# Combinatorial Chemistry and library design

# History

# HISTORICAL DEVELOPMENT

- 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 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.
- For his work on solid-phase synthesis, Bruce Merrifield won the Nobel Prize in chemistry in 1984.
- During this time, automated peptide synthesizer technology was in its infancy, and the preparation of individual peptides was a challenge.
- 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 in and in 1985 Richard Houghten developed a technique for creating peptide libraries in tiny mesh "tea bags" by solid-phase parallel synthesis.

# CONT'D

- 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.
- Through the 80's and into the early 1990's, combinatorial chemistry was focused on peptide synthesis and later oligonucleotide synthesis.
- In the 1990s, the focus of the field changed predominantly to the synthesis of small, drug like Organic com-pounds and many pharmaceutical companies and biotechnology firms now use it in their drug discovery efforts.

# INTRODUCTION

# Introduction

- Combinatorial chemistry has significantly increased the number of molecules that can be synthesized in a modern chemical laboratory.

---

- The invention of combinatorial chemistry in the late 1980s and early 1990s led to the development of a wide variety of automated methods for chemical synthesis in both industrial and academic laboratories.

- Methods range from complex robots designed (synthesis of large combinatorial libraries) to “low-tech” equipment that enables basic functions such as heating or separation to be applied to a small number of samples.

- Common feature : They enable tasks previously applied on a molecule-by-molecule basis to be applied to many molecules simultaneously, greatly increasing the rate at which new chemical entities can be made.

- Compound and library design strategies have played an equally important role in chemoinformatics.

# CONT'D

- Combinatorial chemistry is a laboratory technique in which millions of molecular constructions can be synthesized and tested for biological activity.
- It has generated massive numbers of targeted molecules for testing and the developing techniques of high throughput screening has automated the screening process so larger numbers of biological assays can be done.
- All this together has reduced the discovery-to-market time from what used to be 10-14 years to 5-8 years.
- In a traditional organic synthesis lab, the chemist does the standard reaction

A + B --> C.

# CONT'D

- But with combinatorial chemistry **A is a mixture** of perhaps **5 components** and **B is a mixture** of **10** so instead of getting **one product** the chemist now **gets 50.**
- This collection of molecules synthesized so rapidly is referred to as libraries.
- The library allows for the input of a single generic structure, like benzodiazepine, with the output giving a complete range of benzodiazepine derivatives all in a specified molecular weight range which differ in the composition of the different R groups (figure 1).
- The output of which, as seen in (figure 2), gives different derivatives of benzodiazepine.

fig.1

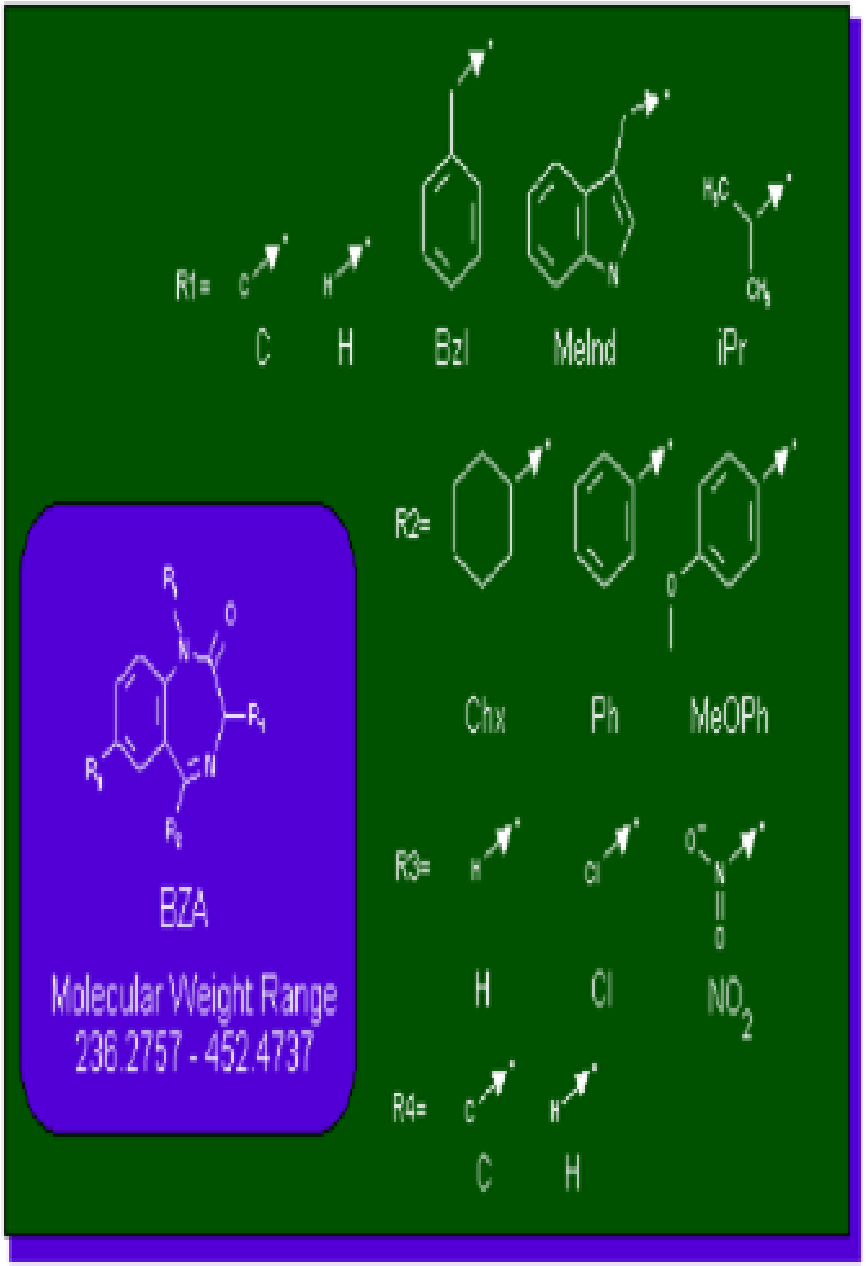
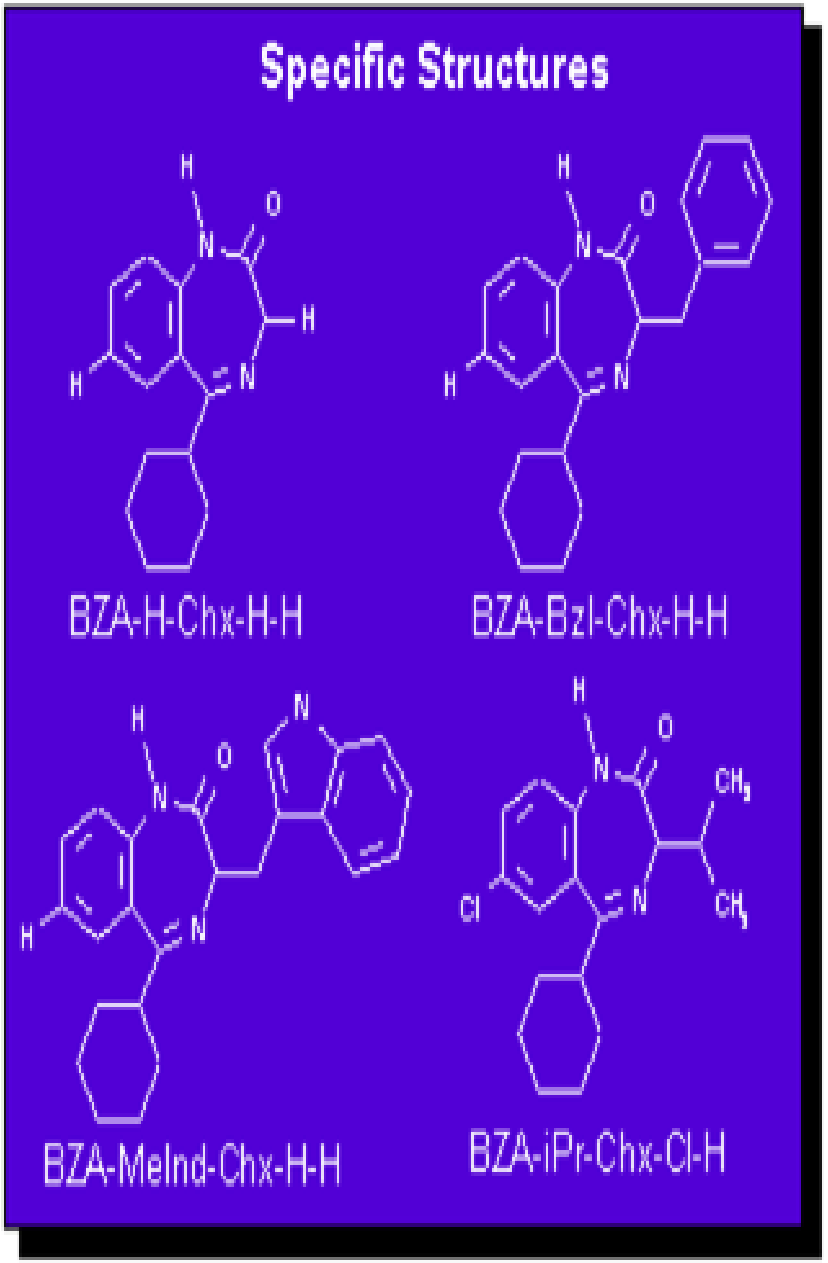

R1=
C
H
Bzl
MeInd
iPr
R2=
Chx
Ph
MeOPh
R3=
H
Cl
NO2
R4=
C
H
BZA
Molecular Weight Range
236.2757 - 452.4737

fig.2


Specific Structures
BZA-H-Chx-H-H
BZA-Bzl-Chx-H-H
BZA-MeInd-Chx-H-H
BZA-iPr-Chx-Cl-H

# CONT'D

- With its slight alterations around a general structure, combinatorial chemistry is an advanced way of finding a needle in a haystack.
- This merely stresses the importance of finding a general structure of biological importance and altering it to find activity, such as a better fit into an active site, rather than picking random molecules to synthesize.
- Therefore, Combinatorial technologies provided a possibility to produce new compounds in practically unlimited number.
- New strategies and technologies have also been developed that made possible to screen very large number of compounds and to identify useful components of mixtures containing millions of different substances.

# CONT'D

- Instead of preparing and examining a single compound, families of new substances are synthesized and screened.
- In addition, combinatorial thinking and practice proved to be useful in areas outside the pharmaceutical research such as search for more effective catalysts and materials research.
- Combinatorial chemistry became an accepted new branch within chemistry.

# CONT'D

- 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.

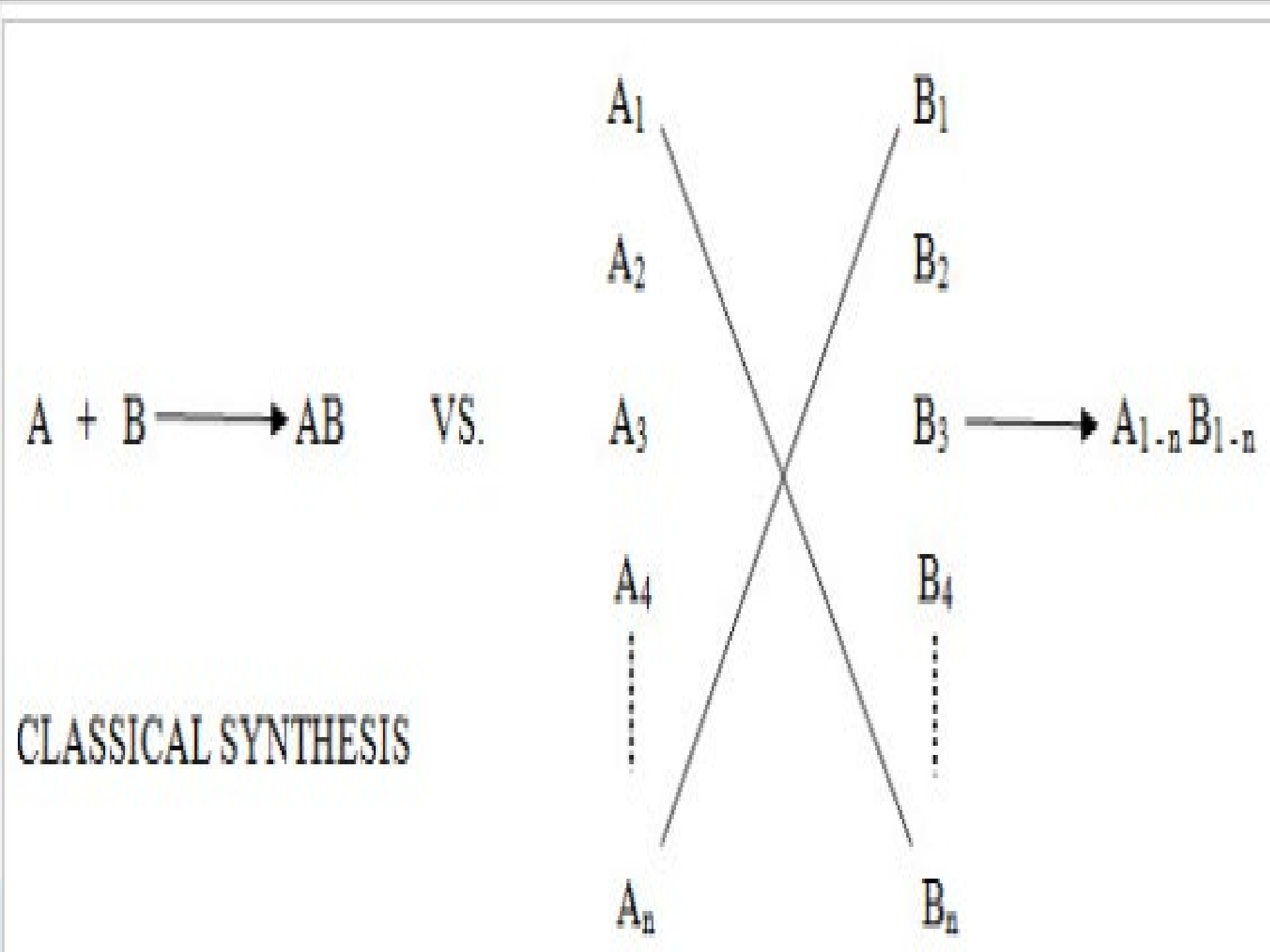
# CONT'D

- 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.

# CONT'D

- 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: $A_{1-n}+B_{1-n}$----->$A_{1-n}$ $B_{1-n}$**


A + B → AB
VS.
A1
A2
A3
A4
An
B1
B2
B3
Bn
A1-n B1-n
CLASSICAL SYNTHESIS

# COMBINATORIAL CHEMISTRY APPROACH

# 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.

# CONT'D

- Thus the combinatorial chemistry approach has two phases/Steps:

1. Making a library.
2. Finding the active compound.

- Then, Screening mixtures for biological activity has been compared to finding a needle in a haystack.

# STRATEGIES

| Conventional | Combinatorial |
| --- | --- |
| - One molecule at a time | - Many molecules at a time |
| - Make → Purity → Test | - Make → Test → Purity |
| - Hundreds of molecules a month | - Thousands of molecules a month |
| - Slower lead generation | - Faster lead generation |
| - High risk of failure | - Low risk of failure |

Synergy

LEAD IDENTIFICATION

# ADVANTAGE & DISADVANTAGE

# PROS & CONS OF COMBINATORIAL CHEMISTRY

- CREATION OF LARGE LIBRARIES OF MOLECULES IN A SHORT TIME.
- COMPOUNDS THAT CANNOT BE SYNTHESIZED USING TRADITIONAL METHODS OF MEDICINAL CHEMISTRY DONE BY COMBI CHEMISTRY.
- 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.
- MORE OPPORTUNITIES TO GENERATE LEAD COMPOUNDS.
- COMBINATORIAL CHEMISTRY SPEEDS UP DRUG DISCOVERY

- NEEDS TO SYNTHESIZE THE RIGHT COMPOUND.
- 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 ISN'T AVAILABLE BECAUSE THE CHEMISTRY AFFECTS THE RESIN.
- THERE IS A GREAT DEAL OF DIVERSITY CREATED, BUT NOT OFTEN A CENTRAL SYNTHETIC IDEA IN THE LIBRARIES.

# TYPES OF COMBINATORIAL LIBRARIES

# LIBRARIES

- Collection of structurally related compounds (peptides, oligonucleotides, oligosaccharides, organic molecules) obtainable by chemical or biological means simultaneously as a mixture and screened for activity as a mixture of compounds, without any isolation protocol step.
- Identification of active compounds derives from the synthesis/production protocol used to generate the library.
- Great acceleration of leads identification since millions of different compounds can be screened simultaneously.

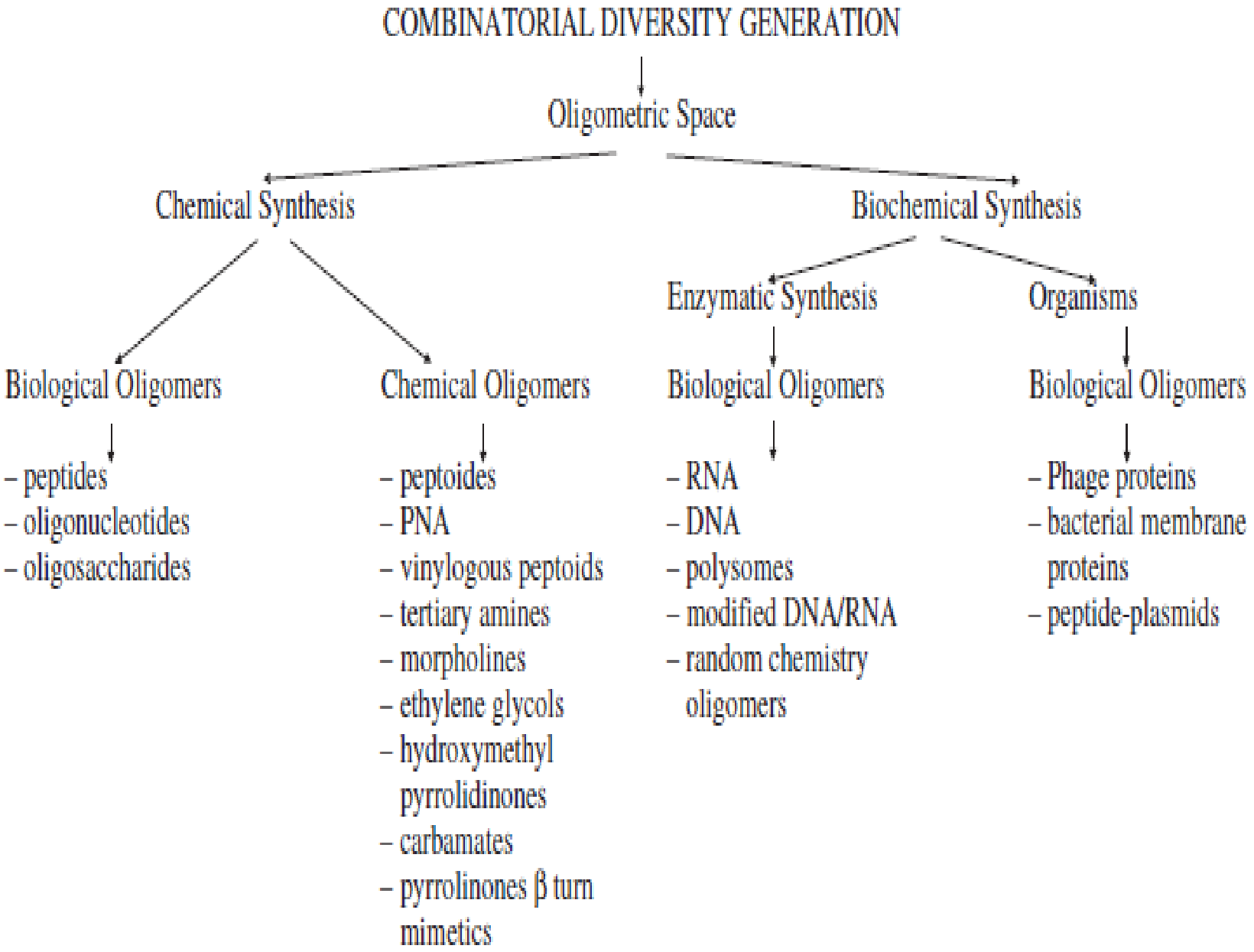



Fig. 3. **Diversity of compounds generated by combinatorial approach**

Types of Combinatorial Library

*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.

**Two approaches to generate libraries are**
**1. Random/Diverse libraries and**
**2. Focused libraries.**

# TYPES OF COMBINATORIAL LIBRARY

## DIVERSED

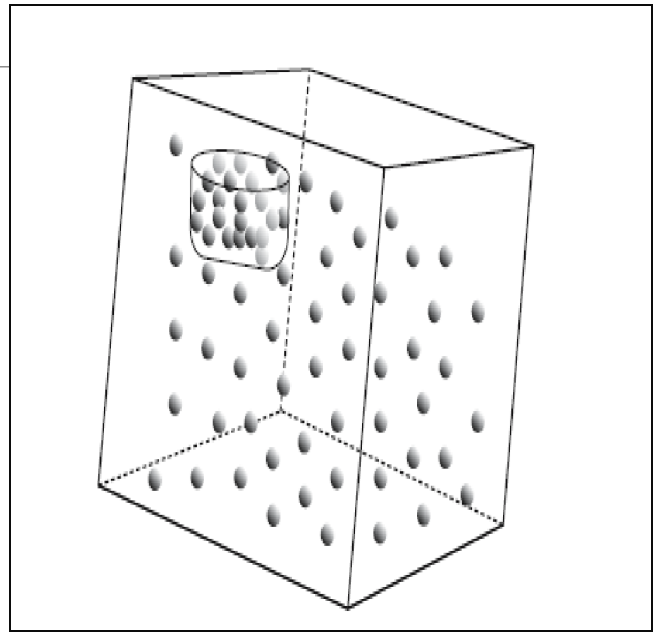
- Synthesis of diverse compounds large no. of molecules more hits(biological assay).
- Little is known about the target more diverse library(primary screening library).

## FOCUSED

- Synthesis of focused compounds small no of molecules.
- Incorporate as much information about the therapeutic target as possible.

# Continued….

- A balance between diversity and focus is needed.
- The amount of diversity α 1/ the amount of information available about the target.
- The difficulty lies in quantifying these factors and achieving the correct balance between them.



**Chemical space representation displaying diverse and focused (boxed) compound sets**

# The Design Of Drug-Like Libraries

- The drug like libraries contain molecules with biological activity and that any hits from such a library will represent more attractive starting points for lead optimization

---

- Most practical implementations of drug-likeness use a computational model which takes as input the molecular structure, together with various properties, and predicts whether the molecule is drug-like or not

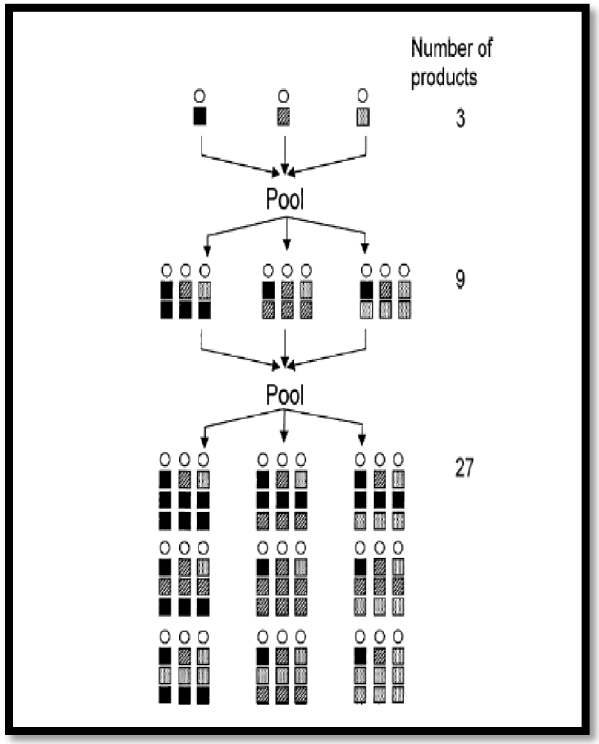
- More sophisticated model may use neural networks or a regression-type equation with coefficients derived using a genetic algorithm to predict drug-likeness from set of molecular properties.

- Filters and drug-likeness models can be extremely valuable in library design and compound selection.

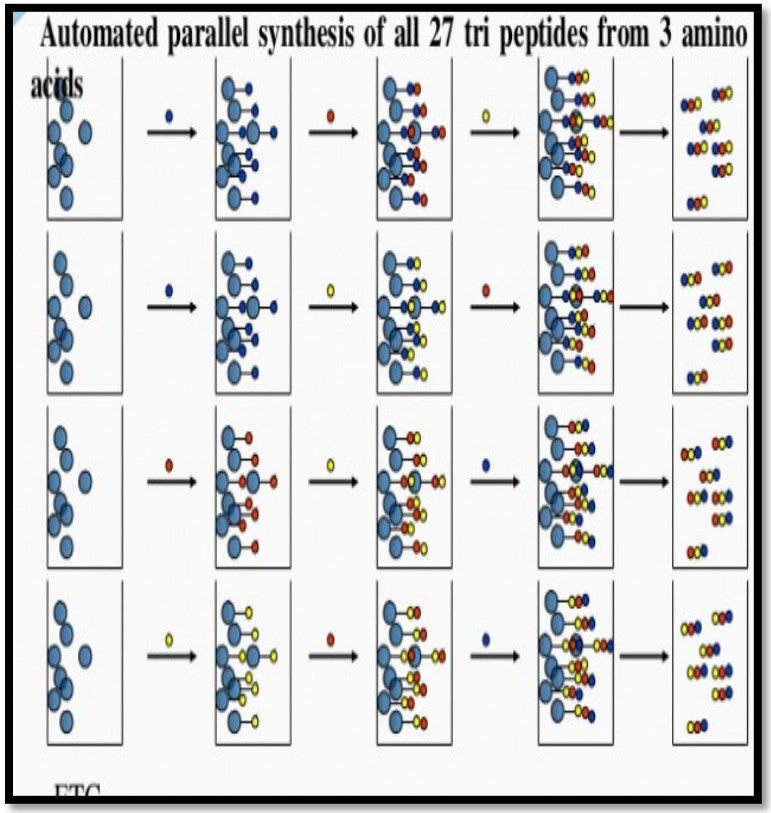
# TECHNIQUE FOR LIBRARIES PREPARATION

# TECHNIQUE

## Split and mix synthesis method




## Parallel synthesis method



# Split - Mix Approach

- Methodology  to generate large numbers of molecules using a scheme.
- In this method, ingredients are assembled on the surface of the beads or micro particles.
- In each step, beads from last steps are partitioned into new building block and several groups are added.
- This leads to the formation of new groups, the different groups of beads are recombined and separated once again.
- Process is continuous with next building block is added until the desired library has been assembled

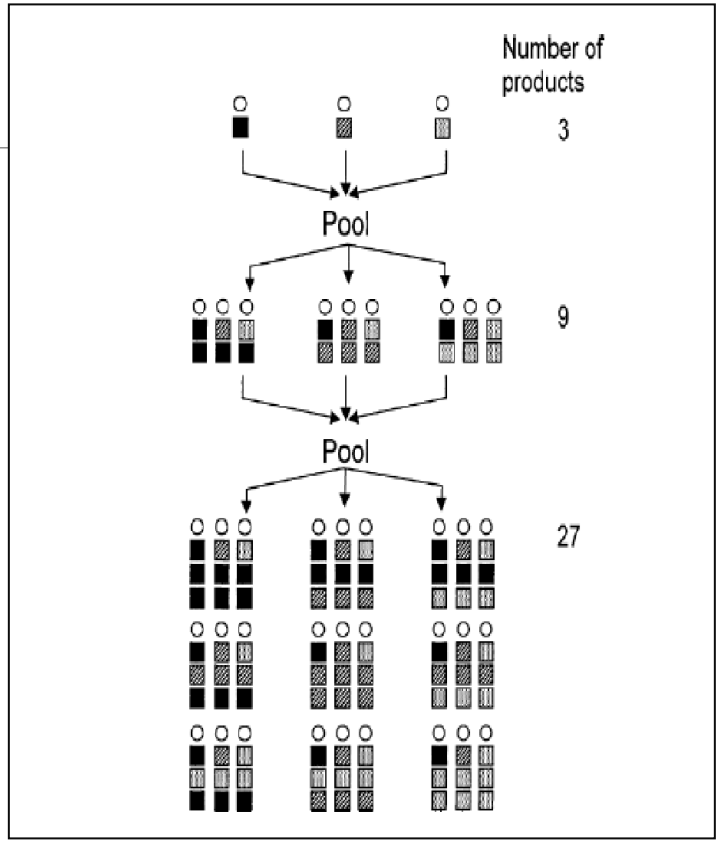



**Illustration of the split–mix approach to combinatorial synthesis, using three sets each containing three monomers**

# CONT'D

**Advantages** :

- Method of having choice for large libraries.
- Less reaction vessels required.

**Disadvantages:**

- Less amounts of the synthesized compounds available.
- Three-fold amount of resin beads necessary

Video: https://slideplayer.com/slide/4366533/

# PARALLEL SYNTHESIS APPROACH

- In separate vessels, different compounds are synthesized (without re-mixing).
- This is not like a split synthesis because it requires a solid support.
- It can be done without solid support or in a solution.
- A 96 well micro titer plate is commonly used format for parallel synthesis.
- Methods of parallel synthesis include Houghton's tea bag procedure and Automated parallel synthesis

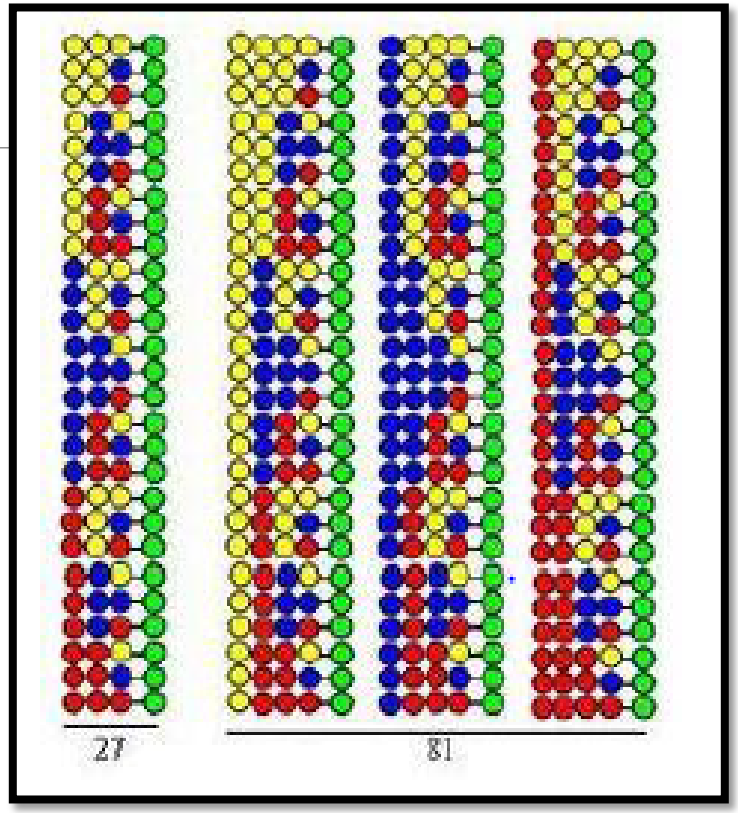



**Illustration of the parallel synthesis approach to combinatorial synthesis.**

# CONT'D

**Advantages:**

- Biological evaluation is easy
- Each compound is substantially pure in its location
- Defined location provides the structure of a certain compound

**Disadvantages:**

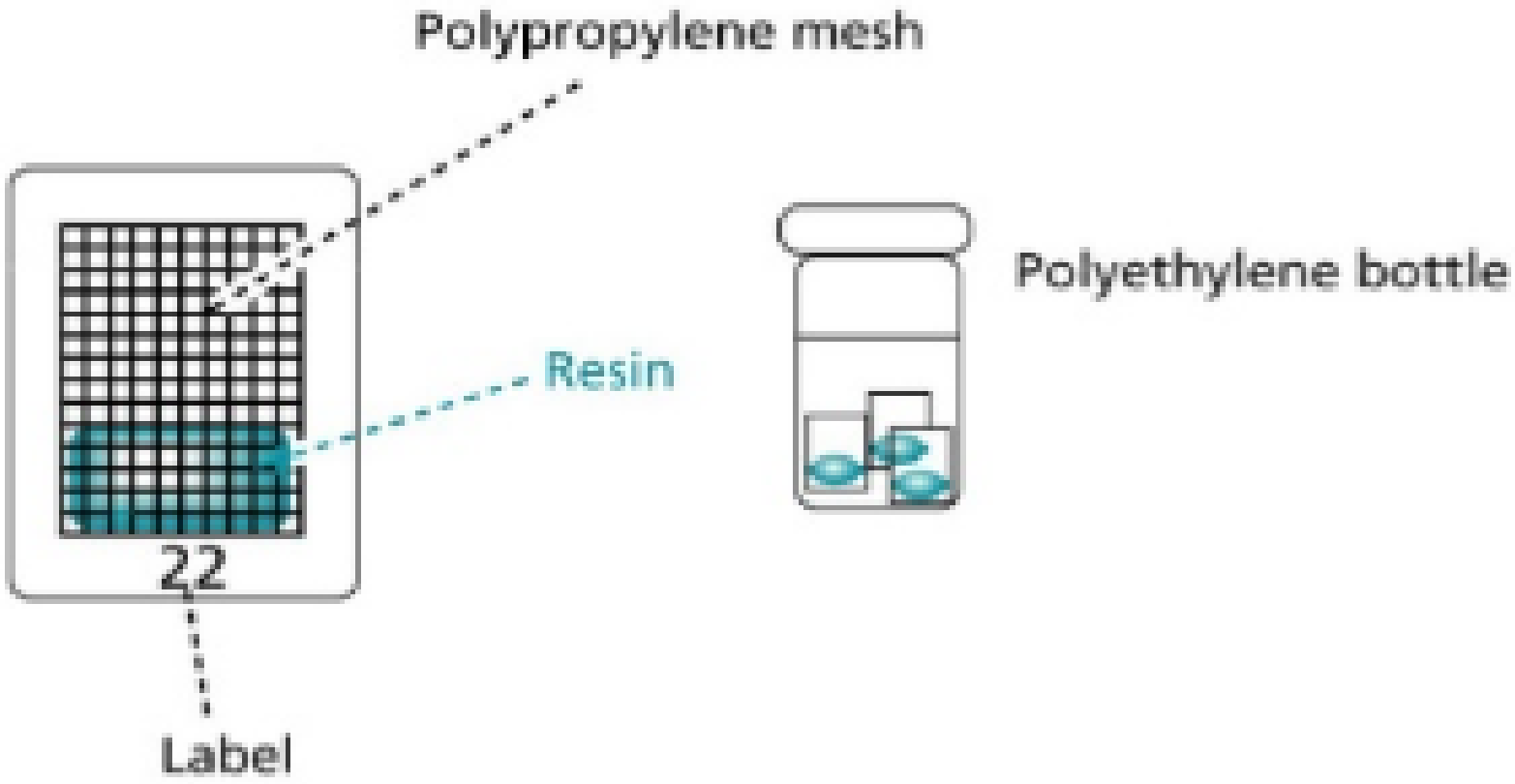
- Applicable only for particular libraries

# Houghton's tea bag procedure Automated parallel synthesis

- 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 (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.

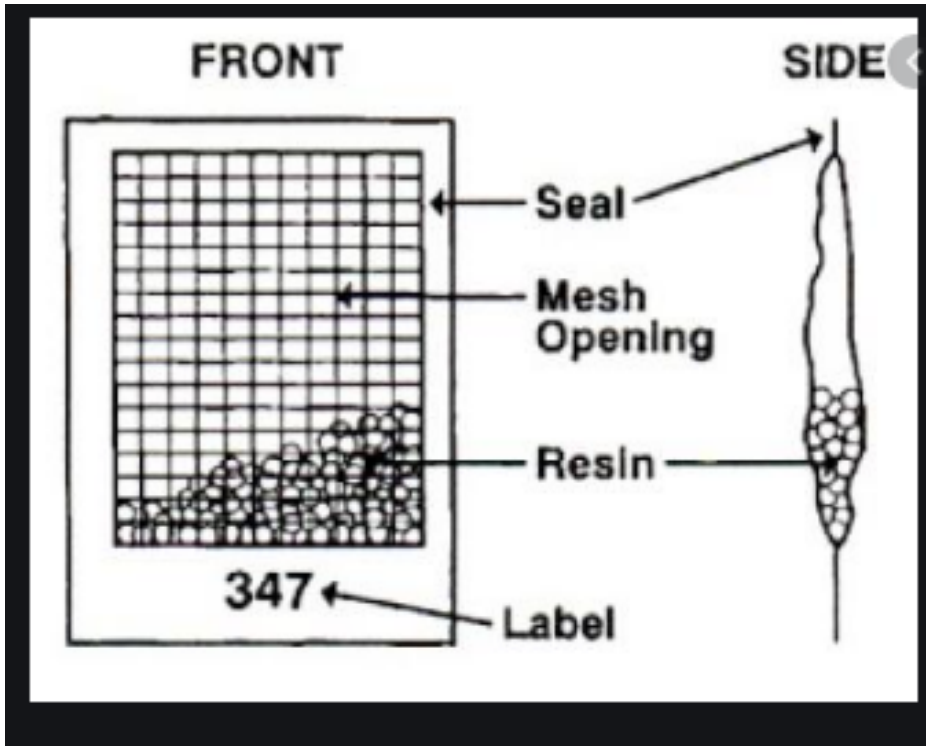
# Automated Parallel synthesis

Houghton's teabag procedure



# CONT'D

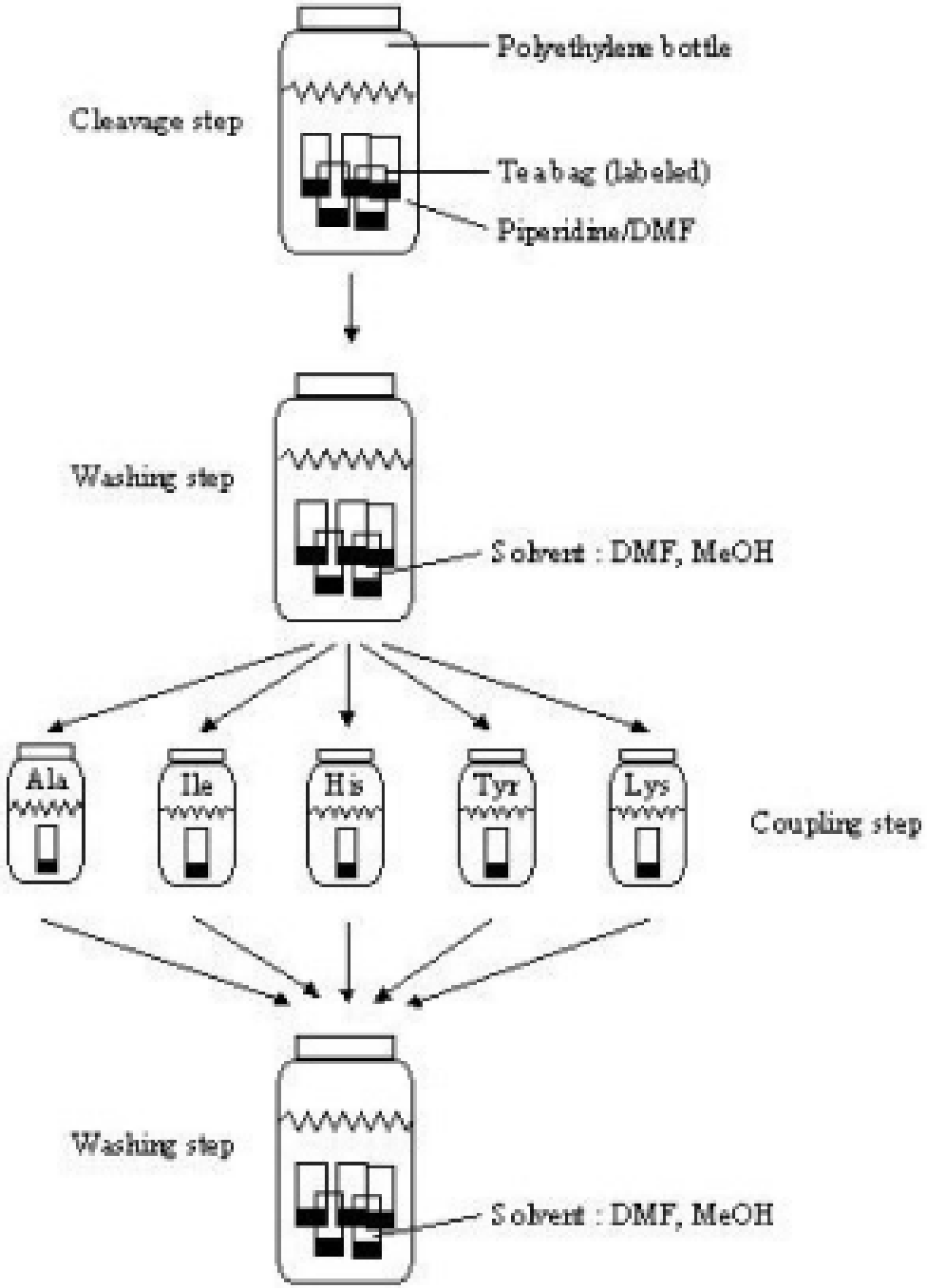
- To save time and work when making many peptides simultaneously, bags could be combined into the same reactors for common chemical steps.



# CONT'D

**For example**,

- In the synthesis of 40 different peptides, all the bags are initially charged with resin beads bearing a Boc (tert-butyloxycarbonyl) 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

Video Link : https://www.youtube.com/watch?v=BVQHw3BG7r0

# CONT'D

**Examples for the use of the tea-bag method**:

1. Characterization of the influenza haemagglutinin protein (HA1) and discovering the amino acid position that is critical important to the binding interaction
2. Production of a small combinatorial library of urea analogues
3. Rapid "tea-bag" peptide synthesis using 9-fluorenylmethoxycarbonyl (Fmoc) protected amino acids applied for antigenic mapping of viral proteins.
4. Studies on the structural requirement for ligand binding to the neuropeptide Y (NPY) receptor from rat cerebral cortex.
5. Peptide and peptidomimetic libraries. Molecular diversity and drug design.
6. The use of tea-bag synthesis with paper discs as the solid phase in epitope mapping studies.
7. Rate of swelling of sodium polyacrylate.

# Identification of Active Ingredient

- 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 ()

2. DNA based encoding (Iffat Khan)

3. Mass encoding (Shalmon)

4. Peptide tag ()

5. Hard tag (Aditi W.)

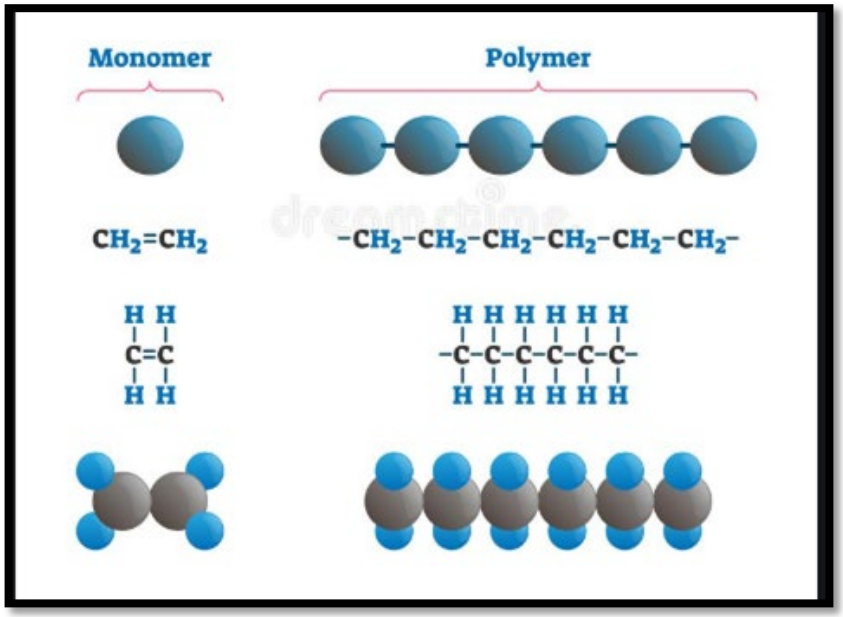
6. Radio frequency encoding ()

**PRESENTED ON_____**

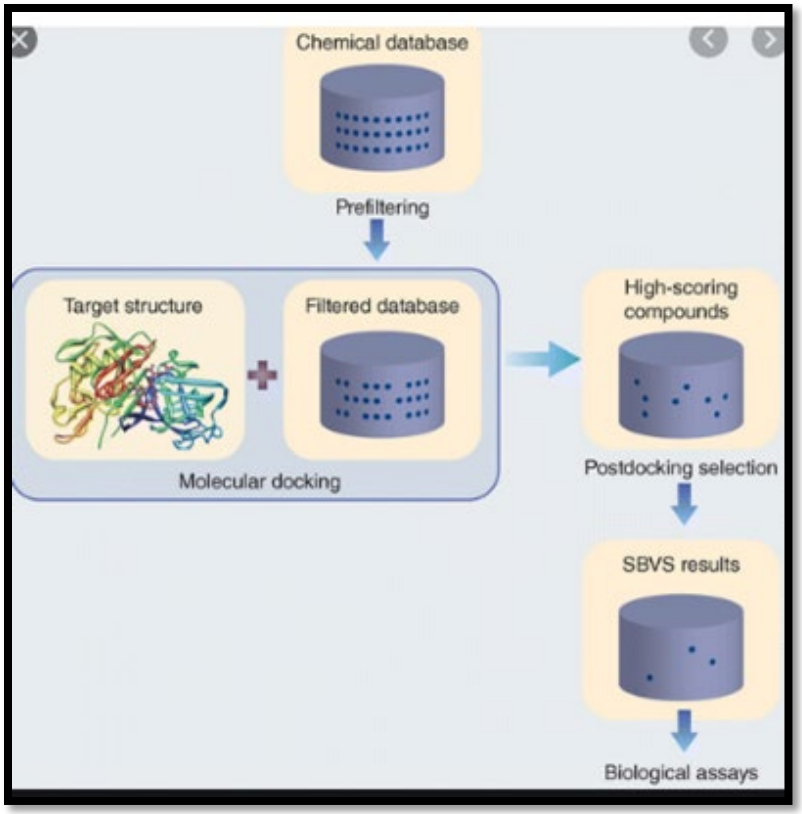
# Strategies FOR Libraries Designing

# Strategies For Libraries Designing

? ***Monomer-based selection***




? ***Product-based selection***



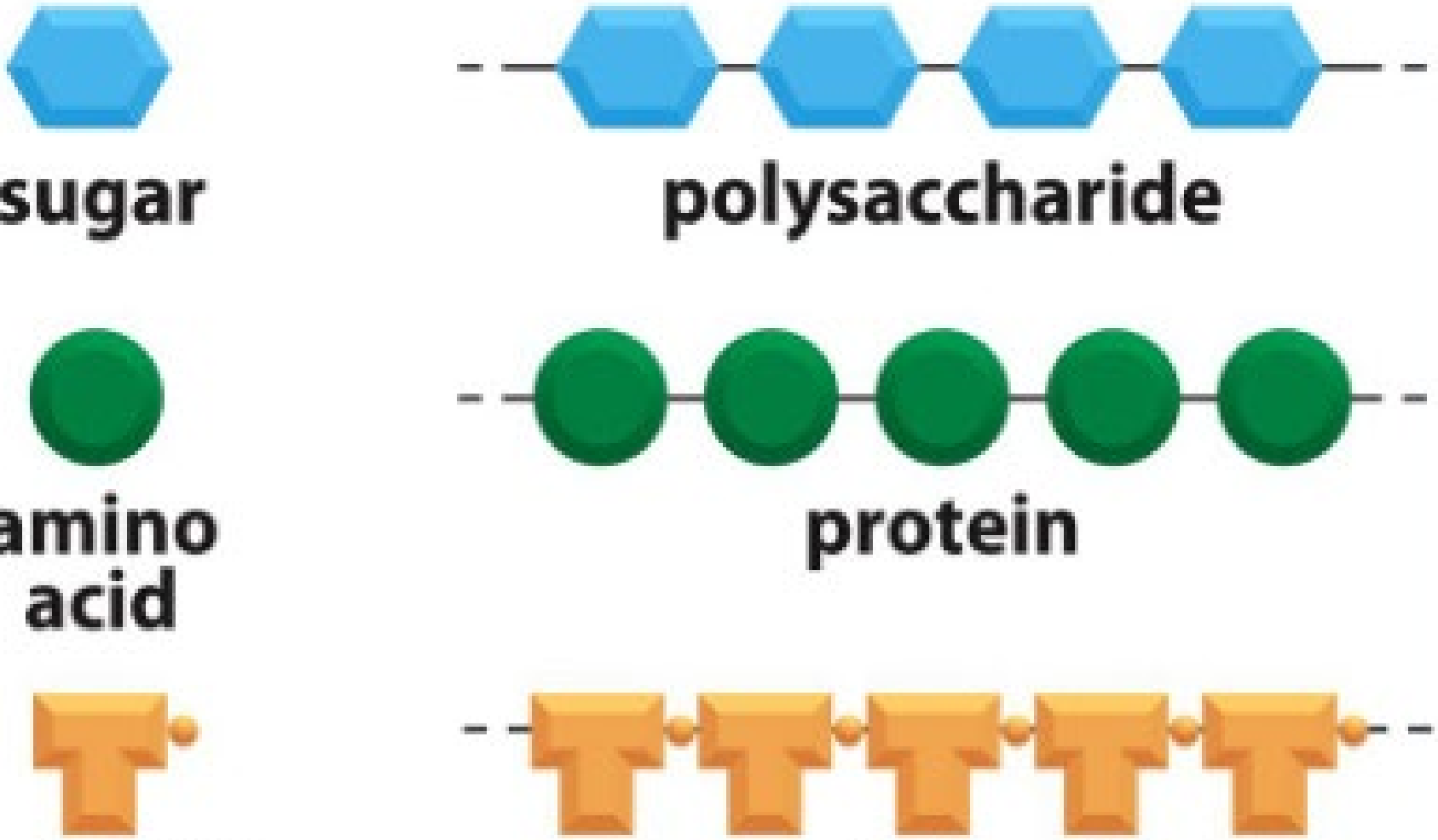
# *Monomer-based selection*

## Monomer

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

–A–A–A–A–A–A–A–A–A–A–A–A–A–A–A–A–A–A–A–

| Monomer | Polymer |
| --- | --- |
| sugar | polysaccharide |
| amino acid | protein |
| nucleotide | nucleic acid |



# CONT'D

- In monomer-based selection **optimized 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 synthesize 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 i.e.** there are **subsets** of **size *n*** contained within a **larger set of *N* compounds.**
- **Eg:** more than 1013 different subsets of size 10 from a pool of 100 monomers

# CONT'D

- It is **not possible to examine all of these**.
- 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.**

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

- **For EG:**

ELSEVIER

# Drug-motif-based diverse monomer selection: Method and application in combinatorial chemistry

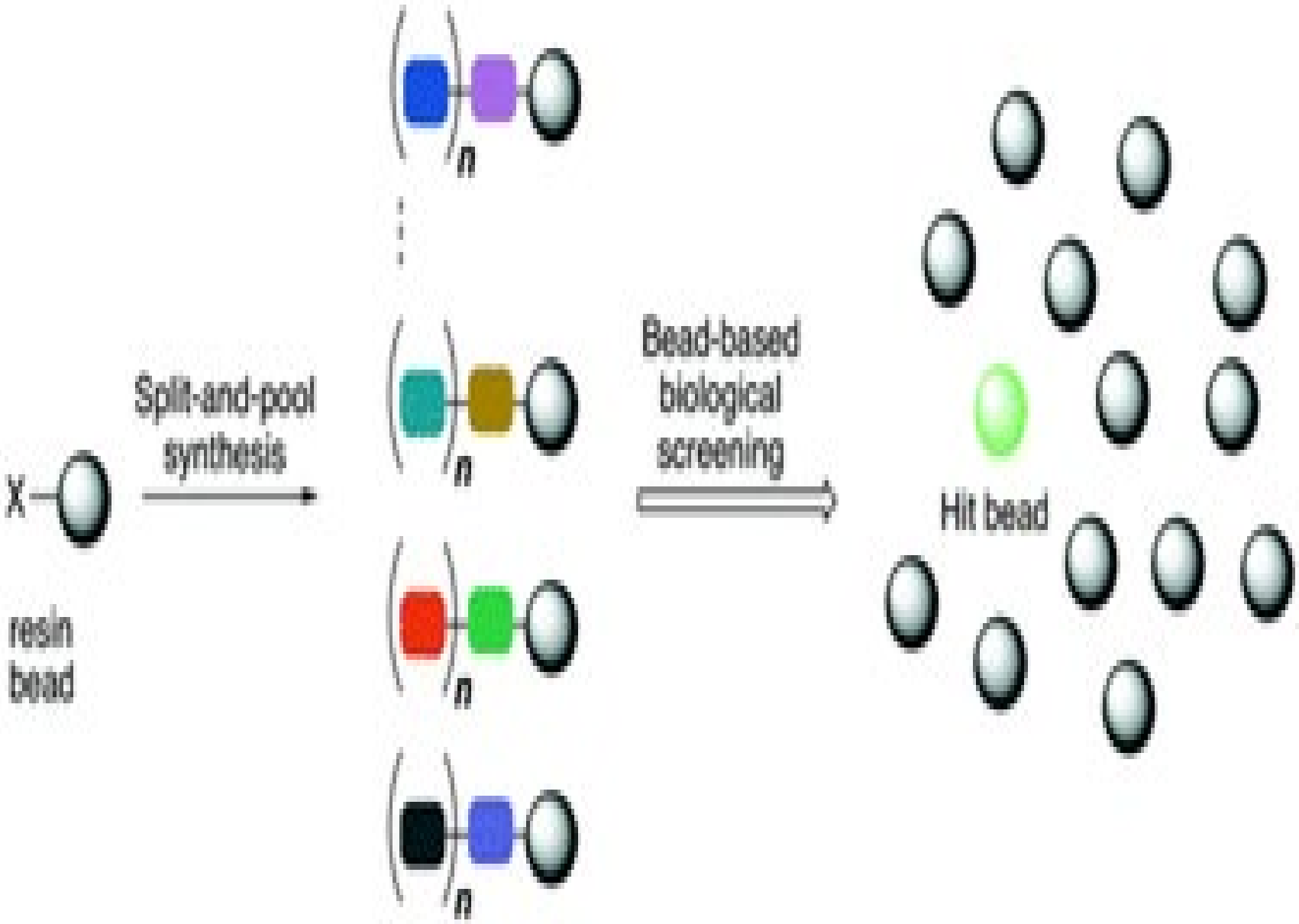
**Xiao Qing Lewell and Ryan Smith**[1]

*Glaxo Wellcome Research and Development, Stevenage, Herts, SG1 2NY, U.K.*

[1]*Currently in his final year at the University of Surrey, Guildford, Surrey GU2 5XH, U.K.*

# *Product-based selection*

- 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 could then be applied.
- This process is generally referred to as ***cherry-picking*** but it is **synthetically inefficient** in so far as combinatorial synthesis is concerned.
- 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**.
- Product-based selection is much **more computationally demanding** than monomer-based selection.


X
resin bead
Split-and-pool synthesis
n
n
n
n
Bead-based biological screening
Hit bead

# CONT'D

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

- 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 synthesized from a **100×100× 100 virtual library**.
- The selection of combinatorial subsets has been tackled using optimization 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 **optimize the properties of a library as a whole**, such as diversity or the distribution of **physicochemical properties.**

# APPROACHES IN PRODUCT-BASED LIBRARY DESIGN

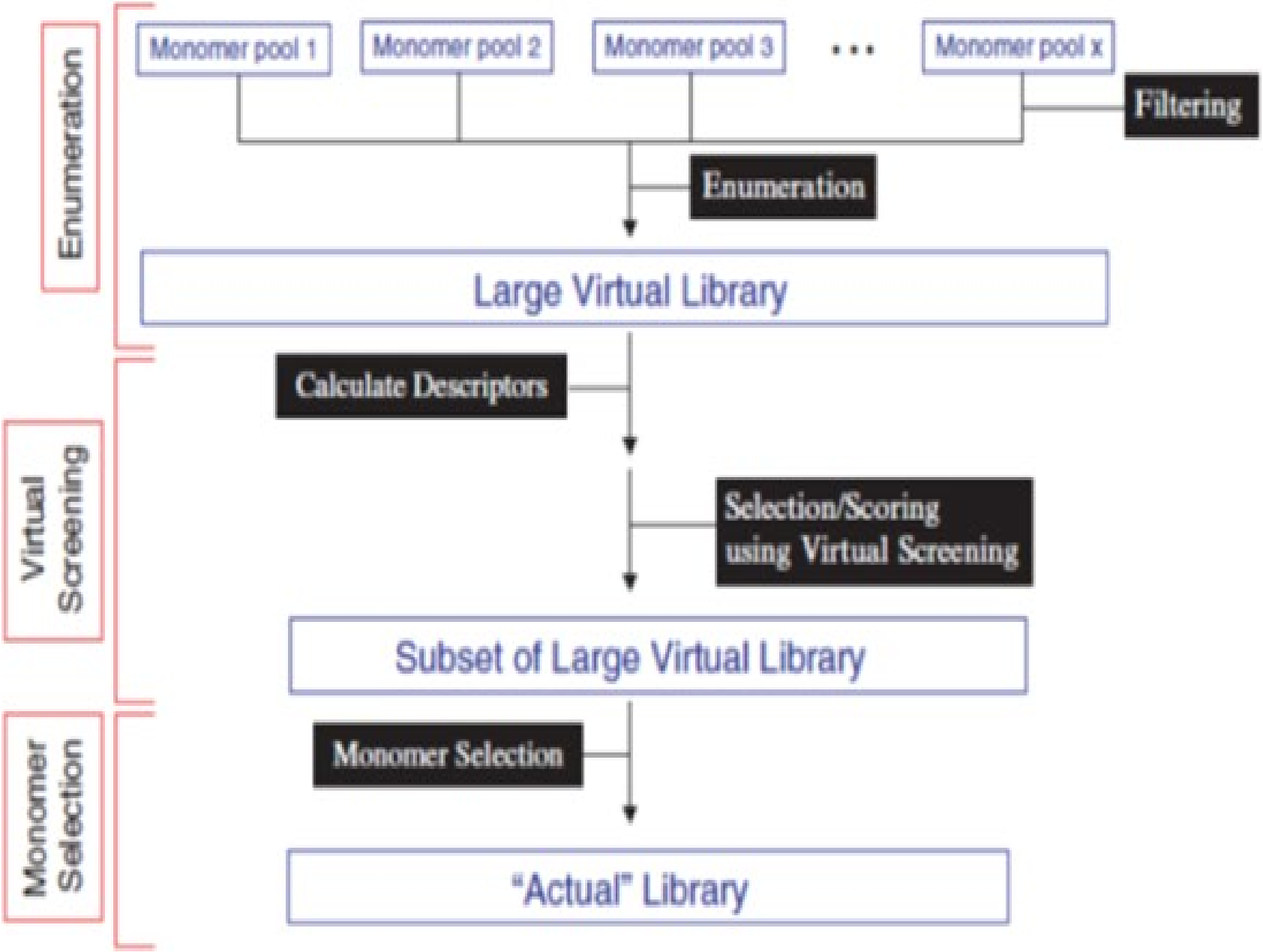
# APPROACHES TO PRODUCT-BASED LIBRARY DESIGN

? A general strategy for product-based library design involves the following three steps:

1. **Lists of potential reagents** are identified (e.g. by searching relevant databases), **filtered as appropriate,** and the **virtual library is enumerated (no of things)**.

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, or the degree of similarity/dissimilarity needed to existing collections.


Enumeration
Monomer pool 1
Monomer pool 2
Monomer pool 3
• • •
Monomer pool x
Filtering
Enumeration
Large Virtual Library
Virtual Screening
Calculate Descriptors
Selection/Scoring using Virtual Screening
Subset of Large Virtual Library
Monomer Selection
Monomer Selection
"Actual" Library

# CONT'D

- 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**.
- EG: 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.
- The **final, monomer selection s**tage is typically implemented using **optimization techniques such as GAs or simulated annealing**.
- EG: The **SELECT program** 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.**

# CONT'D

- The **chromosome of the GA** thus contains ***nA* + *nB* elements**, each **position specifying one possible monomer**.
- 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.
- 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.

# CONT'D

- 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 **focused libraries.**
- 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.
- 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**.

CURRENT OPINION
www.current-opinion.com

## Reagent-based and product-based computational approaches in library design

Eric A Jamois

# Molecular docking analysis of selected natural products from plants for inhibition of SARS-CoV-2 main protease

**Megha Hastantram Sampangi-Ramaiah[1], Ram Vishwakarma[2] and R. Uma Shaanker[1,3,*]**

[1]School of Ecology and Conservation, University of Agricultural Sciences, GKVK, Bengaluru 560 065, India
[2]Indian Institute of Integrative Medicine, Council of Scientific and Industrial Research, Jammu 180 001, India
[3]Department of Crop Physiology, University of Agricultural Sciences, GKVK, Bengaluru 560 065, India

# METHODS OF COMBINATORIAL CHEMISTRY

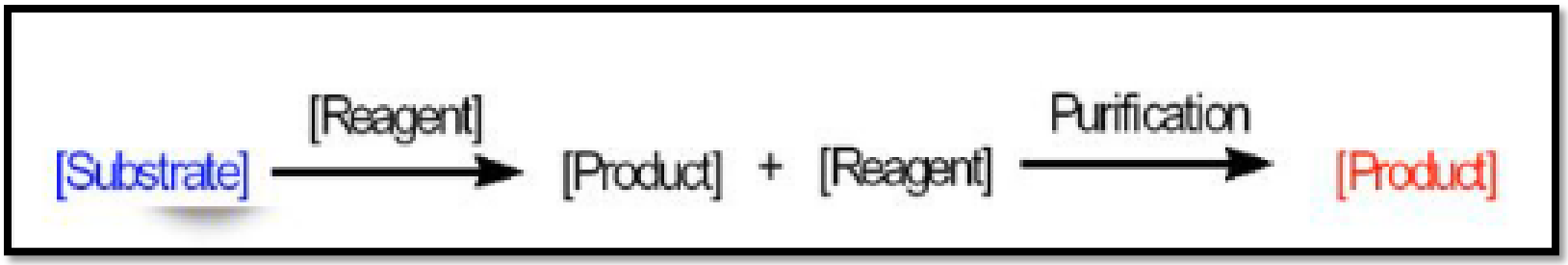
https://www.youtube.com/watch?v=v1qQDF3WWQI

# Combinatorial chemistry is of two types

https://slideplayer.com/slide/11261051/

? **Solid Phase Combinatorial Chemistry**


? **Solution Phase Combinatorial Chemistry**

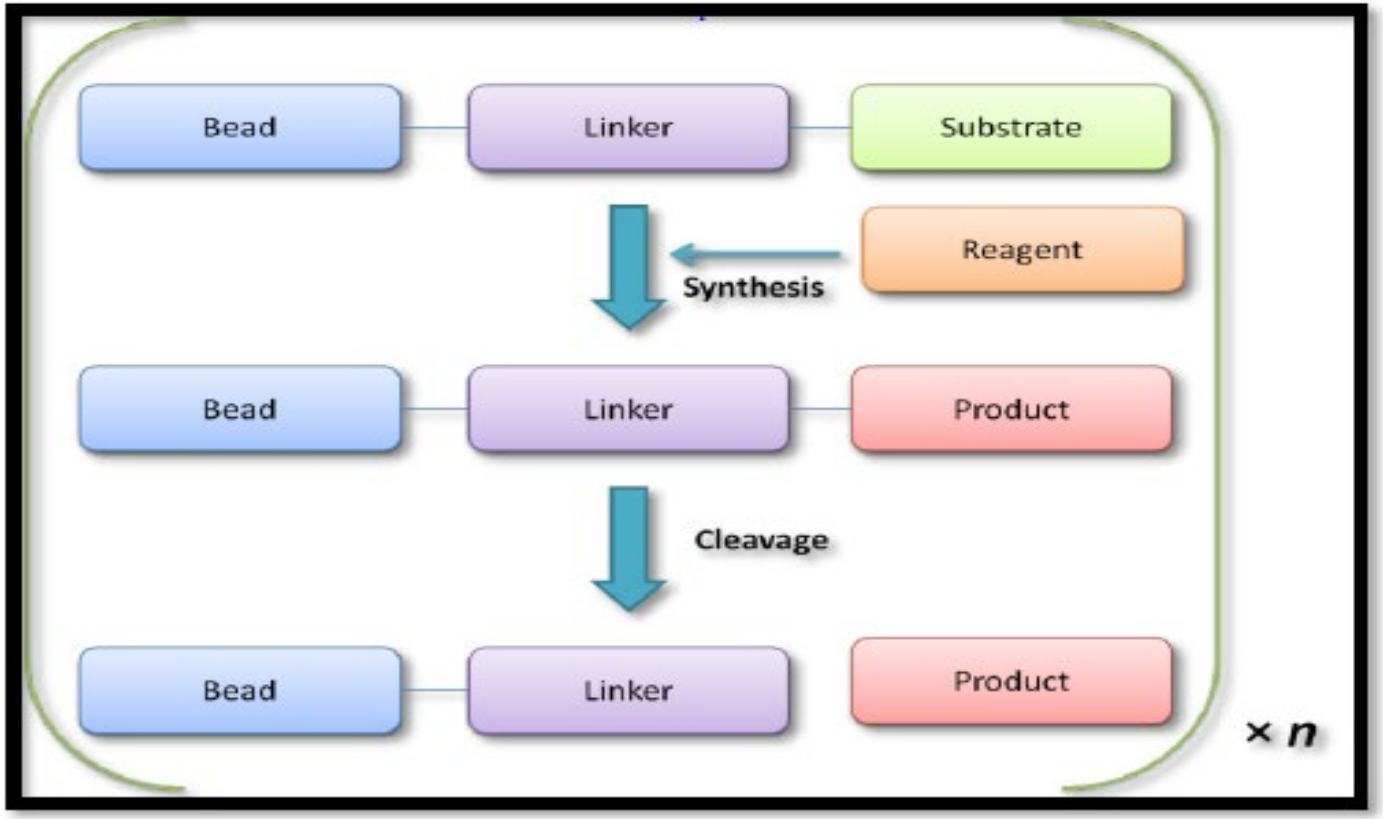


# Solid Phase Combinatorial Chemistry

- In **solid phase combinatorial chemistry**, reagents or products are **attached** to **solid supports** such as **polystyrene beads—is** the most traditional form of phase trafficking.
- In solid-phase organic synthesis, it's easy to purify products by **filtration,** it's possible to do **mix-and-split synthesis** (a technique used to make very large libraries), **excess reagents** can be used to drive reactions to completion, and **syntheses** can be automated easily.
- There are **various types of linkers** which are used for starting compound and are attached to an **insoluble resin bead**.
- **Polystyrene is 1-2% divinyl benzene**, most commonly used in combinatorial chemistry.

# CONT'D

- It's particularly useful for **multi-step reactions,** intermediates resulting in each step can be isolated quickly by this method.



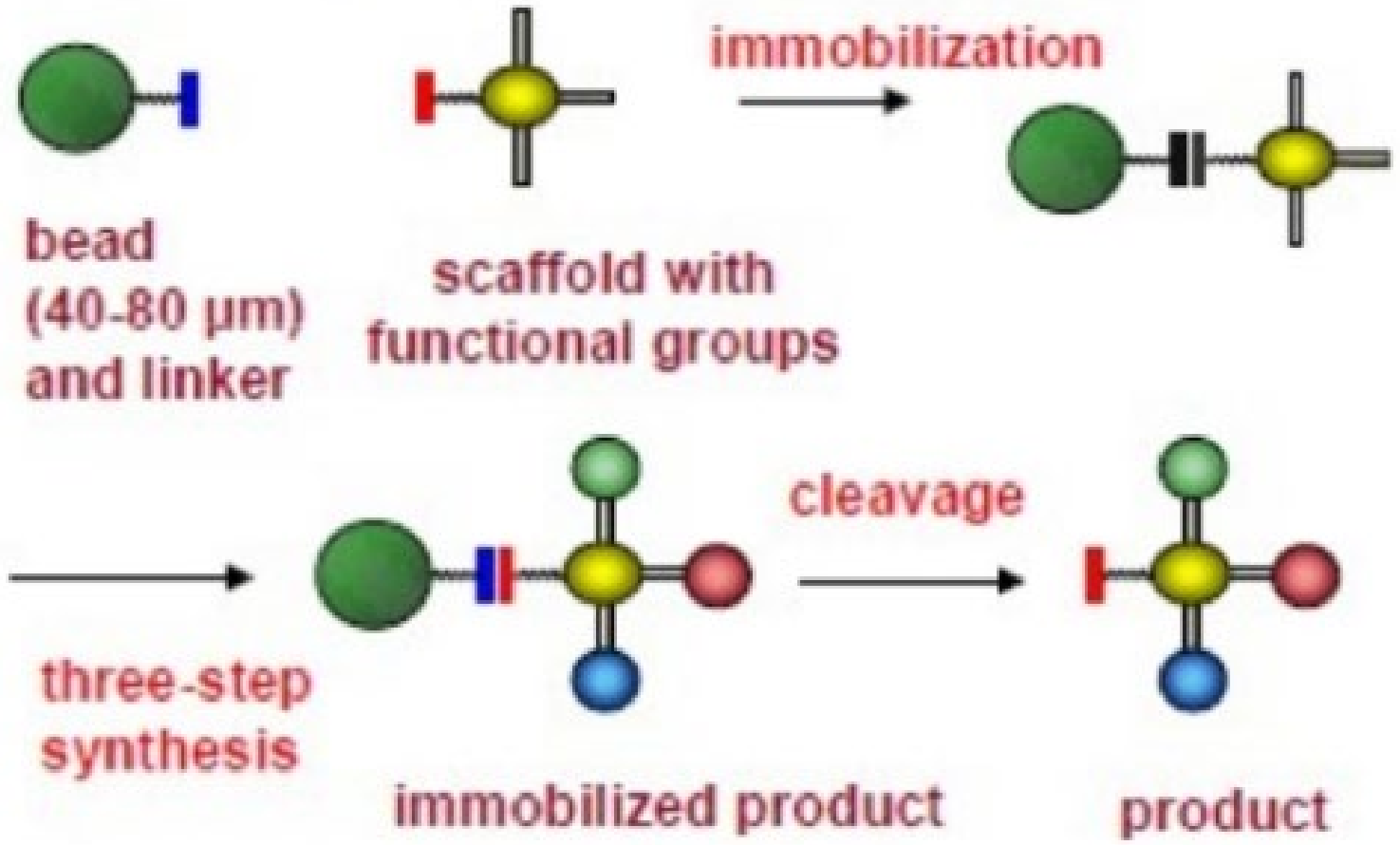
# Advantages

- Specific reactants can be bound to specific beads
- Beads can be mixed and reacted in the same reaction vessel
- Products formed are distinctive for each bead and physically distinct
- Excess reagents can be used to drive reactions to completion
- Excess reagents and by products are easily removed
- Reaction intermediates are attached to bead and do not need to be isolated and purified
- Individual beads can be separated to isolate individual products
- Polymeric support can be regenerated and re-used after cleaving the product
- Automation is possible

# Disadvantages to Solid Phase Synthesis

- All the synthesis can't be done on solid phase.
- Typically, kinetics not the same.
- Unsuitable for solvent assisted chemical reaction.
- High viscosity in reactant system.
- Insufficient purity if reaction steps are incomplete.

# The Principle of Solid Phase Synthesis



## Requirements

- A resin bead or a functionalised surface to act as a solid support
- An anchor or linker
- A bond linking the substrate to the linker. The bond must be stable to the reaction conditions used in the synthesis
- A means of cleaving the product from the linker at the end
- Protecting groups for functional groups not involved in the synthesis

## Resin beads

- Cross-linked, insoluble, solvent swellable polymeric materials, inert to the conditions of synthesis
- 80-200 µm
- Preparation:-

1. Addition and dispersion of an organic phase of monomer and cross-linker in an aqueous solution
2. Dissolvation of a free radical initiator in the organic mixture
3. Raising of temperature starts polymerisation to form resin beads
4. Collection of resin beads by filtration and washing of unreacted monomers and the aqueous phase

# Solid support used in Solid phase synthesis

- Most solid state combinatorial chemistry is conducted by using **polymer beads** ranging from **10 to 750 µm in diameter**.

- The solid support must have the following **characteristics for an efficient solid-phase synthesis:**

1) Physical stability and of the right dimensions to allow for liquid handling and filtration;

2) Chemical inertness to all reagents involved in the synthesis;

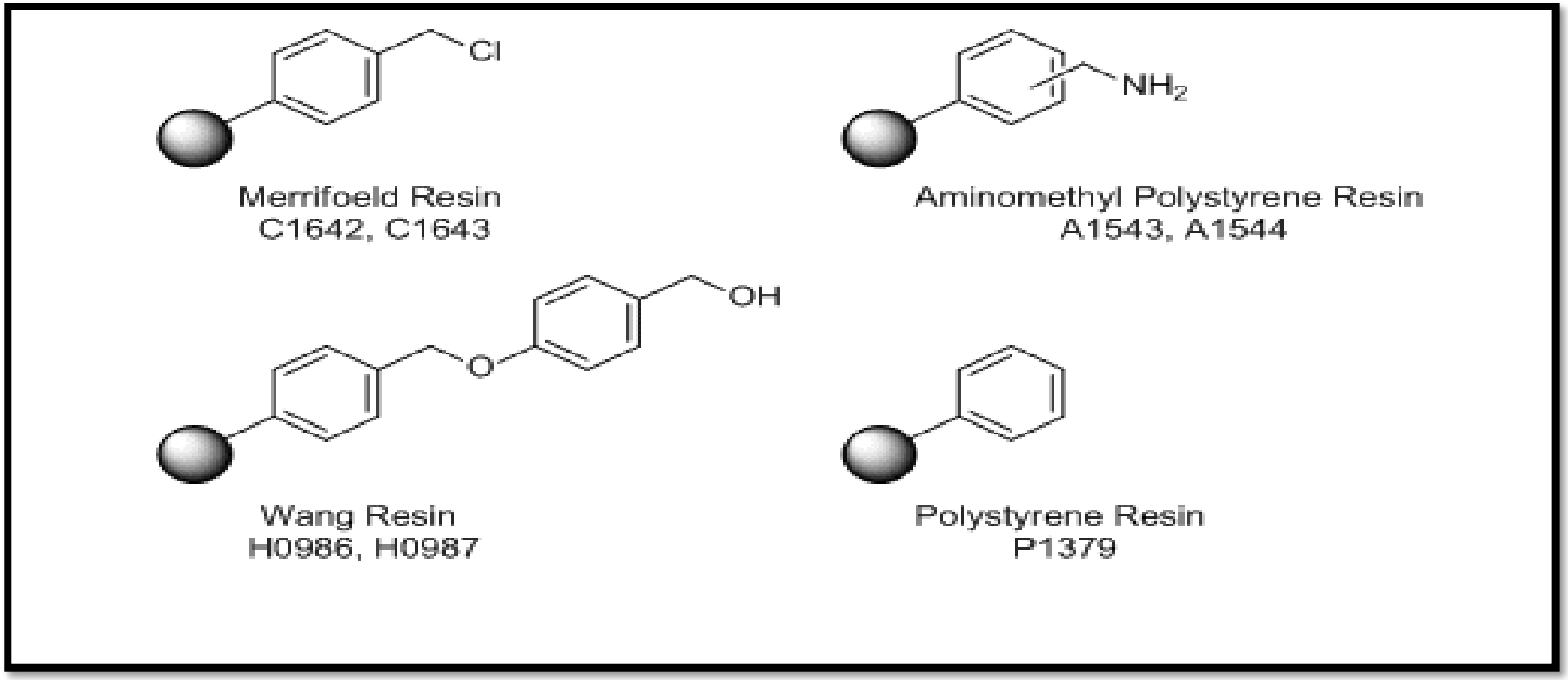
3) An ability to swell while under reaction conditions to allow permeation of solvents and reagents to the reactive sites within the resin;

4) Derivatization with functional groups to allow for the covalent attachment of an appropriate linker or first monomeric unit.

- 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.

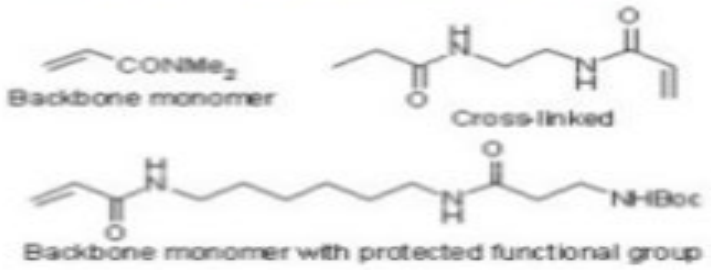
# Types of solid that are used

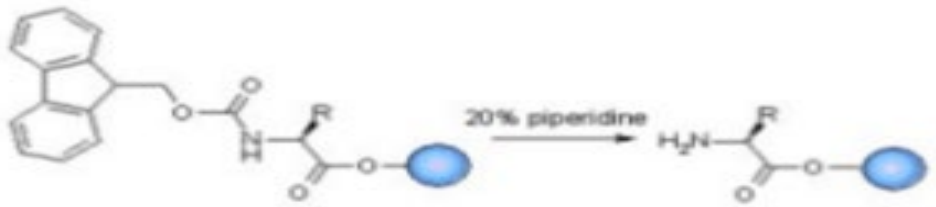
**1. Polystyrene resins:** In this Polystyrene is cross linked with divinyl benzene (about 1% crosslinking) and polystyrene resin are suitable for **nonpolar solvents.**



**Polyacrylamide resins:** like super blue these resin swell better in **polar solvent,** since the contain amide bonds, more closely resemble biological materials.

- N,N-dimethylacrylamide as backbone, cross-linked with N,N`-bisacryloylethylenediamine, functionalised through N-acryloyl-N`-Boc-β-alaninylhexamethylenediamine.
- Swells in polar solvent, limited in less polar solvent.

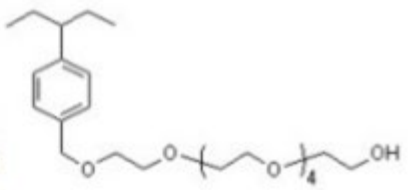



- Sheppard designed polyacrylamide polymers for peptide synthesis as it was expected that these polymers would more closely mimic the properties of the peptide chains themselves and have greatly improved solvation properties in polar, aprotic solvents (e.g. DMF, or N-methyl pyrrolidinone).

**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.**

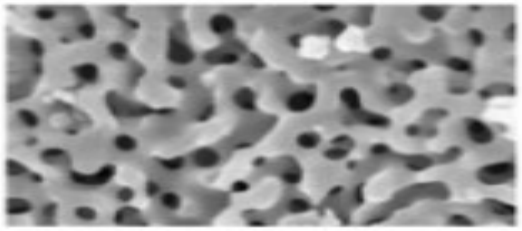
- Polystyrene glycol attached to cross-linked polystyrene through ether link

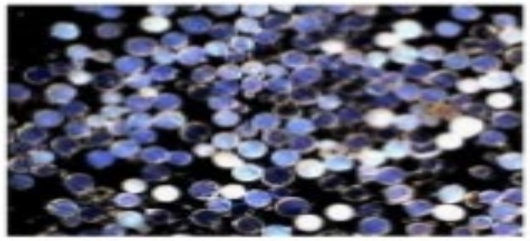



- Prepared by the polymerisation of ethylene oxide on cross-linked polystyrene (derivatised with tetraethylene glycol to give polyethylene glycol chains) long flexible chains that terminate with a reactive site spatially separated from the more rigid polystyrene backbone.
- TGR carry polyethylen glycol chains of about (3 kDa), 70 – 80% of resin weight
- The hydrophilic nature of the resin facilitate release of product in aqueous environment
- Benefits of the soluble polyethylene glycol support with the insolubility and handling characteristics of the polystyrene bead. These beads display relatively uniform swelling in a variety of solvents from medium to high polarity ranging from toluene to water.

**Glass and ceramic beads:** these type of solid supports are used when **high temperature and high pressure reaction** are carried out.

- Macropourous polymers with rigid open pores to permit a ready and continuous solvent flow
- Glass-derived bead material
- Compatible to any type of solvent
- Stable in aggressive reagents and extremes of pressure and temperature.
- Used for combinatorial synthesis of peptides and oligonucleotides.





# Linkers used in solid phase synthesis

- To support the attachment of a synthetic target, the polymer is usually modified by equipping it with a **linker.**
- Linker must be **stable** under the reaction conditions, but they must be **susceptible to a cleavage.**
- Some **specialized linker** have been developed to meet particular reaction or product conditions and this type of linker is known as **traceless linkers**, it can be cleaved from the resin with no residual functionality left.
- This type of linkers allows the **attachment** of **aryl and alkyl products** that do not have **OH or NH functionality ,eg:** silyl group (-Si(CH3)2) that is sensitive to acid and can be cleaved to give unsubstituted phenyl or alkyl product.

# CONT'D

- A **new class of linkers** was developed known as **safety-catch linkers** which is inert to **synthetic condition** and chemically transformed to allow final liberation of the product from the resin.
- Now a **ultraviolet light sensitive protecting groups** are used, like **affymax group** is used in the **synthesis of carboxylic acid and carboxamide products.**
- Some **groups have used linkers** that can only be **cleaved by enzymes.**
- A novel linker possessing **selenocyanate** and **masked carboxylic acid** was developed for the solid-phase synthesis of dehydropeptides.

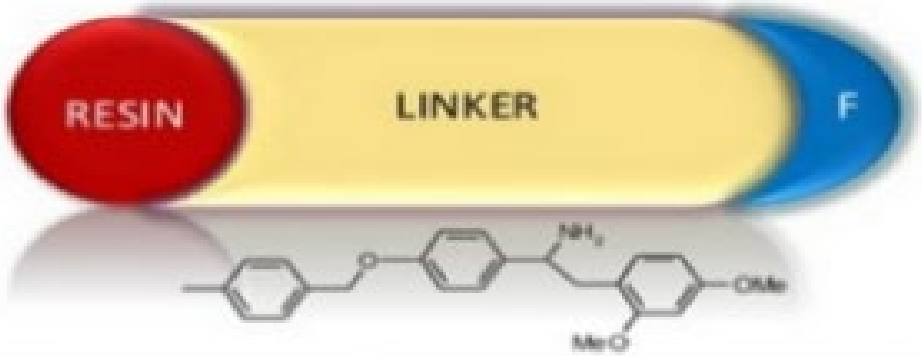
## Anchor or linker

- A molecular moiety which is covalently attached to the solid support, and which contains a reactive functional group
- Allows attachment of the first reactant
- The link must be stable to the reaction conditions in the synthesis but easily cleaved to release the final compound
- Different linkers are available depending on the functional group to be attached and the desired functional group on the product
- Resins are named to define the linker e.g.

Merrifield, Wang, Rink


**WANG RESIN**: linker suitable for attachment & release of carboxylic acids.


**MERRIFIELD RESIN** : linker suitable for peptide products.



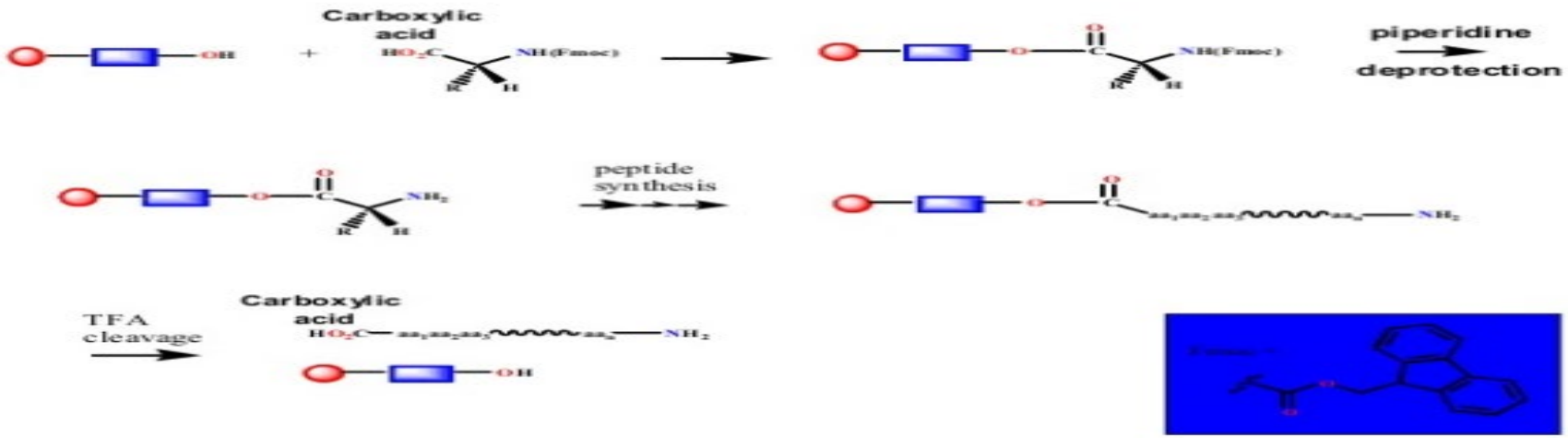

**RINK RESIN** : Linker suitable for attachment & release of carboxamide.

## Wang resin


Wang Resin

||


## Wang resin



# Linkers

- Regenerate the originally linked functionality (-OH or -COOH)
- Convert from one functional group to another (-COOH to $-CONH_2$)
- Totally remove the functionality on cleavage.
- Types:-

A. **Carboxylic acid linker**
B. **Carboxamide linker**
C. **Alcohol linker**
D. **Amine linker**
E. **Traceless linker**
F. **Light cleavable linker**

## Common protecting groups used in solid phase synthesis and their cleavage methods

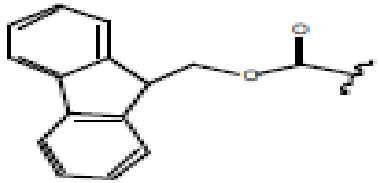
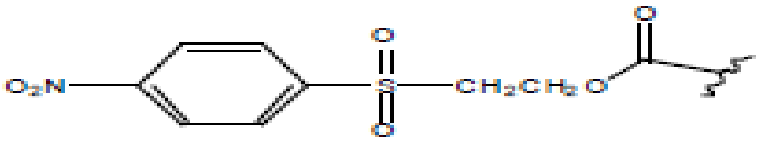
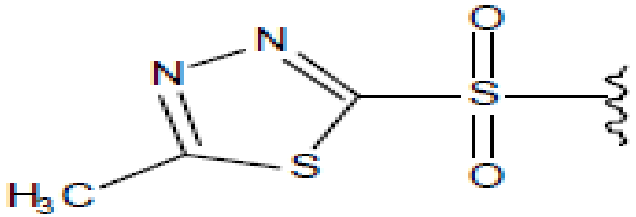
- **Primary function of protecting group** is to protect the portion of the molecule that is **not covalently bound** to the **resin** and must be protected to **avoid subsequent** polymerization of excess monomers in solution.
- The protecting group must be **stable** to the reaction conditions of each coupling.
- After **coupling** is performed, the **protecting group** is removed to expose a **new reactive site** and **synthesis** continues in a repetitive fashion.
- Cleavage conditions are **dictated by the linker used**.

# Solid phase synthesis: protecting groups

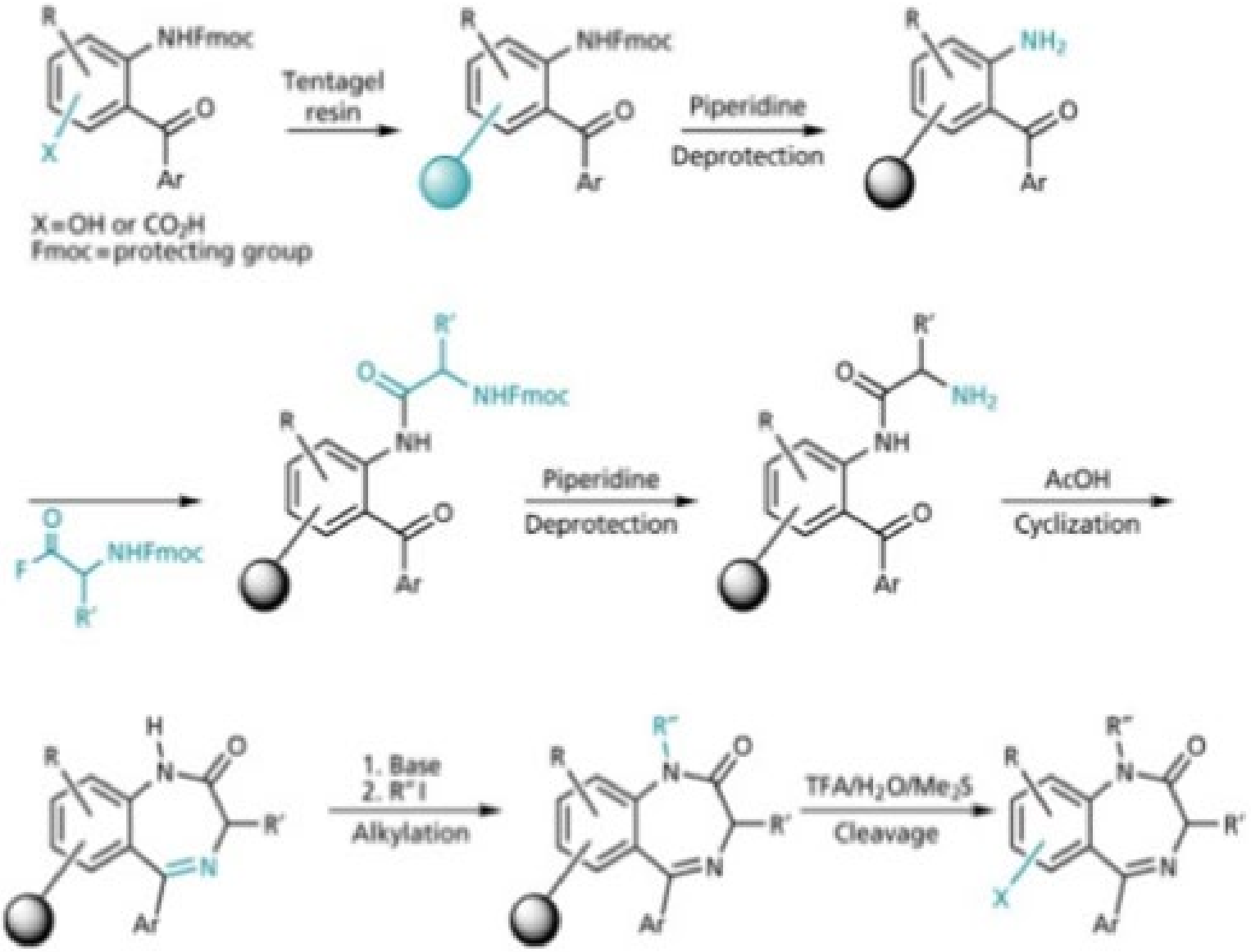
- A few protecting groups used in solid phase synthesis.

- **For amines.**
  - Boc ( t-butoxycarbonyl )
  - Fmoc (9-fluorenylmetoxy carbonyl)
  - Tmsec (2 [ trimethylsilyl ] ethoxycarbonyl)

- **For carboxylic acids.**
  - Tert Bu ester(t-butyl ester)
  - Fm ester(9-fluronyl methyl ester)
  - Tmse ester(2 [trimethylsilyl] ethyl)

| Protecting Group | Structure | Cleavage Method |
|---|---|---|
| **$N^{\alpha}$-Protecting Groups** | | |
| Fluorenylmethoxycarbonyl (Fmoc) |  | Base-catalyzed (20% Piperidine in DMF) |
| 2-(4- nitrophenylsulfonyl) ethoxycarbonyl (Nsc) |  | Base catalyzed (20% piperidine in DMF |
| Allyloxycarbonyl (Alloc) |  | Hydrogenolysis (Pd/C; ethanol) |
| 5-Methyl-1,3,4 thiadiazole-2- sulfonyl (Ths) |  | Zn-Acetic Acid Al-Hg/THF/$H_2O$ |

# Examples: HIV protease inhibitors



# Solution phase synthesis

- Synthetic chemistry takes place in solution phase.
- Solution phase techniques – Used and explored as an alternative to solid-phase chemistry approaches ---------for the **preparation of arrays of compounds** in the drug discovery process.
- This method is -------free from some of the constraints of solid-phase approaches
- Disadvantages: W.r.t **purification**.
- In this, use of soluble polymer is used as support for the product.

  i.e. **PEG** is a common vehicle which is used

(1) Because in solution phase synthesis it can be **liquid or solid at room temperature and** show **varying degrees of solubility in aqueous and organic solvent.**

(2) 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.

# CONT'D

- Another common support which is used in solution phase synthesis is **liquid Teflon** consisting mainly of **long chain** of (-CF2-) 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.**
- **Advantages**
- Handling of material is easy and can be automated.
- **Disadvantages**:
- Purification is hard.
- Quantities produced can be very low for very large libraries.
- Solution phase methods don't always work when compared to the solid phase.
- Characterization of intermediates is difficult as we can't tell if our reaction has worked or not.
- We can't surely detect which compound is attached to any one bead.

# COMPARISON BETWEEN SOLID PHASE & SOLUTION PHASE TECHNIQUE :

| Sr. No. | Parameter | Solid Phase technique | Solution phase Technique |
|---|---|---|---|
| 1 | **Reagent** | Excess | Optimum (unless purification done) |
| 2 | **Purification** | Easy | Can be difficult |
| 3 | **Automation** | Easy | Difficult |
| 4 | **Reaction** | Suitable for few substance | Suitable for any organic reaction |
| 5 | **Scale-up** | Expensive | Easy & inexpensive |
| 6 | **Dependence of reaction development** | Mainly on - - support - Linkers | Time |

# 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)
  - Fragment marking – Central core template and one or more R groups
  - 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

# Combichem Techniques (cont'd)

- 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

# APPROACHES TO PRODUCT BASED LIBRARY DESIGN

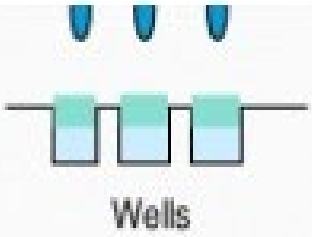
- Identify lists of potential reagents, filter them as needed, and enumerate the virtual library
- Subject virtual library to virtual screening to evaluate and score each structure
- Select reagents from results of virtual screening plus additional criteria (degree of structural diversity required, degree of similarity or dissimilarity to existing collections)
  - Usually done with optimization techniques (e.g., genetic algorithms or simulated annealing)

# Alternatives to Product-Based Library Design

- Molecule-based methods
  - Appropriate for targeted or focused libraries
  - Relatively fast, especially when combined with optimization based on 2D properties

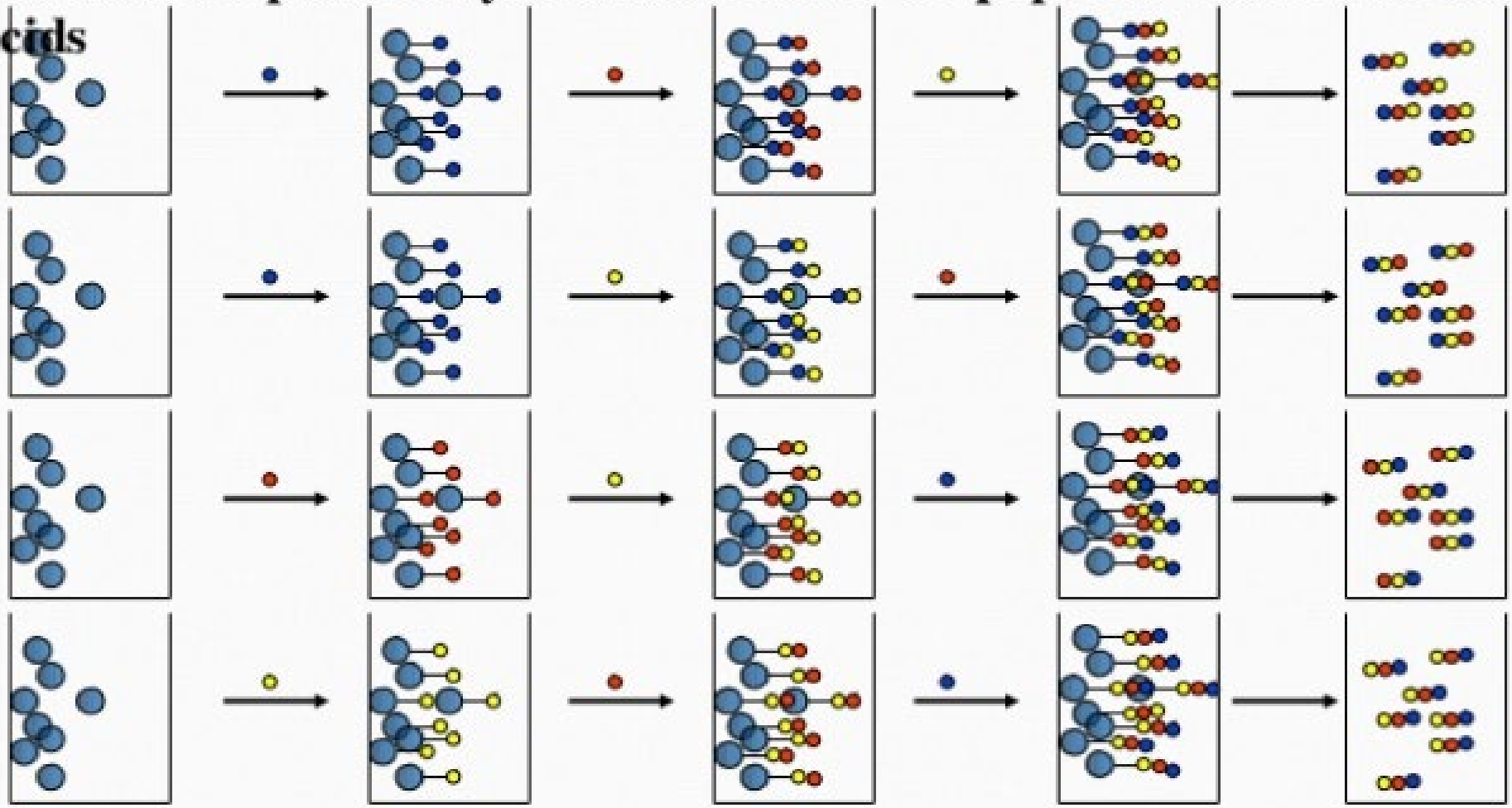
**E.g. Library Enumeration:**
https://jcheminf.biomedcentral.com/articles/10.1186/1758-2946-4-23/figures/2

## Automated parallel synthesis

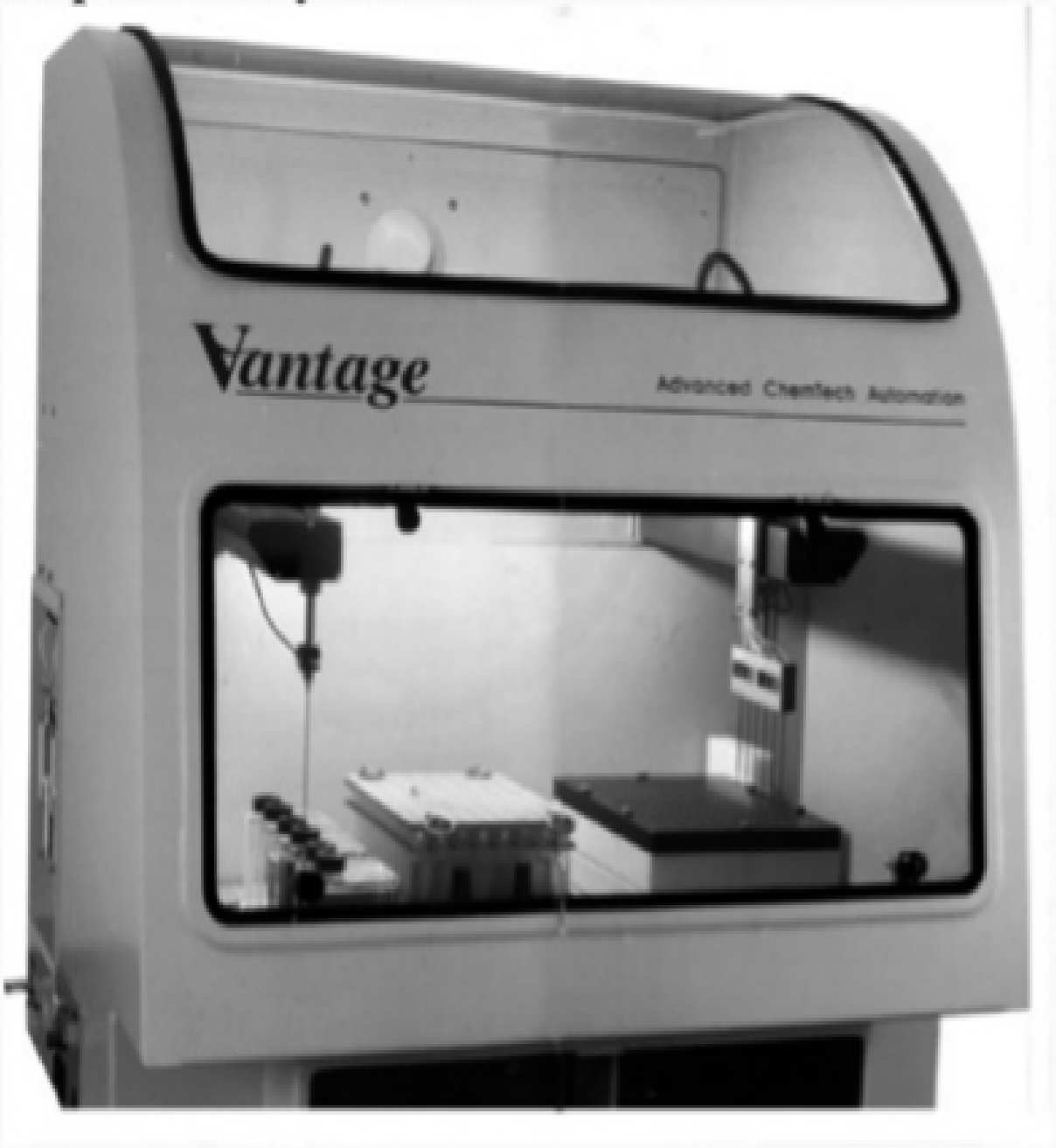



- Automated synthesisers are available with 42, 96 or 144 reaction vessels or wells
- Use beads or pins for solid phase support
- Reactions and work ups are carried out automatically
- Same synthetic route used for each vessel, but different reagents
- Different product obtained per vessel

Automated parallel synthesis of all 27 tri peptides from 3 amino acids



ETC




# AUTOMATED SYNTHETIC MACHINES

# 3. Mixed Combinatorial Synthesis

**Aims**

- 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

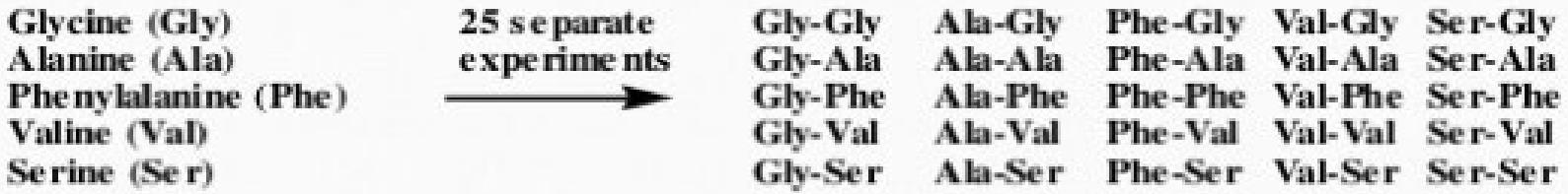
# 3. Mixed Combinatorial Synthesis

## The Mix and Split Method

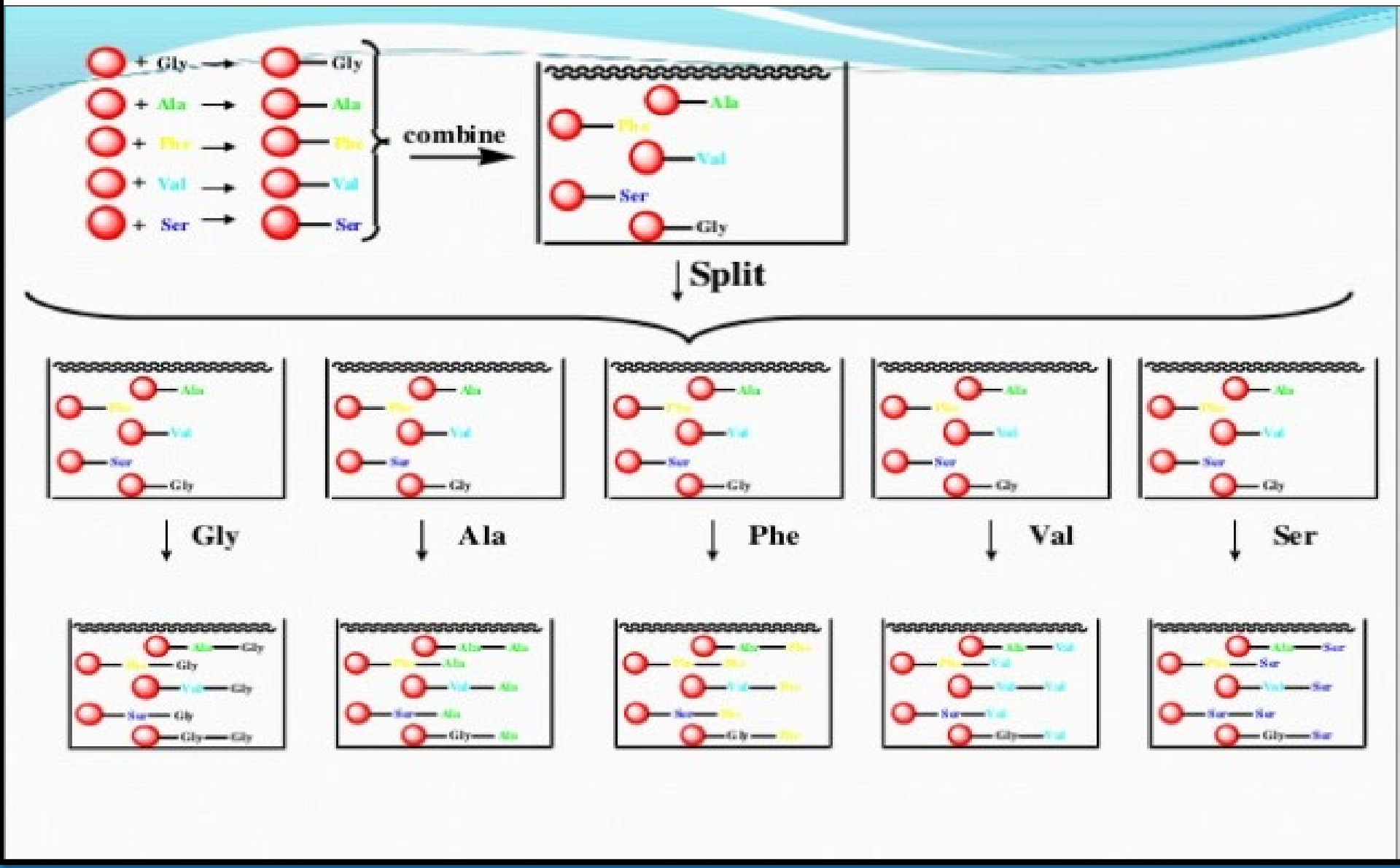
Example

- Synthesis of all possible dipeptides using 5 amino acids

•Standard methods would involve 25 separate syntheses




Combinatorial procedure involves five separate syntheses using a mix and split strategy


+ Gly
+ Ala
+ Phe
+ Val
+ Ser
Gly
Ala
Phe
Val
Ser
combine
Ala
Phe
Val
Ser
Gly
Split
Gly
Ala
Phe
Val
Ser

# Screening of Combinatorial Library

? Can be done in 2 ways: Virtual screening and Experimental real screening.

? **Virtual Screening:**

- Virtual screening uses **computational methods** to predict or simulate how a particular **compound interacts** with a given **target protein**.
- The 3 virtual screening methods used in modern drug discovery include

1. Molecular Docking,
2. Pharmacophore Mapping
3. QSAR/QSPR

**Disadvantages of virtual screening:**

1. Cannot replace real screening
2. Generated hits may be very difficult to chemically synthesize

# CONT'D

? **Experimental real screening:**

- Real screening approaches, such as high-throughput screening (HTS), can test the activity of hundreds of thousands of compounds experimentally, providing real results;

? **Disadvantage:**

1. These methods are far **more expensive** and
2. Slower than virtual screening methods.

# Other Screening Methods

- Most common assay to screen a combinatorial library is to determine the **binding of the library compounds to the target protein**.

---

- Other common assays are **functional assays** such as **biochemical and enzymatic assays, or cell-based assays**.
- **Cell-based assays** can be direct **cytotoxic assays**, **receptor-binding assays, or cell-signaling assays using cell lines with specific genetic reporter systems**.
- **Selection of screening methods** greatly depends on:

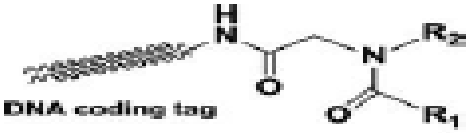
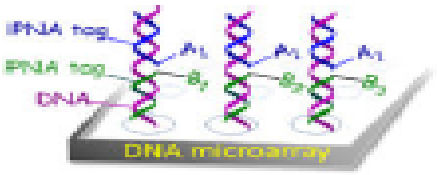
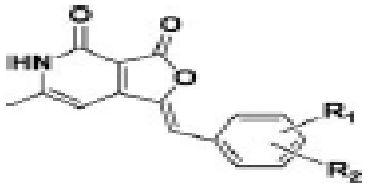
1. The **nature of the combinatorial libraries to be screened**.
2. **Position-addressable soluble libraries** prepared from **parallel synthesis** can be screened with **automated HTS** methods in **96-, 384-, and 1536-well plates**.
3. **Libraries on solid supports** (e.g. OBOC library) can be easily **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**.

# CONT'D

1. **Phage-display peptide libraries** can be screened with **bio-panning** or **limited cell-based functional assays, such as cell-binding and cellular uptake assays.**
2. **Structure-based virtual libraries** are screened **in silico**. Several new screening approaches are as follow:

https://www.researchgate.net/figure/COPA-library-synthesis-screening-structure-elucidation-and-validation **(OBOC-COPA approach Paper)**

**Phytochemical Screening using HPLC and MS.**

| Library type | Library structure | Screening | Note |
|---|---|---|---|
| DECL |   (76,230 compounds) 121 amines x 630 acids | affinity screening | a potent hit compound (X066/Y469) inhibited tankyrase 1 with an $IC_{50}$ of 250 nM |
| PNA-encoded small-molecule library |   (62,500 small-molecules) $A_{125} \times B_{500}$ | affinity screening | one ligand (2a) showed high affinity to Hsp70 with a $K_D$ of 1.58 nM |
| Spatially addressable solution-phase library |  (21 compounds) | plasmid relaxation assay | compound 4 is an inhibitor of LdTop1 with antileishmanial activity ($EC_{50}$ = 4.2 μM). It showed anti-protozoal activity against *Leishmania donovani* promastigote, but non-toxic against normal mammalian COS7 cells |
| OBOC COPA library |  (166,080 compounds) | protein-binding assay | compound 14a is the first non-covalent small-molecule ligand for the wild-type p53 DBD ($K_D$ = 10 μM) |
| OBOC peptidomimetic library |  (1,064 compounds) | cell-binding assay for $\alpha_4\beta_1$ integrin ligand (LLP2A) | LLP2A-aledronate for the treatment of osteoporosis, ongoing Phase I clinical trial |
| OBOC peptoid library |  (29,200 peptoids) | *in situ* releasable assay against *Cryptococcus Neoformans* | one peptoid (AEC5) showed comparable antifungal potency to existing clinical agents, excellent stability, and minimal cytotoxicity in mammalian cells |

# Encoding & Decoding of Chemical Libraries

- 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.
- For **mixture libraries in solution**, such as **positional-scanning libraries**, purification is needed to determine **the identity of the hits**.
- **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..**
- More than **one million codes** can be generated by using combinations of different methods, which are **highly stable** and reliable under **bioassay conditions**.

# Applications

- Application of combinatorial library methods in cancer research and drug discovery
- Building synthetic gene circuits from combinatorial libraries: screening and selection strategies
- Combinatorial library approaches for improving soluble protein expression in *Escherichia coli*
- Combinatorial library-based strategies to optimize proteins
- A Combinatorial Library Strategy for the Rapid Humanization of Anticarcinoma BR96 Fab
- Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery.
- Used in anti viral research

# Summary

- Combinatorial chemistry has significantly increased the number molecules that can be synthesized in a modern chemical laboratory Compound.

- The two main strategies for combinatorial library design are known as ***monomer-based selection*** and ***product-based selection***.
- A general strategy for product-based library design involves 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.

# SUMMARY

- Combinatorial chemistry has accelerated the development of a whole set of **combinatorial tools** comprising **combinatorial library design, efficient synthetic methods, reagents for library synthesis (including solid supported reagents), linkers, bilayer beads, library encoding and decoding strategies, HTS methods and equipment, etc.**
- A methodology could be used to generate large numbers of molecules using a scheme known as ***split–mix as well as parallel synthesis.***
- The two types of Combinatorial libraries such as Diverse as well as Focused libraries has helped to derived millions of compound.
- Many investigators, particularly in the pharmaceutical industry, are now working on **smaller target-focused solution-phase libraries** of compounds with **drug-like properties**, and incorporating **ADMET filters** and **structure-based drug design** approaches into library development .

# CONT'D

- However, for **novel lead discovery** against a large number of therapeutic targets, particularly for those targets with **little structural information**, the various **high diversity library methods** outlined in this mini-review will undoubtedly be invaluable.
- Another promising method in combinatorial chemistry is the use of nature's highly stable peptides, such as **cyclotides**, as scaffolds for library design.
- Random peptide loops can be grafted, chemically or recombinantly , into cysteine knots to form cyclotide libraries.
- Although **the initial high expectations of combinatorial chemistry** for drug discovery have yet to be realized, much has been learned over **the last 30 years.**
- Many **new chemical, biological, computational, and screening tools have been developed** with which the **limitations and strengths** of combinatorial chemistry are better understood and are in better position to truly leverage the power of **combinatorial technologies** for the discovery and development of **next-generation drugs**.

# References


- Molecular Modelling. Principles and Applications (2nd Edition) - Andrew R. Leach
- An Introduction to Chemoinformatics- Andrew. Leach, V. Gillet
- Chemoinformatics -Theory Practice and Products - Barry Bunin
- Chemoinformatics - Concepts, Methods and Tools for Drug Discovery- Bajorath
- Designing Combinatorial Libraries Optimized on Multiple Objectives- Valerie J.Gillet
- Combinatorial Cheimistry: https://www.slideshare.net/naresh2231989/combinatorial-chemistry-1
- https://www.slideshare.net/Touhidkhan5/combinatorial-chemistry-2
- https://slideplayer.com/slide/11261051/
- Fundamentals of Medicinal Chemistry by Gareth Thomas (From page no 116-127)