product and not an impurity. This is especially important if one of the starting materials has biological activity in the particular assay. This may hamper the reliability of screening.

Enumeration inevitable:

Enumeration, identification of the 'hit', is inevitable when combinatorial mixtures are synthesized. And when in this additional step, the ratio of compounds to the ratio of basic lead increases, time factor associated with enumeration also increases. This is more conspicuous in very large combinatorial libraries.
Data management:

Data management presents major problem for a synthetic organic chemist who is dealing with large combinatorial libraries, mainly because the automation equipment must be integrated into data management infrastructure. This is often the most difficult obstacle. Surprisingly it is easier to make libraries than to keep track of what you have made and where it ended up. The correct structure of each of thousand products must be stored with its physical location in format that makes it easy to make sense of screening research.

Difficulties in analytical technique:

The advent of larger numbers of combinatorial drug candidates-compared with the number that used to be produced by conventional synthesis-makes it extremely difficult for analytical research groups to keep up. Such groups cannot simply increase resources 10-fold to assist in the evaluation of new drug candidates using traditional protocols. Doing 100,000 NMR runs although possible, will take far too much time and money.

Diversity concept:

What one is looking for, in combinatorial jargon is "diversity". But about which 'diversity'? is a key question. A few researchers in the field have expressed a skeptical view about the 'diversity' concept in combinatorial chemistry. According to them, it's biological relevant chemical diversity where combinatorial chemistry should focus. Just 'chemical' diversity, irrelevant or even semi-relevant with the desired 'biology' may attenuate the power of combinatorial chemistry.

Strategic problems of the combinatorial approach:

a) Although a combinatorial library is a fountainhead of substituent, position and frequency variations, it is practically not possible to apply a concept of homology variation to such libraries. Because this will make open ended (i.e. an infinite) set of compounds and this clearly rules out enumeration as a technique for dealing with them. There can be no combinatorial approach for homology variation.

b) Although, extension of combinatorial chemistry to biological system (combinatorial biology) is said to be one of its applications, this may not work, caution some molecular biologists. According to them, the number of genes available that directly affect structures of pharmacological compounds is limited. Thus, the diversity is also limited. This exposes the disadvantage of combinatorial approach. Where there is limited diversity, combinatorial approach may not do better.

Looking at so many chinks in the technology, reality sets in and the original euphoria from overzealous claims and expectations from combinatorial chemistry, fizzes out a little bit. But the question is, are these limitations going to persist in the coming years? And the obvious answer is no. Many researchers are trying to overcome these problems. New analytical procedures for combinatorial drug profiling are underway which will eliminate the need for repetitive analyses of single compounds. To increase the participation from academic research scientists, many universities in the US are building their own combinatorial chemistry units. Advancements in software for data-management, improved screening procedures, advanced synthesizes technologies are trying to eliminate most of the limitations of the technology.

CONCLUSION

Bringing a drug to market costs around $500 million and takes approximately 12-15 years. Only one in every three drugs is profitable, pressuring pharmaceutical companies to develop blockbuster drugs to offset the costs of the other drugs in their portfolios. With such obstacles, how can companies develop innovative drug that are profitable in today's lower-cost, managed-care environment? Combinatorial chemistry could revolutionize drug research by enabling a dramatic reduction in development time (four to seven years) by speeding up the synthesis and identification of lead compounds. One of the assessments of financial implications of combinatorial chemistry has revealed that if combinatorial chemistry can reduce the time required for drug discovery from its current average of 4 years to 1 year, then the peak market size required for a new drug could be reduced from approximately $500 million to $400 million. Similarly, if the development time for a new chemical could be reduced from 5 years to 3 years, the maximum required market size could be reduced from $120 million to $90 million. It seems that combinatorial chemistry has the potential to achieve these numbers. But whether this 'unorthodox' approach becomes a norm in future drug discovery or whether combinatorial chemistry has passed its heyday, only time will tell.
ACKNOWLEDGEMENTS

I sincerely thank Krishna Iyer (Bombay College of Pharmacy) for his valuable suggestions in the preparation of manuscript.

REFERENCES

43. Barnard, J.M., Downs G.M., Poster presentation in the 'Knowledge-Based Library Design’ (KBLD), session at the First Electronic Molecular Graphics and Modelling Society Conference, October 7-18, 1996. (Also available on www.bci1.demon.co.uk)