REVIEW ARTICLE

Accepted 19 May 1999
Received 27 April 1998
Indian J .Pharm. Sci., 1999,61(5) 255-258

Combinatorial Chemistry

K. G. BOTHARA*, AJIT D. GAIKWAD AND ATMAJA SAXENA
A.I.S.S.M.S. College of Pharmacy
Kennedy Road, Pune - 411 001

A promising new approach to drug discovery concerns with the synthesis and screening of combinatorial libraries in order to identify new compounds that express high affinity and specificity for a pharmacologically relevant, biomolecular target. Advances in molecular biology, automated chemical synthesis and robotics have facilitated the formulation of vast libraries of structurally related molecules. An essential aspect of screening large combinatorial libraries is the ability to identify the active components in these complex mixtures, which is usually based on the strength of binding to a selected target macromolecule.

Combinatorial Chemistry is a new subfield of Chemistry with the goal of synthesizing very large number of chemical entities by condensing a small number of reagents together in all combinations defined by a given reaction sequence[1]. It is also referred 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*1*1=1 product as the result of 1+1+1=3 total reactions. While 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'*. Combinatorial Chemistry or 'molecular diversity' was first popularized by Merrifield in the 1960s who received a Noble prize for his work on solid-phase peptide synthesis (SPPS). The 1970s visualized the emergence of solid phase non-peptide synthesis. It was in the 1980s, primarily peptides or oligonucleotides and recombinant protein or nucleic acid based technologies developed. The need for Combinatorial Chemistry is underscored by the high cost, long time frame and high rate of failure in research and development for new drug. Today, the preferred methods for building libraries are parallel synthesis, split synthesis and a combination of the two. Parallel synthesis is in essence an automated form of the traditional 'one at a time' approach. It can use either solution or solid phase Chemistry. For split synthesis a range of synthons or reactants are used at each step, in principle, giving a library in which every possible compound from every possible combination of serial step using these synthons, is created.

When designing a combinatorial mixture library for lead identification, it is desirable that the compounds within the library be as diverse as possible, to fully explore the scope of activity against the target. However, the design of a library should take into account many other factors, not least among them is the effort needed to deconvolute the mixtures, once hits are obtained. Factors such as cost and availability of reagents or the range of physical properties of library products may also require optimization[2]. Furthermore, in the design of a mixtures, the combinatorial constraint always applies, that is, every substituent at each position will occur in combination with every substituent at all other positions. Each of these additional considerations may mean that a certain amount of possible diversity in a library has to be sacrificed.

*For correspondence

LIBRARY SIZE, DESIGN AND SELECTION

Compound libraries need to serve two distinct functions in the drug discovery process, lead identification and lead optimization, which dictate their size and composition. Thus, when a lead has been identified from screening, rapid analog synthesis is performed to provide optimization of both potency and selectivity[3]. Ideal libraries to address this lead optimization phase will invariable be smaller in size than libraries used for lead identification and consists of individual components (≤ 1000 discretes), and can be prepared by parallel synthesis around the lead.

The realization that analog library synthesis impacts the time for the lead optimization phase of drug discovery, is now being extensively exploited by medicinal chemists. Industries such as Pfizer and Lilly have described their capability to optimize lead structure in 6-12 month time frame[3]. Pfizer has reported an analog synthesis of an antiatherosclerotic drug with 100 fold increase in potency, while Lilly has described the optimization of a 5 HT agonist, for development of a potential new antimigraine drug.

Another function of compound libraries is to provide a range of leads, preferable of different structural types, to initiate the drug discovery process. One simple approach is to focus on structural motifs of molecules of proven efficacy. A good example is the benzodizepine libraries with diverse biological activities were discovered and this encouraged many groups to devise libraries around other structural motifs which potentially offer other activity profiles[4].

Synthetic libraries that have been targeted more specifically at a particular family of receptor or enzyme, by incorporating a key recognition element for binding, are also available. One such an example is the hydroxamate libraries of metalloproteinase inhibitors.

For novel targets, there is the requirement to device libraries of fundamentally distinct structures which offer the best chance of finding novel screening hits. The synthetic hurdles to such libraries are not inconsiderable and it is not surprising, therefore, that much more emphasis has been given over the past two years to applying software design techniques to enhance the content of such libraries, while constraining the size. One of the elements of the design that reflects library composition is molecular diversity[4].

In a new approach, the concept of libraries for information was proposed, where in the informatory molecules, used to provide finger print of the structural requirement of the target, have been designed on promiscuity rather than diversity[5]. The overall description of a promiscuous molecule is its proven ability to interact with a wide range of targets. Subsequent libraries are designed and synthesized to allow the separation and identification of the specific binding characteristics that are relevant to a particular target under study and this is achieved by an iterative but convergent approach using experimental design and testing.

Library Diversity: One approach to produce diverse library products has been to ensure that the set of precursors used to construct the library are as diverse as possible. If all the suitable precursors for a given diverse site are grouped on the basis of some desirable features, such as distribution of potential pharmacophoric points, then selecting no more than one from each group should ensure that the set is diverse[6]. To produce a diverse library, it is therefore desirable to assume that every precursor is taken from different group.

A recent study has suggested that maximizing diversity among the precursor set may not necessarily give the most diverse possible set of library products[7]. An alternative is to consider the diversity among the library products themselves. This may be measured by enumerating the library products and either clustering them and attempting to pick as few compounds as possible from each cluster or using a cell-based partitioning method attempting to pick as few compounds as possible from each cell.

Deconvolution: A number of solutions have been suggested to the problem of deconvolution. These include tagging beads with various types of chemically and spectroscopically readable labels or producing libraries on silicon chips whose identities can later be determined by radiofrequency scanning[7]. Once this information is known, selective synthesis and testing of all the compounds in the library with that molecular weight will be required. Since it is desirable to keep this work to a minimum, it is sensible to design a combinatorial library to have the smallest possible number of compounds of any one molecular weight, given the constraints of the required size of the library and the availability of the precursor[7]. For a given number of library products all having the same molecular weight, deconvolution will be simpli-

fied further if substituents used at the diversity sites have different molecular weights. It is therefore desirable to minimize the substituent molecular weight redundancies for each given product molecular weight. In a 'mix and split' strategy for combinatorial synthesis, the pools of compounds that result from the addition of the final diversity sites are often not mixed. Since these pools are screened individually the deconvolution problem applies to each subpool separately[7]. A library design strategy must therefore be able to suggest pools within each of which there is the minimum products and substituent molecular weight redundancy.

Problem Size : The number of precursors that are suitable for use in combinatorial mixture library is often larger than the number that can be reasonably be used. Hence typical mixture library design problems have huge search spaces, the size of which are best illustrated by an example[8].

In a library, two sites of diversity were available, there were 360 commercially available precursors compatible with the chemistry for R_1 and 259 for R_2. There are therefore, 92, 240 possible library products compounds. The design of this library requires production of 10,000 compounds by combining 100 R_1s with 100 R_2s. The number of ways of selecting k objects from n is:

$$^{n}C_{k} = n\ !\ /(n-k)!k!$$

And so the number of libraries is

$$^{360}C_{100} \cdot {}^{259}C_{100} = 2.5 \times 10^{164}$$

Advances in compound library production:

Library design is complex problem, requiring the optimization of a number of often competing factors, over a vast search space. Genetic algorithms[9] have been successfully applied to a wide range of such problems in both chemical and non-chemical domains. A genetic algorithm is a computational technique that mimics the process of Darwinian evolution. A potential solution to a problem is encoded in a representation termed as chromosome. This is typically a string of bits, integers, real numbers or symbols each of which is termed as gene. A genetic algorithm operates on a population of these chromosomes that are generated by assigning values to the genes in chromosomes, often at random. A fitness function measures how well adapted each chromosome is to its environment.

There are two pre-requisites to being able to apply a genetic algorithm to a problem. The first is to choose a representation that allows every possible solution to the problem to be encoded in a chromosome. The second is that it must be possible to write a fitness function to decode the chromosome and produce a score that reflects the quality of that solution.

A most promising new approach to drug discovery concerns the synthesis in one-pot reaction, without isolation or purification and the reaction mixture is screened using a competitive binding assay based on pulsed ultrafiltration, electrospray mass spectroscopy[10] (PUF/ESMS) which tentatively identify those derivatives having the highest affinity for the target receptors. As a model system to test this approach, a synthetic scheme designed to prepare a series of analogs of the adenosine deaminase inhibitor, erythro-9-(2-hydroxy-3-nonyl) adenine(EHNA), as diastereomeric mixtures, was carried out. Pulsed ultrafiltration screening of the crude reaction mixtures against controls without protein, detected protonated molecules corresponding to EHNA-type derivatives and three of its linear, alkyl homologues. It did not show protonated molecules for an isobutyl or benzylic EHNA derivative, suggesting the latter was inactive.

An important feature of combinatorial chemistry is the synthesis of compounds on solid support allowing "split and pool" methodology to be employed for library construction. The method involves the use of an appropriate linker to tether the initial starting substrate to solid support. The linker needs to be stable during the synthesis phase, but capable of facile cleavage to free up the final product. The ideal linker would be one capable of product release with formation of a carbon-hydrogen bond in place of resin attachment, thus leaving behind no memory of the site of attachment on the solid phase support[10].

Classes of potential drugs have been synthesized recently by using solid phase techniques are 1,4-dihydropyridines and polyisoxazolines. A more general separation technique, which introduces the concept of third phase, the so called fluorous phase[10] relies on the preferential partitioning of heavily fluorinated substrates into fluorinated solvents such as FC-72, which can then form a third phase separable from both, the aqueous phase and common organic solvents.

In library synthesis, either reactant or product can constitute the pre-fluorinated substrate which then is sepa-

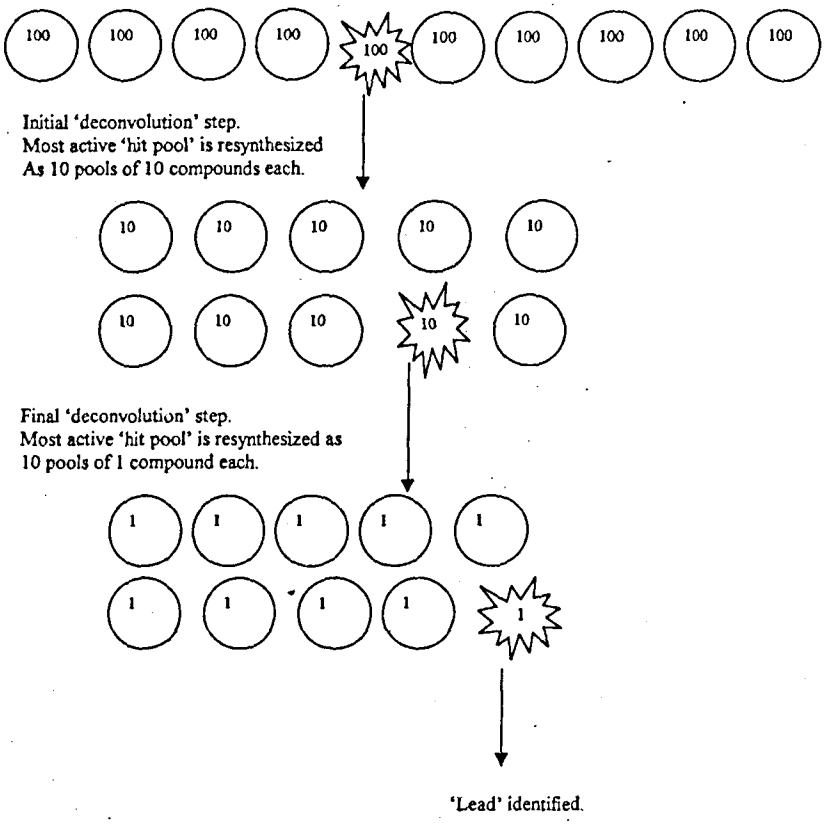

Fig. 1 : Example of deconvolution process; repeated resynthesis of subsets of hit pools results in the identification of the active compound(s) in a mixture

rated from organic or inorganic contaminants by liquid-liquid extraction via the fluorinated solvent. Thus spectroscopy and chromatography can be used to both monitor analytical purity and as a preparative tool to isolate purified product.

In conclusion, the synergy of structure-based design with combinatorial synthesis is an obvious marriage in enhancing all technologies. As can be seen, there are many design strategies for lead generation libraries. All share certain aspects in common, notably the desire to reduce the physical size of the library while maintaining or enhancing the information content.

REFERENCES

1. Czarni, A.W. and DeWitt, S.H., In; A practical guide to combinatorial chemistry, American Chemical Society, Washington DC, 1997, 1.
2. Shao, Y. and Still, W.C., **J. Org. Chem.** 1997, 61, 6086.
3. Myers, P.L., In; Current opinion in biotechnology, Current biology publications, London 1997, 8, 701.
4. Garden, M.F., Patel, D.V. and Gorden, E.M., **J. Org. Chem.** 1996, 61, 924.
5. Janda, K.D., Lo, L.C Lo, C.h, Sim, M., Wong, R.C. and Lerner, R.A., **Science,** 1997, 275, 945.
6. Marzinik, A.L. and Felder, E.R., **J. Org. Chem.** 1998, 63, 723.
7. Homes C.P. and Jones, D.G., **J. Org. Chem.** 1995, 60, 2318.
8. Brown, R.D. and Martin, Y.C. **J. Med. Chem.** 1997, 40, 2304.
9. Zhao, Y.Z. Breemen, R.B.V. Nicolic, D. Huang, C. Woodburg, C.P., Schilling, A. and Venton, D.L. **J. Org. Chem.** 1997, 40, 4006.
10. Kuth, M.J., Randall, L.A.A. and Takennouchi, K. **J. Org. Chem.** 1996, 61, 8755.