



Received on 14 March, 2013; received in revised form, 14 April, 2013; accepted, 21 June, 2013

COMBINATORIAL CHEMISTRY: A REVIEW

Anas Rasheed* and Rumana Farhat

Active Pharma Labs, Raja Enclave, # 404, Bhagyanagar Colony, Opp: R.S. Brothers, Beside K.S. Baker's, K.P.H.B. Colony, Hyderabad-72, Andhra Pradesh, India

Keywords:

Combinatorial chemistry, Libraries, Rapid synthesis, Computer simulation, CADD (Computer aided drug design)

Correspondence to Author:

Anas Rasheed

Active Pharma Labs, Raja Enclave, # 404, Bhagyanagar Colony, Opp: R.S. Brothers, Beside K.S. Baker's, K.P.H.B. Colony, Hyderabad-72, Andhra Pradesh, India

E-mail: kmehmoodi@yahoo.in

ABSTRACT: Combinatorial chemistry is a new methodology by which we can simultaneously synthesize a number of possible compounds that could produce simultaneously a very large number of compounds, called libraries. Combinatorial chemistry involves the rapid synthesis or the computer simulation of a large number of different but often structurally related molecules or materials. Combinatorial chemistry is especially common in CADD (Computer aided drug design) and can be done online with web based software, such as Molinspiration. In the past, chemists have traditionally made one compound at a time. For example compound A would have been reacted with compound B to give product AB, which would have been isolated after reaction work up and purification through crystallization, distillation, or chromatography. In contrast to this approach, combinatorial chemistry offers the potential to make every combination of compound A1 to Am with compound B1 to Bn. Although combinatorial chemistry has only really been taken up by industry since the 1990s, its roots can be seen as far back as the 1960s when a researcher at Rockefeller University, Bruce Merrifield, started investigating the solid-state synthesis of peptides.

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

Synthesis of molecules in a combinatorial fashion can quickly lead to large numbers of molecules. For example, a molecule with three points of diversity (R_1 , R_2 , and R_3) can generate $N_{R_1} \times N_{R_2} \times N_{R_3}$ possible structures, where N_{R_1} , N_{R_2} , and N_{R_3} are the numbers of different substituents utilized.

Although combinatorial chemistry has only really been taken up by industry since the 1990s, its roots can be seen as far back as the 1960s when a researcher at Rockefeller University, Bruce Merrifield, started investigating the solid-phase synthesis of peptides.

Professor Pieczenik, a colleague of Nobel Laureate Merrifield, synthesized the first combinatorial library. US Patent 5,866,363. In the 1980s researcher H. Mario Geysen developed this technique further,

QUICK RESPONSE CODE 	DOI: 10.13040/IJPSR.0975-8232.4(7).2502-16
	Article can be accessed online on: www.ijpsr.com

creating arrays of different peptides on separate supports, but not a combinatorial library based on random synthesis¹.

In its modern form, combinatorial chemistry has probably had its biggest impact in the pharmaceutical industry. Researchers attempting to optimize the activity profile of a compound create a 'library' of many different but related compounds. Advances in robotics have led to an industrial approach to combinatorial synthesis, enabling companies to routinely produce over 100,000 new and unique compounds per year¹⁷.

In order to handle the vast number of structural possibilities, researchers often create a 'virtual library', a computational enumeration of all possible structures of a given pharmacophore with all available reactants. Such a library can consist of thousands to millions of 'virtual' compounds. The researcher will select a subset of the 'virtual library' for actual synthesis, based upon various calculations and criteria.

Finding of novel drug is a complex process. Historically, the main source of biologically active compounds used in drug discovery programs has been natural products, isolated from plant, animal or fermentation sources.

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 drugs¹⁸.

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 combi-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 discovery².

History of Combinatorial Chemistry: Combinatorial chemistry was first conceived about 15 years ago - although it wasn't called that until the early 1990s.

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

Although combinatorial chemistry has only really been taken up by industry since the 1990s, its roots can be seen as far back as the 1960s when a researcher at Rockefeller University, Bruce Merrifield, started investigating the solid-state synthesis of peptides¹⁹.

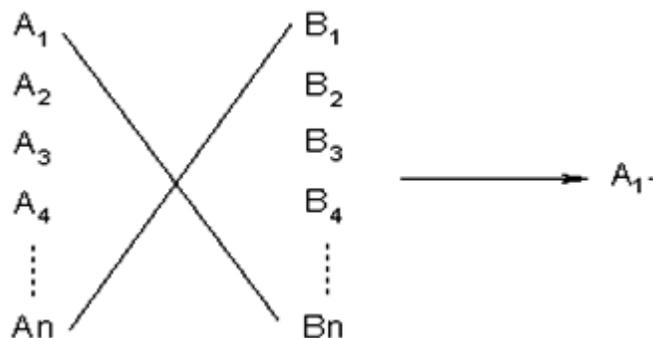
In the past decade there has been a lot of research and development in combinatorial chemistry applied to the discovery of new compounds and materials. This work was pioneered by P.G. Schultz et al. in the mid-nineties (*Science*, 1995, 268: 1738-1740) in the context of luminescent materials obtained by co-deposition of elements on a silicon substrate. Since then the work has been pioneered by several academic groups as well as industries with large R&D programs (Symyx Technologies, GE, etc)³.

Principle of Combinatorial Chemistry: Combinatorial chemistry is a technique by which large numbers of structurally distinct molecules may be synthesized in a time and submitted for pharmacological assay. The key of combinatorial chemistry is that a large range of analogues is synthesized using the same reaction conditions, the same reaction vessels. In this way, the chemist can synthesize many hundreds or thousands of compounds in one time instead of preparing only a few by simple methodology⁴.

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



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



The range of combinatorial techniques is highly diverse, and these products could be made individually in a parallel or in mixtures, using either solution or solid phase techniques. Whatever the technique used the common denominator is that productivity has been amplified beyond the levels that have been routine for the last hundred years⁵.

Combinatorial chemistry (or CombiChem) is an innovative method of synthesizing many different substances quickly and at the same time. Combinatorial chemistry contrasts with the time-consuming and labor intensive methods of traditional chemistry where compounds are synthesized individually, one at a time. While combinatorial chemistry is primarily used by organic chemists who are seeking new drugs, chemists are also now

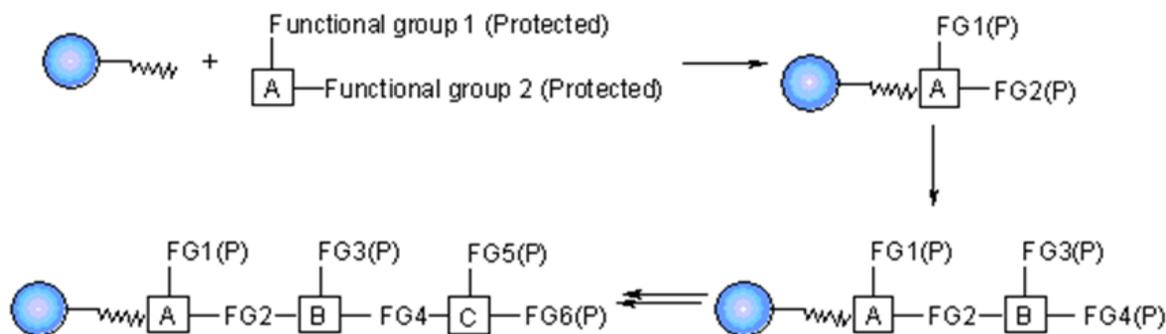
applying combinatorial chemistry to other fields such as semiconductors, superconductors, catalysts and polymers²⁰.

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

Over the last few years, the combinatorial chemistry has emerged as an exciting new paradigm for the drug discovery. In a very short time the topic has become the focus of considerable scientific interest and research efforts²¹.

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

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



- 1) A cross linked, insoluble polymeric material that is inert to the condition of synthesis;
- 2) Some means of linking the substrate to this solid phase that permits selective cleavage of

some or all of the product from the solid support during synthesis for analysis of the extent of reaction(s), and ultimately to give the final product of interest;

- 3) A chemical protection strategy to allow selective protection and deprotection of reactive groups.

Merrifield developed a series of chemical reactions that can be used to synthesise proteins. The direction of synthesis is opposite to that used in the cell. The intended carboxy terminal amino acid is anchored to a solid support. Then, the next amino acid is coupled to the first one. In order to prevent further chain growth at this point, the amino acid, which is added, has its amino group blocked. After the coupling step, the block is removed from the primary amino group and the coupling reaction is repeated with the next amino acid²².

The process continues until the peptide or protein is completed. Then, the molecule is cleaved from the solid support and any groups protecting amino acid side chains are removed. Finally, the peptide or protein is purified to remove partial products and products containing errors⁶.

Synthesis of Combinatorial Library:

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

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

A simple example of a 3 x 3 x 3 library gives all 27 possible combinations of trimeric products. X, Y and Z could be amino acids, in which case the final products would be tripeptides, but more generally

they could be any type of monomeric unit or chemical precursor. It can be seen that the mix and split procedure finally gives three mixtures each consisting of nine compounds each, and there are several ways of progressing these compounds to biological screening. Although the compounds can be tested whilst still attached to the bead, a favored method is to test the compounds as a mixture following cleavage from the solid phase. Activity in any given mixture reveals the partial structure of active compounds within the library, as the residue coupled last (usually the N-terminal residue) is unique to each mixture. Identification of the most active compound relies on deconvoluting the active mixtures in the library through further synthesis and screening.²³

In the example where the active structure is YXY, the mixture with Y at the terminal position will appear as the most active. Having retained samples of the intermediate dimers on resin (so-called "recursive" deconvolution) addition of Y to each of the three mixtures will give all nine compounds with Y at the terminal position, and the second position defined by the mixture. The most active mixture here defines the middle position of the most active trimer to be residue X. Finally, the three individual compounds can be independently resynthesized and tested to reveal both the most potent compound and also some structure activity relationship data.²⁴

In contrast, Lam et al. tested a family of peptides whilst still attached to the resin bead solid phase. Nineteen amino acids were incorporated into pentapeptides to generate a library of almost two and a half million compounds. By using a colorimetric assay, beads bearing peptide sequences that bound tightly to the protein streptavidin or to an antibody raised against β -endorphin were revealed by visual inspection. Bead picking using micromanipulation isolated the beads, and the active peptide structures were determined by microsequencing²⁵.

A modification of this method has allowed screening of such libraries in solution. Linkers have been devised that allow several copies of the library compounds to be released sequentially. Using this method it is possible to identify an active mixture using a solution assay, and then return to the beads that produced these compounds, and redistribute them into smaller mixtures for retest.

By repeatedly reducing the mixture size, ultimately to single compounds, the bead containing the most

potent sequence may be identified and the peptide product sequenced⁷.

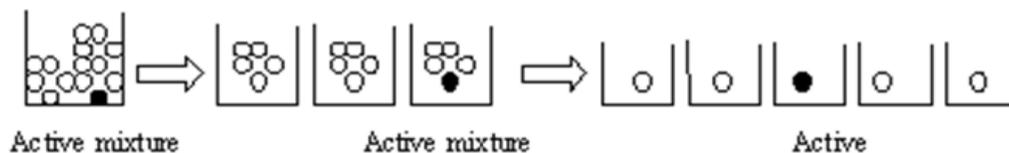


FIG. 4: PEPTIDE POTENT SEQUENCE

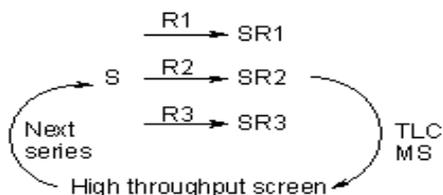
Combinatorial synthesis in Solution: Despite the focus on the use of solid-phase techniques for the synthesis of combinatorial libraries, there have been few examples where libraries have successfully been made and screened in solution.

The benefit of preparing libraries on resin beads has been explained as offering advantages in handling, especially where a need to separate excess reagents from the reaction products is attached to the resin. In most of case a simple filtration effects a rapid purification and the product are ready to further synthetic transformation. But it should be remember that using solid phase chemistry brings several disadvantages as well. Clearly the range of chemistry available on solid phase is limited and it is difficult to monitor the progress of reaction when the substrate and product are attached to the solid phase²⁶.

Indeed some groups have expressed a preference for solution libraries because there is no prior requirement to develop workable solid phase coupling and linking techniques. The difficulty is purifying large number of compounds without sophisticated automated processes⁸.

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

Parallel analogue Synthesis and Screening



Orthodox Analogue Synthesis and Screening

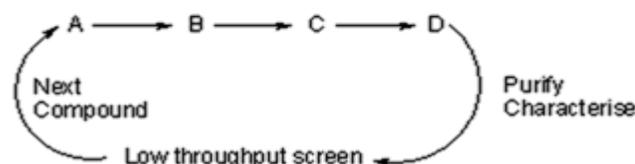


FIG. 5: PARALLEL AND ORTHODOX SYNTHESIS

Orthodox synthesis usually involves a multistep sequence, e.g. from A through to the final product D, which is purified and fully characterized before screening. The next analogue is then designed, guided by the biological activity of the previous compound, prepared, and then screened. This process is repeated to optimise both activity and selectivity²⁷.

In contrast parallel analogue synthesis involves reaction of a substrate S with multiple reactants, R1, R2, R3 ... Rn, to produce a compound library of n individual products SR1, SR2, SR3 ... SRn. The library is screened, usually without purification, and with only minimal characterization of the individual compounds, using a rapid throughput screening technique⁹.

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

Resins for Solid Phase synthesis: In solid phase support synthesis, the solid support is generally based on a polystyrene resin. The most commonly used resin supports for SPS include spherical beads of lightly cross linked gel type polystyrene (1–2% divinylbenzene) and poly(styrene-oxethylene) graft copolymers which are functionalised to allow attachment of linkers and substrate molecules.

Each of these materials has advantages and disadvantages depending on the particular application.

- **Cross-linked Polystyrene:** Lightly cross-linked gel type polystyrene (GPS) (Figure) has been most widely used due to its common availability and inexpensive cost. GPS beads which are functionalised with chloromethyl-, aminomethyl-, and a variety of linkers are commercially available from a variety of sources. A prominent characteristic of GPS beads is their ability to absorb large relative volumes of certain organic solvents (swelling). This swelling causes a phase change of the bead from a solid to a solvent-swollen gel, and therefore, the reactive sites are accessed by diffusion of reactants through a solvent-swollen gel network.

In solvents, which swell the polymer well, the gel network consists of mostly solvent with only a small fraction of the total mass being polymer backbone. This allows relatively rapid diffusional access of reagents to reactive sites within the swollen bead. In solvents, which do not swell the polymer, the cross-linked network does not expand and the diffusion of reagents into the interior of the bead is impeded²⁸.

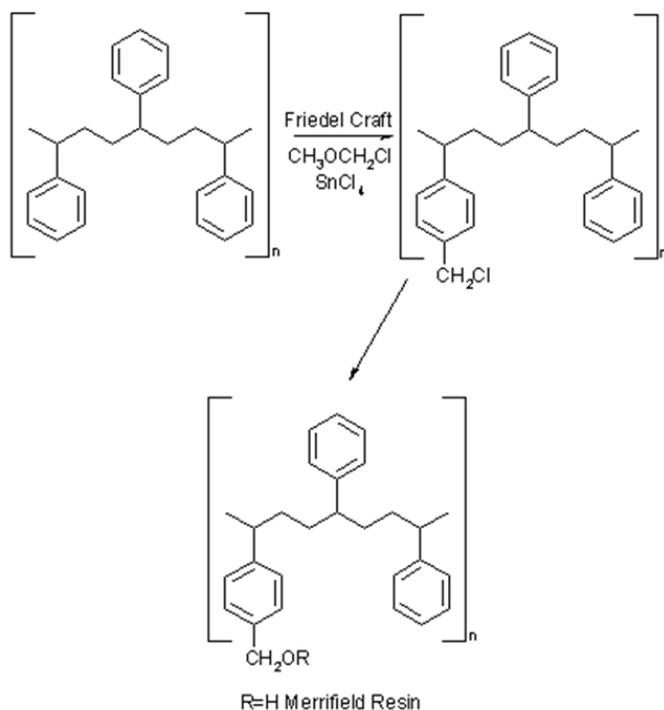


FIG. 6: DIFFUSION OF REACTANTS

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

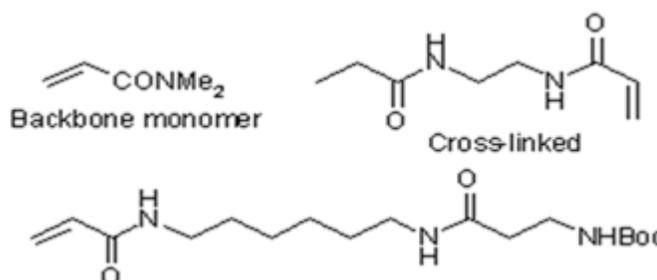


FIG. 7 BACKBONE MONOMER WITH FUNCTIONAL GROUPS

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

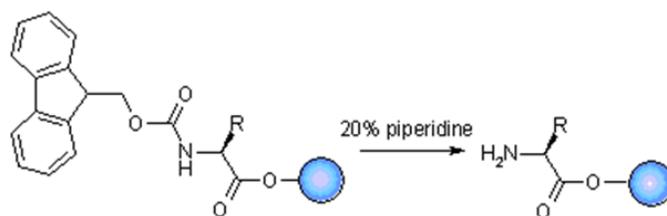


FIG. 8: PIPERIDINE BASE USED FOR PROTECTION

- **Linkers:** The group that joins the substrate to the resin bead is an essential part of solid phase synthesis. The linker is a specialised protecting group, in that much of the time, the linker will tie up a functional group, only for it to reappear at the end of the synthesis. The linker must not be affected by the chemistry used to modify or extend the attached compound. And finally the cleavage step should proceed readily and in a good yield. The best linker must allow attachment and cleavage in quantitative yield¹⁰.

Combinatorial Libraries: Two groups have recently disclosed solution libraries prepared in mixtures. In each case the groups from Glaxo and Pirrung have synthesised dimeric compounds using amide, ester or carbamate bond-forming reactions.

Every library compound was prepared twice in mixtures of different composition. Testing all of these mixtures allows identification of likely active compounds without the need to resynthesise every compound in an active mixture.

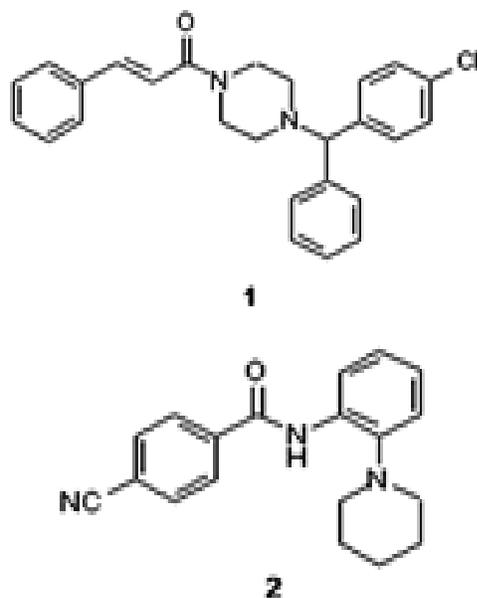


FIG. 9: CHEMICAL GROUPS USE IN BOND FORMING REACTIONS

In the glaxo example, 40 acid chlorides were reacted with 40 amines or alcohols to give amides or esters respectively in two sets. In the first set, each acid chloride (A) was reacted with a stoichiometric amount of an equimolar mixture of all 40 nucleophiles (N1-40). In the second set each amine or alcohol (N) was reacted with an equimolar mixture of the acid chlorides (A1-40). The 80 mixtures of 40 components each were screened against a wide variety of pharmacological targets, and a positive result from any sample identified half of the structure of a likely active dimeric compound. Weak leads against the neurokinin-3-receptor 1 and matrix metalloproteinase - 1 and 2 were detected¹¹.

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 resynthesis and rescreening offers a solution, but as it can be slow and requires a further dedication of synthetic and screening

resource, there have been a number of new methods devised where information concerning the active compound may be carried on the bead in the form of a "tag".

The synthetic efficiency of the split synthesis technique can be contrasted with the technical difficulties encountered when analysing the resulting libraries. For example, the simple split synthesis scenario outlined above results in a library consisting of 10 pools of 1 000 compounds each. These compounds can be cleaved into solution and screened as soluble pools, or the ligands can remain attached to the beads and screened in immobilised form. Neither scenario is ideal for several reasons. Because of limitations on solubility, the concentration of the individual compounds present in soluble pools must be correspondingly diminished as the pool size increases – 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 assay. Because of this, biologically active pools are often iteratively resynthesised and reassayed as increasingly smaller subsets until activity data are obtained on homogenous compounds²⁹.

This process of iterative resynthesis is time consuming, requires multiple bioassays, and the deconvolution of a single pool to its individual constituents typically require more synthetic steps than were required to prepare the parent library. When multiple pools are active, the deconvolution process becomes additively complex if each active subset is chosen for resynthesis.

In addition to being inefficient, positive selection strategies such as iterative deconvolution ignore negative biological information, the knowledge of which is often important in the design of subsequent libraries. In some instances, bead-based split synthesis libraries can be successfully assayed with the ligands still 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 labelled receptors, or anti-receptors antibodies similarly labelled with a reporter molecule, that can be employed to "label" active beads. Beads thus marked are physically removed and 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.¹²

- **DNA – Based Encoding:** One of the first reported successful ligand encoding strategies exploited oligo-deoxyribonucleic acid (DNA) as the surrogate analyte. This DNA encoding concept had in fact been demonstrated in some of the first combinatorial library preparation methods ever reported – those utilising

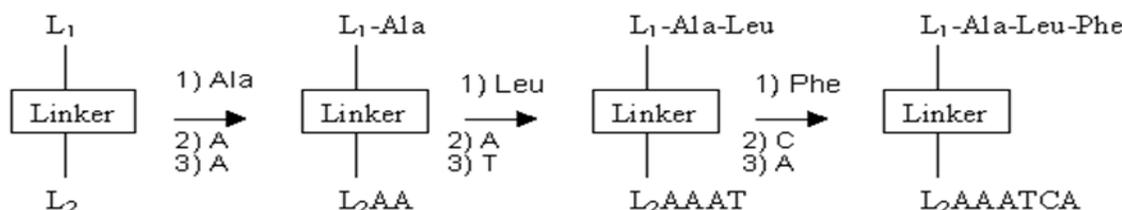


FIG. 10: SEQUENCING OF ENCODED DNAs

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

- **Peptide Tag:** Zuckermann *et al.*, at Chiron recognised that peptides could be employed as tag since their information content could be extracted with high sensitivity via Edman degradation and sequencing. Since the Edman degradation requires a free N-terminus, this peptide as code strategy could also be used to

filamentous phage particles. In this approach, libraries of peptides are prepared biochemically from the cloning and expression of random sequence oligonucleotides. Pools of oligonucleotides encoding the peptides of interest are interested into an appropriate expression system, where upon translation the resulting peptides are 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.

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.

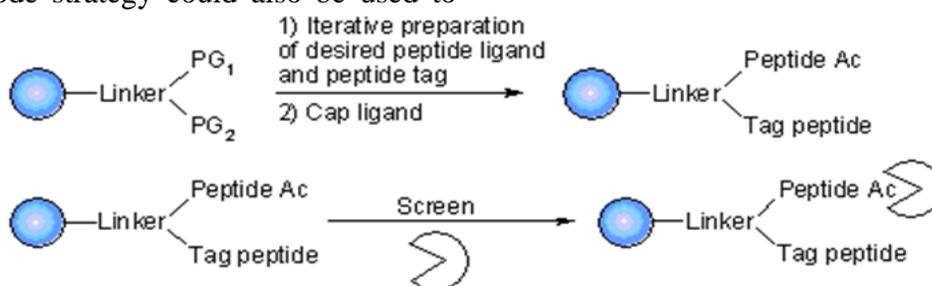


FIG. 11: BINDING AND CODING OF PEPTIDES

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

The above peptide and DNA encoding techniques are not ideal because of the chemical lability of these oligomers. This places a severe restriction on the scope of the synthetic techniques that may be applied during library synthesis, and restricts the synthesis of more pharmaceutically attractive small organic molecules³⁰.

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

It can be seen that in each of these cases above, the use of a tagging group allows the synthesis of any type of compound within the library. The tagging molecules can encode for any building block and any synthetic transformation. Furthermore, given the uncertainties of much synthetic chemistry, the tag may be looked upon as not so much encoding a specific compound structure, but encoding instead a synthetic procedure. Thus, even if the intended compound was not made but biological activity was detected, the tagging system facilitates a replication of the synthetic steps employed in producing the active compound, and thus aids structure determination³¹.

Drug Discovery: Drug discovery and development is an expensive process due to the high costs of R&D and human clinical tests. The average total cost per drug development varies from US\$ 897 million to US\$ 1.9 billion. The typical development time is 10-15 years.

R&D of a new drug involves the identification of a target (e.g. protein) and the discovery of some suitable drug candidates that can block or activate the target.

Clinical testing is the most extensive and expensive phase in drug development and is done in order to obtain the necessary governmental approvals. In the US drugs must be approved by the Food and Drug Administration (FDA).

- **R&D – Finding the Drug:** One of the most successful ways to find promising drug candidates is to investigate how the target protein interacts with randomly chosen compounds, which are usually a part of compound libraries. This testing is often done in so called high-throughput screening (HTS) facilities. Compound libraries are commercially available in sizes of up to several millions of compounds. The most promising compounds obtained from the screening are called hits – these are the compounds that show binding activity towards the target. Some of these hits are then promoted to lead compounds – candidate structures which are further refined and modified in order to achieve more favorable interactions and less side-effect.
- **Drug Discovery Methods:** The following are methods for finding a drug candidate, along with their pros and cons:
 1. Virtual screening (VS) based on the computationally inferred or simulated real screening;

The main advantages of this method compared to laboratory experiments are:

- Low costs, no compounds have to be purchased externally or synthesized by a chemist;
- It is possible to investigate compounds that have not been synthesized yet;
- Conducting HTS experiments is expensive and VS can be used to reduce the initial number of compounds before using HTS methods;
- Huge amount of chemicals to search from. The number of possible virtual molecules available for VS is exceedingly higher than the number of compounds presently available for HTS;

The disadvantage of virtual screening is that it cannot substitute the real screening.

2. The real screening, such as high-throughput screening (HTS), can experimentally test the activity of hundreds of thousands of compounds against the target a day. This method provides real results that are used for drug discovery. However, it is highly expensive.³²

- **Virtual Screening in Drug Discovery:** Computational methods can be used to predict or simulate how a particular compound interacts with a given protein target. They can be used to assist in building hypotheses about desirable chemical properties when designing the drug and, moreover, they can be used to refine and modify drug candidates. The following three virtual screening or computational methods are used in the modern drug discovery process: Molecular Docking, Quantitative Structure-Activity Relationships (QSAR) and Pharmacopoeia Mapping.
- **Quantitative Structure – Activity:**
 - **Relationships (QSAR):** As mentioned in the previous paragraph it is necessary to know the geometrical structure of both the ligand and the target protein in order to use molecular docking methods. QSAR (Quantitative Structure-Activity Relationships) is an example of a method which can be applied regardless of whether the structure is known or not.

QSAR formalizes what is experimentally known about how a given protein interacts with some tested compounds. As an example, it may be known from previous experiments that the protein under investigation shows signs of activity against one group of compounds, but not against another group.

In terms of the lock and key metaphor, we do not know what the lock looks like, but we do know which keys work, and which do not. In order to build a QSAR model for deciding why some compounds show sign of activity and others do not, a set of descriptors are chosen. These are assumed to influence whether a given compound will succeed or fail in binding to a given target. Typical descriptors are parameters such as molecular weight, molecular volume, and electrical and thermodynamical properties. QSAR models are used for virtual screening of

compounds to investigate their appropriate drug candidates descriptors for the target¹³.

Screening:

- **Solid Support Combinatorial Chemistry In Lead:**
 - **Discovery and SAR Optimization:** The widespread acceptance and use of high throughput screening technologies for the purposes of drug discovery and development has created an unprecedented demand for small organic molecules. The requirements for;
 - (i) Large numbers of diverse and novel chemical entities and
 - (ii) Methods to rapidly optimize the compounds or 'hits' found by screening may not be met by medicinal chemistry teams employing traditional synthetic methods.

Alternatively, combinatorial chemistry in solution or on solid support, is being developed to increase the efficiency of organic syntheses. Furthermore, successful applications of such methods leading to the discovery of therapeutic candidates have been reported.

- **The Ontogen approach:** Hardware and software platforms have been designed and developed to significantly increase the number of compounds that a synthetic organic/ medicinal chemist can prepare in a given period of time. Thus, libraries of compounds can be created for biological screening and perform medicinal chemistry optimization strategies ultimately leading to compounds for human clinical trials

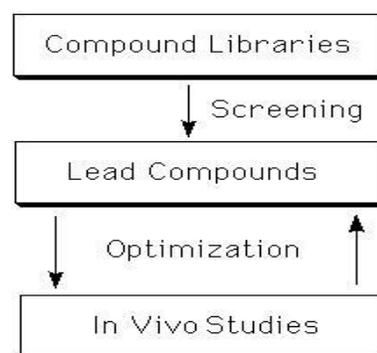


FIG. 12: STRATEGIES LEADING FOR CLINICAL TRIALS

The synthesis of complex small molecules on solid support using different organic reactions such as multi-step sequential substitution reactions, multi-component condensation reactions and pharmacophore modifying reactions has been accomplished.

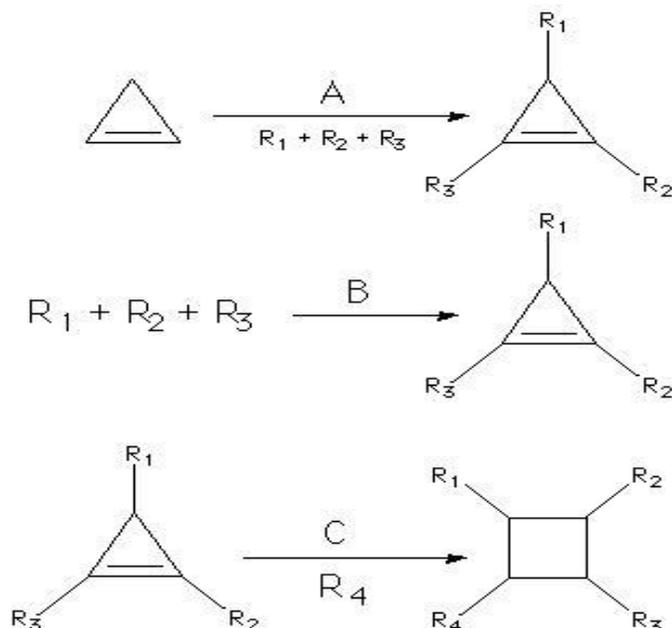


FIG. 12A): SEQUENTIAL SUBSTITUTION, B) MULTI-COMPONENT CONDENSATION ARRAY (MCCA), C) PHARMACOPHORE TRANSFORMATION

In this fashion complex, diverse, non-peptide, chemical compound libraries such as:

Beta-lactams; hydantoin imides and thioimides; imidazoles; N-acyl-alpha-amino amides, esters, acids; oxazoles; phosphonates (alpha-hydroxy, alpha-amino, alpha-acylamino); phosphinates; pyrroles; tetra-substituted 5 membered ring lactams; tetra-substituted 6 membered ring lactams and tetrazoles are synthesized on solid support using a wide range of organic transformations including: acylations; aldol condensations; alkylations; Claisen couplings; Heck reactions; heterocycle forming reactions such as condensations, dipolar cycloadditions, annulations, etc.; Michael additions; Mitsunobu couplings; multicomponent condensation reactions and reductions.

The final products are cleaved into a standard 96 well microtiter plate, one compound per well. Each plate can be directly submitted for high throughput screening as well as quantitative and semi-quantitative analysis in order to assess purity, identity and yield of each compound synthesized¹⁴.

- **Design of Pharmacophore:** The design of the pharmacophore basis of a particular library is driven by the nature of the biological target of interest. The following types of information are considered, if available: The biology of the target enzyme or receptor;

The nature of substrate; the mechanism of target-substrate interaction; related literature information; 3-D structural information.

In general, the method of synthesis is designed to allow full control over each of the individual substituents. This is accomplished through the selection of the starting materials or inputs (charge, electron withdrawing/donating, hydrogen bond donor/acceptor, hydrophobicity, steric bulk, etc.). In general the inputs are chosen to be commercially available. On occasion, inputs are synthesized for specific cases, fully aware that input synthesis has the potential to dramatically reduce the efficiencies of the combinatorial approach.

- **High-Throughput Screening (HTS):** High-throughput screening (HTS) is a method for scientific experimentation especially used in drug discovery and relevant to the fields of biology and chemistry. Using robotics, data processing and control software, liquid handling devices, and sensitive detectors, High-Throughput Screening allows a researcher to quickly conduct millions of chemical, genetic or pharmacological tests. Through this process one can rapidly identify active compounds, antibodies or genes which modulate a particular biomolecular pathway. The results of these experiments provide starting points for drug design and for understanding the interaction or role of a particular biochemical process in biology.
- **Assay Plate Preparation:** The key labware or testing vessel of HTS is the microtiter plate: a small container, usually disposable and made of plastic that features a grid of small, open divots called wells. Modern (circa 2008) microplates for HTS generally have either 384, 1536, or 3456 wells. These are all multiples of 96, reflecting the original 96 well microplate with 8 x 12 9mm spaced wells. Most of the wells contain experimentally useful matter, depending

on the nature of the experiment. This could be an aqueous solution of dimethyl sulfoxide (DMSO) and some other chemical compound, the latter of which is different for each well across the plate. It could also contain cells or enzymes of some type (The other wells may be empty, intended for use as optional experimental controls.)

A screening facility typically holds a library of stock plates, whose contents are carefully catalogued, and each of which may have been created by the lab or obtained from a commercial source. These stock plates themselves are not directly used in experiments; instead, separate assay plates are created as needed. An assay plate is simply a copy of a stock plate, created by pipetteing a small amount of liquid (often measured in nanoliters) from the wells of a stock plate to the corresponding wells of a completely empty plate¹⁵.

- **Reaction observation:** To prepare for an assay, the researcher fills each well of the plate with some logical entity that he or she wishes to conduct the experiment upon, such as a protein, or an animal embryo. After some incubation time has passed to allow the biological matter to absorb, bind to, or otherwise react (or fail to react) with the compounds in the wells, measurements are taken across all the plate's wells, either manually or by a machine. Manual measurements are often necessary when the researcher is using microscopy to (for example) seek changes or defects in embryonic development caused by the wells' compounds, looking for effects that a computer could not easily determine by itself.

Otherwise, a specialized automated analysis machine can run a number of experiments on the wells (such as shining polarized light on them and measuring reflectivity, which can be an indication of protein binding). In this case, the machine outputs the result of each experiment as a grid of numeric values, with each number mapping to the value obtained from a single well. A high-capacity analysis machine can measure dozens of plates in the space of a few minutes like this, generating thousands of experimental datapoints very quickly.

Depending on the results of this first assay, the researcher can perform follow up assays within the same screen by "cherrypicking" liquid from the source wells that gave interesting results (known as "hits") into new assay plates, and then re-running the experiment to collect further data on this narrowed set, confirming and refining observations.

- **Automation Systems:** Automation is an important element in HTS's usefulness. Typically, an integrated robot system consisting of one or more robots transports assay-microplates from station to station for sample and reagent addition, mixing, incubation, and finally readout or detection. An HTS system can usually prepare, incubate, and analyze many plates simultaneously, further speeding the data-collection process. HTS robots currently exist which can test up to 100,000 compounds per day. Automatic colony pickers pick thousands of microbial colonies for high throughput genetic screening. The term uHTS or ultra-high throughput screening refers to screening in excess of 100,000 compounds per day.
 - **Experimental Design and Data Analysis:** With the ability of rapid screening of diverse compounds (such as small molecules or siRNAs) to identify active compounds, HTS has led to an explosion in the rate of data generated in recent years. Consequently, one of the most fundamental challenges in HTS experiments is to glean biochemical significance from mounds of data, which relies on the development and adoption of appropriate experimental designs and analytic methods for both quality control and hit selection.
- HTS research is one of the fields which have a feature described by Eisenstein as follows: soon, if a scientist does not understand some statistics or rudimentary data-handling technologies, he or she may not be considered to be a true molecular biologist and thus will simply become a dinosaur.
- **Quality Control:** High-quality HTS assays are critical in HTS experiments. The development of high-quality HTS assays requires the integration of both experimental and computational approaches for quality control (QC).

Three important means of QC are (i) good plate design, (ii) the selection of effective positive and negative chemical/biological controls, and (iii) the development of effective QC metrics to measure the degree of differentiation so that assays with inferior data quality can be identified.

A good plate design helps to identify systematic errors (especially those linked with well position) and determine what normalization should be used to remove/reduce the impact of systematic errors on both QC and hit selection.

Effective analytic QC methods serve as a gatekeeper for excellent quality assays. In a typical HTS experiment, a clear distinction between a positive control and a negative reference such as a negative control is an index for good quality. Many quality assessment measures have been proposed to measure the degree of differentiation between a positive control and a negative reference. Signal-to-background ratio, signal-to-noise ratio, signal window, assay variability ratio, and Z-factor have been adopted to evaluate data quality. Strictly Standardized Mean Difference (SSMD) has recently been proposed for assessing data quality in HTS assays.

- **Hit Selection:** A compound with a desired size of effects in an HTS screen is called a hit. The process of selecting hits is called hit selection. The analytic methods for hit selection in screens without replicates (usually in primary screens) differ from those with replicates (usually in confirmatory screens). For example, the z-score method is suitable for screens without replicates whereas the t-statistic is suitable for screens with replicate. The calculation of SSMD for screens without replicates also differs from that for screens with replicates.

For hit selection in primary screens without replicates, the easily interpretable ones are average fold change, mean difference, percent inhibition, and percent activity. However, they do not capture data variability effectively. The z-score method or SSMD, which can capture data variability based on an assumption that every compound has the same variability as a negative reference in the screens. However, outliers are

common in HTS experiments, and methods such as z-score are sensitive to outliers and can be problematic. Consequently, robust methods such as the z*-score method, SSMD*, B-score method, and quantile-based method have been proposed and adopted for hit selection.

In a screen with replicates, we can directly estimate variability for each compound; consequently, we should use SSMD or t-statistic that does not rely on the strong assumption that the z-score and z*-score rely on. One issue with the use of t-statistic and associated p-values is that they are affected by both sample size and effect size. They come from testing for no mean difference, thus are not designed to measure the size of compound effects. For hit selection, the major interest is the size of effect in a tested compound. SSMD directly assesses the size of effects. SSMD has also been shown to be better than other commonly used effect.

The population value of SSMD is comparable across experiments and thus we can use the same cutoff for the population value of SSMD to measure the size of compound effects.

Techniques for increased throughput and efficiency unique distributions of compounds across one or many plates can be employed to increase either the number of assays per plate, or to reduce the variance of assay results, or both. The simplifying assumption made in this approach is that any N compounds in the same well will not typically interact with each other, or the assay target, in a manner that fundamentally changes the ability of the assay to detect true hits.

For example, imagine a plate where compound A is in wells 1-2-3, compound B is in wells 2-3-4, and compound C is in wells 3-4-5. In an assay of this plate against a given target, a hit in wells 2, 3, and 4 would indicate that compound B is the most likely agent, while also providing three measurements of compound B's efficacy against the specified target. Commercial applications of this approach involve combinations in which no two compounds ever share more than one well, to reduce the (second-order) possibility of interference between pairs of compounds being screened.

- **Recent advances:** In March 2010 research was published demonstrating an HTS process allowing 1,000 times faster screening (100 million reactions in 10 hours) at 1 millionth the cost (using 10–7 times the reagent volume) than conventional techniques using drop-based microfluidics. Drops of fluid separated by oil replace microplate wells and allow analysis and hit sorting while reagents are flowing through channels.

In 2010 researchers developed a silicon sheet of lenses that can be placed over microfluidic arrays to allow the fluorescence measurement of 64 different output channels simultaneously with a single camera. This process can analyze 200,000 drops per second.¹⁹

- **Increasing Lab Utilization of HTS:** HTS is a relatively recent innovation, made lately feasible through modern advances in robotics and high-speed computer technology. It still takes a highly specialized and expensive screening lab to run an HTS operation, so in many cases a small-to-moderately sized research institution will use the services of an existing HTS facility rather than set up one for it. There is a trend in academia to be their own drug discovery enterprise (High-throughput screening goes to school). These facilities, which normally are only found in industry, are now increasingly be found as well at universities. UCLA for example, features an HTS laboratory (Molecular Screening Shared Resources (MSSR, UCLA) which can screen more than 100,000 compounds a day on a routine basis.

The University of Illinois also has a facility for HTS, as does the University of Minnesota. The Rockefeller University has an open access (infrastructure) HTS Resource Center HTSRC (The Rockefeller University, HTSRC) which offers a library of over 165,000 compounds. Northwestern University's High Throughput Analysis Laboratory supports target identification, validation, assay development, and compound screening.

In the United States, the National Institute of Health or NIH has created a nationwide consortium of small molecule screening centers that has been recently funded to produce

innovative chemical tools for use in biological research. The Molecular Libraries Screening Center Network or MLSCN performs HTS on assays provided by the research community, against a large library of small molecules maintained in a central molecule repository¹⁶.

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

It is a method for reacting a small number of chemicals to produce simultaneously a very large number of compounds, called libraries, which are screened to identify useful products such as drug candidates and a method in which very large numbers of chemical entities are synthesized by condensing a small number of reagents together in all combinations defined by a small set of reactions.

REFERENCES:

1. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D, 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science* 251:767-73. PMID 1990438.
2. E. V.Gordeeva et al. "COMPASS program - an original semi-empirical approach to computer-assisted synthesis" *Tetrahedron*, 48 (1992) 3789.
3. X. -D. Xiang et al. "A Combinatorial Approach to Materials Discovery" *Science* 268 (1995) 1738.
4. J.J. Hanak, J. Mater. Sci, *Combinatorial Characterization*, 1970, 5, 964-971.
5. *Combinatorial methods for development of sensing materials*, Springer, 2009. ISBN 978-0-387-73712-6.
6. V. M. Mirsky, V. Kulikov, Q. Hao, O. S. Wolfbeis. Multiparameter High Throughput Characterization of Combinatorial Chemical Microarrays of Chemosensitive Polymers. *Macromolec. Rap. Comm.* 2004, 25, 253-258.
7. H. Koinuma et al. "Combinatorial solid state materials science and technology" *Sci. Technol. Adv. Mater.* 1 (2000).
8. Andrei Ionut Mardare et al. "Combinatorial solid state materials science and technology" *Sci. Technol. Adv. Mater.* 9 (2008) 035009.
9. *Applied Catalysis A*, Volume 254, Issue 1, Pages 1-170 (10 November 2003).
10. J. N. Cawse *et al.*, *Progress in Organic Coatings*, Volume 47, Issue 2, August 2003, Pages 128-135.
11. *Combinatorial Methods for High-Throughput Materials Science*, MRS Proceedings Volume 1024E, Fall 2007.

12. Combinatorial and Artificial Intelligence Methods in Materials Science II, MRS Proceedings Volume 804, Fall 2004.
13. QSAR and Combinatorial Science, 24, Number 1 (February 2005).
14. J. N. Cawse, Ed., Experimental Design for Combinatorial and High Throughput Materials Development, John Wiley and Sons, 2002.
15. D. Newman and G. Cragg "Natural Products as Sources of New Drugs over the Last 25 Years" *J Nat Prod* 70 (2007) 461.
16. M. Feher and J. M. Schmidt "Property Distributions: Differences between Drugs, Natural Products, and Molecules from Combinatorial Chemistry" *J. Chem. Inf. Comput. Sci.*, 43 (2003) 218.
17. E. Campian, J. Chou, M. L. Peterson, H. H. Saneii, A. Furka, R. Ramage, R. Epton (Eds) In Peptides 1996, 1998, Mayflower Scientific Ltd. England, 131.
18. A. Furka, F. Sebestyen, J. Gulyás, Computer made electrophoretic peptide maps. Proc. 2nd Int. Conf. Biochem. Separations, Keszthely, Hungary, pp. 35-42 (1988).
19. Lehn, J.-M.; Ramstrom, O. Generation and screening of a dynamic combinatorial library. PCT. Int. Appl. WO 20010164605, 2001.
20. Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J.-L.; Sanders, J. K. M.; Otto, S. (Sep 2006). "Dynamic combinatorial chemistry". *Chem. Rev.* 106 (9): 3652–3711.
21. K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski, R. J. Knapp *Nature* 1991, 354, 82; and its correction: K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski, R. J. Knapp *Nature* 1992, 360, 768.
22. H. M. Geysen, R. H. Meloen, S. J. Barteling *Proc. Natl. Acad. Sci. USA* 1984, 81, 3998.
23. M. H. J. Ohlmeyer, R. N. Swanson, L. W. Dillard, J. C. Reader, G. Asouline, R. Kobayashi, M. Wigler, W. C. Still *Proc. Natl. Acad. Sci. USA* 1993, 90, 10922.
24. E. Campian, M. Peterson, H. H. Saneii, A. Furka *Bioorg. & Med. Chem. Letters* 1998, 8, 2357.
25. *Applied Catalysis A*, Volume 254, Issue 1, Pages 1-170 (10 November 2003).
26. T. Carell, E. A. Winter, J. Rebek Jr. *Angew. Chem. Int. Ed. Engl.* 1994, 33, 2061.
27. V. Nikolaiev, A. Stierandova, V. Krchnak, B. Seligman, K. S. Lam, S. E. Salmon, M. Lebl *Pept. Res.* 1993, 6, 161.
28. A. Stierandova, V. Krchnak, B. Seligman, K. S. Lam, S. E. Salmon, M. Lebl *Pept. Res.* 1996, 7, 191.
29. D. Newman and G. Cragg "Natural Products as Sources of New Drugs over the Last 25 Years" *J Nat Prod* 70 (2007) 461.
30. Leeson, P. D. et al. (2007). "The influence of drug-like concepts on decision-making in medicinal chemistry". *Nat. Rev. Drug Disc.* 6 (11): 881–890.
31. John Faulkner D, Newman DJ, Cragg GM (February 2004). "Investigations of the marine flora and fauna of the Islands of Palau". *Nat Prod Rep* 21 (1): 50–76.
32. Hopkins, A. L., Groom, C. R. and Alexander, A. (2004). "Ligand efficiency: a useful metric for lead selection". *Drug Discovery Today* 9(10): 430–431.

How to cite this article:

Rasheed A and Farhat R: Combinatorial Chemistry. *Int J Pharm Sci Res* 2013; 4(7); 2502-2516. doi: 10.13040/IJPSR.0975-8232.4(7).2502-16

All © 2013 are reserved by International Journal of Pharmaceutical Sciences and Research. This Journal licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.