REVIEW ARTICLE

Accepted 24 April 2001
Revised 18 April 2001
Received 22 June 2000
Indian J. Pharm. Sci., 2001, 63(4) 279-285

Combinatorial Chemistry- A Novel Tool for Drug Discovery

P. MURALIDHAR AND B. S. SASTRY*
Department of Pharmaceutical Sciences,
Andhra University, Visakhapatnam-530 003.

Combinatorial Chemistry is a relatively recent concept adapted very quickly into a time and cost effective technology, which imparted a totally new orientation to medicinal chemistry and drug discovery research. The conceptual and technical approaches that encompass combinatorial chemistry are reviewed in this article.

Traditionally, new medicinal chemical lead structures have originated from the isolation of natural products from microbiological fermentations, plant extracts, and animal sources; from screening of pharmaceutical company databases; and more recently through the application of both mechanism-based and structure-based approaches to rational drug design[1]. Several studies have shown that the average cost of creating a new molecular entity by the above methods in a major pharmaceutical company is around $7500/compound. The advent of high-throughput screening has made possible the robotized screening of in excess of hundreds of thousands of individual compounds per year, per drug target. The availability of this capacity, combined with a major worldwide shift in emphasis by the drug industry toward more cost-effective pharmaceutical products, has exacerbated the need for a continuous flow of huge numbers of novel molecules.

To produce huge numbers of novel molecules and to address inefficiencies inherent in the contemporary new lead discovery process, researchers have turned to the concept of using combinatorial chemistry. The Journal Science recognized combinatorial chemistry as one of the nine breakthroughs in scientific research for 1998. Recently, Galande[2] reviewed the advantages and limitations of combinatorial technologies.

* For correspondence

Combinatorial chemistry:

Combinatorial chemistry is a type of synthetic strategy, which leads to large chemical libraries. Chemical library is defined as a set of compounds or collections of different molecules prepared either synthetically or bio-synthetically. Conceptually it involves the systematic and repetitive, covalent connection of a set of different "building blocks" of varying structures to each other to yield a large array of diverse molecular entities[3]. The idea of combinatorial chemistry is, to make a large number of chemical variants all at one time, test them for bioactivity, and then to isolate and identify the most promising compounds for further development. Combinatorial chemistry has focused on technologies that have the potential to make large numbers of products, either through preparation of many single compounds in parallel, or many compounds simultaneously in mixtures. For example, if coupling monomer A with monomer B gives the product A-B, combinatorial synthesis can take a range of monomers A_1 to A_n, and react those with B_1 to B_n, and make many product combinations (fig.1).

An essential starting point for the generation of molecular diversity is an assortment of small, reactive molecules, which may be considered as chemical building blocks. Theoretically, the number of possible different individual compounds, N, prepared by an ideal combinatorial synthesis is determined by two factors[3,4], the number of blocks available for each step "b", and the

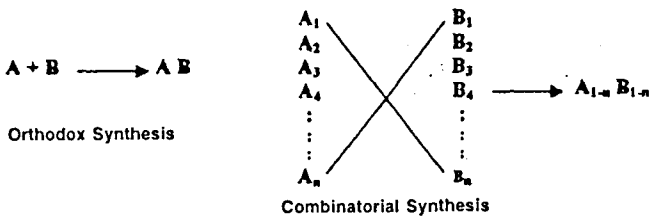


Fig. 1: Comparison of Orthodox and Combinatorial synthesis

Orthodox synthesis generally produces only one compound at a time, while combinatorial methods provide a range of products.

number of synthetic steps in the reaction scheme, "x". If an equal number of building blocks are used in each step, then $N = b^x$. If the number of building blocks for each step varies (e.g., b, c, d in a 3-step synthesis), then N = bcd.

In the short span of less than a decade, combinatorial chemistry has emerged as an important approach for generation and optimization of pharmaceutical leads to produce drug candidates[5]. It is estimated that while it needs about $30,000 to generate about 4 compounds by an orthodox method, directed against a particular target, application of combinatorial chemistry can give rise to about 3,300 compounds at a cost of only about $40,000.

Historically, these methods made their first appearance in the development of peptide libraries. The ready availability of a large and structurally diverse range of amino acid building blocks, a highly refined generic coupling chemistry, and the fact that small peptides are biologically and pharmaceutically key molecules, focused early efforts on peptide chemistry as a useful vehicle for exploring the power and conceptual issues attached to combinatorial ligand discovery. Generalization of the combinatorial strategy has led to the construction of collections of other natural polymers (e.g., oligonucleotides) and synthetic polymeric libraries. Currently there is an acute interest in the generation of small molecule non-peptide libraries.

Types of methods in combinatorial synthesis; Solution-phase synthesis:

Combinatorial compounds are created either in solution-phase or in solid-phase. At its simplest level, the solution-phase synthesis involves conducting solution-phase chemical reactions simultaneously preferably in well-ordered sets (arrays) of reaction vessels[6]. We can illustrate this by the preparation of a small array of amides. The process consists of placing a different acid chloride A_n and amine B_n in each of a matrix reaction vessel (along with a tertiary amine to neutralize liberated hydrochloric acid); incubating to form the amide; performing a liquid-liquid extraction to remove excess reagents and evaporating the solvent and testing the crude amides directly in a biological assay.

Solid-phase synthesis:

The majority of the compound libraries that have been prepared to date have been synthesized on a solid support (a solid support is an insoluble material to which compounds are covalently attached during a synthesis sequence). In this method, the first step involves the attachment of the starting material of interest to an insoluble support by a linker that can be cleaved under specific conditions. The reaction is then driven to completion by adding a large excess of reagent, which is subsequently removed by filtering the insoluble support. After extensive washing, the next reaction is carried out similarly. Several reactions later, the linker is cleaved with the appropriate reagent, to give the target compound.

There are two advantages to this strategy. First, isolation of support-bound reaction product is accomplished simply by washing away reagents from the support bound material, and therefore reactions can be driven to completion by the use of excess reagents. Second, innovative methods are available for the manipulation of discrete compounds and for "tracking" the identity of compounds when compounds are attached to a solid support. Hermkens and co-workers[7] recently reviewed the chemical reactions that are adaptable to the solid-phase synthesis.

Solid Supports and Linkers:

Polystyrene cross-linked with 2% divinyl benzene was demonstrated by Merrifield as useful support for solid-phase peptide synthesis in 1963, and polystyrene cross-linked with 1-2% divinyl benzene continues to be one of the most commonly used supports for solid-phase organic synthesis. Even though these beads are cheap and have high loading levels, these resin beads are not well solvated in protic solvents resulting in poor reaction site accessibility and diminished reaction rates. Another type of solid support known as Tentagel resin is prepared by means of anionic polymerization of ethylene oxide to

attach polyethylene glycol chains of controlled lengths on cross-linked polystyrene beads containing hydroxyl groups. This resin is well solvated in protic solvents and because polyethylene glycol chains are not cross-linked, the reaction sites are highly accessible resulting in greater reaction rates[8].

Other materials used for solid-phase synthesis include cellulose in the form of Perloza beads, Sheppard's polyamide resin, paper, cotton and glass[4]. The group that joins the substrate to the resin bead, known as linker, is an essential part of solid-phase synthesis[9]. Many linkers have been developed for this purpose. Some of the commonly used linkers were given in fig. 2.

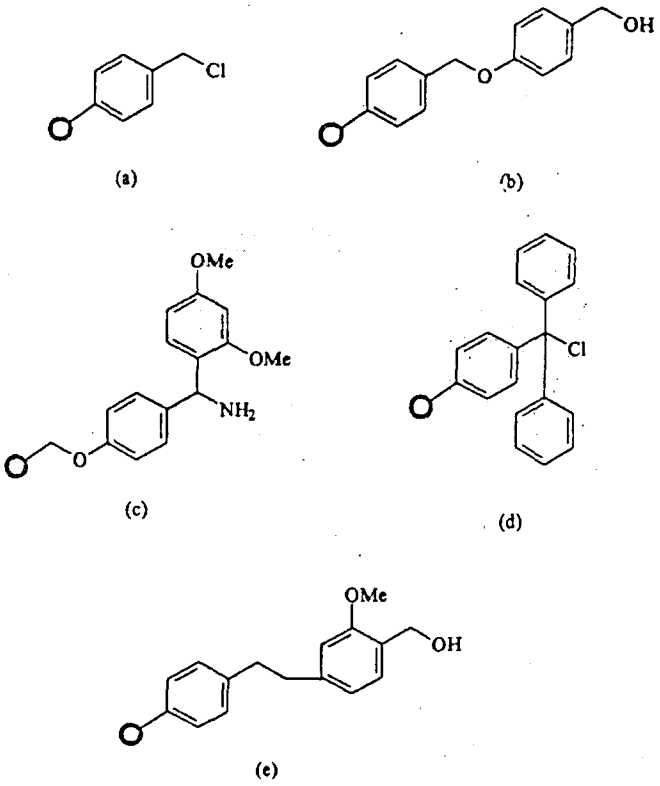


Fig. 2: Stuctures of commonly used linkers.

(a) Merrifield resin linker, (b) Wang resin linker, (c) Rink resin linker, (d) Trityl resin linker, (e) Sasrin resin linker

Sequence of steps:

All combinatorial approaches involve three essential steps.

(i) Creation of chemical library
(ii) Screening of the library components, and
(iii) Determination of the chemical structures of active compounds.

Creation of chemical library; Split and Mix Synthesis[10]:

Combinatorial chemists use different techniques for the creation of chemical libraries. Split and mix method is particularly employed for solid-phase synthesis. In this method, compounds are assembled on the surface of micro-particles or beads. In each step, beads from previous steps are divided into several groups and a new building block is added. The different groups of beads are then recombined and separated once again to form new groups. The next building block is added and the process continues until the desired library has been assembled. This method gives rise to all possible combinations of the building blocks. It can produce very large numbers of molecules and is useful in random screening, particularly for targets of unknown structure. A schematic representation of this method is shown in fig. 3.

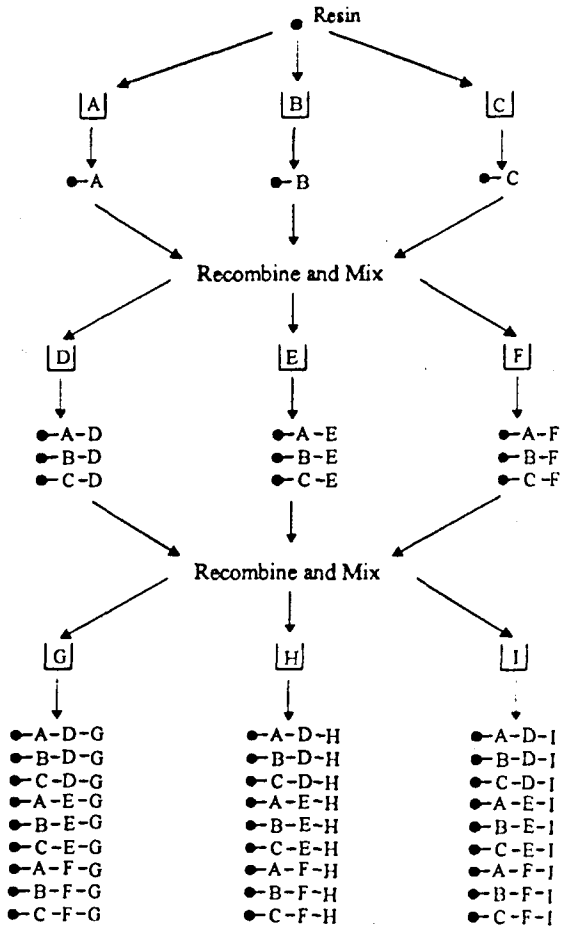


Fig. 3: Schematic representation of 'Split and Mix synthesis'.

A, B, C, D, E, F, G, H and I are different building blocks.

Spatially separate parallel synthesis[11]:

This method involves performing modular chemical reactions in parallel, using discrete reaction chambers laid in a spatially addressable format (such as a 96-well microtitre plate). Parallel synthesis can be done either on a solid support or in solution.

The parallel synthesis approach can be illustrated with hydantoin library synthesis. Any parallel synthesis benefits by starting with a set of reactions that work efficiently with a wide range of reagents. In this scheme these reactions are, (i) amine deprotection, (ii) reaction of an amine with an isocyanate to give a urea; and (iii) acid-catalyzed cyclisation of the urea to give a hydantoin. Thus, a 40-vessel synthesis of hydantoins can be performed as follows. Eight resins containing different protected amino acids are each placed in five reaction vessels. Deprotection (i) affords the corresponding charged-charged resins, now possessing free amine groups. Each resin-bound charged is then reacted with five different isocyanates (ii), to give a possible total of 40 different ureas. In the final step (iii), treating all of the reaction vessels with 6 M HCl results in cyclative cleavage and release of the resulting hydantoins from the resin. These are separately dissolved in methanol, concentrated and analyzed.

The parallel synthesis method generates fewer molecules, but the spatially addressable format provides structure-activity data immediately and simplifies follow-up production of large amounts of material. This method is useful for lead generation, particularly for target of known structure, and is highly useful for lead optimization and drug development.

Screening of the library components:

A reliable high-throughput assay is essential to successfully screen a combinatorial library. Both solid-phase and solution-phase assays have been developed for the combinatorial libraries. In the solid-phase assays, the ligands are still attached to the solid support and the assays are two types: (i) direct binding of molecular target to the bead-bound ligand[12]; this binding can be detected by direct visualization (e.g., a colour target such as a dye), or by using a reporter group such as an enzyme, a radio nuclide or a fluorescent probe. And (ii) detection of functional properties of the bead-bound ligand such as identifying phosphorylation or proteolytic substrates[13,14]

Solution-phase assays, usually in the 96-well plate format, have been used in mass screening for most drug discovery programmes. Synthetic compounds or natural products are usually added in a soluble form into each individual well for biological testing. There are many solution-phase assays available, e.g. competitive receptor binding, assays with radio labeled ligands, various enzymatic assays, cell-based signal transduction assays, antibacterial, antiviral and anticancer assays. All these assays can be adapted to combinatorial chemistry. There are two general approaches to screen compound library with solution-phase assays: (i) the 96-well two-stage releasable assays[15] and (ii) the *in situ*-releasable solution-phase assay with immobilized beads[16-18]. In both approaches, ligands are attached to the solid support via cleavable linkers. The ligands are then released from each bead into solution-phase when the biological assays take place. The bead-of-origin of the positive releasate can subsequently identified, and isolated for structure determination.

Determination of Chemical Structure of Active Ligand:

An essential element of the combinatorial discovery process is that one must be able to extract the information made available by library screening. Put another way, creating large quantities of molecular diversity for ligand discovery is insufficient, unless there is a format at hand to capture the information, which in this case is the chemical structure of active compounds[3]. Primarily three methods are used for structural deconvolution of active compounds: iterative deconvolution[18], positional scanning deconvolution[19], and encoding[20].

Iterative Deconvolution Method:

In this approach (fig. 4), pools of compounds are prepared such that each separate pool has defined building blocks at either one or two positions, and at the remaining positions all combinations of building blocks are incorporated. The optimal building block(s) at the defined position(s) is selected by determining which pool(s) has the greatest biological activity. A second round of synthesis is then performed with the selected building block(s) in place at the initial defined position(s) in order to prepare pools where the next defined position is introduced. Each pool is evaluated for biological activities in order to select the optimal building block at the additional define position. This deconvolution process of iterative resynthesis and evaluation is then repeated until all of the positions are defined.

For example, in the deconvolution sequence illustrated in fig. 4, three pools are produced that each contain a defined building block at the last position (G, H, or I) and all possible combinations of building blocks at the first and second positions. For deconvolution, the pools are assayed and building block H is selected at the last position since the pool with H at the last position has the greatest activity. Three pools are now prepared in the second round of synthesis that incorporate building block mixtures at the first position, a defined building block at the second position (D,E or F), and the selected building block H at the last position. Each pool is evaluated for biological activity and building block D is selected for the second to last position. Three final pools are synthesized with defined building blocks at each position; building blocks A, B or C at the first position, building block D at the second position, and building block H at the last position. Evaluation of these pools results in the identification of compound C-D-H.

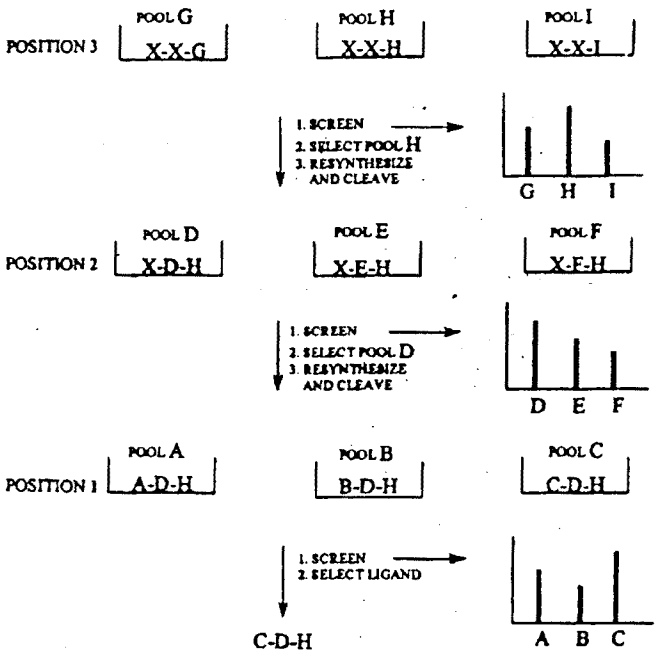


Fig. 4: Deconvolution through iterative resynthesis.

A to I are specific monomeric building blocks while X representing anyone of them. The most useful building block at each of the positions is decided by iterative resynthesis and screening as shown above.

Positional Scanning Deconvolution Method:

The positional scanning (PS) approach is illustrated schematically in fig. 5. It involves the screening of

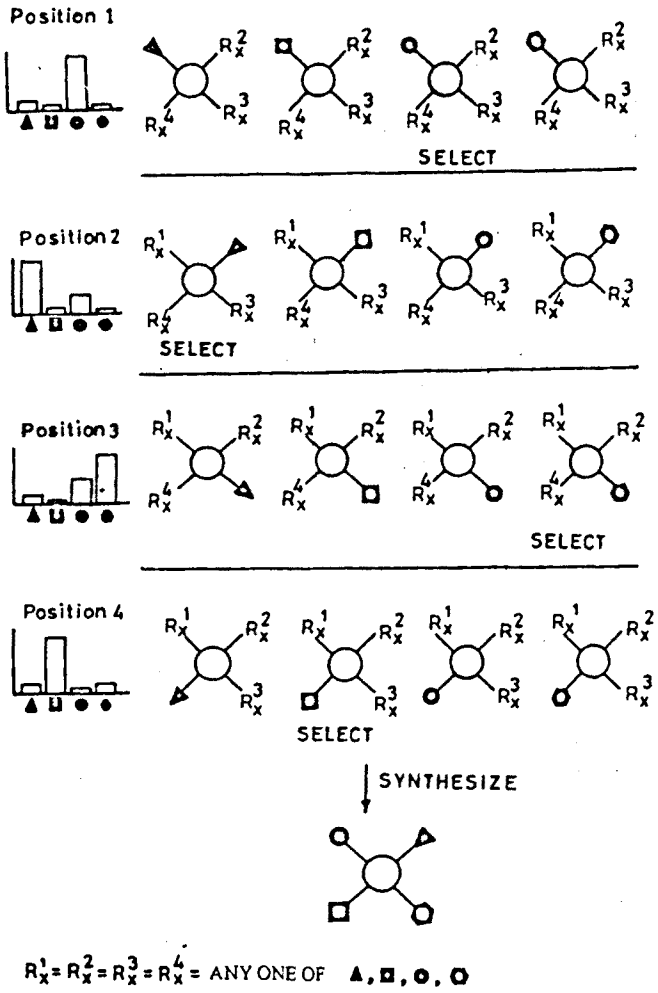


Fig. 5: Deconvolution through Positional scanning.

Deconvolution involves defining which of the variables is the best fit for a given position by constructing and screening pools of compounds. (Four pools, each pool containing 64 compounds for a four positions- four variables situation as depicted here.) The histogram on the side shows the activity profile of each defined pool.

separate, single defined position synthetic combinatorial libraries (SCLs) to individually identify the most important functionalities at each position of diversity within a library. A complete PS – SCL having four positions of diversity consists of four sublibraries (designated OXXX, XOXX, XXOX, XXXO), each of which has a single defined functionality at one position and a mixture of functionalities at each of the other three positions. The generic library shown contains 256 compounds (4 x 4 x 4 x 4 = 256). The pooling of each sublibrary, which contains the same 256 compounds (4 mixtures of 64 compounds), would

vary on the basis of functionality at the defined position of that sublibrary. The structure of individual compounds can be determined from such a screening since each compound is present in only one mixture of each sublibrary. In theory, if only one compound was active in the library, activity corresponding to that compound would be found in the one mixture of each sublibrary containing that compound. When considered in concert, the defined functionality in each mixture can then be used to identify the individual active compound responsible for the activity. In reality, the same result is seen, but the activity is generally due to the sum of more than one active compound. Anomalous results are seen if the activity is due to the sum of many weakly active compounds.

Encoding:

In the above methods the structural determination of active compounds in the library is achieved either by a systematic iterative resynthesis and rescreening of specific mixtures and compounds, or by microsequencing of a peptide sequence. However, many investigators may not wish to embark on the resynthesis of library components, and if the library constituents are not peptides, microsequencing can no longer be applied. Thus, a major area of investigation has been the development of other methods for the structure determination of active compounds within libraries. One such method is 'encoding' or 'tagging'.

This method employs a readable chemical tag that is simultaneously attached to the individual bead for each step in the synthesis of the actual molecule on the bead (fig. 6). The identity of active compounds are determined by analysis of tags cleaved either by UV irradiation of a photolabile linker followed by gas chromatography (GC) or by acid hydrolysis followed by GC. The tagging molecules used are halophenoxy derivatives of aliphatic alcohols, diazomethane derivatives and secondary amines.

Illustrations of combinatorial synthesis:

Even though combinatorial chemistry started with the synthesis of peptide libraries, currently the major effort in the field is focused on combinatorial libraries of non-peptide like small molecules. The rationale for this is two-fold. First, the character of peptide backbone limits the structural diversity of peptides. Secondly, other classes of compounds have been shown to be structurally and chemically diverse and have characteristics not present in peptides that are important for drug candidates such as oral bioavailability and resistance to protease degradation.

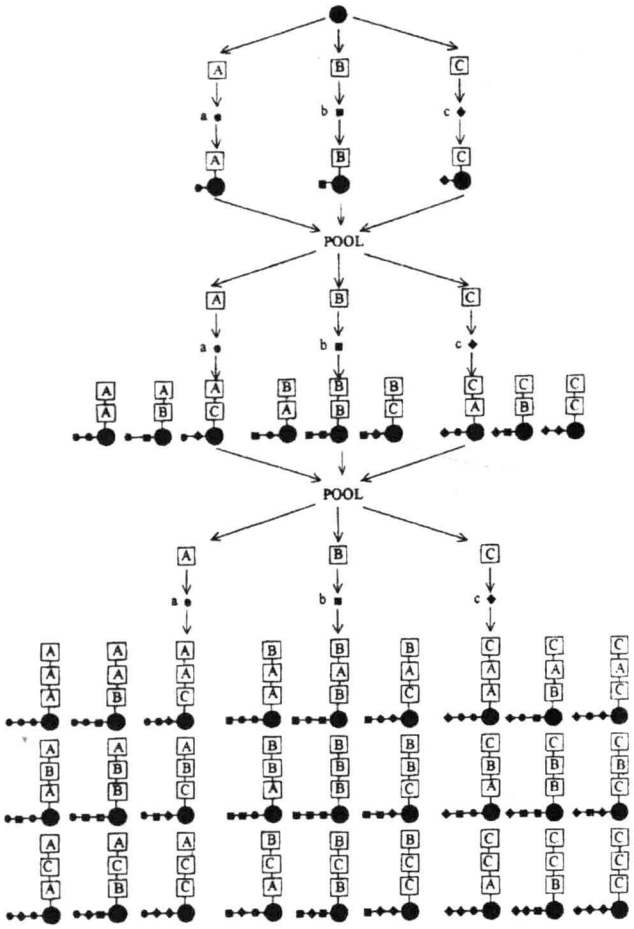


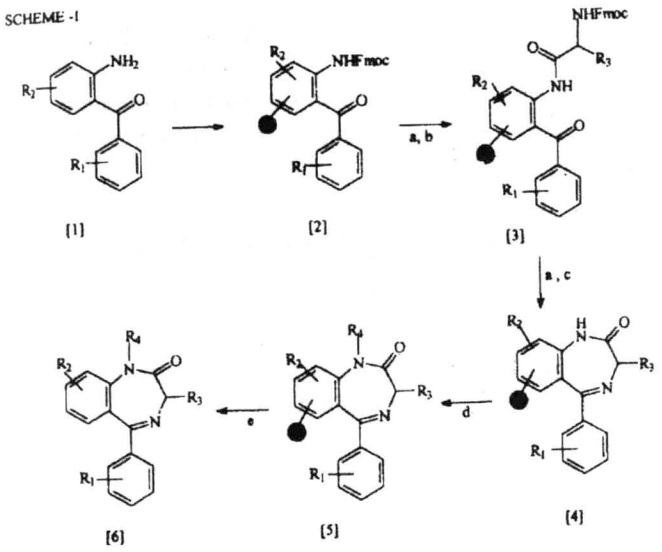
Fig. 6: Deconvolution through Encoding.

A,B and C are building blocks. Shapes represented by a, b and c are chemical tags.The structure of active compound can be identified through its specific tag.

The current chemical literature is almost flooded with description of polymer-supported chemistries that provide a variety of heterocyclic compounds, amenable for the combinatorial synthesis of libraries. Nefzi *et al.*[21] have reviewed heterocyclic combinatorial libraries. One example for heterocyclic combinatorial synthesis is given below.

Benzodiazepine library synthesis:

The first example of combinatorial solid-phase synthesis of heterocyclic compounds, 1,4-benzodiazepines, was described by Bunin and Ellman[22] (Scheme I). In the synthesis, a range of independently synthesized 2-aminobenzophenones is linked to Tentagel



Scheme I: 1,4 – Benzodiazepine library synthesis

a- 20% piperidine in DMF ; b : N- Fmoc- aminoacid flouride, 4- methyl- 2,6- di-tert-butyl pyridine; c: 5% aceticacid in DMF, 60 C; d: lithiated 5-(phenyl methyl)-2- oxazolidinone in THF, -78 C, followed by alkylating agent and DMF; e: TFA/ H_2O/ Me_2S (95:5:10).

resin through a phenol or acid residue and acylated with a set of Fmoc-protected a-amino acids. Deprotection and acid-catalyzed cyclisation gives representative benzodiazepines [6] on resin that are further functionalised by amide N-alkylation. Acid-catalyzed cleavage liberates the final benzodiazepine in a relatively pure state for screening.

CONCLUSIONS

Combinatorial chemistry is now considered one of the most important recent developments in medicinal chemistry. The molecular biology revolution of the last two decades enables researchers to routinely clone and express biological receptors, enzymes, and proteins of pharmacological interest. Many new drug targets for various diseases have now been identified. Relying on traditional drug discovery methodology to screen this many targets would almost be an impossible task. The development of combinatorial chemistry is timely and undoubtedly will contribute to the discovery of new drugs that can benefit mankind. Combinatorial chemistry is especially useful when applied in conjunction with modern computational chemistry and molecular modeling techniques since many leads can often be generated from a single combinatorial library screen. In addition to the discovery of initial leads, combinatorial chemistry can also be applied to the optimization of drug leads.

The most likely developments in the near future in this field include new assay formats for rapid screening of compound libraries, new coupling chemistries and new building blocks and cleavable linkers that are compatible with a variety of organic syntheses and assay systems. In the coming years, combinatorial chemistry can be expected to rapidly lead to more active, more specific, safer, and less expensive therapeutics to human kind.

REFERENCES

1. Patchett, A.A., **J. Med. Chem.,** 1993, 36, 2051.
2. Galande, A.K., **Indian J. Pharm. Sci.,** 2000, 62, 84.
3. Gallop, M.A., Barrett, R.W., Dower, W.J., Fodor, S.D.A, and Gordon, E.H., **J. Med. Chem.**, 1994, 37, 1233.
4. Terrett, N.K., Gardener, M., Gordon, D.W. and Steele, J., **Tetrahedron**, 1995, 51, 8135.
5. Houghten, R.A., Pinilla, C., Appel, J.R., Blondelle, S.E., Dooley, C.T., Eicher, J., Nefzi, A. and Ostresh, J.M., **J. Med. Chem**, 1999, 42, 3743.
6. Floyd, C.D., Lewis, C.N. and Whittaker, M., **Chem. Brit.**, 1996, 31.
7. Hermkens, P.H.H., Ottenheijim, H.C.J. and Ress, D., **Tetrahedron**, 1996, 52, 4527.
8. Thomson, L.A., Ellman, J.A., **Chem. Rev.**, 1996, 96, 555.
9. Lam, K.S., Lebl, M. and Krchnak, V., **Chem. Rev**., 1997, 97, 411.
10. Furka, A., Sebestyen, F., Asgedom, M. and Dibo, G., **Int. J. Pept. Protein Res**., 1991, 37, 487.
11. Czarnik, A.W. and DeWitt, S.H., **Chem. Brit.**, 1996, 43.
12. Lam, K.S. and Lebl, M., **Methods (San Diego)**, 1994, 6, 372.
13. Wu, J., Ma, Q.N. and Lam, K.S., **Biochemistry-USA**, 1994, 33, 14825.
14. Meldal M., Svendsen, I., Breddam, K. and Auzanneaue, F.I., **Proc. Natl. Acad. Sci. USA**, 1994, 91, 3314.
15. Salmon, S.E., Lam, K.S., Lebl, M., Kondala, A., Khattri, P.S., Wade, S., Patck, M. and Felder, S., **Proc. Natl. Acad. Sci. USA**, 1993, 90, 11708.
16. Jayawickreme, C.K., Graminski, G.F., Quillan, J.M. and Lerner, M.R., **Proc. Natl. Acad. Sci. USA**, 1994, 91, 1615.
17. Salmon, S.E., Liu – Stevens, R.H., Zhao, Y., Lebl, M., Krchnak, V., Wertman, K.I., Sepetov, N. and Lam, K.S., **Mol. Divers.**, 1996, 2, 57.
18. Houghten R.A., Pinilla, C., Blondelle, S.E., Appeal, J.R., Dooley, C.T. and Cvervo, J.H., **Nature**, 1991, 354, 84.
19. Pinilla, C., Appel, J.R., Blanc, P. and Houghten, R.A., **Biotechniques**, 1992, 13, 901.
20. Janda, K.D., **Proc. Natl. Acad. Sci., USA**, 1994, 91, 10772.
21. Nefzi, A., Ostresh, J.M. and Houghten, R.A., **Chem. Rev**., 1997, 97, 449.
22. Bunin. B.A. and Ellman, J.A., **J. Amer. Chem. Soc.**, 1992, 114, 10997.