**M.Sc. BIOINFORMATICS SEMESTER III**

**ASSIGNMENT**

**PAPER II- Cheminformatics & Drug Designing, GNKPSBI302 ,**

**UNIT 4 - Combinatorial chemistry**

Q.1 Describe a brief history of Combinatorial Chemistry.

Combinatorial chemistry is a very young science, having only been around for approximately 20 years. It has been applied to drug design for an even shorter period of time. The roots of combinatorial chemistry can be dated back to the 1960’s, however. It was at this time that Bruce Merrifield developed a method for solid phase synthesis of peptides. He won the 1984 Nobel Prize in Chemistry for this work. In the 1980’s, the first combinatorial techniques were developed.  H. Mario Geyson attached resins to a surface, an idea which resembled having individual peptide synthesis reactions on the head of a pin.  These pins were reacted in individual wells, as seen below, and were a way to positionally map the products being synthesized.

Also contributing to this growing field in the mid 1980’s was Richard Houghton, who introduced a technique involving placing resins in individual, porous bags similar to tea bags. A big developed in this time period was by Furka, who took the ideas of Merrifield a step further. He applied the solid phase synthesis of peptides in a combinatorial way, developing the ‘split and mix’ technique which will be discussed in more detail below. Through the 80’s and into the early 1990’s, combinatorial chemistry was focused on peptide synthesis and later oligonucleotide synthesis. This all changed in 1992 with the publication of a paper by Bunin and Ellman.  This paper described the application of combinatorial methods to the synthesis of a benzodiazepine library. This was the first example of small molecule synthesis by combinatorial chemistry. The possibilities of using these techniques in drug design were now evident, and this branch of science began to really take off with the adoption of its techniques by the pharmaceutical and medicinal chemistry industries.

Combinatorial chemistry involves the rapid synthesis or the computer simulation of a large number of different but often structurally related molecules or materials. In a combinatorial synthesis, the number of compounds made increases exponentially with the number of chemical steps. In a binary light-directed synthesis, 2n compounds can be made in n chemical steps. Combinatorial chemistry is especially common in CADD (Computer aided drug design) and can be done online with web-based software, such as Molinspiration.

Combinatorial chemistry comprises [chemical synthetic methods](https://en.wikipedia.org/wiki/Chemical_synthesis) that make it possible to prepare a large number (tens to thousands or even millions) of compounds in a single process. These [compound libraries](https://en.wikipedia.org/wiki/Compound_library) can be made as mixtures, sets of individual compounds or chemical structures generated by computer software. Combinatorial chemistry can be used for the synthesis of small molecules and for peptides. Strategies that allow identification of useful components of the libraries are also part of combinatorial chemistry. The methods used in combinatorial chemistry are applied outside chemistry, too.

Combinatorial chemistry had been invented by Furka Á (Eötvös Loránd University Budapest Hungary) who described the principle of it, the combinatorial synthesis and a deconvolution procedure in a document that was notarized in 1982. The principle of the combinatorial method is: synthesize a multi-component compound mixture (combinatorial library) in a single stepwise procedure and screen it to find drug candidates or other kinds of useful compounds also in a single process. The most important innovation of the combinatorial method is to use mixtures in the synthesis and screening that ensures the high productivity of the process. Motivations that led to the invention had been published in 2002. **:**Combinatorial chemistry was first conceived about 15 years ago - although it was not called that until the early 1990s.

Initially, the field focused primarily on the synthesis of peptide and oligonucleotide libraries. H. Mario Geysen, distinguished research scientist at Glaxo Wellcome Inc., Research Triangle Park, N.C., helped jump-start the field in 1984 when his group developed a technique for synthesizing peptides on pin-shaped solid supports. At the Coronado conference, Geysen reported on his group's recent development of an encoding strategy in which molecular tags are attached to beads or linker groups used in solid-phase synthesis. After the products have been assayed, the tags are cleaved and determined by mass spectrometry (MS) to identify potential lead compounds.

Although combinatorial chemistry has only really been taken up by industry since the 1990s, its roots can be seen as far back as the 1960s when a researcher at Rockefeller University, Bruce Merrifield, started investigating the solid-state synthesis of peptides. In the past decade there has been a lot of research and development in combinatorial chemistry applied to the discovery of new compounds and materials. This work was pioneered by P.G. Schultz et al. in the mid-nineties (Science, 1995, 268: 1738-1740) in the context of luminescent materials obtained by co-deposition of elements on a silicon substrate. Since then the work has been pioneered by several academic groups as well as industries with large R&D programs (Symyx Technologies, GE, etc)

Q.2 Explain in detail Combinatorial Chemistry study.

Combinatorial chemistry is a collection of techniques which allow for the synthesis of multiple compounds at the same time.

This nascent technology already produced more new compounds in just a few years than the pharmaceutical industry did in its entire history. Combinatorial chemistry has turned traditional chemistry upside down. It required chemists not to think in terms of synthesizing single, well-characterized compounds but in terms of simultaneously synthesizing large populations of compounds.

Combinatorial chemistry is one of the important new methodologies developed by researchers in the pharmaceutical industry to reduce the time and costs associated with producing effective and competitive new drugs.

By accelerating the process of chemical synthesis, this method is having a profound effect on all branches of chemistry, but especially on drug discovery. Through the rapidly evolving technology of combinatorial chemistry, it is now possible to produce compound libraries to screen for novel bioactivities. This powerful new technology has begun to help pharmaceutical companies to find new drug candidates quickly, save significant money in preclinical development costs and ultimately change their fundamental approach to drug discovery.

This branch of chemistry is very young, but in this short time it has had profound effects. This can be seen by its impact on medicinal chemistry and, in particular, the drug design process. Traditionally, potential lead compounds were synthesized one at a time. The biological activity of this compound was assayed, and the results would be reflected in the next round of design. This traditional method was useful, but time consuming and expensive. Computational chemistry led to more rational design of compounds to be tested, and high throughput screening led to quick in vitro assays. Synthesis of one compound at a time could no longer keep up, and thus became the rate limiting step in the process. Combinatorial chemistry was the solution to this problem. In combinatorial approach, one can cover many combinations An x Bn in one reaction Instead of doing multiple A x B type reactions.

Conventional Reaction:   A   +   B------------------->AB

Combinatorial Chemistry: A1- n     +     B1- n--------->A1- n B1- n

Hence, In combinatorial chemistry, large numbers of compounds are made at the same time in small amounts, forming libraries which can be assayed for desired properties all at once. Finally, the active compound is identified and made in quantity as a single compound.

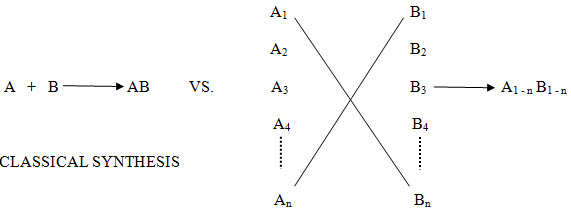
**COMBINATORIAL CHEMISTRY APPROACH**  
Combinatorial chemistry may be defined as the systematic and repetitive, covalent connection of a set of different “building blocks” of varying structures to each other to yield a large array of diverse molecular entities.

Combinatorial chemistry encompasses many strategies and processes for the rapid synthesis of large, organized collections of compounds called libraries. The collection is then tested for the biological activity. Finally, the active compound is identified and made in quantity as a single compound.

Thus, the combinatorial chemistry approach has two phases:   
1. Making a library.  
2. Finding the active compound. Screening mixtures for biological activity has been compared to finding a needle in a haystack.

In the past, chemists have traditionally made one compound at a time. For example, compound ‘A’ would have been reacted with compound ‘B’ to give product ‘AB’, which would have been isolated after reaction work up and purification through crystallization, distillation or chromatography.

In contrast to conventional approach, combinatorial chemistry offers the potential to make every combination of compound ‘A1’ to ‘An’  with compound ‘B1’ to ‘Bn’



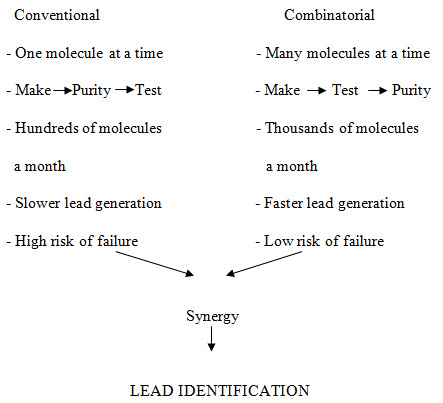
**Figure -1.** Comparison of Tradition and Combinatorial Strategies in Drug Design

The range of combinatorial techniques is highly diverse, and these products could be made individually in a parallel or in mixtures, using either solution or solid phase techniques. Whatever the technique used the common denominator is that productivity has been amplified beyond the levels that have been routine for the last hundred years.

Combinatorial chemistry-a technology for creating molecules en masse and testing them rapidly for desirable properties-continues to branch out rapidly. Compared with conventional one-molecule-at-a-time discovery strategies, many researchers see combinatorial chemistry as a better way to discover new drugs, catalysts and materials.

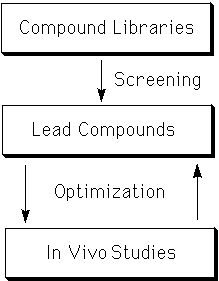
The development of new processes for the generation of collection of structurally related compounds (libraries) with the introduction of combinatorial approaches has revitalized random screening as a paradigm for drug discovery and has raised enormous excitement about the possibility of finding new and valuable drugs in short times and at reasonable costs.

**STRATEGIES**

****

**Figure -2**. Principle characteristics of conventional vs. combinatorial strategy of drug discovery

In contrast to this approach, combinatorial chemistry offers the potential to make every combination of compound A1 to An with compound B1 to Bn.



**Figure 3:** Combinatorial Synthesis

The range of combinatorial techniques is highly diverse, and these products could be made individually in a parallel or in mixtures, using either solution or solid phase techniques. Whatever the technique used the common denominator is that productivity has been amplified beyond the levels that have been routine for the last hundred years.

Combinatorial chemistry is an innovative method of synthesizing many different substances quickly and at the same time. Combinatorial chemistry contrasts with the time-consuming and labor intensive methods of traditional chemistry where compounds are synthesized individually, one at a time. While combinatorial chemistry is primarily used by organic chemists who are seeking new drugs, chemists are also now applying combinatorial chemistry to other fields such as semiconductors, superconductors, catalysts and polymers.

Combinatorial Chemistry is used to synthesize large number of chemical compounds by combining sets of building blocks. Each newly synthesized compound's composition is slightly different from the previous one. A traditional chemist can synthesize 100-200 compounds per year. A combinatorial robotic system can produce in a year thousands or millions compounds which can be tested for potential drug candidates in a high-throughput screening process. Over the last few years, the combinatorial chemistry has emerged as an exciting new paradigm for the drug discovery.

Q.3 Comment on Pros & Cons of Combinatorial Chemistry study.

**ADVANTAGES**

(1) The creation of large libraries of molecules in a short time is the main advantage of combinatorial chemistry over traditional.

(2)Compounds that cannot be synthesized using traditional methods of medicinal chemistry can be synthesized using combinatorial techniques.

(3) Thecost of combinatorial chemistry library generation and analysis of said library is very high, but when considered on a per compound basis the price is significantly lower when compared to the cost of individual synthesis.  
(4) More opportunities to generate lead compounds.

(5)Combinatorial chemistry speeds up drug discovery.

The creation of large libraries of molecules in a short amount of time is the main advantage of combinatorial chemistry over traditional. The cost of combinatorial chemistry’s library generation and analysis of said library is very high, but when considered on a per compound basis the price is significantly lower when compared to the cost of individual synthesis.

**DISADVANTAGES**

(1)  Though combinatorial chemistry would solve all the problems associated with drug discovery, one still needs to synthesize the right compound.

(2) There is a limit to the chemistry you can do when using solid phase synthesis. The resin you use is often affected by the reaction types available and care must be taken so that the attachment of the reagent to the substrate and bead are unaffected. Each reaction step has to be carefully planned, and often a reaction is not available because the chemistry affects the resin.

(3) While a large number of compounds are created, the libraries created are often not focused enough to generate a sufficient number of hits (Library components whose activity exceeds a predefined, statistically relevant threshold) during an assay for biological activity. There is a great deal of diversity created, but not often a central synthetic idea in the libraries. One can argue that there should be a focus on the type of molecule developed in order to maximize hits.

There is a limit to the chemistry you can do when using solid phase synthesis. The resin you use is often affected by the reaction types available and care must be taken so that the attachment of the reagent to the substrate and bead are unaffected. Each reaction step has to be carefully planned, and often a reaction is not available because the chemistry affects the resin.

While a large number of compounds are created, the libraries created are often not focused enough to generate a sufficient number of hits during an assay for biological activity. There is a great deal of diversity created, but not often a central synthetic idea in the libraries. One can argue that there should be a focus on the type of molecule developed in order to maximize hits.

Q.4 Comment on types of Combinatorial library.

***SCAFFOLD/BACKBONE BASED LIBRARY***

The basic structural component that is the starting point for the production of

a chemical library is termed the ‘scaffold’. It is a fixed part of the library on

which functional groups may be attached or substituted, and it is common to

all the compounds in the particular series. The scaffolds employed generally

are ‘privileged structures’; molecules that are capable of acting as ligands

for a wide array of receptors.

For a molecule to be successful as a drug, it has to be both selective and

potent. These attributes are determined by the compound’s stereochemistry

and rigidity. The stereochemistry depends on the number of chiral centers

present in the molecule, which make the molecule selective through

stereospecific binding on the target. Rigidity improves binding affinity as

a consequence of the lower entropic loss on binding. Natural products,

because of their functional role, are found with a large number of chiral

centers and show greater rigidity compared to their synthetic counterparts,

for these reasons, majority of scaffolds in use today are based on natural

products.

Ideal properties of a scaffold:

• The scaffold should be easily synthesizable by simple reactions and

inexpensive reagents.

• It should be readily available if the product is to be obtained from

natural sources.

• It should be amenable to chemical modification and introduction of

different functional groups.

• It should not have any highly reactive groups.

• It should be free from structural elements that could engender toxicity,

such as alkylating agents or planar three ring aromatic systems.

• It should abide by Lipinski’s rule of five, that is,

i) the molecular weight should be less than 500.

ii) log P value should be less than 5.

iii) there should be less than 5 H-bond donors (calculated as the sum

of O-H’s and N-H’s).

***RANDOM/DIVERSE LIBRARIES***

A diverse chemical library is a collection of compounds

which aims to act at multiple targets. Since these compound collections

are much larger as opposed to focused libraries, they have the obvious

disadvantage of cost of reagents and screening. However, they may generate

leads which may not be amenable by other methods because of the added

chemical space they encompass. The logic is that greater the molecular

diversity higher is the chance for a ligand to be recognized by a given receptor

during screening.

***VIRTUAL LIBRARIES***

A virtual library contains all possible molecules that may be synthesized,

taking into consideration the constraints of the chemistry being used and reagents that are commercially available and compatible with the chemistry

being employed. Care should be taken in the selection of a scaffold

during the initial stages. Small, flexible scaffolds such as guanidine or

lysine are preferred over hydantoins or benzodiazepines owing to the

limitations in conformational space spanned by these ring structures.

Commonly applied filters for scaffolds and substituents include hydrophobicity,

molecular weight, hazardous or toxic features, price and biological

relevance. Another commonly employed filter is ‘drug-likeness’ in that if

they are to be found active against the target in question, then they should

less likely show problems in their absorption, distribution, metabolism,

excretion or toxicology, so as to shorten the drug discovery and development

cycle.

***TARGETED OR FOCUSED LIBRARIES***

The common practice in drug discovery over the past few years is the creation

of ‘diverse’ chemical libraries; compounds collections which aim to

act at multiple targets or proteins (Harris et al., 2011) The logic behind such

a design is that if more number of compounds are synthesized, greater will

be the chemical space that is covered and hence greater the chances of generating

a lead compound (Drewry and Young, 1999). But unfortunately, the

synthesis and screening of such a large collection of compounds has proved

to be a costly affair. As a result, the focus has now shifted towards building

‘targeted’ or focused’ libraries. These are libraries that are smaller, containing

higher quality compounds that have been designed keeping in mind a

specific target or protein. The obvious advantage of designing such a library

is that they eliminate compounds which are unlikely to bind to the target;

as a result, fewer number of compounds have to be screened to identify

leads. Also, the hit rates observed with such a library are greater compared

to diverse libraries, thus reducing the hit-to-lead timescale.

An understanding of the target is an essential prerequisite for designing

a focused library. The design can be based on the structural information

available about the target (e.g., kinase receptors, where an abundance of

crystallographic data is available) or sequence or mutagenesis data (e.g.,

ion channels). An alternative approach can be based on known ligands for

the target in question provided high quality structural information or prior

knowledge of the binding site interactions is available. Targeted libraries are

usually obtained as a subset of much larger collections using computational techniques. These libraries commonly employ a single core entity with

attachment points to which specific substituents, or side chains can be

appended to arrive at the desired compounds. When selecting the scaffold

and the substituents, synthetic feasibility must be given careful consideration.

Generally, all possible combinations of compounds are considered

and a subset is chosen for synthesis. Thus the process of selection of compounds

out of all possibilities is very important (Harris et al., 2011). This

selection process is governed by the chemical space of interest for the target

in question, followed by finalization of the set of molecules which will

represent that space. A filter may be the introduction of a constraint that the

molecules must be ‘drug-like’, that is, they must have properties which give

them a reasonably good chance of having good oral bioavailability and a

decent pharmacokinetic profile.

Q.5 Elaborate on approaches used to generate libraries (Diverse Library & Focused Library)

The initial step in library design is an idea of a new synthetic route or an innovative scaffold. Such a new scaffold might be synthetically accessible through a novel, elegant route, or it is of medicinal chemistry relevance, such as protein secondary structure motifs, turn mimetics, benzodiazepines, pyrimidines and many other examples. It must be clearly evaluated, which building blocks are compatible with a particular synthetic scheme. This reaction validation phase is the key step for a successful library design and realization, as computer-assisted design is encompassing the rules derived from these systematic studies. Any library design procedure needs to make sure that every new compound is unique and adds useful information to the entire library. To design a generic library for lead finding, the consideration of two very similar compounds does not enhance the ability to find different types of biological activity; indeed, it would be better to replace one of these compounds with a more dissimilar one. This concept is the key motivating factor for the design of optimally diverse compound libraries. **A** major component is the definition of an exclusion region around each compound (similarity radius), where it is undesirable to have overlapping candidates. **As** suggested by validation studies, we typically choose a Tanimoto coefficient of 0.85 for 2D fingerprints to represent the boundary between similarity/dissimilarity in terms of biological activity for product structures.

The library itself should exhibit a wide coverage of the physico-chemical property space. New compounds should be true drug-like compounds with biological relevance. There have been some very interesting approaches to estimate the ‘drug-likeness’ of organic compounds, either by an empirical set of rules, a detailed substructure analysis using a genetic algorithm approach, their match to frequently occurring molecular frameworks or the use of scoring schemes to discriminate between drugs and non-drugs. This latter promising method is based on simple structural parameters like extended atom types or short selected fragments, which are subjected to data analysis using neural networks to account for non-linearity. These approaches are effective in providing a reasonable ranking of compounds in a virtual or existing library, so that synthetic and screening resources can be focused on compound subsets of general biological interest.

The initial emphasis in combinatorial chemistry was on the synthesis of as many “diverse” compounds as possible, in the expectation that a larger number of molecules would inevitably lead to more hits in the biological assays. This was not observed in practice; “historical” compounds built up over many years generally performed much better than the new libraries. Moreover, the hits produced from combinatorial chemistry often had properties that made them unsuitable as lead compounds (e.g. they were too large, too insoluble, or contained inappropriate functional groups). This is illustrated in Figure 9-2 which shows the molecular weight profile of an aminothiazole library designed to be “diverse” superimposed on the corresponding profile of compounds in the WDI. It can be seen that the molecules in the library tend to have higher molecular weights than is seen in known drug molecules.

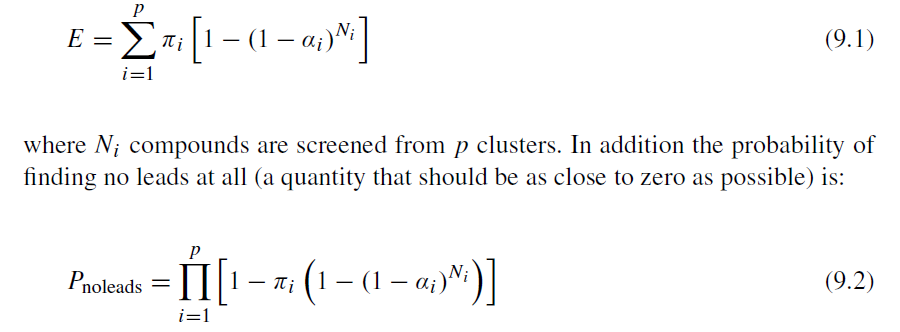
Since the early days of combinatorial chemistry, two trends have emerged. The first is towards the design of smaller, more focussed libraries that incorporate as much information about the therapeutic target as possible. Of course, the amount of information that is available will vary from one drug discovery project to another At one extreme, when the 3D structure of the target protein is available, combinatorial libraries (and screening libraries in general) may be designed to fit the active site using docking methods. If some active compounds are known then libraries may be designed using pharmacophoric approaches, 2D similarity or physicochemical properties. A recent trend is to focus libraries on a family of targets, such as proteases or kinases, in which case the design should incorporate properties that are known to be important for all targets in the family.When little is known about the target then more diverse libraries may be relevant. Such libraries are sometimes referred to as primary screening libraries and are designed to give a broad coverage of chemistry space so that they can be screened against a range of structural targets. In general, a balance between diversity and focus is needed, with the amount of diversity required being inversely related to the amount of information available about the target [Baldwin 1996; Hann and Green 1999]. The difficulty lies in quantifying these factors and achieving the correct balance between them.

The second trend to emerge is based on the realisation that compounds within

combinatorial libraries, whether they are diverse or focussed, should have “druglike” physicochemical properties so that they constitute good start points for further optimisation.

A framework to help to achieve this balance between diversity and focus when designing libraries or when constructing sets of compounds or indeed whole screening collections has been provided by the mathematical model developed by Harper et al. This model starts from two assumptions about screening in drug discovery. The first is that once a hit has been found then structurally similar molecules will be selected using substructure and/or similarity searching methods for further testing in order to provide evidence of a structure–activity relationship. The second assumption is that the overall goal is to find multiple lead series from the screen.

The chemical space is assumed to be clustered or partitioned in some way into structural clusters. Two parameters are introduced into the model. These are the probability that a randomly selected molecule from some cluster *i* will test active *(αi )* and the probability that the cluster will contain a lead molecule *(πi )*. A lead molecule is more than just a “hit”; it is a compound with the correct combination of activity, physicochemical, ADMET properties and novelty to provide a good chance of becoming a marketed drug. Based on these two parameters it is possible to derive a number of useful quantities, including the expected number of lead containing clusters in which leads are found, *E*:



The consequences of different ways to construct a compound collection can then be derived. For example, if one uses values for *πi* of 0.00001 (a value typical for a set of “diverse” compounds screened against tractable drug targets) and *αi* of 0.3 [Martin et al. 2002] then for a set of 1 million compounds there is a probability of finding no leads greater than 0.9 if the set comprises clusters of size 100. However, if the collection has two compounds per cluster then the probability of finding no leads drops to less than 0.08. The model enables the optimal design of various types of compound sets to be determined, be they screening collections, focussed sets or combinatorial libraries. Moreover, the model enables a diversity score to be derived that conforms to all but one of the criteria proposed by Waldman.

Q.6 Explain in detail Split & Mix technique in library preparation along with an appropriate example.

This technique was pioneered by Furka and co-workers in 1988 for the synthesis of large peptide libraries. This approach is termed divide couple and recombine synthesis by other workers.

In this technique, the starting material is split in ‘n’ portions, reacted with ‘n’ building blocks, and recombined in one flask. For the second step, this procedure is repeated. This method is particularly employed for solid phase synthesis.

In theory, if one wanted to make a library of the compounds A-B1, A-B2, A-B3, and A-B4, it could be accomplished by adding the reagents B1, B2, B3, and B4 to a solution of resin-bound A. All four of the desired compounds should be produced. However, there is a problem here. The reaction kinetics will not be the same for each reagent added. Therefore, the resulting mixture will not be equimolar, and will likely contain some dominating products. As well, individual beads will contain a mixture of products, not just a single one.  In combinatorial chemistry, equimolar mixtures of products and one product per bead are desired. The Split and Pool technique, also known as the Split and Mix technique, is the solution to this.

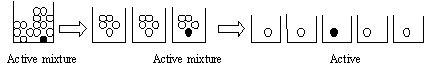
In this method, resins are divided equally into wells. The number of pools needed is dictated by the number of reagents they will be used during the reaction. As well, the overall number of beads used must be at least equal to the total number of compounds which will be generated by the process. Each individual well is reacted with a different reagent. The products from these reactions are isolated via filtration, and the pooled together to give an equimolar mixture. This mixture is then equally divided amongst the wells and reacted again with reagents. The process is repeated as many times as necessary. The end result is equimolar library of compounds representing all possible combinations of reagents. Only a very small amount of each product is produced (often a single bead represents one product), but only a small amount is needed for biological testing. When an active compound is discovered, it would need to be produced on a larger scale for further assays regardless. However, since the majority of compounds formed will not have any biological activity, there is no point than making any more of it than is needed to determine this.

A simple example of a 3 x 3 x 3 library gives all 27 possible combinations of trimeric products. X, Y and Z could be amino acids, in which case the final products would be tripeptides, but more generally they could be any type of monomeric unit or chemical precursor. It can be seen that the mix and split procedure finally gives three mixtures each consisting of nine compounds each, and there are several ways of progressing these compounds to biological screening. Although the compounds can be tested whilst still attached to the bead, a favoured method is to test the compounds as a mixture following cleavage from the solid phase. Activity in any given mixture reveals the partial structure of active compounds within the library, as the residue coupled last (usually the N-terminal residue) is unique to each mixture. Identification of the most active compound relies on deconvoluting the active mixtures in the library through further synthesis and screening.

In the example where the active structure is YXY, the mixture with Y at the terminal position will appear as the most active. Having retained samples of the intermediate dimers on resin (so-called "recursive" deconvolution ) addition of Y to each of the three mixtures will give all nine compounds with Y at the terminal position, and the second position defined by the mixture. The most active mixture here defines the middle position of the most active trimer to be residue X. Finally, the three individual compounds can be independently resynthesised and tested to reveal both the most potent compound and also some structure activity relationship data.

In contrast, Lam et al. tested a family of peptides whilst still attached to the resin bead solid phase. Nineteen amino, acids were incorporated into pentapeptides to generate a library of almost two and a half million compounds. By using a colourimetric assay, beads bearing peptide sequences that bound tightly to the protein streptavidin or to an antibody raised against β-endorphin were revealed by visual inspection. Bead picking using micromanipulation isolated the beads, and the active peptide structures were determined by microsequencing.

A modification of this method has allowed screening of such libraries in solution. Linkers have been devised that allow several copies of the library compounds to be released sequentially. Using this method, it is possible to identify an active mixture using a solution assay, and then return to the beads that produced these compounds, and redistribute them into smaller mixtures for retest. By repeatedly reducing the mixture size, ultimately to single compounds, the bead containing the most potent sequence may be identified and the peptide product sequenced.



**Advantages**:  
a) Only few reaction vessels required  
(b) Large libraries can be quickly generated (up to 105 compounds)   
**Disadvantages:**  
(a) Threefold amount of resin beads necessary  
(b) The amount of synthesized product is very small.  
(c) Complex mixtures are formed.   
(d) Deconvolution or tagging is required.  
(e) Synergistic effects may be observed during screening, leading to false positives.

Q.7 Comment on Houghton Tea bag approach in library preparation along with an example.

A polypropylene mesh bag, with dimensions of approximately 15 x 20 mm, filled with resin beads, sealed and labelled for a later identification, is known as a tea-bag, designed by Houghten (1985). The “tea-bag” mesh size is too small to allow resin beads to escape, but solvents and soluble reagents could readily enter. The principles of its use are to make multimilligram (up to 500 μmoles) quantities of a single peptide sequence in each packet, which is sufficient for full characterization and screening. To save time and work when making many peptides simultaneously, bags could be combined into the same reactors for common chemical steps.

For example, in the synthesis of 40 different peptides, all the bags are initially charged with resin beads bearing a Boc-protected amino acid, and the packets are combined for resin deprotection, washing, and neutralization steps. Then the bags are sorted into groups for the addition of the next amino acid. Then the bags could be combined again for deprotection, washing, and neutralization. After an appropriate number coupling steps, all the bags can then be treated with HF/anisole to cleave the peptides from the beads. As the first intention was to speed up peptide synthesis, nowadays the tea-bag method is a classic example for combinatorial synthesis, its speed, and effectiveness.

Schematic overview of a typical group of steps carried out using the tea-bag procedure:

Some examples for the use of the tea-bag method:

• Characterization of the influenza haemagglutinin protein (HA1) and discovering the amino acid position that is critical important to the binding interaction

• Production of a small combinatorial library of urea analogues

• Rapid "tea-bag" peptide synthesis using 9-fluorenylmethoxycarbonyl (Fmoc) protected amino acids applied for antigenic mapping of viral proteins

• Studies on the structural requirement for ligand binding to the neuropeptide Y (NPY) receptor from rat cerebral cortex

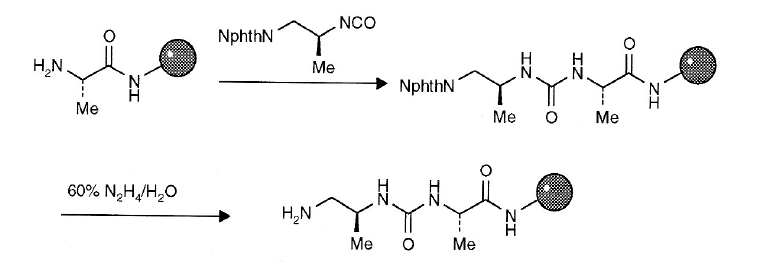
• Peptide and peptidomimetic libraries. Molecular diversity and drug design

• The use of tea-bag synthesis with paper discs as the solid phase in epitope mapping studies

• Rate of swelling of sodium polyacrylate

**Structures:**

Scheme 1: The Burgess Two-step cycle for the addition of a monomer in the synthesis of oligoureas – using “tea-bags”



Q.8 Explain monomer-based selection approach for library designing with an appropriate example. **(If required, case study can also be discussed in form of answer )**

In most combinatorial reaction schemes there are many more monomers available than can be handled in practice. For example, consider a benzodiazepine



library constructed from three monomer pools, as shown in Figure 9-5. It is easy to find many monomers for each position of variability by searching a database of available reagents such as the ACD [ACD]. The virtual libraries in such cases can easily contain millions if not billions of products; much larger than can actually be synthesised in practice. Thus, a key issue in combinatorial library design is the selection of monomers such that the resultant library is of a manageable size and has the desired properties, whether it is to be diverse or focussed or should exhibit a balance between the two.

The two main strategies for combinatorial library design are known as

*monomer-based selection* and *product-based selection*.

In monomer-based selection optimised subsets of monomers are selected

without consideration of the products that will result. Consider a hypothetical

three-component library with 100 monomers available at each position of

variability, where the aim is to synthesise a diverse 10 × 10 × 10 combinatorial library. In monomer-based selection this would involve selecting the 10 most diverse monomers from each set of monomers. In general, there are

*N* !

*n*!*(N* − *n)*!

subsets of size *n* contained within a larger set of *N* compounds. For example,

there are more than 1013 different subsets of size 10 from a pool of 100

monomers. It is not possible to examine all of these.

Selecting compounds for screening where the techniques of dissimilarity-based compound selection, clustering and partitioning were introduced, together with related optimisation methods. Any of these techniques can also be used in the monomer-based selection approach.

An early example of monomer-based design is described by Martin and

colleagues and used experimental design. Diverse subsets of monomers were

selected for the synthesis of peptoid libraries (peptoids are synthetic oligomers

with a peptide backbone but with the side chain attached to the nitrogen atom

rather than the alpha-carbon atom). The diversity measure included features such as lipophilicity, shape and chemical functionality and uses a variety of descriptors including ClogP, topological indices, Daylight fingerprints and atom layer properties based on receptor recognition descriptors. Diverse subsets of monomers were selected using D-optimal design.

The monomers are subjected to cluster analysis, and a representative (termed the *centroid)* from each cluster is selected. Each of the centroid monomers is compared with the 30 000 SDF molecules in several ways:

1. Whole-molecule comparisons: in which the monomer as compared to each of the SDF molecules and the highest similarity score is recorded.

2. Exact substructure occurrence of the monomer in SDF molecules: in which an exact substructure search is carried out and the number of occurrences recorded.

3. Similar substructure occurrence of the monomer in SDF molecules: in which part of an SDF molecule is similar to the monomer as measured by a similarity score. The occurrence of similar substructures is recorded. The whole-monomer/whole-molecule similarity measure in option 1 will give monomers that are close in size to the SDF molecules. However, since the library products usually are formed from at least two monomers, it is more meaningful to detect substructures contained in SDF molecules that are similar to the monomer in question, so that the product molecules formed will contain these substructural (or “druglike”) motifs. These centroid monomers that are either highly similar as determined by analysis 1 or occur frequently as determined by analysis 2 or 3, are then selected to initiate chemistry. In this way, diversity is covered by these centroid monomers as the starting set and similarity criteria will select from this diverse set monomers that are also “druglike” compared to SDF molecules.

Molecules are represented by a 1024-bit fingerprint, using the DAYLIGHT fingerprint.\* Two similarity measures are used. In the case of monomer and SDF whole-molecule comparison, the traditional Tanimoto index expressed by

Tanimoto similarity score = (A &B)/(A + B - A & B)

is used, where *(A & B)* denotes the number of common bits turned on in both molecules, *A* represents those turned on in molecule *A, and B* denotes those turned on in molecule *B.* In the case of monomer and SDF substructure comparison, the substructure similarity index Substructure similarity score = (A & B)/A is used, where *A* denotes the number of bits turned on in monomer *A,* and *(A & B)* denotes the number of common bits turned on in both monomer *A* and SDF molecule *B.* The occurrence of the substructure is recorded if this similarity score is greater than 0.9.

When comparing monomers with substructures of the SDF molecules, similar substructure occurrence is counted if the substructure similarity index is greater than 0.9. The cutoff value is determined by looking at substructures thus selected and is therefore empirical and somewhat arbitrary. Exact substructure occurrence is a specific case of the substructural

similarity when the substructural similarity index is I .O. A trend similar to that observed in monomer and whole SDF molecule comparison was also observed here, i.e., in general, as the monomers become singleton-like, the occurrences of similar substructures in SDF become fewer.

In fact, a general correlation between the whole-molecule similarity and substructural similarity occurrence exists, as one would expect. Thus, using whole-molecule similarity as a measure of closeness to druglike motifs may be a reasonable approximation.

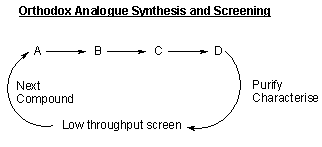
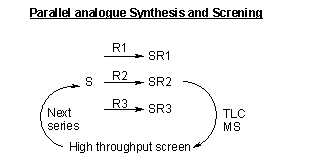
Q.9 Explain in detail Parallel synthesis technique in library preparation along with an appropriate example.

Parallel synthesis involves running multiple reactions, each in a separate vessel, at once instead of in series. Like the split and pool method, it results in the production of multiple compounds at the same time. However, unlike split an pool, parallel synthesis gives individual compounds, not a mixture. Thus, deconvolution is not an issue in this method.

Advances in automation have made parallel synthesis possible. Many different apparatuses have been developed to either fully or partially automate this process. Synthesizers are machines in which the process is completely automated. The complete reaction, including all addition and separation steps as well as cleavage of the final product from the resin are done within the machine, allowing the reaction to occur unattended. Workstations allow for an assembly line-like set up. There are individual stations for reagent addition, product purification, ect. The group of reaction vessels, most commonly set up as a 96 well plate, are moved manually from station to station. However, as one vessel is moved to the next station, another can replace it.

Parallel synthesis does allow for the production of a large number of individual compounds, but other methods (ie-split and pool) are capable of higher numbers in less time, and thus are better for the earlier stages of drug discovery. Parallel synthesis is better suited to stages of lead optimization.

Manual or automated approaches can be used for the parallel preparation of tens to hundreds of analogues of a biologically active substrate. The products are synthesised using reliable coupling and functional group interconversion chemistry and are progressed to screening after removal of solvent and volatile by-products. Parallel and orthodox synthesis is compared below.

Orthodox synthesis usually involves a multistep sequence, e.g. from A through to the final product D, which is purified and fully characterised before screening. The next analogue is then designed, guided by the biological activity of the previous compound, prepared, and then screened. This process is repeated to optimise both activity and selectivity.

In contrast parallel analogue synthesis involves reaction of a substrate S with multiple reactants, R1, R2, R3 … Rn, to produce a compound library of n individual products SR1, SR2, SR3 … SRn. The library is screened, usually without purification, and with only minimal characterisation of the individual compounds, using a rapid throughput screening technique.

Panlabs have recently disclosed an interest in making large number of compounds as individual components using parallel, reliable solution chemistry. Reactions are pushed to completion by the use of excess quantities of the reactive reagent, and are isolated by solvent - solvent extraction. There is no further purification, and thus they prefer to describe these samples as "reaction products".

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 theri 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 HCI 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 follow­up 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.

Q.10 Explain product-based selection approach for library designing with an

appropriate example.

In product-based selection, the properties of the resulting product molecules

are taken into account when selecting the monomers. Having enumerated the

virtual library any of the subset selection methods introduced in Chapter 6 could

then in principle be applied. This process is generally referred to as *cherry-picking* but it is synthetically inefficient insofar as combinatorial synthesis is concerned.

Synthetic efficiency is maximised by taking the *combinatorial constraint* into

account and selecting a combinatorial subset such that every reagent selected

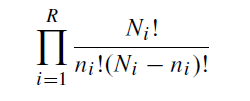
at each point of variation reacts with every other reagent selected at the other

positions.

Product-based selection is much more computationally demanding than

monomer-based selection. The number of combinatorial subsets in this case is

given by the following equation:

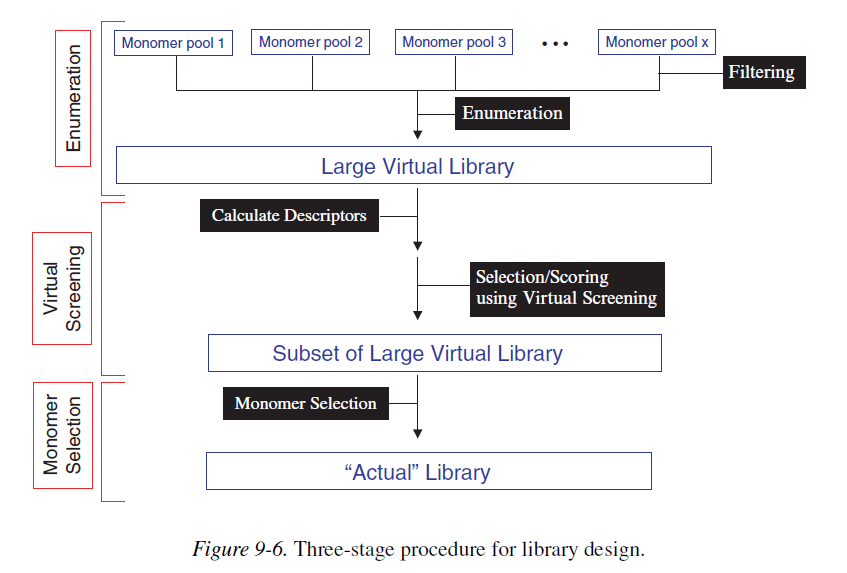


where *R* is the number of positions of variability and there are *ni* monomers to be selected from a possible *Ni* at each substitution position. Thus, there are almost 1040 different 10×10×10 libraries that could be synthesised from a 100×100× 100 virtual library. The selection of combinatorial subsets has been tackled using optimisation techniques such as simulated annealing and genetic algorithms.

Despite the greater computational complexity of performing product-based

selection compared to monomer-based selection it can be a more effective method when the aim is to optimise the properties of a library as a whole, such as diversity or the distribution of physicochemical properties. For example, it has been shown that more diverse libraries result if selection is performed in product space rather than in monomer-space. In addition, product-based selection is usually more appropriate for focussed libraries which require consideration of the properties of the product structures A general strategy for product-based library design involves the following three steps. First, lists of potential reagents are identified (e.g. by searching relevant databases), filtered as appropriate, and the virtual library is enumerated. In the second step the virtual library is subjected to virtual screening to evaluate and score each of the structures. In the third stage the reagents to be used in the actual library for synthesis are selected using the results from the virtual screening together with any additional criteria such as the degree of structural diversity required, or the degree of similarity/dissimilarity needed to existing collections. This three-stage process is illustrated in Figure 9-6.

It is important to note that it may be possible to reduce significantly the size of the virtual library by eliminating from consideration monomers that can be unambiguously identified as being inappropriate. For example, monomers that contain functionality known to be incompatible with the proposed reaction scheme can be eliminated, as can monomers that are unavailable or cannot be purchased in time. This is the initial “filtering” step in Figure 9-6.



The final, monomer selection stage is typically implemented using optimisation techniques such as GAs or simulated annealing. By way of example, the SELECT is based on a GA in which each chromosome encodes one possible combinatorial subset. Assume a two component combinatorial synthesis in which *nA* of a possible *NA* first monomers are to be reacted with *nB* of a possible *NB* second monomers. The chromosome of the GA thus contains *nA* + *nB* elements, each position specifying one possible monomer. The fitness function quantifies the “goodness” of the combinatorial subset encoded in the chromosome and the GA evolves new potential subsets in an attempt to maximise this quantity. In some cases the virtual library is too large to allow full enumeration and descriptor calculation, making product-based combinatorial subset selection unfeasible. A number of methods have been proposed to try to overcome this problem. For example, the Markush-based enumeration strategy described above enables certain types of product descriptors to be calculated very rapidly without actually constructing the products themselves [Barnard et al. 2000]. In the first phase, descriptors are calculated for each of the monomers in the library. In the second phase the monomer descriptors are combined together to generate the descriptor for the product. Such an approach is applicable to any properties that are additive, such as molecular weight, counts of features such as donors and acceptors, calculated log*P* and even 2D fingerprints. An alternative approach is to use random sampling techniques to derive a statistical model of the property under consideration; in addition to the estimation of molecular properties such an approach may be used for monomer selection and to prioritise different library scaffolds.

Alternative approaches to product-based library design have been developed that do not require enumeration of the entire virtual library. These methods have been termed *molecule-based methods* to distinguish them from library based methods and they are appropriate for the design of targeted or focussed libraries. For example, the method developed by Sheridan and Kearsley uses a GA in which each chromosome encodes a product molecule (rather than a combinatorial library). The fitness function measures the similarity of the molecule encoded by a chromosome to a target molecule.

When the GA has terminated the entire population of chromosomes is analysed to identify monomers that occur frequently across all the molecules in the population. The frequently occurring monomers could then be used to construct a combinatorial library. The method was tested on a tripeptoid library with

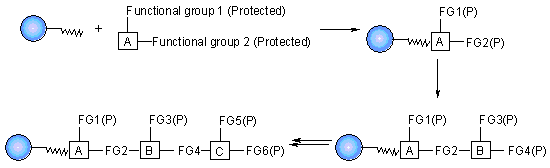
three positions of variability; 2,507 amines were available for two of the substitution positions and 3,312 for the third position. This gives a virtual library of approximately 20 billion potential tripeptoids. The method was able to find molecules that were very similar to given target molecules after exploring only a very small fraction of the total search space.

The molecule-based method is a relatively fast procedure, especially when optimisation is based on 2D properties, since the fitness function involves a pairwise molecular comparison rather than the analysis of an entire library, as is the case in library-based methods. In these approaches, however, there is no guarantee that building libraries from frequently occurring monomers will result in optimised libraries, nor is it possible to optimise properties of the library as a whole.

Q.11 Explain Solid phase combinatorial chemistry method in detail (should include details about principle, linkers, resin beads, advantages & disadvantages).

Since [Merrifield](http://en.wikipedia.org/wiki/Bruce_Merrifield) pioneered solid phase synthesis back in 1963, work, which earns him a Nobel Prize, the subject, has changed radically. Merrifield’s Solid Phase synthesis concept, first developed for biopolymer, has spread in every field where organic synthesis is involved. Many laboratories and companies focused on the development of technologies and chemistry suitable to SPS. This resulted in the spectacular outburst of combinatorial chemistry, which profoundly changed the approach for new drugs, new catalyst or new natural discovery.

The use of solid support for organic synthesis relies on three interconnected requirements:



1) A cross linked, insoluble polymeric material that is inert to the condition of synthesis;  
2) Some means of linking the substrate to this solid phase that permits selective cleavage of some or all of the product from the solid support during synthesis for analysis of the extent of reaction(s), and ultimately to give the final product of interest;  
3) A chemical protection strategy to allow selective protection and deprotection of reactive groups.

Merrifield developed a series of chemical reactions that can be used to synthesise proteins. The direction of synthesis is opposite to that used in the cell. The intended carboxy terminal amino acid is anchored to a solid support. Then, the next amino acid is coupled to the first one. In order to prevent further chain growth at this point, the amino acid, which is added, has its amino group blocked. After the coupling step, the block is removed from the primary amino group and the coupling reaction is repeated with the next amino acid. The process continues until the peptide or protein is completed. Then, the molecule is cleaved from the solid support and any groups protecting amino acid side chains are removed. Finally, the peptide or protein is purified to remove partial products and products containing errors.

Combinatorial synthesis on solid phase can generate very large numbers of products, using a method described as mix and split synthesis. This technique was pioneered by Furka and has been enthusiastically exploited by many others since its first disclosure. For example, Houghten has used mix and split on a macro scale in a "tea bag" approach for the generation of large libraries of peptides.

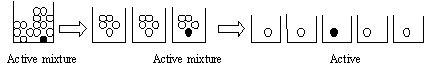
The method works as follows: a sample of resin support material is divided into a number of equal portions (x) and each of these are individually reacted with a single different reagent. After completion of the reactions, and subsequent washing to remove excess reagents, the individual portions are recombined, the whole is thoroughly mixed, and may then be divided again into portions. Reaction with a further set of activated reagents gives the complete set of possible dimeric: units as mixtures and this whole process may then be repeated as necessary (for a total of n times). The number of compounds obtained arises from the geometric increase in potential products; in this case x to the power of n.

A simple example of a 3 x 3 x 3 library gives all 27 possible combinations of trimeric products. X, Y and Z could be amino acids, in which case the final products would be tripeptides, but more generally they could be any type of monomeric unit or chemical precursor. It can be seen that the mix and split procedure finally gives three mixtures each consisting of nine compounds each, and there are several ways of progressing these compounds to biological screening. Although the compounds can be tested whilst still attached to the bead, a favoured method is to test the compounds as a mixture following cleavage from the solid phase. Activity in any given mixture reveals the partial structure of active compounds within the library, as the residue coupled last (usually the N-terminal residue) is unique to each mixture. Identification of the most active compound relies on deconvoluting the active mixtures in the library through further synthesis and screening.

In the example where the active structure is YXY, the mixture with Y at the terminal position will appear as the most active. Having retained samples of the intermediate dimers on resin (so-called "recursive" deconvolution ) addition of Y to each of the three mixtures will give all nine compounds with Y at the terminal position, and the second position defined by the mixture. The most active mixture here defines the middle position of the most active trimer to be residue X. Finally, the three individual compounds can be independently resynthesised and tested to reveal both the most potent compound and also some structure activity relationship data.

In contrast, Lam et al. tested a family of peptides whilst still attached to the resin bead solid phase. Nineteen amino, acids were incorporated into pentapeptides to generate a library of almost two and a half million compounds. By using a colourimetric assay, beads bearing peptide sequences that bound tightly to the protein streptavidin or to an antibody raised against β-endorphin were revealed by visual inspection. Bead picking using micromanipulation isolated the beads, and the active peptide structures were determined by microsequencing.

A modification of this method has allowed screening of such libraries in solution. Linkers have been devised that allow several copies of the library compounds to be released sequentially. Using this method it is possible to identify an active mixture using a solution assay, and then return to the beads that produced these compounds, and redistribute them into smaller mixtures for retest. By repeatedly reducing the mixture size, ultimately to single compounds, the bead containing the most potent sequence may be identified and the peptide product sequenced.

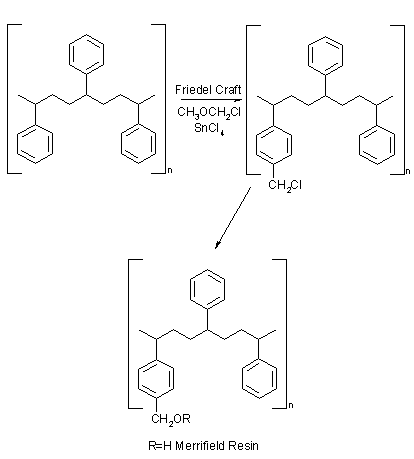


**Resins for Solid Phase Synthesis**

In solid phase support synthesis, the solid support is generally based on a polystyrene resin. The most commonly used resin supports for SPS include spherical beads of lightly cross linked gel type polystyrene (1–2% divinylbenzene) and poly(styrene-oxyethylene) graft copolymers which are functionalised to allow attachment of linkers and substrate molecules. Each of these materials has advantages and disadvantages depending on the particular application.

**Crosslinked Polystyrene**

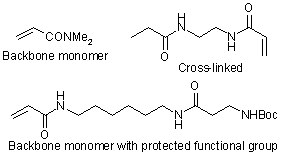
Lightly cross-linked gel type polystyrene (GPS) (Figure) has been most widely used due to its common availability and inexpensive cost. GPS beads which are functionalised with chloromethyl-, aminomethyl-, and a variety of linkers are commercially available from a variety of sources. A prominent characteristic of GPS beads is their ability to absorb large relative volumes of certain organic solvents (swelling). This swelling causes a phase change of the bead from a solid to a solvent-swollen gel, and therefore, the reactive sites are accessed by diffusion of reactants through a solvent-swollen gel network. In solvents, which swell the polymer well, the gel network consists of mostly solvent with only a small fraction of the total mass being polymer backbone. This allows relatively rapid diffusional access of reagents to reactive sites within the swollen bead. In solvents, which do not swell the polymer, the cross-linked network does not expand and the diffusion of reagents into the interior of the bead is impeded.



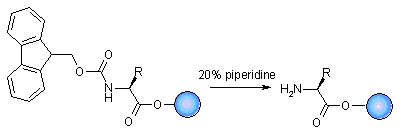
GPS has good swelling characteristics in solvents of low to medium polarity ranging from aliphatic hydrocarbons to dichloromethane. Polar, protic solvents, such as alcohols and water, do not swell GPS resins, and accessibility to all reaction sites may be compromised. Hence GPS supports are most suitable for chemistry performed in solvents of low to medium polarity.

**Polyamide Resins**

Sheppard designed polyacrylamide polymers for peptide synthesis as it was expected that these polymers would more closely mimic the properties of the peptide chains themselves and have greatly improved solvation properties in polar, aprotic solvents (e.g. DMF, or N-methyl pyrrolidinone).

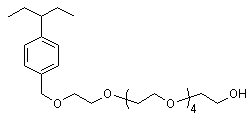


Sheppard also proposed the use of a new protection and linking strategy. The Merrifield approach depended on a benzyl ester linkage and Boc protection. But a more mild protection/deprotection were sought. The protecting group finally chosen was the fluorenylmethoxycarbonyl (Fmoc) which can be removed by base (usually piperidine).



**TentaGel Resins**

Poly (styrene-oxyethylene) graft copolymers, first reported by Bayer and Rapp, are another class of widely used supports for organic synthesis. As with the polyacrylamide resins, in order to produce a polar reaction milieu that is closer to the solvents generally used by solution synthetic chemists, grafted polymer beads have been prepared. The most pre-eminent of these is TentaGel resin which consists of polyethylene glycol attached to cross-linked polystyrene through an ether link, and combines the benefits of the soluble polyethylene glycol support with the insolubility and handling characteristics of the polystyrene bead. The resin was originally prepared by the polymerisation of ethylene oxide on cross-linked polystyrene already derives with tetraethylene glycol to give polyethylene glycol chains. (Figure) Poly (styrene-oxyethylene) graft copolymers beads display relatively uniform swelling in a variety of solvents from medium to high polarity ranging from toluene to water. The polymers are produced by grafting ethylene oxide from the polystyrene backbone creating long flexible chains that terminate with a reactive site spatially separated from the more rigid polystyrene backbone.



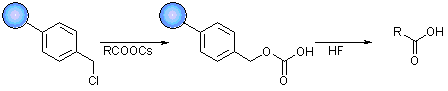
Some disadvantages of Poly(styrene-oxyethylene) graft copolymers supports are:  
- Relatively low functional group loading compared with GPS; the potential for the PEG chains to complex Lewis acids;  
- The potential instability of PEG;  
- The presence of linear PEG impurities found in the small molecule products after cleavage from the resin;  
- The tendency for resins to become sticky and difficult to handle as the synthesis progress.

**Linkers**

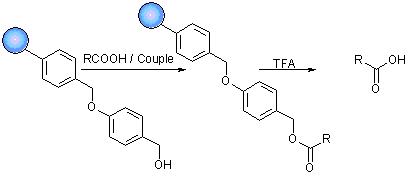
The group that joins the substrate to the resin bead is an essential part of solid phase synthesis. The linker is a specialised protecting group, in that much of the time, the linker will tie up a functional group, only for it to reappear at the end of the synthesis. The linker must not be affected by the chemistry used to modify or extend the attached compound. And finally the cleavage step should proceed readily and in a good yield. The best linker must allow attachment and cleavage in quantitative yield.

**Carboxylic acid linkers**

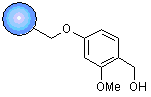
The first linking group used for peptide synthesis bears the name of the father of solid phase synthesis. Merrifield resin is cross-linked polystyrene functionalised with a chloromethyl group. The carbonyl group is attached by the nucleophilic displacement of the chloride with a cesium carboxylate salt in DMF. Cleavage to regenerate the carboxylic acid is usually achieved by hydrogen fluoride.



The second class of linker used for carboxylic acid is the Wang linker. This linker is generally attached to cross-linked polystyrene, TentaGel and polyacrylamide to form Wang resin. It was designed for the synthesis of peptide carboxylic acids using the Fmoc-protection strategy, and due to the activated benzyl alcohol design, the carboxylic acid product can be cleaved with TFA.



A more acid-labile form of the Wang resin has been developed. The SASRIN resin has the same structure as the Wang linker but with the addition of a methoxy group to stabilise the carbonium ion formed during acid catalysed cleavage.

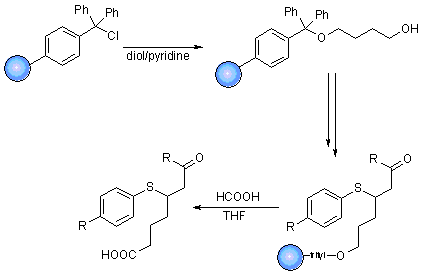


**Alcohol Linkers**

A hydroxyl linker based on the tetrahydropyranyl (THP) protecting group has been developed by Thompson and Ellmann. All type of alcohols readily add to dihydropyran and the resulting THP protecting group is stable to strong base, but easily cleaved with acid. This linker is attached to a Merrifield resin.

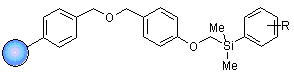
The generation and cleavage of the THP linker for alcohol

The trityl group is a good acid-labile protecting group for a lot of heteroatoms. The trityl group has been used to anchor alcohols in the synthesis of a library of β-mercaptoketones.



**Traceless Linkers**

In some case, the starting materials are loaded onto the resin in one form, such as carboxylic acid, and cleaved in another form; a carboxamide for example. This is perfectly acceptable if the target compound requires the released function. (Peptides invariably contain a carboxylic acid or carboxamide.) However, the growth in interest in combinatorial libraries of low molecular weight non-peptides has elicited a need in new types of linker. These linkers show non-specific function after cleavage. Traceless linkers are so called because an examination of the final compound reveals no trace of the point of linkage to the solid phase.



The first and most widely explored of this traceless linker is the silyl linker. Silicon attached to a phenyl group can undergo a protodesilylation reaction cleaving the silicon – aryl bond when treated with acid.

Advantages and Disadvantages

The ease of this isolation is one of the biggest advantages of solid phase synthesis. It makes it particularly useful for multi-step reactions, as the intermediates resulting in each step can be isolated quickly by this method. Another related advantage is that since removal of unreacted reagents is possible, large excesses can be used to drive the reaction to completion. However, optimal reaction conditions for solid phase synthesis can be difficult to determine, and developing these are far more time consuming than the actual reactions will be. The attachment to the resin puts limits on the chemistry that can be performed in this phase. Assessment of the purity of the resin attached intermediates is also difficult.  The best options today are NMR or FTIR. Purifying the final product after cleavage from the resin also proves to be a challenge.

The use of solution phase synthesis is on the rise, with multiple techniques available. The biggest challenge with solution phase synthesis is isolation of the product, and ways to automate this.  Some progress has been made in automating liquid-liquid extraction.  However, several new techniques have been introduced to make the extraction process easier.  Ion exhange resins are currently in use.  These remove byproducts and therefore eliminate the need for an aqueous work up.  Fluorous phase-chemistry is another method which eases extraction of the final product.  The starting compound has a perflourinated group attached to it. After the reaction has been completed, the product can be isolated by using a fluorocarbon solvent, which the compound will preferentially by extracted by. Removal of the perfluorinated group gives the desired final product

Q.12 With the help of case study explain how product-based library designing is helpful in drug discovery study.

Product-based library design involves a more complex optimization procedure that we term ‘combinatorial optimization’ where the reagent selection is optimized against the properties of the corresponding products. In this scheme, the combinatorial nature of the sub-library is maintained through combinatorial constraints whereas evaluation of diversity, focusing or other criteria is performed on the products. Procedures using either a Monte Carlo optimization or genetic algorithms have been reported. These procedures are more computationally intensive than simpler reagent based considerations because conventional techniques require full enumeration of the products, descriptor calculation for the entire library and optimization of the subset . The combinatorial optimization process described above attempts to identify a selection of reagents, which provides the desired product properties If we take the example of a virtual library consisting of a 100 x 100 x 100 array (1 000 000 possible products) for R1 x R2 x R3 and seek to isolate subsets of 10 x 10 x 10 (1000 com-pounds), the total number of possible solutions is C10010 x C10010 x C10010 = 5 x 1039.

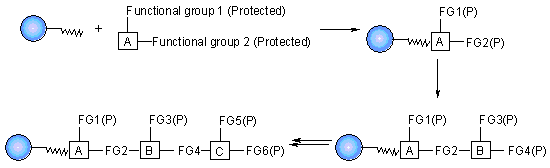
This is a formidable number, which makes it generally impossible to systematically investigate every possible subset. So we rely on the optimization procedure to provide a near optimal solution. Related studies using genetic algorithms or the Monte-Carlo procedure described above suggest that the subsets obtained with such procedures are only slightly sub-optimal. Several conditions have been reported for the Monte Carlo optimization using a large number of steps at a given temperature or a simulated annealing procedure. We have previously ascertained the consistent quality of the subsets returned, although the solutions themselves may be different. Several molecular descriptors have been used in library design and in the characterization of molecular diversity or focusing. Similar descriptors to those described earlier in the reagent-based design section are typically used in the analysis of products. Cell-based approaches typically provide fast measurements of product space coverage. However, a low dimensionality space is required. Several techniques are routinely used for dealing with the high dimensional problems. In the case of physicochemical descriptors, principal component analysis is routinely applied. In the case of fingerprint descriptors (MDL ISIS Keys, Daylight Fingerprints or BCI Fingerprints), diversity metrics such as those counting the number of on-bits can successfully be applied. For focused designs, the Tanimoto distance to a given lead compound can be used as a basis for optimization. Pharmacophore fingerprints have also been applied in library design scenarios. Whereas early computational methods for library subsetting concentrated on either diversity or similarity to known leads, other aspects such as drug-likeness are also part of the library-design process. In this fashion, we can produce libraries whose hits can be more easily optimized into successful drug candidates. This approach requires that we pursue several objectives simultaneously; diversity (or similarity) and drug-likeness for example. It has been shown that drug-likeness can be achieved with minimal impact on the diversity of the compounds selected. So we can obtain sets that are almost equally diverse but different in their drug-like character. Such an outcome is not entirely surprising when we look at the vast number of possible sub-libraries that can be generated. The large ensemble of solutions, at first perceived as a liability, may be turned into an asset where we now have the flexibility to provide libraries that are both drug-like and diverse.

Ideally, we would like to apply product-based design criteria for properties that are not adequately represented by the reagents (such as diversity and drug-like character) and reagent-based design criteria for practical considerations such availability, cost or ease of handling. For example, in addition to diversity and drug-like character, we would like to introduce a bias so that reagent selection is directed towards those preferred by chemists. Such a process involves simultaneous optimization against several criteria that need to be balanced appropriately. If the criteria involved in the optimization are normalized, then we can ensure the proper balance between them and reach an acceptable compromise solution. In this fashion, we can provide sub-libraries that are combinatorial, reasonably diverse, drug-like and use mostly ‘desirable’ reagents. A procedure involving such conditions has been reported recently. As mentioned earlier, product-based library design introduces considerable computational complexity, involving library enumeration, descriptor calculation and the optimization procedure itself. The latter can be performed rather efficiently so only the former two steps remain as bottlenecks. Novel approaches have appeared in the literature involving partial enumeration, stochastic sampling and GA-based optimization of subsets. For descriptor calculation, Markush-based representations have been used to handle a number of 2D descriptors with order of magnitude speedup compared to conventional approaches .We recently experimented with an ‘on the fly optimization’ workflow. This modified workflow combines the optimization of subsets using Monte-Carlo or GA with a Markush-based descriptor calculation. Results suggest that suitable solutions may be obtained while sampling only a very small fraction of the complete virtual library, and confirm earlier findings. These breakthroughs make product-based library design more efficient, especially when dealing with very large virtual libraries.

Q.13 Describe various types of solid support as well as linkers used in solid phase combinatorial chemistry study.

Since [Merrifield](http://en.wikipedia.org/wiki/Bruce_Merrifield) pioneered solid phase synthesis back in 1963, work, which earns him a Nobel Prize, the subject, has changed radically. Merrifield’s Solid Phase synthesis concept, first developed for biopolymer, has spread in every field where organic synthesis is involved. Many laboratories and companies focused on the development of technologies and chemistry suitable to SPS. This resulted in the spectacular outburst of combinatorial chemistry, which profoundly changed the approach for new drugs, new catalyst or new natural discovery.

The use of solid support for organic synthesis relies on three interconnected requirements:



1) A cross linked, insoluble polymeric material that is inert to the condition of synthesis;  
2) Some means of linking the substrate to this solid phase that permits selective cleavage of some or all of the product from the solid support during synthesis for analysis of the extent of reaction(s), and ultimately to give the final product of interest;  
3) A chemical protection strategy to allow selective protection and deprotection of reactive groups.

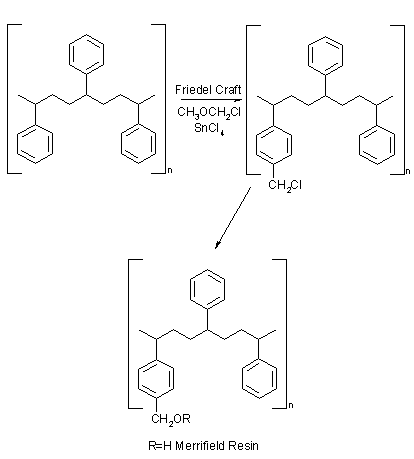
Merrifield developed a series of chemical reactions that can be used to synthesise proteins. The direction of synthesis is opposite to that used in the cell. The intended carboxy terminal amino acid is anchored to a solid support. Then, the next amino acid is coupled to the first one. In order to prevent further chain growth at this point, the amino acid, which is added, has its amino group blocked. After the coupling step, the block is removed from the primary amino group and the coupling reaction is repeated with the next amino acid. The process continues until the peptide or protein is completed. Then, the molecule is cleaved from the solid support and any groups protecting amino acid side chains are removed. Finally, the peptide or protein is purified to remove partial products and products containing errors.

**Resins for Solid Phase Synthesis**

In solid phase support synthesis, the solid support is generally based on a polystyrene resin. The most commonly used resin supports for SPS include spherical beads of lightly cross linked gel type polystyrene (1–2% divinylbenzene) and poly(styrene-oxyethylene) graft copolymers which are functionalised to allow attachment of linkers and substrate molecules. Each of these materials has advantages and disadvantages depending on the particular application.

**Crosslinked Polystyrene**

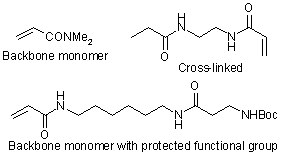
Lightly cross-linked gel type polystyrene (GPS) (Figure) has been most widely used due to its common availability and inexpensive cost. GPS beads which are functionalised with chloromethyl-, aminomethyl-, and a variety of linkers are commercially available from a variety of sources. A prominent characteristic of GPS beads is their ability to absorb large relative volumes of certain organic solvents (swelling). This swelling causes a phase change of the bead from a solid to a solvent-swollen gel, and therefore, the reactive sites are accessed by diffusion of reactants through a solvent-swollen gel network. In solvents, which swell the polymer well, the gel network consists of mostly solvent with only a small fraction of the total mass being polymer backbone. This allows relatively rapid diffusional access of reagents to reactive sites within the swollen bead. In solvents, which do not swell the polymer, the cross-linked network does not expand and the diffusion of reagents into the interior of the bead is impeded.



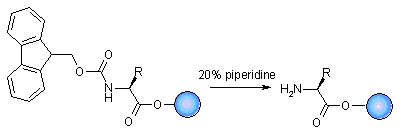
GPS has good swelling characteristics in solvents of low to medium polarity ranging from aliphatic hydrocarbons to dichloromethane. Polar, protic solvents, such as alcohols and water, do not swell GPS resins, and accessibility to all reaction sites may be compromised. Hence GPS supports are most suitable for chemistry performed in solvents of low to medium polarity.

**Polyamide Resins**

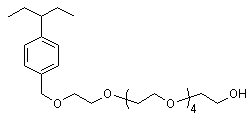
Sheppard designed polyacrylamide polymers for peptide synthesis as it was expected that these polymers would more closely mimic the properties of the peptide chains themselves and have greatly improved solvation properties in polar, aprotic solvents (e.g. DMF, or N-methyl pyrrolidinone).



Sheppard also proposed the use of a new protection and linking strategy. The Merrifield approach depended on a benzyl ester linkage and Boc protection. But a more mild protection/deprotection were sought. The protecting group finally chosen was the fluorenylmethoxycarbonyl (Fmoc) which can be removed by base (usually piperidine).



Poly (styrene-oxyethylene) graft copolymers, first reported by Bayer and Rapp, are another class of widely used supports for organic synthesis. As with the polyacrylamide resins, in order to produce a polar reaction milieu that is closer to the solvents generally used by solution synthetic chemists, grafted polymer beads have been prepared. The most pre-eminent of these is TentaGel resin which consists of polyethylene glycol attached to cross-linked polystyrene through an ether link, and combines the benefits of the soluble polyethylene glycol support with the insolubility and handling characteristics of the polystyrene bead. The resin was originally prepared by the polymerisation of ethylene oxide on cross-linked polystyrene already derives with tetraethylene glycol to give polyethylene glycol chains. (Figure) Poly (styrene-oxyethylene) graft copolymers beads display relatively uniform swelling in a variety of solvents from medium to high polarity ranging from toluene to water. The polymers are produced by grafting ethylene oxide from the polystyrene backbone creating long flexible chains that terminate with a reactive site spatially separated from the more rigid polystyrene backbone.

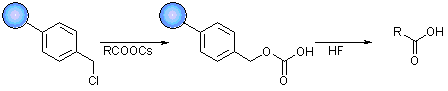


Some disadvantages of Poly(styrene-oxyethylene) graft copolymers supports are:  
- Relatively low functional group loading compared with GPS; the potential for the PEG chains to complex Lewis acids;  
- The potential instability of PEG;  
- The presence of linear PEG impurities found in the small molecule products after cleavage from the resin;  
- The tendency for resins to become sticky and difficult to handle as the synthesis progress.

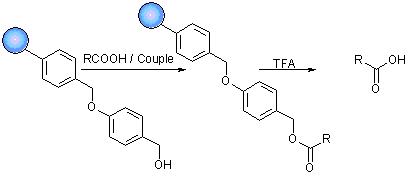
The group that joins the substrate to the resin bead is an essential part of solid phase synthesis. The linker is a specialised protecting group, in that much of the time, the linker will tie up a functional group, only for it to reappear at the end of the synthesis. The linker must not be affected by the chemistry used to modify or extend the attached compound. And finally the cleavage step should proceed readily and in a good yield. The best linker must allow attachment and cleavage in quantitative yield.

**Carboxylic acid linkers**

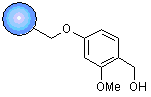
The first linking group used for peptide synthesis bears the name of the father of solid phase synthesis. Merrifield resin is cross-linked polystyrene functionalised with a chloromethyl group. The carbonyl group is attached by the nucleophilic displacement of the chloride with a cesium carboxylate salt in DMF. Cleavage to regenerate the carboxylic acid is usually achieved by hydrogen fluoride.



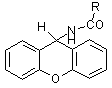
The second class of linker used for carboxylic acid is the Wang linker. This linker is generally attached to cross-linked polystyrene, TentaGel and polyacrylamide to form Wang resin. It was designed for the synthesis of peptide carboxylic acids using the Fmoc-protection strategy, and due to the activated benzyl alcohol design, the carboxylic acid product can be cleaved with TFA.



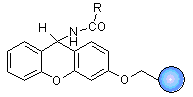
A more acid-labile form of the Wang resin has been developed. The SASRIN resin has the same structure as the Wang linker but with the addition of a methoxy group to stabilise the carbonium ion formed during acid catalysed cleavage.



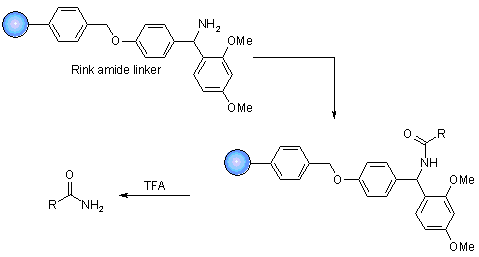
The Wang ester linker can be cleaved with ammonia to generate primary carboxamide, but this is a difficult reaction, that is very slow with sterically hundred amino acids such as valine. A prolonged treatment with ammonia could lead to a racemisation of chiral peptides. Chemists developed a linking group that would generate carboxamide in mild acidic conditions. The first developed was the methylbenzhydrylamine (MBHA) linker on polystyrene for improved synthesis of peptides using the Boc protection strategy. Sieber developed a new linker with greater acid lability. He used a xanthenyl derivate. The acid-labile 9-xanthenyl group has been used to protect the amide group of asparagine and glutamine.



As the xanthenyl group is less acid-labile than Boc, an additional –OCH2- group was introduced between the anchor and polystyrene to increase acid-lability.

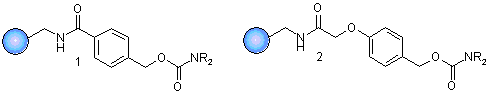


An other acid-labile linker was developed in the same time. The rink linker is now preferred for generating primary carboxamide on solid phase. The greater acid sensitivity in this linker is a consequence of the two additional electron donating methoxy group. In the generation of primary carboxamide, the starting material is attached to the linker as a carboxylic acid and after synthetic modification is cleaved from the resin with TFA.

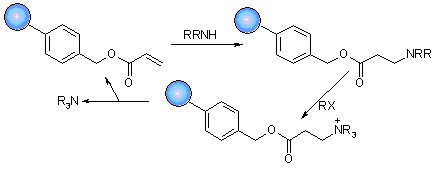


The use of Rink resin to produce carboxamide following TFA-catalysed cleavage.

Carbamates linker has been used for the synthesis of a combinatorial library of 576 polyamines prepared in the search of inhibitors of trypanosomal parasitic infections. Two linkers were investigated. One based on hydroxymethylbenzoic acid 1, and the other one, an electron-donating group has been added 2. The last one allowed cleavage by TFA while the first one could be cleaved with strong acidic conditions.



A very useful linker has been recently developed for the generation of tertiary amine. (Tertiary amines are commonly used in drug molecules.) Primary and secondary amines are introduced to the linker by Michael addition. The amine may be alkylated to gives a resin-bound quaternary ammonium ion. In mildly basic condition, Hoffmann elimination occurs to give a tertiary amines of high purity.



Q.14 Explain Solution phase combinatorial chemistry method in detail.

Combinatorial compounds are created either in solution-phase or in solid-phase. At its simplest level, the solution-phase synthesis involves conducting solution­phase chemical reactions simultaneously preferably in well-ordered sets (arrays) of reaction vessels6 We can illustrate this by the preparation of a small array of amides. The process consists of placing a diiferent acid chloride An and amine Bn in each of a matrix reaction vessel (along with a tertiary amine to neutralize liberated hydrochloric acid); incubating to form the amide; performing a liquid­liquid extraction to remove excess reagents and evaporating the solvent and testing the crude amides directly in a biological assay.

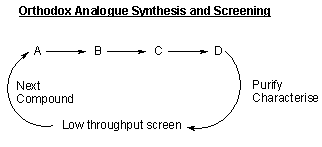
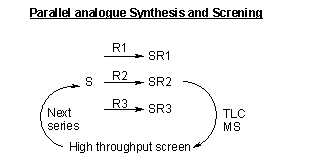
A reliable high-throughput assay is essential to successfully screen a combinatorial library. Both solid­phase and solution-phase assays have been developed for the combinatorial libraries. In the solid-phase assays, the ligands are still attached to tha solid support and the assays are two types: (i) direct binding of molecular target to the bead-bound ligand'2; 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 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) tha in situ-relea able solution-phase assay with immobilized beads. 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.

Despite the focus on the use of solid-phase techniques for the synthesis of combinatorial libraries, there have been few examples where libraries have successfully been made and screened in solution.

The benefit of preparing libraries on resin beads has been explained as offering advantages in handling, especially where a need to separate excess reagents from the reaction products is attached to the resin. In most of case a simple filtration effects a rapid purification and the product are ready to further synthetic transformation. But it should be remember that using solid phase chemistry brings several disadvantages as well. Clearly the range of chemistry available on solid phase is limited and it is difficult to monitor the progress of reaction when the substrate and product are attached to the solid phase.

Indeed some groups have expressed a preference for solution libraries because there is no prior requirement to develop workable solid phase coupling and linking techniques. The difficulty in purifying large number of compounds without sophisticated automated processes.

Manual or automated approaches can be used for the parallel preparation of tens to hundreds of analogues of a biologically active substrate. The products are synthesised using reliable coupling and functional group interconversion chemistry and are progressed to screening after removal of solvent and volatile by-products. Parallel and orthodox synthesis is compared below.

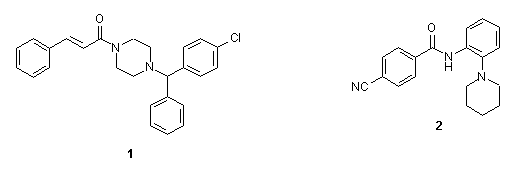
 

Orthodox synthesis usually involves a multistep sequence, e.g. from A through to the final product D, which is purified and fully characterised before screening. The next analogue is then designed, guided by the biological activity of the previous compound, prepared, and then screened. This process is repeated to optimise both activity and selectivity.

In contrast parallel analogue synthesis involves reaction of a substrate S with multiple reactants, R1, R2, R3 … Rn, to produce a compound library of n individual products SR1, SR2, SR3 … SRn. The library is screened, usually without purification, and with only minimal characterisation of the individual compounds, using a rapid throughput screening technique.

Panlabs have recently disclosed an interest in making large number of compounds as individual components using parallel, reliable solution chemistry. Reactions are pushed to completion by the use of excess quantities of the reactive reagent, and are isolated by solvent - solvent extraction. There is no further purification, and thus they prefer to describe these samples as "reaction products".

Two groups have recently disclosed solution libraries prepared in mixtures. In each case the groups from Glaxo and Pirrung have synthesised dimeric compounds using amide, ester or carbamate bond-forming reactions. Every library compound was prepared twice in mixtures of different composition. Testing all of these mixtures allows identification of likely active compounds without the need to resynthesise every compound in an active mixture.



In the glaxo example 40 acid chlorides were reacted with 40 amines or alcohols to gives amides or ester respectively in two sets. In the first set, each acid chloride (A) was reacted with a stoichiometric amount of an equimolar mixture of all 40 nucleophiles (N1-40). In the second set each amine or alcohol (N) was reacted with an equimolar mixture of the acid chlorides (A1-40). The 80 mixtures of 40 components each were screened against a wide variety of pharmacological targets, and a positive result from any sample identified half of the structure of a likely active dimeric compound. Weak leads against the neurokinin-3-receptor 1 and matrix metalloproteinase - 1 and 2 were detected.

Q.15 How does Library Enumeration method helpful in Combinatorial chemistry study. Explain with an appropriate example.

The analysis of a combinatorial library obviously requires computer

representations of the molecules contained within it. Given the large numbers

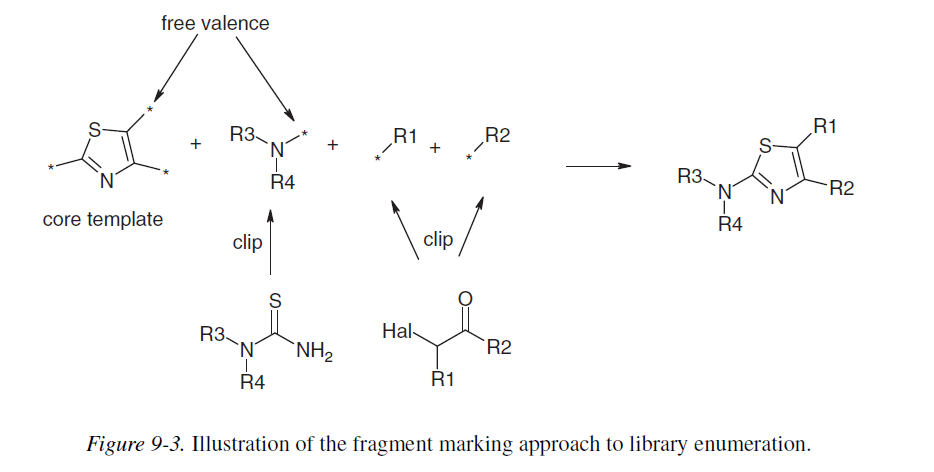
of compounds that are usually involved, it is not practical to input each product

molecule individually. *Enumeration* refers to the process by which the molecular

graphs (e.g. connection tables or SMILES strings) of the product molecules are

generated automatically from lists of reagents. Enumeration may be required for

both virtual libraries and for libraries that have actually been synthesised.



There are two general approaches to enumeration. The first of these is called

*fragment marking* and considers the library to be composed of a central core

template with one or more *R groups*. The core template must be common to

all members of the library; different product structures correspond to different

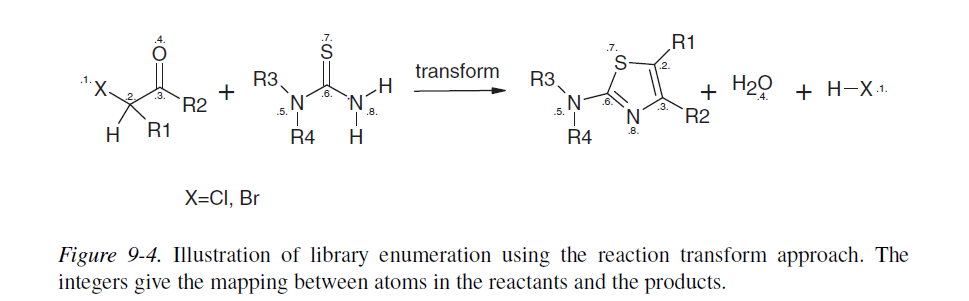
combinations of R groups. In order to enumerate the library, bonds are created

between the template and the appropriate R groups. The first task is to generate

the R group structures from the corresponding reagents. This is most easily

achieved by replacing the appropriate reactive group by a “free valence” to give a “clipped” reagent. This procedure is illustrated in Figure 9-3 for the enumeration of aminothiazoles from α-haloketones and thioureas in which a central core template corresponding to the thiazole ring is constructed with three attachment points. The three sets of clipped fragments would then be generated from the reagent structures for combination with the core template to produce complete product structures. The alternative to fragment marking is the reaction transform approach. C

Central to this method is the transform itself, which is a computer-readable representation of the reaction mechanism. The enumeration program takes as input the reagent structures themselves and applies the transform to generate the products. The reaction transform approach thus more closely mimics the actual synthetic process.



As can be seen from the example in Figure 9-4 it is necessary to ensure that the

transform is defined so that the appropriate atoms are matched between reagents

and products (i.e. according to the reaction mechanism). This may be achieved

using *atom mapping*, where pairs of corresponding atoms in the reagents and

products are labelled.

Both the fragment marking and the reaction transform approaches have

advantages and disadvantages. Fragment marking generally enables the

enumeration to be performed very quickly once the core template and the R

group fragments have been defined. It is particularly useful when the library

can be naturally defined as a “core plus R group” and when the reagents can

be easily clipped using some form of automated procedure. In other cases,

however, it can be difficult to define a common core and it may prove difficult

to generate the fragments automatically. Certain types of reactions (e.g. Diels–

Alder reactions) may not be handled properly by a simple fragment marker, giving

rise to extraneous and incorrect structures. The reaction transform method by

contrast more closely corresponds to the actual chemical synthesis and so it may

prove more intuitive to bench scientists, especially if delivered via an easy-touse

interface such as the web [Leach et al. 1999]. Some expertise is required

to compose efficient transforms, but once done they can be reused many times

without further alteration.

A more recent development is the use of Markush-based approaches to

enumeration, building upon research into the representation and search of the

Markush structures found in patents (see Chapter 1). This approach is ideally

suited to combinatorial libraries, particularly where a common core can be

identified. The Markush approach [Downs and Barnard 1997] recognises that

certain subsets of the product structures in the library may have features in

common (e.g. a subset of the R groups at a particular position may contain a

benzoic acid). The method constructs a compact computational representation of

the library from which the structures of the enumerated library can be generated

very rapidly.

Q.16 Discuss mixed combinatorial chemistry method and how it is different.

This technique was pioneered by Furka and co-workers in 1988 for the synthesis of large peptide libraries. This approach is termed divide couple and recombine synthesis by other workers.

In this technique, the starting material is split in ‘n’ portions, reacted with ‘n’ building blocks, and recombined in one flask. For the second step, this procedure is repeated. This method is particularly employed for solid phase synthesis.

In theory, if one wanted to make a library of the compounds A-B1, A-B2, A-B3, and A-B4, it could be accomplished by adding the reagents B1, B2, B3, and B4 to a solution of resin-bound A. All four of the desired compounds should be produced. However, there is a problem here. The reaction kinetics will not be the same for each reagent added. Therefore, the resulting mixture will not be equimolar, and will likely contain some dominating products. As well, individual beads will contain a mixture of products, not just a single one.  In combinatorial chemistry, equimolar mixtures of products and one product per bead are desired. The Split and Pool technique, also known as the Split and Mix technique, is the solution to this.

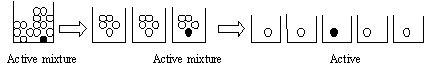
In this method, resins are divided equally into wells. The number of pools needed is dictated by the number of reagents they will be used during the reaction. As well, the overall number of beads used must be at least equal to the total number of compounds which will be generated by the process. Each individual well is reacted with a different reagent. The products from these reactions are isolated via filtration, and the pooled together to give an equimolar mixture. This mixture is then equally divided amongst the wells and reacted again with reagents. The process is repeated as many times as necessary. The end result is equimolar library of compounds representing all possible combinations of reagents. Only a very small amount of each product is produced (often a single bead represents one product), but only a small amount is needed for biological testing. When an active compound is discovered, it would need to be produced on a larger scale for further assays regardless. However, since the majority of compounds formed will not have any biological activity, there is no point than making any more of it than is needed to determine this.

A simple example of a 3 x 3 x 3 library gives all 27 possible combinations of trimeric products. X, Y and Z could be amino acids, in which case the final products would be tripeptides, but more generally they could be any type of monomeric unit or chemical precursor. It can be seen that the mix and split procedure finally gives three mixtures each consisting of nine compounds each, and there are several ways of progressing these compounds to biological screening. Although the compounds can be tested whilst still attached to the bead, a favoured method is to test the compounds as a mixture following cleavage from the solid phase. Activity in any given mixture reveals the partial structure of active compounds within the library, as the residue coupled last (usually the N-terminal residue) is unique to each mixture. Identification of the most active compound relies on deconvoluting the active mixtures in the library through further synthesis and screening.

In the example where the active structure is YXY, the mixture with Y at the terminal position will appear as the most active. Having retained samples of the intermediate dimers on resin (so-called "recursive" deconvolution ) addition of Y to each of the three mixtures will give all nine compounds with Y at the terminal position, and the second position defined by the mixture. The most active mixture here defines the middle position of the most active trimer to be residue X. Finally, the three individual compounds can be independently resynthesised and tested to reveal both the most potent compound and also some structure activity relationship data.

In contrast, Lam et al. tested a family of peptides whilst still attached to the resin bead solid phase. Nineteen amino, acids were incorporated into pentapeptides to generate a library of almost two and a half million compounds. By using a colourimetric assay, beads bearing peptide sequences that bound tightly to the protein streptavidin or to an antibody raised against β-endorphin were revealed by visual inspection. Bead picking using micromanipulation isolated the beads, and the active peptide structures were determined by microsequencing.

A modification of this method has allowed screening of such libraries in solution. Linkers have been devised that allow several copies of the library compounds to be released sequentially. Using this method it is possible to identify an active mixture using a solution assay, and then return to the beads that produced these compounds, and redistribute them into smaller mixtures for retest. By repeatedly reducing the mixture size, ultimately to single compounds, the bead containing the most potent sequence may be identified and the peptide product sequenced.



To use a standard synthetic route to produce a large variety of different analogues where each reaction vessel or tube contains a mixture of products

• The identities of the structures in each vessel are not known with certainty

• Useful for finding a lead compound

• Capable of synthesising large numbers of compounds quickly

• Each mixture is tested for activity as the mixture.

• Inactive mixtures are stored in combinatorial libraries

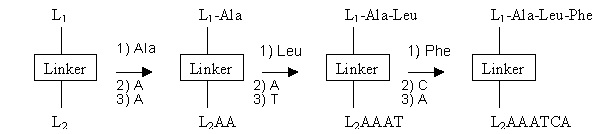
• Active mixtures are studied further to identify active component

**Advantages**:  
a) Only few reaction vessels required  
(b) Large libraries can be quickly generated (up to 105 compounds)   
**Disadvantages:**  
(a) Threefold amount of resin beads necessary  
(b) The amount of synthesized product is very small.  
(c) Complex mixtures are formed.   
(d) Deconvolution or tagging is required.  
(e) Synergistic effects may be observed during screening, leading to false positives.

Q.17 How does screening method help in detection of active compounds? (Use any 2 methods listed below)

a. DNA based encoding

One of the first reported successful ligand encoding strategies exploited oligo-deoxyribonucleic acid (DNA) as the surrogate analyte. This DNA encoding concept had in fact been demonstrated in some of the first combinatorial library preparation methods ever reported – those utilising filamentous phage particles. In this approach, libraries of peptides are prepared biochemically from the cloning and expression of random sequence oligonucleotides. Pools of oligonucleotides encoding the peptides of interest are interested into an appropriate expression system, where upon translation the resulting peptides are synthesised as fusion proteins. One of the common expression systems fuses these sequences to the gene III or the gene VIII coat protein of filamentous phage particles. Each viral particle contains a unique DNA sequence that encodes only a simple peptide. After screening a library in a given biological system, any viral particles displaying active peptides are isolated and the structure of the active peptides is elucidated by sequencing their encoding DNAs. A distinct disadvantage with this approach is that the molecular diversity of such systems is limited to peptides, and amino acids that compose these peptides are restricted to the 20 encoded by genes.



DNA encoded peptide prepared in a 1:1 correspondence on a linker capable of anchoring the synthesis of both oligomers. The structure of the peptides are determined by sequencing their accompanying unique DNA sequence.

b. Molecular Docking

When the structure of the target is available, usually from X-ray crystallography, the most commonly used virtual screening method is molecular docking. Molecular docking can also be used to test possible hypotheses before conducting costly laboratory experiments. Molecular docking programs predict how a drug candidate binds to a protein target. This software consists of two core components:

1. A search algorithm, sometimes called an optimisation algorithm. The search algorithm is responsible for finding the best conformations of the ligand, a small drug-like molecule and protein system. A conformation is the position and orientation of the ligand relative to the protein. In flexible docking the conformation also contains information about the internal flexible structure of the ligand – and in some cases about the internal flexible structure of the protein. Since the number of possible conformations is extremely large, it is not possible to test all of them. Therefore, sophisticated search techniques have to be applied. Examples of some commonly used methods are Genetic Algorithms and Monte Carlo simulations.

2. An evaluation function, sometimes called a score function. This is a function providing a measure of how strongly a given ligand will interact with a particular protein. Energy force fields are often used as evaluation functions. These force fields calculate the energy contribution from different terms such as the known electrostatic forces between the atoms in the ligand and in the protein, forces arising from deformation of the ligand, pure electron-shell repulsion between atoms and effect from the solvent in which the interaction takes place.

It is not possible to guarantee that the search algorithm will find the same solution as the true natural process, but more efficient search algorithms will be more likely to find the true solution if the evaluation functions properly reflect the natural processes.

Metaphorically, the active site of the protein can be viewed as a lock, and the ligand can be thought of as a key. Molecular docking is the process of testing whether a given key fits a particular lock. This description is slightly oversimplified due to the fact that neither the ligand nor the proteins are completely rigid structures. Their shapes are somewhat flexible and may adapt to each other.

c. Analytical Technique

The resin bead mix and split method can be used to generate hundreds, thousands or even millions of different products. As an example, a four-step synthesis employing 10 building blocks at each step would afford 10 000 different compounds in only 10\*4 chemical steps. Although synthesis is rapid, the power of combinatorial libraries is only evident if structural information on active components may be easily obtained. The iterative resynthesis and rescreening offers a solution, but as it can be slow and requires a further dedication of synthetic and screening resource, there have been a number of new methods devised where information concerning the active compound may be carried on the bead in the form of a "tag".

The synthetic efficiency of the split synthesis technique can be contrasted with the technical difficulties encountered when analysing the resulting libraries. For example, the simple split synthesis scenario outline above results in a library consisting of 10 pools of 1 000 compounds each. These compounds can be cleaved into solution and screened as soluble pools, or the ligands can remain attached to the beads and screened in immobilised form. Neither scenario is ideal for several reasons. Because of limitations on solubility, the concentration of the individual compounds present in soluble pools must be correspondingly diminished as the pool size increase – perhaps below a desirable threshold for screening. Biological screens performed on such large mixtures of soluble compounds can be ambiguous since the observed activity could be due to a single compound or due to a collection of compounds acting either collectively or synergistically. The subsequent identification of specific biologically active members is challenging, since the number of compounds present in the pools and their often-limited concentration deter their isolation and erase. Because of this, biologically active pools are often iteratively resynthesized and reassayed as increasingly smaller subsets until activity data are obtained on homogenous compounds.

This process of iterative resynthesis is time consuming, requires multiple bioassays, and the deconvolution of a single pool to its individual constituents typically require more synthetic step than were required to prepare the parent library. When multiple pool is active, the deconvolution process becomes additively complex if each active subset is chosen for resynthesis. In addition to begin inefficient, positive selection strategies such as iterative deconvolution ignore negative biological information, the knowledge of which is often important in the design of subsequent libraries.

In some instances, bead-based split synthesis libraries can be successfully assayed with the ligands still immobilised to the beads. In this process, a reporter system is employed in the biological assay such that beads displaying active ligands can be physically distinguished from those displaying inactive compounds. Suitable reporter system includes the use of fluorescently labelled receptors, or anti-receptors antibodies similarly labelled with a reporter molecule, that can be employed to "label" active beads. Beads thus marked are physically removed and analysed to identify the attached ligand. This technique is limited by the capacity of the biological screen to detect immobilised ligands, as well as the sensitivity of the analytical methods employed to unambiguously identify the attached compounds.

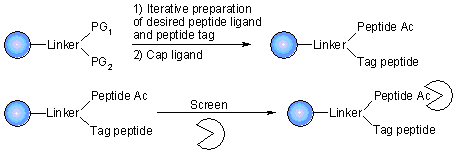
d. Mass Encoding **(after presentation need to be incorporated)**

The entire reported single bead encoding schemes require the cosynthesis of a suitable tagging moiety to record the synthetic history of each compound prepared in the library. This is inherently inefficient, since each unique compound could encode for itself if appropriate analytical techniques such as 1H, 13C NMR could be used to assign structures to ligands present in the amounts provided by single beads.

It can be seen that in each of these cases above, the use of a tagging group allows the synthesis of any type of compound within the library. The tagging molecules can encode for any building block and any synthetic transformation. Furthermore, given the uncertainties of much synthetic chemistry, the tag may be looked upon as not so much encoding a specific compound structure, but encoding instead a synthetic procedure. Thus, even if the intended compound was not made but biological activity was detected, the tagging system facilitates a replication of the synthetic steps employed in producing the active compound, and thus aids structure determination.

e. Peptide Tag

Zuckermann et al. at Chiron recognised that peptides could be employed as tag since their information content could be extracted with high sensitivity via Edman degradation and sequencing. Since the Edman degradation requires a free N-terminus, this peptide as code strategy could also be used to encode other peptide by acylating the N-terminus of the binding peptide strand and leaving a free amine at the coding peptide terminus. To accommodate the parallel synthesis of both binding and coding peptides, an orthogonally protected bifunctional linker was employed that contained both acid and base sensitive protecting groups. This bifunctional linker resided on the cleavable Rink amide linker, such that peptide-encoded peptide conjugates would be released into solution upon treatment of the Rink linker with 95% TFA.

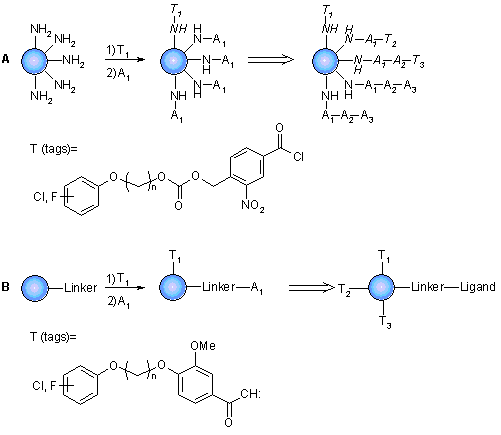


The ligand and its associated tag are synthesised on a 1:1 correspondence on a cleavable linker and realised into free solution. Affinity selection techniques are employed to isolate conjugates that bind to the receptor, enzyme, or antibody target of interest.

The above peptide and DNA encoding techniques are not ideal because of the chemical lability of these oligomers. This places a severe restriction on the scope of the synthetic techniques that may be applied during library synthesis and restricts the synthesis of more pharmaceutically attractive small organic molecules.

f. Hard Tag **(after presentation need to be incorporated)**

Still et al. reported the first encoding method utilising such chemically stable tagging moieties. The tag consisted of haloaromatic reagent linker to a carboxylic acid though an internal photochemically cleavable linker. Amide bond chemistry served to attach the tag to the beads. These haloaromatic reagents acylated the same synthesis sites used for ligand synthesis (Figure), but due to the sensitivity of tag detection this competition could be minimised. Once the haloaromatic analyte was attached to the bead it could be selectively detached into solution upon photolysis with UV light. The liberated tag could then be resolved and detected as subpicomole concentrations using electron capture capillary gas chromatography EC-GC.



Chemically "hard" haloaromatic tag suitable for encoding applications where the beads will be exposed to rigorous synthetic conditions the tag are released photochemically and then detected via EC-GC.  
**A**: haloaromatic tags incorporated via amine bond chemistry at the expense of the ligand synthesis sites.  
**B**: tags incorporated via carbene insertion.  
In both A & B tag concentration are minimised to prevent chemically derivation of the encoded ligands or the quenching of their synthesis sites.

While hard tagging strategies have been successfully used to encode a variety of different synthetic chemistries, a common limitation remains – the requirement for parallel synthesis (ligand and encoding tags). Since the robust preparation of a large combinatorial library is frequently a difficult synthesis challenge, it would be desirable to obviate the need for tag cosynthesis and instead delineate individual compounds by other physical means. Recent approaches that can replace or minimise the need to cosynthesise a surrogate are described in the following methods: Radio Frequency Encoding & Mass Encoding.

g. Radio Frequency Encoding

Radio frequency (RF) encoding techniques physically encapsulate an RF encodable microchip with the synthesis resin, such that the RF transponder can be scanned post-synthesis to identify its associated product. RF encoding successfully avoids the need to cosynthesise surrogate analytes, and also permits the larger scale synthesis of compounds since each microcapsule can hold tens milligrams of synthesis beads.

Q.18 Describe various applications of Combinatorial chemistry study.

Applications of combinatorial chemistry are very wide Scientists use combinatorial chemistry to create large populations of molecules that can be screened efficiently. By producing larger, more diverse compound libraries, companies increase the probability that they will find novel compounds of significant therapeutic and commercial value. Provides a stimulus for robot-controlled and immobilization strategies that allow high-thrughput and multiple parallel approaches to drug discovery.

Combinatorial chemistry is not refined completely yet – it is still developing.

1. The first stage is finding promising lead compounds, which is facilitated greatly with the compilation of the combinatorial libraries.  The library can also take the lead compound and derivatize it to give it pharmacological properties that are necessary for a potential drug.

2. The second step - lead development – is where combinatorial chemistry begins to fall short of ideal.  As the lead molecule becomes more complex, problems that can arise are difficulties with detaching the structurally diverse molecules from the solid support, unfavourable side reactions, and harmful effects that often occur in the harsh reaction conditions employed by combinatorial syntheses.  These problems magnify if the natural product happens to be sensitive and fragile.  Solutions to these conditions are being studied in combinatorial biocatalysis which employs enzymatic conditions and microbial transformations on an existing lead instead of the more harsh organic reaction conditions

Analytical methods, like GC/MS, LC/MS, MS/MS, etc. for small scale reactions are constantly improving so that our synthetic characterization is optimized.  Combinatorial method is also a relatively current innovation, and thus with practice comes perfect.  So, if combinatorial methods were not perceived to be valuable, then companies would not have made the investments that they have to advance combinatorial chemistry.

The combinatorial chemistry first shows its presence in synthesis of peptide libraries. The peptide plays varying role in body. By the use of combinatorial chemistry, we can generate vast peptide, which may be active. Biologically active peptide hormones play an important role in regulating a multitude of human physiological response and many low molecular weight bioactive peptides can act as a hormone receptor against or antagonists. In addition, peptide structure commonly is found in molecules designed to inhibit enzymes that catalyse proteolysis, phosphorylation and other past translational protein modification that may play important role in pathologies of various disease states.

----------------------------------------------------------------------