

## (UNIT 4)

### Introduction to Combinatorial Chemistry and Library Design

#### ❖ Combinatorial chemistry:

- 1) Combinatorial chemistry is a collection of techniques which allow for the synthesis of multiple compounds at the same time.
- 2) It 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.
- 3) It aims to mimic the natural sources to produce pool of chemicals out of which one of them may be proved as lead compound.

#### Principle:

- 1) The basic principle of combinatorial chemistry is, to prepare a large number of different compounds at the same time instead of synthesizing compounds in a conventional one at a time manner and then to identify the most promising compound for further development by high throughput screening.
- 2) The characteristics of combinatorial synthesis is that, different compounds are generated simultaneously under identical reaction conditions (i.e. using the same reaction conditions and the same reaction vessels.) in a systematic manner, so that ideally the products of all possible combinations of a given set of starting materials (termed building blocks) will be obtained at once.
- 3) The collection of these finally synthesized compounds is referred to as a combinatorial library.
- 4) The library is then screened for useful properties and the active compounds are identified.

#### The combinatorial chemistry approach has two phases:

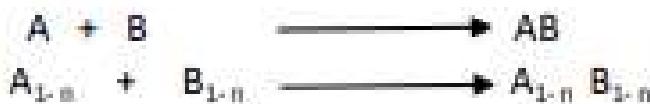
- i. Making a library.
  - ii. Finding the active compound. Screening mixtures for biological activity has been compared to finding a needle in a haystack
- 1) The combinatorial libraries can be structurally related by a central core structure, termed scaffold (i.e. all compounds of library have a common core structure), or by a common backbone. In both the cases, the accessible dissimilarities of compounds within the library depend on the building blocks which are used for the construction.
  - 2) 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.

- 3) By accelerating the process of chemical synthesis, this method is having a profound effect on all branches of chemistry, but especially on drug discovery.

**Example:**

- 1) In a conventional synthesis, one starting material A reacts with one reagent B resulting in one product AB.
- 2) In a combinatorial synthesis, building blocks of type A ( $A_1-A_n$ ) are treated simultaneously with different building blocks of type B ( $B_1-B_n$ ) according to combinatorial principles, each starting material A reacts separately with all reagents B resulting in a combinatorial library  $A_{1-n} \cdot B_{1-n}$ .
- 3) Therefore, in combinatorial approach one can cover many combinations  $A^n \times B^n$  in one reaction Instead of doing multiple  $A \times B$  type reactions.

**Conventional Reaction:**



**Combinatorial Chemistry:**



❖ **Need for Combinatorial Chemistry:**

- Earlier, there were problems with Traditional/Conventional Synthesis:
  - 1) The chemist would make only one molecule at a time.
  - 2) Each synthesis was very time consuming.
  - 3) Multistep synthesis has loss at each step.
  - 4) Purification of products very time-consuming between steps.
  - 5) Yields can be low
  - 6) This lead to the production of very few molecules at a time for testing which results in slower lead generation.
  - 7) Hundreds of molecules are generated in a month.
  - 8) There is a high risk of failure.
  - 9) Therefore, to reduce the time and cost, combinatorial chemistry approach was developed by the researchers in the pharmaceutical industry to aid in producing effective and competitive new drugs.
  - 10) One chemist would make multiple molecules at a time.
  - 11) The time and cost associated with the generation and analysis of each individual molecule is significantly less when compared to the time and cost of an individual synthesis.
  - 12) Yields can be high and produces many molecules at a time for testing.
  - 13) Thus leads to faster lead generation and thousands of molecules are generated in a month.
  - 14) Also, there was a low risk of failure.

❖ **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) The cost of combinatorial chemistry library generation and analysis 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) Yields can be high and produces many molecules at a time for testing. This leads to faster lead generation.
- 5) Thousands of molecules in a month are generated.
- 6) There is a low risk of failure.
- 7) Multiple molecules synthesized at a time.
- 8) Combinatorial chemistry speeds up the drug discovery process.

❖ **Disadvantages:**

- 1) Though combinatorial chemistry would solve all the problems associated with drug discovery, one still needs to synthesize the right compound.
- 2) 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.

❖ **History of Combinatorial chemistry:**

- 1) The origins of combinatorial chemistry can be traced back at least as far as 1963, when biochemistry professor R. Bruce Merrifield of Rockefeller University, New York City, developed a way to make peptides by solid-phase synthesis.
- 2) For his work on solid-phase synthesis, Bruce Merrifield won the Nobel Prize in chemistry in 1984 for his work on solid-phase synthesis.
- 3) During this time, automated peptide synthesizer technology was in its infancy, and the preparation of individual peptides was a challenge.
- 4) The field in its modern dimensions only began to take shape in the 1980s, when in 1984 research scientist H. Mario Geysen, now at Glaxo Wellcome, Research Triangle Park, N.C., developed a technique to synthesize arrays of peptides on pin-shaped solid supports and in 1985, Richard Houghten developed a technique for creating peptide libraries in tiny mesh "tea bags" by solid-phase parallel synthesis.

- 5) Another early pioneer was Dr. Árpád Furka who introduced the commonly used split-and-pool method in 1988, which is used to prepare millions of new peptides in only a couple of days and also for synthesizing organic libraries.
- 6) Through the 80's and into the early 1990's, combinatorial chemistry was focused on peptide synthesis and later oligonucleotide synthesis.
- 7) In the 1990s, the focus of the field changed predominantly to the synthesis of small, drug like organic compounds and many pharmaceutical companies and biotechnology firms now use it in their drug discovery efforts.

#### ❖ Pros and Cons of Combinatorial Chemistry:

<ul style="list-style-type: none"> <li><input type="checkbox"/> Creation of large libraries of molecules in a short time.</li> <li><input type="checkbox"/> Compounds that cannot be synthesized using traditional methods of medicinal chemistry done by Combi Chemistry.</li> <li><input type="checkbox"/> Cost 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.</li> <li><input type="checkbox"/> More opportunities to generate lead compounds.</li> <li><input type="checkbox"/> Combinatorial chemistry speeds up drug discovery</li> </ul>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Needs to synthesize the right compound.</li> <li><input type="checkbox"/> 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.</li> <li><input type="checkbox"/> Each reaction step has to be carefully planned, and often a reaction isn't available because the chemistry affects the resin.</li> <li><input type="checkbox"/> There is a great deal of diversity created, but not often a central synthetic idea in the libraries.</li> </ul>
---	---

#### ❖ Types of Combinatorial chemistry:

- 1) The range of combinatorial techniques is highly diverse, and these products could be made individually in parallel or in mixtures, using either solution or solid phase techniques.
- 2) Combinatorial chemistry is of two types:
  - i. Solid phase combinatorial chemistry (The compound library has been synthesized on solid phase such as resin bead).
  - ii. Solution phase combinatorial chemistry (The compound library has been synthesized in solvent in the reaction flask).

## **1) SOLID PHASE COMBINATORIAL CHEMISTRY:**

(Compound library synthesized on solid phase such as resin bead)

### **Steps:**

- 1) In solid phase combinatorial chemistry, the starting compound is attached to an inert solid/resin bead.
- 2) Reagents are added to the solution in excess.
- 3) Separation of products (attached to resin beads) by simple filtration.
- 4) Cleavage and isolated of products from the beads.

### **Requirements:**

- 1) A cross-linked insoluble polymeric support which is inert to the synthetic conditions (e.g. a resin bead);
- 2) An anchor or linker covalently linked to the resin—the anchor has a reactive functional group that can be used to attach a substrate;
- 3) A bond linking the substrate to the linker, which will be stable to the reaction conditions used in the synthesis;
- 4) A means of cleaving the product or the intermediates from the linker;
- 5) Protecting groups for functional groups not involved in the synthetic route.

### **Example of Solid supports:**

- 1) Partially cross-linked polystyrene beads: Polystyrene is cross linked with divinyl benzene, hydrophobic in nature, causes problems in peptide synthesis due to peptide folding.
- 2) Sheppard's polyamide resin – more polar
- 3) Tentagel resin- similar environment to ether
- 4) Beads, pins and functionalized glass surfaces.

### **Characteristics of Solid supports:**

- 1) Beads must be able to swell in the solvent used, and remain stable.
- 2) Most reactions occur in the bead interior.

### **Advantages:**

- 1) Since, the reaction is carried out on a solid support such as resin beads, a range of different starting materials are available that can be bound to separate resin beads, which are mixed together, such that all the starting material can be treated with another reagent in a single experiment. Therefore, it is possible to do multi-step synthesis and mix-and split synthesis (a technique used to make large number of libraries).

- 2) Since, the products are bound to solid support, excess reagents or by-products can be easily removed by washing with appropriate solvent. Hence, large excesses of reagents can be used to drive reactions to completion.
- 3) Intermediates in a reaction sequences are bound to the bead and need not be purified.
- 4) Individual beads can be separated at the end of reaction to get individual products.
- 5) Polymeric support can be regenerated and reutilized if appropriate cleavage conditions and suitable anchor/linker group are chosen.
- 6) Automation is possible.

**Disadvantages:**

- Not all synthesis can be done on solid phase.
  - 1) Some molecules don't attach well to beads.
  - 2) Some chemistry just doesn't work in this fashion.
  - 3) Removal of product from bead, can be damaging to product if not careful.
- Typically, kinetics is not the same.
  - 1) Reaction rates can be slower.
- It is difficult to monitor the progress of reaction when the substrate and product are attached to the solid phase.
- Assessment of the purity of the resin attached intermediates is also difficult.
- Purifying the final product after cleavage from the resin also proves to be a challenge.



## 2) SOLUTION PHASE COMBINATORIAL CHEMISTRY:

(Compound library synthesized in solvent in the reaction flask)

- 1) All chemical reactions are conducted simultaneously, preferably in well-ordered sets (arrays) of reaction vessels in solution.
- 2) Soluble polymer are used as support for the product.
- 3) Most ordinary synthetic chemistry takes place in solution phase.
- 4) The use of solution phase techniques has been explored as an alternative to solid-phase chemistry approaches for the preparation of arrays of compounds in the drug discovery process.

**Advantages:**

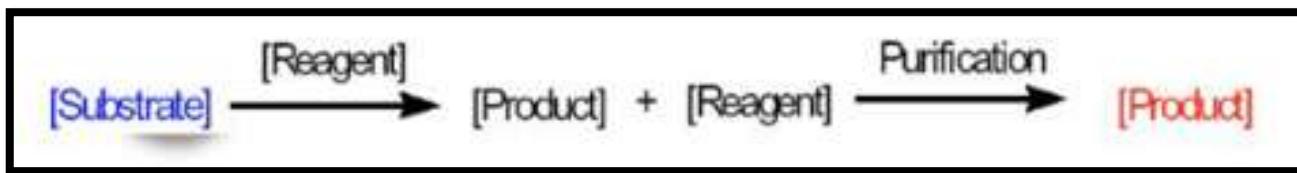
- 1) Handling of material is easy and can be automated.

#### **Disadvantages:**

- 1) Solution phase work is free from some of the constraints of solid-phase approaches but has disadvantages with respect to purification.
- 2) In solution phase synthesis we use soluble polymer as support for the product.
- 3) PEG is a common vehicle which is used in solution phase synthesis it can be liquid or solid at room temperature and show varying degrees of solubility in aqueous and organic solvent.
- 4) By converting one OH group of PEG to methyl ether (MeO-PEG-OH) it is possible to attached a carboxylic acid to the free OH and use in solution phase combinatorial synthesis.
- 5) Another common support which is used in solution phase synthesis is liquid Teflon consisting mainly of long chain of (-CF<sub>2</sub> -) groups attached to a silicon atom. When these phases are used as a soluble support for synthesis the resulting product can be easily separated from any organic solvent.
- 6) The main disadvantage of this method is when number of reagents are taken together in a solution, it can result in several side reactions and may lead to polymerization giving a tarry mass.

#### **Limitations:**

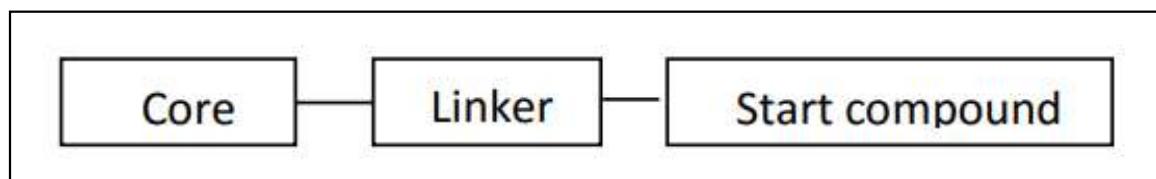
- 1) When numbers of reagents are taken together in a solution.
  - It can result in several side reactions
  - It leads to polymerization giving a tarry mass.



#### **❖ Resin (Solid support) used in Solid phase synthesis:**

- 1) Most solid state combinatorial chemistry is conducted by using polymer beads ranging from 10 to 750 µm in diameter.
- 2) The solid support must have the following characteristics for an efficient solid phase synthesis:
  - a. Physical stability and of the right dimensions to allow for liquid handling and filtration;
  - b. Chemical inertness to all reagents involved in the synthesis;
  - c. An ability to swell under reaction conditions to allow permeation of solvents and reagents to the reactive sites within the resin;
  - d. Derivatization with functional groups to allow for the covalent attachment of an appropriate linker or first monomeric unit.

- 3) The solid supports are usually composed of two parts: the core and the linker.
- 4) The starting Compound of the synthesis is attached to the support via the linker.



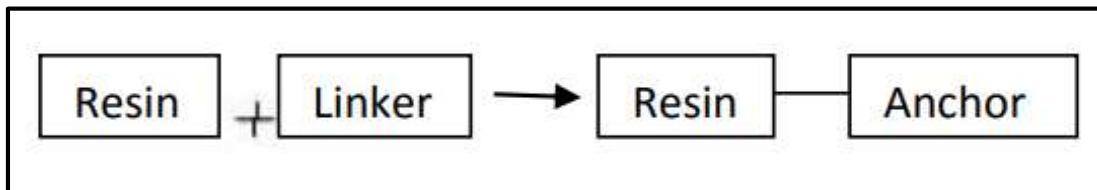
- 5) The compounds to be synthesized are not attached directly to the polymer molecules but attached by using a **linker moiety** that enables attachment in a way that can be easily reversed without destroying the molecule that is being synthesized and allow some room for rotational freedom of the molecules attach to the polymer.
- 6) The core ensures the insolubility of the support and determines the swelling properties, while the linker provides the functional group for attachment of the start compound and determines the reaction conditions for the cleavage of the product.
- 7) The linker itself and the covalent bond formed with the start compound must be stable under the reaction conditions of the synthesis.
- 8) The bead should be capable of swelling in solvent, yet remain stable.
- 9) Swelling is important because most of the reactions involved in Solid Phase Synthesis takes place in the interior of the bead rather on the surface.
- 10) Although beads are the common shape for the solid support, a range of other shapes such as pins have been designed to maximize the surface area available for reaction and hence maximize the amount of compound linked to the solid support.
- 11) Functionalized glass surfaces have also been used and are suitable for oligonucleotide synthesis

#### ❖ Types of Solid supports used:

- 1) **Polystyrene resins:** Polystyrene is cross linked with divinyl benzene (about 1% cross linking). Polystyrene resin is suitable for non-polar solvents.
- 2) **Tenta Gel resins:** Polystyrene in which some of the phenyl groups have polyethylene glycol (PEG) groups attached in the para position. The free OH containing resins are suitable for use in polar solvents.**Poly acrylamide resins:** This resin swell better in polar solvent, since they contain amide bonds that more closely resemble biological materials.
- 3) **Glass and ceramic beads:** This type of solid supports is used when high temperature and high pressure reaction are carried out.

#### ❖ Linkers / anchors used in solid phase synthesis:

- 1) The initial building block of the compound to be prepared by solid phase synthesis is covalently attached to the solid support via the linker.
- 2) A molecular moiety which is covalently attached to the solid support and which contains a reactive functional group.
- 3) Allows attachment of the first reactant.
- 4) The bond formed between the linker and substrate must be stable to the reaction conditions used throughout the synthesis and it should be easily cleaved to release the final compound after the synthesis is completed.
- 5) It is a bi-functional molecule, one functional group for irreversible attachment to the core resin and a second functional group for forming a reversible covalent bond with the initial building block of the product and the linker remains after cleavage at the resin.
- 6) Different linkers are available depending on the functional group to be attached and the desired functional group on the product.



- 7) A series of selected examples are found below. Resins are named to define the linker e.g.,
  - **Merrifield resin:** The Merrifield resin can be used to attach carboxylic acids to the resin. The product can be cleaved from the resin in carboxylic acid form using HF.
  - **Wang resin:** The resin is used to bind carboxylic acids. The ester linkage formed has a good stability during the solid phase reactions but its cleavage conditions are milder than that of the Merrifield resin. Usually 95% TFA is applied. It is frequently used in peptide synthesis.
  - **Rink resin** The Rink resin is designed to bind carboxylic acids and cleave the product in carboxamide form under mild conditions. The amino group in the resin is usually present in protected form.
  - **Hydroxymethyl resin:** The resin can be applied for attachment of activated carboxylic acids and the cleavage conditions resemble that of the Merrifield resin.
  - **Photolabile anchors:** Photolabile anchors have been developed that allow cleavage of the product from the support by irradiation without using any chemical reagents. Such anchors, like the 2-nitrobenzhydrylamine resin below, usually contain nitro group that absorbs UV light.

→ **Traceless anchors:** The initial building block of a multi-step solid phase synthesis needs to have one functional group (in addition to others) for its attachment to the solid support. It may happen that in the end product, this group is unnecessary and needs to be removed. For this reason anchors have been developed that can be cleaved without leaving any functionality in the end product at the cleavage site. These traceless anchors usually contain silicon based linkers.

❖ **Protecting groups used in Solid phase synthesis:**

- 1) Primary function of protecting group is to protect the portion of the molecule that is not covalently bound to the resin and must be protected to avoid subsequent polymerization of excess monomers in solution.
- 2) A protecting group is reversibly attached to the functional group to convert it to a less reactive form.
- 3) When the protection is no longer needed, the protecting group is cleaved and the original functionality is restored.
- 4) A large number of protecting groups were developed for use in peptide synthesis since the amino acids are multifunctional compounds.
- 5) It is an important requirement for a protecting group to be stable under the expected reaction conditions of each coupling.
- 6) After coupling is performed, the protecting group is removed to expose a new reactive site and synthesis continues in a repetitive fashion.
- 7) Cleavage conditions are dictated by the linker used.
- 8) Two protecting groups are said to be orthogonal if either of them can be removed without affecting the stability of the other one.
- 9) Some of the protecting groups most widely used in peptide synthesis are described below. Protection of amino groups:
  - a) Benzyl carbonyl (Z) group.
  - b) t-butoxy carbonyl (Boc) group.
  - c) 9-fluorenyl methoxy carbonyl (9-Fmoc) group.

❖ **Characteristics of solid phase and solution phase combinatorial chemistry:**

SOLID PHASE	SOLUTION PHASE
Make a mixture of products	Makes only one product
Small amounts of products formed	Large amounts of products formed
Simple isolation of product by filtration	Work-up and purification more difficult
Requires two extra reaction steps: linkage & cleavage	No extra steps for attachment & cleavage needed
Limits to chemistry which can be performed	Wide range of reactions can be utilized
Automation possible	Automation difficult
Large excesses of reagent can be used to drive the reaction to completion	Large excesses of reagent cannot be used as it causes subsequent separation problem.
Longer reaction time than in solution phase	Less reaction time
Monitoring of reaction very difficult	Monitoring of reaction easy
Split and mix technique as well as parallel synthesis can be applied	Split and mix strategy not possible Parallel synthesis can be applied

❖ DIFFERENCE BETWEEN SOLID PHASE AND SOLUTION PHASE SYNTHESIS:

On a solid support	In solution
Reagents can be used in excess in order to drive the reaction to completion	Reagents cannot be used in excess, unless addition purification is carried out
Purification is easy: simply wash the support	Purification can be difficult
Automation is easy	Automation is difficult
Fewer suitable reactions	In theory any organic reaction can be used
Scale-up relatively expensive	Scale-up is easy and relatively inexpensive
Not well documented and time will be required to find a suitable support and linker for a specific synthesis	Only requires time for the development of the chemistry

#### ❖ TYPES OF COMBINATORIAL LIBRARIES:

##### 1) Scaffold-based Libraries:

- Core-structure, which is common to all compounds of the library.
- Several single building blocks can consist of Scaffold.
- Example- Amino acid and Amino Benzophenone.

##### 2) Backbone-based Libraries

- Example- Nucleic acid and Carbohydrate.

- 2 approaches to generate libraries are as follows:

##### 1) Random/diverse libraries:

- Synthesis of diverse compounds – large number of molecules – more hits (biological assay).
- Little is known about the target – more diverse library (primary screening library).

##### 2) Focused libraries:

- Synthesis of focused compounds – small number of molecules.
- Incorporate as much information about the therapeutic target as possible.

#### ❖ TECHNIQUES FOR LIBRARY PREPARATION:

1) There are two methods, which used for synthesis of compounds in combinatorial chemistry.

2) They include:

- a) Split and mix synthesis or Split and pool synthesis or Portioning – Mixing (PM) synthesis (one bead-one compound library).
- b) Parallel synthesis (one vessel-one compound library)

#### A. SPLIT AND MIX SYNTHESIS ‘OR’ PORTIONING– MIXING SYNTHESIS: (One bead-one compound library)

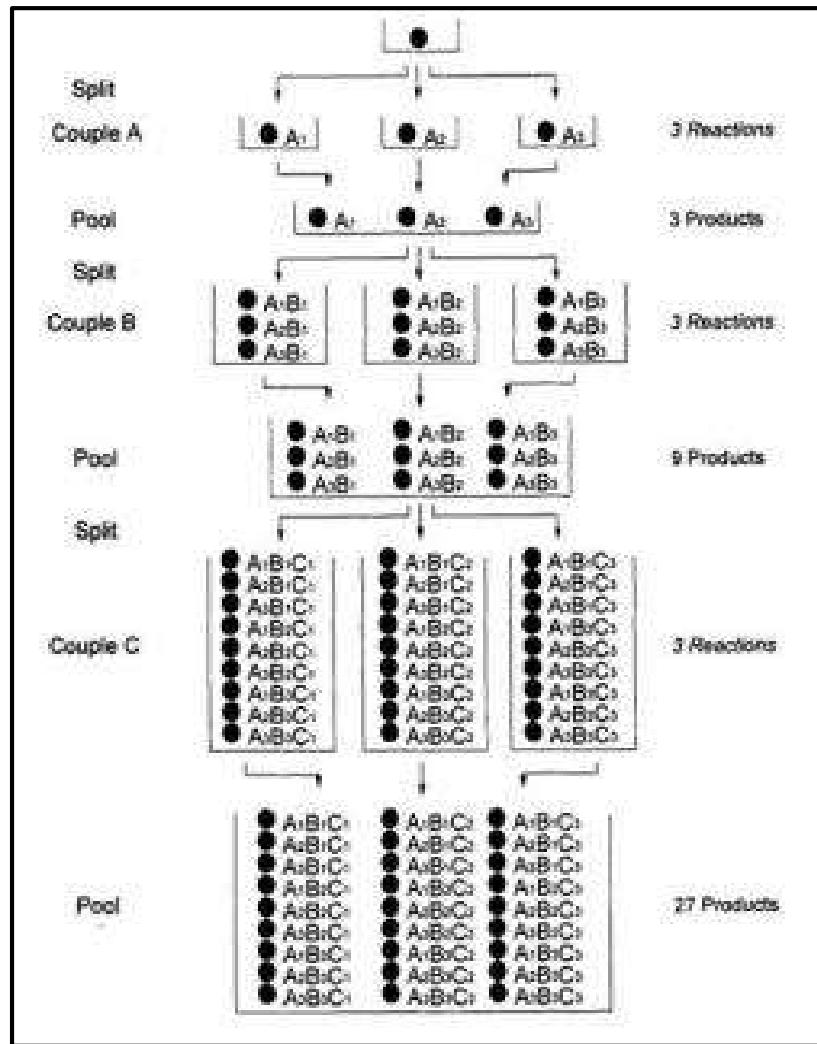
- 1) This technique was pioneered by Dr. Árpád Furka and co-workers in 1988 for the synthesis of large peptide libraries.

2) This approach was termed divide couple and recombines synthesis by other workers.

**Steps:**

- 1) In this method, ingredients are assembled on the surface of the beads or micro particles.
- 2) In each step, beads from last steps are partitioned into new building block and several groups are added.
- 3) This leads to the formation of new groups, the different groups of beads are recombined and separated once again.
- 4) Process is continuous with next building block is added until the desired library has been assembled.
- 5) After a Split-Pool synthesis, just one single compound is bound to each resin bead.
- 6) Split-Pool procedure requires a solid support.
- 7) Therefore, this method is particularly employed for solid phase synthesis.

**Example:** In following figure spheres represents resin beads, A, B & C represent the sets of building block and borders represents the reaction vessels. In the case, when three building blocks are used, in each coupling step after three stages (ie. divide, couple & recombine), a total number of 27 different compounds, one on each resin bead, are formed using 9 individual reactions (ignoring deprotection). On the resulting products from split and pool synthesis, bioassays is performed and active mixture is discovered. Once an active mixture has been discovered, the next task is discovering which individual compound(s) in that mixture are active. The process of determining these active compounds is known as deconvolution.



### Advantages:

- Only few reaction vessels required.
- Large libraries can be quickly generated (up to  $10^5$  compounds).

### Disadvantages:

- Threefold amount of resin beads necessary.
- The amount of synthesized product is very small.
- Complex mixtures are formed.

## B. PARALLEL SYNTHESIS:

(One vessel-one compound library)

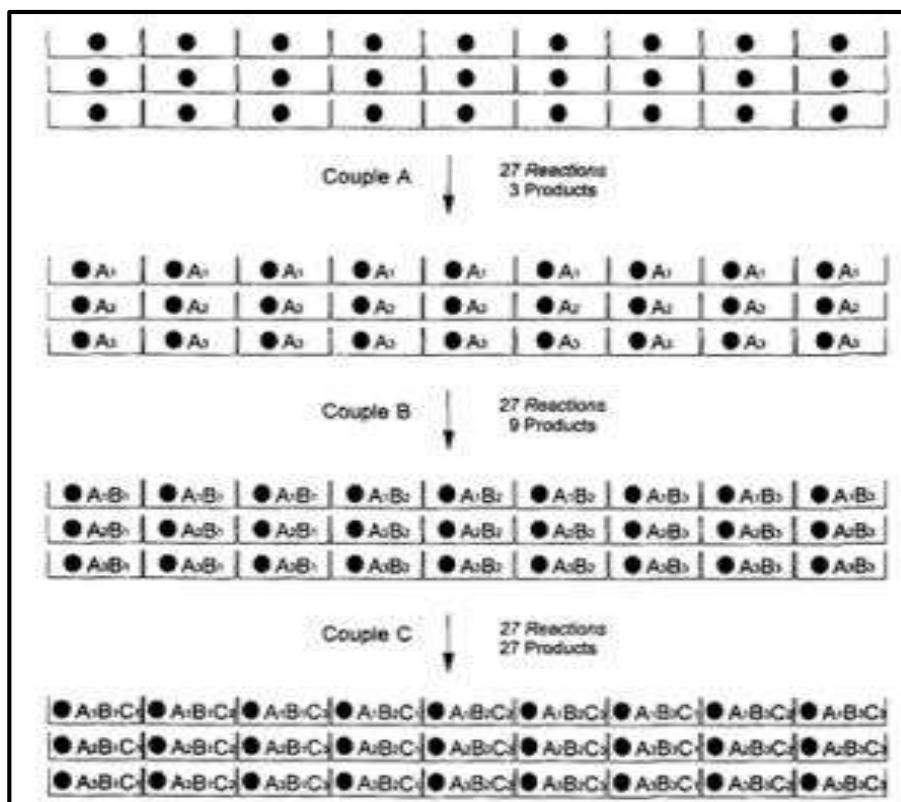
- It involves multiple reactions, at once instead of in series, each in a separate vessel.
- A single product is obtained in each different reaction vessel.

### Steps:

- Each compound is synthesized in specific reaction vessel.

- 2) Each starting material is reacted with each building block separately.
- 3) Then, product is spilt into portions, reacted with different building block separately again.
- 3) Methods of parallel synthesis include Houghton's tea bag procedure and Automated parallel synthesis.

**Example:** In following figure spheres represents resin beads, A, B & C represent the sets of building block and borders represents the reaction vessels. In the case, when three building blocks are used, in each coupling step after three stages, a total number of 27 different compounds, one on each resin bead, are formed using 9 individual reactions (ignoring deprotection).



#### Advantages:

- a) It creates the compounds individually and in their own vessel. Thus the identity of the product is already known.
- b) No deconvolution is required.
- c) Each compound is substantially pure in its location.
- d) Defined location provides the structure of a certain compound.
- e) Biological evaluation is easy.

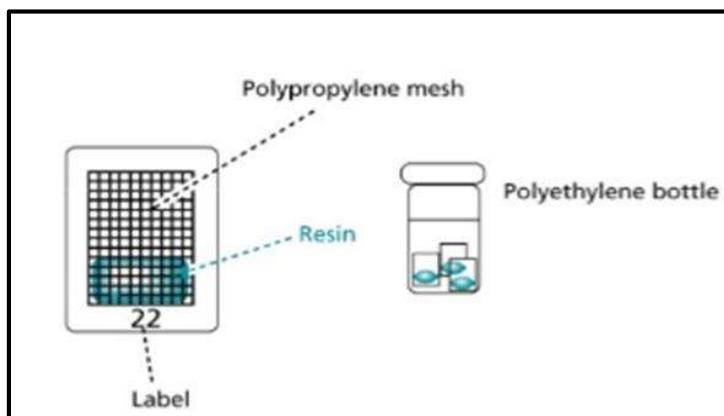
#### Disadvantage:

- a) Applicable only for medium libraries (several thousand compounds).

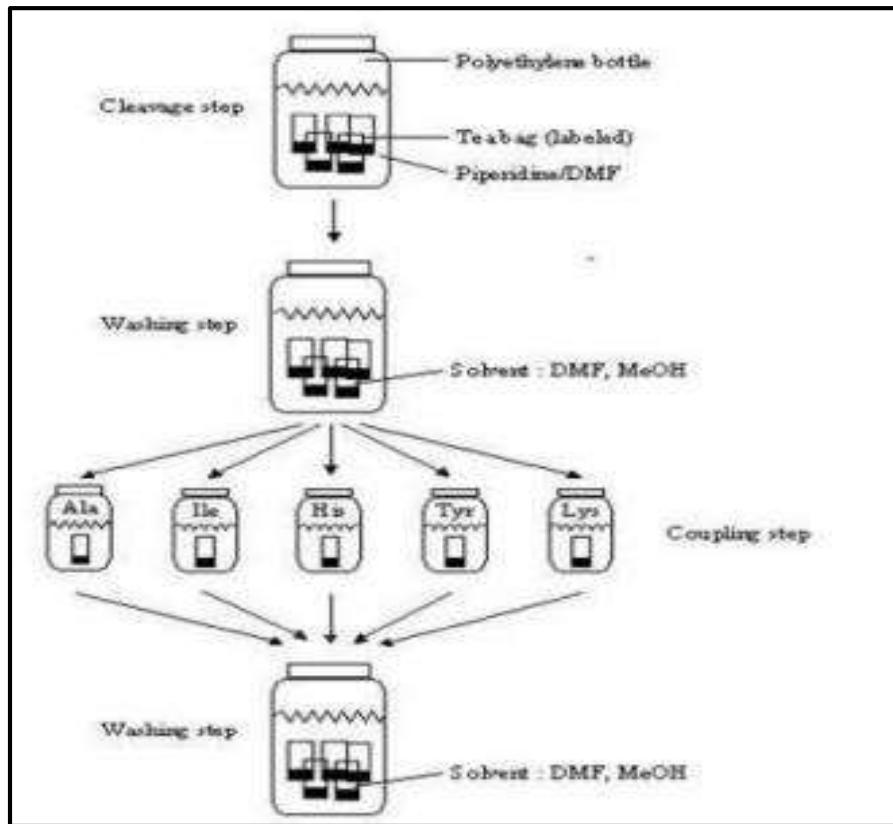
- b) Large amount of vessels are required.
- c) Large number of reactions is to be performed.

❖ **METHOD FOR PARALLEL SYNTHESIS:**

- **Houghton Teabag method:**



- 1) **Definition:** A polypropylene mesh bag, with dimensions of approximately 15 x 20 mm, filled with resin beads, sealed and labeled for a later identification, is known as a tea-bag, designed by Houghten in 1985.
- 2) The “tea-bag” mesh size is too small to allow resin beads to escape, but solvents and soluble reagents could readily enter.
- 3) 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.
- 4) It is a manual approach to parallel synthesis.
- 5) To save time and work while making many peptides simultaneously, bags could be combined into the same reactors for common chemical steps.
- 6) **Example:**
- 7) 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.
- 8) Then the bags are sorted into groups for the addition of the next amino acid.
- 9) Then the bags could be combined again for deprotection, washing, and neutralization.
- 10) After an appropriate number coupling steps, all the bags can then be treated with HF/anisole to cleave the peptides from the beads.
- 11) 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.
- 12) **Schematic overview of a typical group of steps carried out using the tea-bag procedure:**



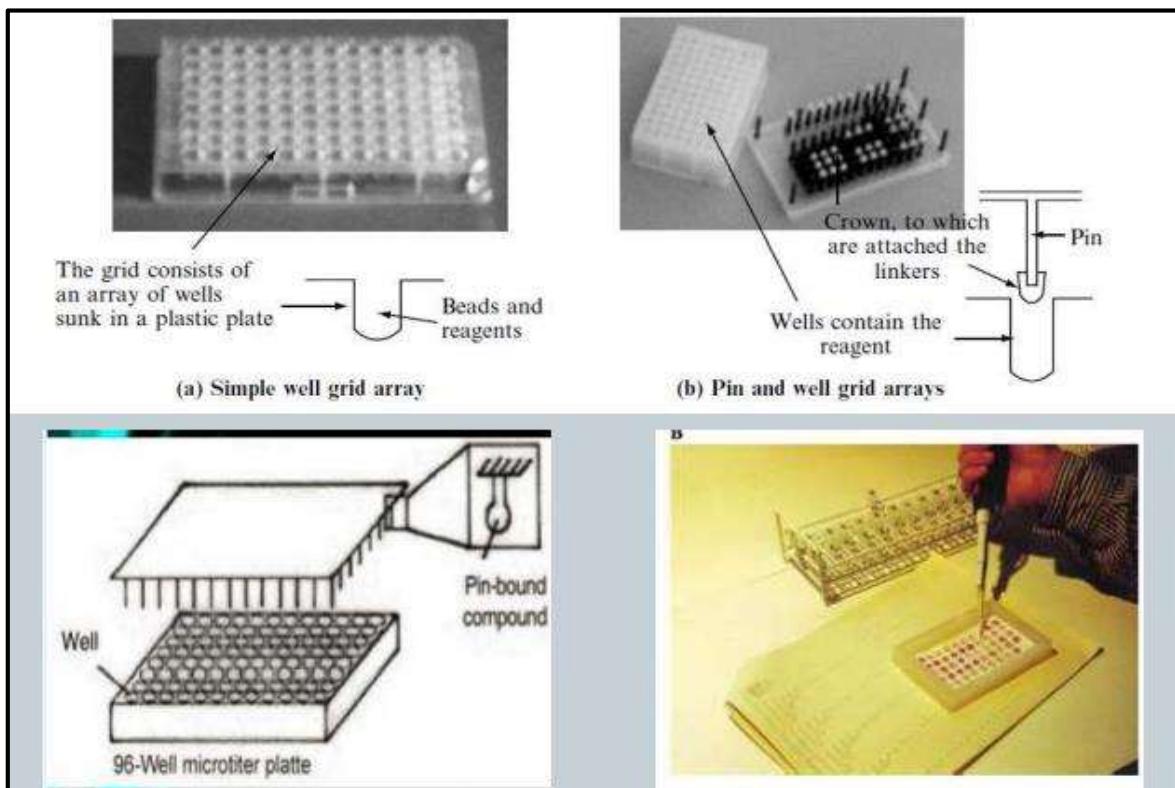
#### → Advantages –

- Easy to identify active hit as its position (X, Y coordinate) in the array encodes the reagents and thus structure of the product.
- New equipment, such as 'personal synthesizers' and 'multi vial apparatus,' allows parallel synthesis of many compounds simply and quickly by one chemist.
- Robotized technology.
- Greater quantity of each compound is available at once (structural characterization).
- Labeling of the tea bags leads to easier identification of each compound.

#### → Disadvantages –

- Maximum impurities can occur unless the reactions are very clean.
- Most useful for one to three step reactions only.
- Can only be used for making smaller (more focused) libraries.

- **Automated Parallel synthesis:**



- 1) Parallel synthesis is possible when it advances in automation.
- 2) Automated synthesizers are available with 42, 96, 144 reaction vessels or wells.
- 3) Beads or pins are used for solid support.
- 4) Reactions and work ups are carried out automatically.
- 5) Same synthetic route for each vessel, but different reagents.
- 6) Different product obtained per vessel.

**Steps:**

- 1) In this method, each starting material is reacted with each building block separately (i.e. in separate vessel), without remixing.
- 2) This is not like a split synthesis because it requires a solid support.
- 3) It can be done without solid support or in a solution.
- 4) A 96 well micro titer plate is commonly used format for parallel synthesis.
- 5) After each reaction step the product is split into 'n' portions before it is reacted with n new building blocks.
- 6) In following figure spheres represent resin beads, A, B & C represent the sets of building block and borders represents the reaction vessels. In the case, when three building blocks are used, in

each coupling step after three stages, a total number of 27 different compounds, one on each resin bead, are formed using 9 individual reactions (ignoring deprotection).

- 7) Like the split and pool method, it results in the production of multiple compounds at the same time.
- 8) However, unlike split and pool, parallel synthesis gives individual compounds, not a mixture. Thus deconvolution is not an issue in this method.

### C. **MIXED COMBINATORIAL SYNTHESIS:**

- 1) The aim is 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.
- 2) The identities of the structures in each vessel are not known with certainty.
- 3) It is useful for finding a lead compound.
- 4) It is capable of synthesizing large numbers of compounds quickly.
- 5) Each mixture is tested for activity as the mixture.
- 6) Inactive mixtures are stored in combinatorial libraries.
- 7) Inactive mixtures are studied further to identify active component.

#### ❖ **Screening of Combinatorial Library:**

- 1) Can be done in 2 ways: Virtual screening and Experimental real screening.

##### **Virtual screening:**

- 2) Virtual screening uses computational methods to predict or simulate how a particular compound interacts with a given target protein.
- 3) The 3 virtual screening methods used in modern drug discovery include
  - Molecular Docking,
  - Pharmacophore Mapping
  - QSAR/QSPR
- 4) Disadvantages of virtual screening:
  - Cannot replace real screening
  - Generated hits may be very difficult to chemically synthesize

##### **Experimental real screening:**

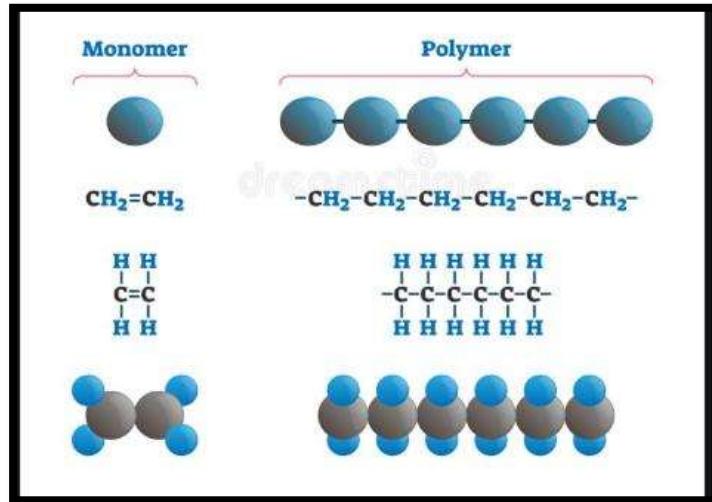
- 5) Real screening approaches, such as high-throughput screening (HTS), can test the activity of hundreds of thousands of compounds experimentally, providing real results;
- 6) Disadvantage:
  - These methods are far more expensive and
  - Slower than virtual screening methods.

- 7) Most common assay to screen a combinatorial library is to determine the binding of the library compounds to the target protein.
  - 8) Other common assays are functional assays such as biochemical and enzymatic assays, or cell-based assays.
  - 9) Cell-based assays can be direct cytotoxic assays, receptor-binding assays, or cell-signaling assays using cell lines with specific genetic reporter systems.
  - 10) Selection of screening methods greatly depends on:
    - The nature of the combinatorial libraries to be screened.
    - Position-addressable soluble libraries prepared from parallel synthesis can be screened with automated HTS methods in 96-, 384-, and 1536-well plates.
    - Libraries on solid supports (e.g. OBOC library) can be easily screened against a variety of biological targets (proteins, cells, viruses, etc.) for binding or functional activities or released in situ for solution phase functional assays.
  - 11) Phage-display peptide libraries can be screened with bio-panning or limited cell-based functional assays, such as cell-binding and cellular uptake assays.
  - 12) Structure-based virtual libraries are screened in silico.
- ❖ **Applications of Combinatorial Chemistry:**
- 1) Application of combinatorial library methods in cancer research and drug discovery
  - 2) Building synthetic gene circuits from combinatorial libraries: screening and selection strategies
  - 3) Combinatorial library approaches for improving soluble protein expression in *Escherichia coli*
  - 4) Combinatorial library-based strategies to optimize proteins
  - 5) A Combinatorial Library Strategy for the Rapid Humanization of Anticarcinoma BR96 Fab
  - 6) Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery.
  - 7) Used in anti-viral research.

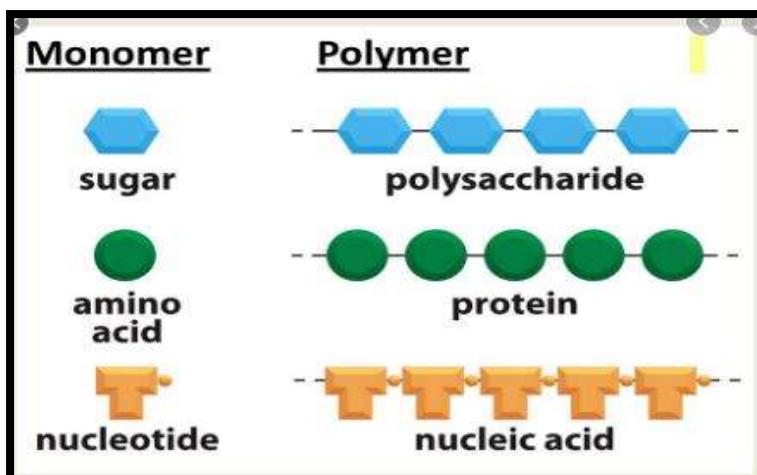
#### ❖ Strategies for Library Designing:

## Two main strategies:

### **1. Monomer-based selection:**



- 1) The small individual repeating units/molecules are known as monomers (means single part).
  - 2) Imagine that a monomer can be represented by the letter A. Then a polymer made of that monomer would have the structure.

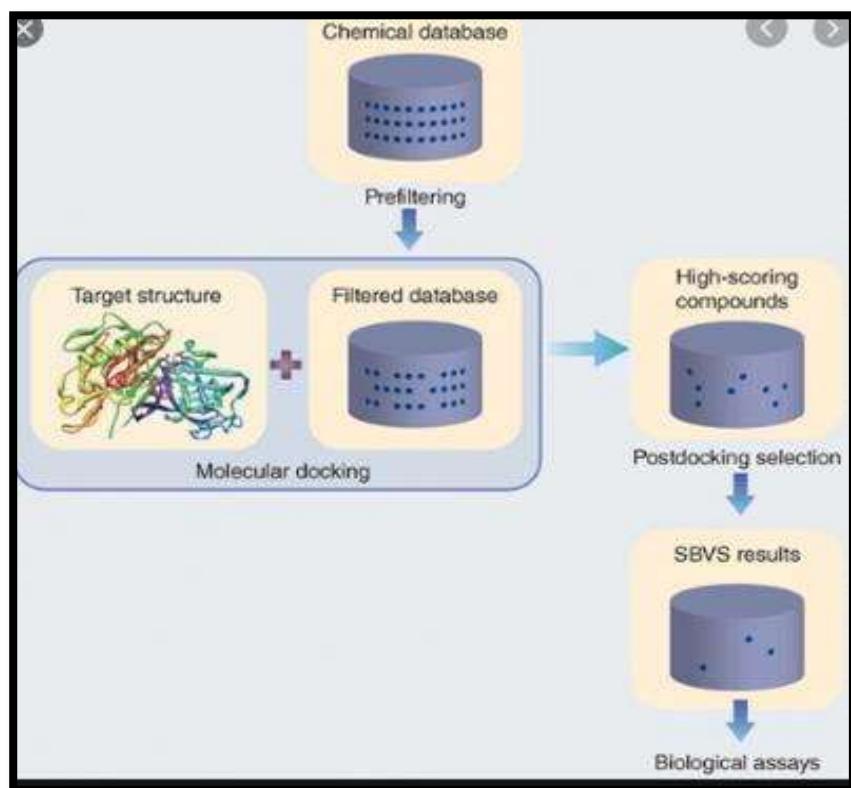


- 3) In monomer-based selection optimized subsets of monomers are selected without consideration of the products that will result.
  - 4) Consider a hypothetical three-component library with 100 monomers available at each position of variability, where the aim is to synthesize a diverse  $10 \times 10 \times 10$  combinatorial library.

- 5) In monomer-based selection this would involve selecting the 10 most diverse monomers from each set of monomers i.e. there are subsets of size  $n$  contained within a larger set of  $N$  compounds.
- 6) Eg: more than  $10^{13}$  different subsets of size 10 from a pool of 100 monomers.
- 7) It is not possible to examine all of these.
- 8) The subset selection problem can be solved in the context of selecting compounds for screening where the techniques of dissimilarity-based compound selection, clustering and partitioning were introduced, together with related optimization methods.

$$\frac{N!}{n!(N-n)!}$$

## 2. Product-based selection:



- 1) 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.
- 2) In product-based selection, the properties of the resulting product molecules are taken into account when selecting the monomers.

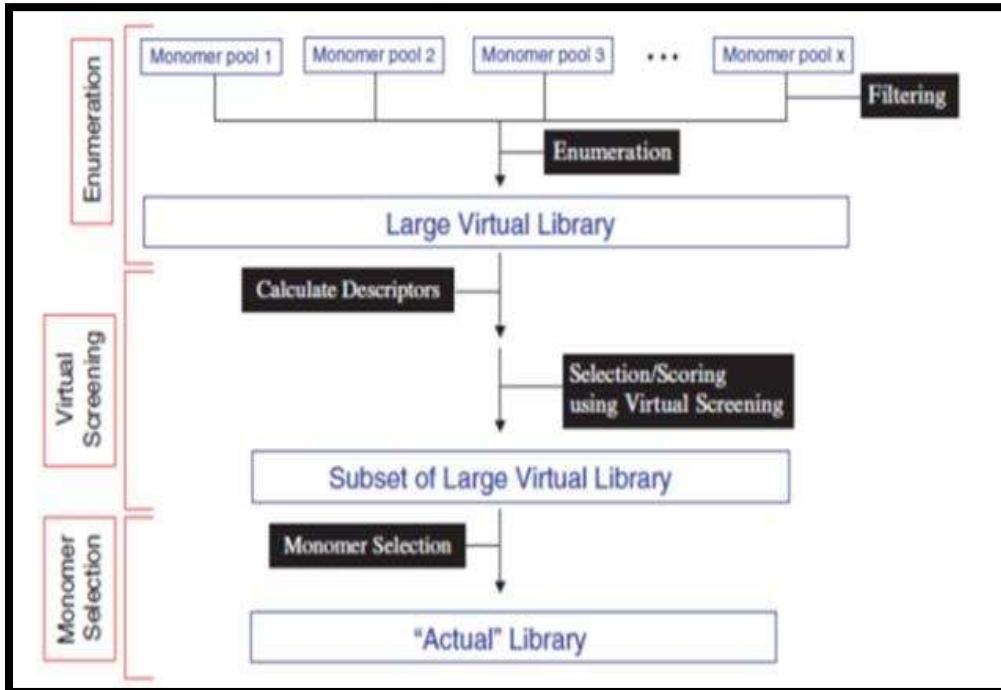
- 3) Having enumerated the virtual library any of the subset selection methods could then be applied.
- 4) This process is generally referred to as cherry-picking but it is synthetically inefficient in so far as combinatorial synthesis is concerned.
- 5) Synthetic efficiency is maximized by taking the 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.
- 6) It is much more computationally demanding than monomer-based selection, but can be more effective while optimizing the properties of a library as a whole.
- 7) 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  $n_i$  monomers to be selected from a possible  $N_i$  at each substitution position.
- 8) Thus, there are almost 1040 different  $10 \times 10 \times 10$  libraries that could be synthesized from a  $100 \times 100 \times 100$  virtual library.
- 9) The selection of combinatorial subsets has been tackled using optimization techniques such as simulated annealing and genetic algorithms.
- 10) 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 optimize the properties of a library as a whole, such as diversity or the distribution of physicochemical properties.

$$\prod_{i=1}^R \frac{N_i!}{n_i!(N_i - n_i)!}$$

### **Approaches to Product-based Library Design:**

- A general strategy for product-based library design involves the following 3 steps. They are as follows:

- 1) Lists of potential reagents are identified (e.g., by searching relevant databases), filtered them as needed, and the virtual library is enumerated.
- 2) The virtual library is subjected to virtual screening to evaluate and score each of the structures.
- 3) 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, degree of similarity or dissimilarity to existing collections).



- 4) 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.
- 5) The final, monomer selection stage is typically implemented using optimization techniques such as GAs or simulated annealing.
- 6) 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.
- 7) The chromosome of the GA thus contains  $nA + nB$  elements, each position specifying one possible monomer.
- 8) Then, 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 maximize this quantity.
- 9) In some cases the virtual library is too large to allow full enumeration and descriptor calculation, making product-based combinatorial subset selection unfeasible.
- 10) A number of methods have been proposed to try to overcome this problem.
- 11) Alternative approaches to product-based library design have been developed that do not require enumeration of the entire virtual library.
- 12) 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 focused libraries.

- 13) The molecule-based method is a relatively fast procedure, especially when optimization 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.
- 14) In these approaches, however, there is no guarantee that building libraries from frequently occurring monomers will result in optimized libraries, nor is it possible to optimize properties of the library as a whole.

❖ **Encoding Combinatorial Libraries:**

- 1) By the process of screening the number of libraries that has “desirable properties” are sorted out.
- 2) It is now very important to learn the identity of “winning” library member.
- 3) The process of identification of active compound in a mixture of compounds is known as Encoding.
- 4) Chemical structure of individual compounds in conventional addressable combinatorial libraries or planar microarray libraries are known, there is no need to encode and decode the chemical hits.
- 5) For mixture libraries in solution, such as positional-scanning libraries, purification is needed to determine the identity of the hits.
- 6) Biological-displayed peptide libraries (e.g., phage, yeast or mRNA-display) are genetically encoded and can be decoded with PCR, DNA barcoding, DNA sequencing, Edman microsequencing , NGS, mass spectroscopy of released coding tags, fluorescence-based encoding method ,etc..
- 7) More than one million codes can be generated by using combinations of different methods, which are highly stable and reliable under bioassay conditions.
- 8) For identification of active compound following types of encoding methods are used:
  - a. **Positional encoding or deconvolution (iterative resynthesis and rescreening).**
  - b. **Chemical encoding (Tagging)**
  - c. **Electronic encoding**
    - a. **Positional encoding or deconvolution (iterative resynthesis and rescreening):**
      - ➔ In this method, the resynthesis and rescreening is carried out to know the identity of the active compound. In other terms, it is a process of optimizing an activity of interest by fractionating (normally by resynthesis, or by elaborating a partial library) a pool with some level of the desired activity to give a set of smaller pools.
      - ➔ Repeating this strategy leads to single members with (ideally) a high level of activity and is termed iterative deconvolution.
    - b. **Chemical encoding (Tagging):**
      - ➔ The most common approach to encoding solid phase libraries is to attach a chemical tag to the resin beads as the target molecule gets synthesized.

- ➔ Typically, at each step in the reaction, a tag is attached that is unique for the given step.
- ➔ For example, if we are creating a tripeptide and we have 10 possible amino acids at each position, we need to attach either a single tag that says “the tripeptide on this bead has amino acid Ala at position 1, Phe at position 2 and Gly at position 3” or we need to attach three different tags, one for each position.

**c. Electronic encoding:**

- ➔ This technique uses a micro electronic device called a radio frequency (RF) memory tag.
- ➔ The tag measuring  $13 \times 3$  mm is encased in heavy walled glass and contains the following:
  - A silicon chip ,onto which laser etched a binary code,
  - A rectifying circuit with which absorbed RF energy is converted to D.C. electrical energy,
  - A transmitter/receiver circuit,
  - An antenna, through which energy is received and RF signals are both received and sent.

**6) Library Enumeration:**

- ➔ Process by which the molecular graphs of the product molecules are generated automatically from lists of reagents (using connection tables or SMILES strings).

**1) Fragment marking**

- Central core template and one or more R groups.

**2) Reaction transform approach**

- Transform is a computer-readable representation of the reaction mechanism: atom mapping.

**Advantages / Disadvantages:**

- Fragment marking generally a very fast enumeration once core template and R group fragments are defined.
- May be difficult to generate the core and to generate fragments automatically.

**3) Markush-based approaches to enumeration:**

- Ideally suited when a common core can be identified.
- Certain subsets of the product structures may have features in common.

**7) Identification of Active Ingredient Major Challenge in developing library of compounds:**

- ➔ Major challenge in developing library of compounds is screening the library for the activity of the chemical species responsible.
- ➔ The goal of producing molecular libraries is to discover compounds that have some desired properties to serve as a drug.

## **1. Analytical techniques:**

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 re-synthesis 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 analyzing 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 immobilized 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 reuse. Because of this, biologically active pools are often iteratively re-synthesized and re-assayed as increasingly smaller subsets until activity data are obtained on homogenous compounds.

In some instances, bead-based split synthesis libraries can be successfully assayed with the ligands still immobilized 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 labeled receptors, or anti-receptors antibodies similarly labeled with a reporter molecule, that can be employed to "label" active beads.

Beads thus marked are physically removed and analyzed to identify the attached ligand.

This technique is limited by the capacity of the biological screen to detect immobilized ligands, as well as the sensitivity of the analytical methods employed to unambiguously identify the attached compounds.

## **2. 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 utilizing 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 inserted into an appropriate expression system, where upon translation the resulting peptides are synthesized 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 is determined by sequencing their accompanying unique DNA sequence.

## **3. Mass encoding:**

The entire reported single bead encoding schemes require the co-synthesis 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  $^1\text{H}$ ,  $^{13}\text{C}$  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.

## **4. Peptide tag:**

It has been 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 synthesized on a 1:1 correspondence on a cleavable linker and realized 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.

##### **5. Hard tag (Haloaromatic tag):**

the first encoding method utilizing 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.

#### **6. 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.

---