# Concepts of Combinatorial Chemistry and Combinatorial Technologies

**Article** *in* Chemicke Listy · January 2000

CITATIONS
15

READS
2,587

**3 authors:**

Stanislav Miertus
International Centre For Applied Research And Sustainable Technology
**218** PUBLICATIONS **12,021** CITATIONS
SEE PROFILE

Giorgio Fassina
Xeptagen
**169** PUBLICATIONS **3,182** CITATIONS
SEE PROFILE

Pierfausto Seneci
University of Milan
**115** PUBLICATIONS **1,609** CITATIONS
SEE PROFILE

# CONCEPTS OF COMBINATORIAL CHEMISTRY AND COMBINATORIAL TECHNOLOGIES

**STANISLAV MIERTUS[a], GIORGIO FASSINA[b], and P. F. SENECI[c]**

*[a]International Centre for Science and High Technology of the United Nations Industrial Development Organization, ICS--UNIDO, Area Science Park, Building L2, Padriciano 99, 34012, Trieste, [b]Biopharmaceuticals, TECNOGEN S.C.p.A, Parco Scientifico, 81015 Piana di Monte Verona (CE), [c]Lead Discovery Department, GlaxoWellcome Medicines Research Centre, Via Fleming 4, 37100 Verona, Italy*

Received January 21, 2000

Key words: combinatorial chemistry, combinatorial technologies

## Introduction

Combinatorial Chemistry and Combinatorial Technology (CC/CT) are a new interdisciplinary field joining computer-assisted combinatorial chemistry with automated parallel synthesis of chemical "libraries" followed by automated screening.

The main purpose of computer assisted combinatorial chemistry is to generate thousands structurally diverse compounds as libraries, maximising their diversity, which are then considered in an experimental parallel automated synthesis and screening on the basis of their properties. The key issue is to integrate all important steps of CC/CT in a single, multidisciplinary approach.

This nascent technology already produced more new compounds in just a few years than the pharmaceutical industry did in its entire history.

Combinatorial chemistry has turned traditional chemistry upside down. It required chemists not to think in terms of synthesizing single, well-characterized compounds but in terms of simultaneously synthesizing large populations of compounds. It also required that those people involved with information management and computational chemistry systems address these same issues as the chemists.

While CC/CT had the greatest impact on the drug industry, combinatorial methods need not be restricted to pharmaceutical applications. Whenever high numbers of compounds have to be prepared for testing, this technique can be used. For example, agricultural research, new materials and new catalysts research and development stand to gain from this technology. However, for the time being, main emphasis is still on pharmaceutical research. Most major pharmaceutical companies are active in the field. It is generally accepted that the method has a great potential for finding leads in the drug discovery process where the technology is expected to contribute to the reduction in time and cost.

Many countries have emphasized the urgent need to get acquainted with Combinatorial Technologies in order to enable local enterprises to remain competitive and economically viable in the coming decades and gain expertise on application practices of combinatorial technology. In view of global competition, CC/CT together with molecular modeling may be considered as powerful tools for the implementation and/or the increase of a country's capabilities in drug design, agrochemistry, new materials and new catalysts. The above considerations become even more significant if it is taken into account that many countries have abundant natural resources which are presently well below their proper exploitation. Combinatorial chemistry can enhance the potential of these resources.

## Combinatorial Technologies

Combinatorial approaches were originally based on the premise that the probability of finding a molecule in a random screening process is proportional to the number of molecules subjected to the screening process. In its earliest expression, the primary objective of combinatorial chemistry focused on the simultaneous generation of large numbers of molecules and on the simultaneous screening of their activity. Following this approach, the success rate to identify new leads is greatly enhanced, while the time required is considerably reduced.

The development of new processes for the generation of collection of structurally related compounds (libraries) with the introduction of combinatorial approaches has revitalized random screening as a paradigm for drug discovery and has raised enormous excitement about the possibility of finding new and valuable drugs in short times and at reasonable costs (Figure 1).

However the advent of this new field in drug discovery did not obscure the importance of "classical" medicinal chemistry approaches, such as computer-aided rational drug design and QSAR for example, but catalyzed instead their evolution to complement and to be integrated with combinatorial technologies.

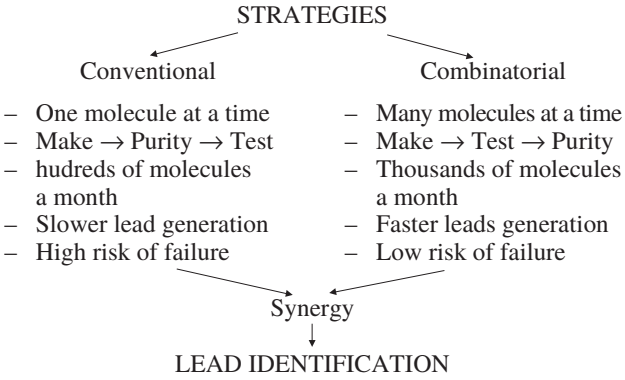


Fig. 1. **Principal characteristics of conventional vs. combinatorial strategy of drug discovery**

PEPTIDES CHEMICAL DIVERSITY

Given a linear amino acid sequence of n residues
$X_1 - X_2 - X_3 - X_4 - \ldots - X_n$
the total number of different peptides obtainable equals to:

$y^n$

n = peptide length
y = number of different amino acids used in the synthesis (usually 18)

| | | | |
|---|---|---|---|
| n = 3 | 5.832 peptides | n = 5 | 1.889.568 peptides |
| n = 4 | 104.976 peptides | n = 6 | 34.012.224 peptides |

Fig. 2. **Number of compounds (peptides) generated by combinatorial approach**

The word "combinatorial" appeared in the scientific literature at the beginning of the '90, but the generation of the first combinatorial libraries can be dated back to the beginning of the '80. The first reports dealt with the simultaneous production of collection of chemically synthesized peptides, produced by solid phase methods on solid supports[2-6]. Peptides were particularly suited for combinatorial synthesis given the well established synthetic protocols available, the great number of different molecules attainable, and the potential to generate leads of biological and pharmaceutical value (Figure 2).

The use of peptide libraries was greatly accelerated by the introduction of biological methods for library preparation, such as the phage display technology, which provided interesting advantages over the synthetic counterpart[7,8]. At the same time, the first papers on the generation of oligonucleotide libraries appeared in the literature[9,10], suggesting thus the possibility to extend the applicability of combinatorial approaches even to other classes of synthetic or natural oligomeric compounds, such as carbohydrates. There are many important biologically active glyco-conjugate drugs whose carbohydrate constituents are associated with the molecular mechanism by which these drugs exhibit their effect. Because of this, the exploration of carbohydrate molecular diversity has the potential for identifying novel agents with enhanced potency (Figure 3).

## The Need for Combinatorial Technologies

Drug discovery in the past has been based traditionally about the random screening of collections of chemically synthesized compounds or extracts derived from natural sources, such as microorganisms, bacteria, fungi, plants, of terrestrial or marine origin or by modifications of chemicals with known physiological activities (Figure 4).

SOURCES OF MOLECULAR DIVERSITY

- Plant extracts
- Microbial extracts
- Collection of chemical compounds (synthetic)
- Oligonucleotide libraries (biological or synthetic)
- Oligosaccharide libraries
- Chemical compounds libraries (synthetic)
- Peptide libraries (biological or synthetic)

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. 4. **The basic sources of molecular diversity and definition of libraries**

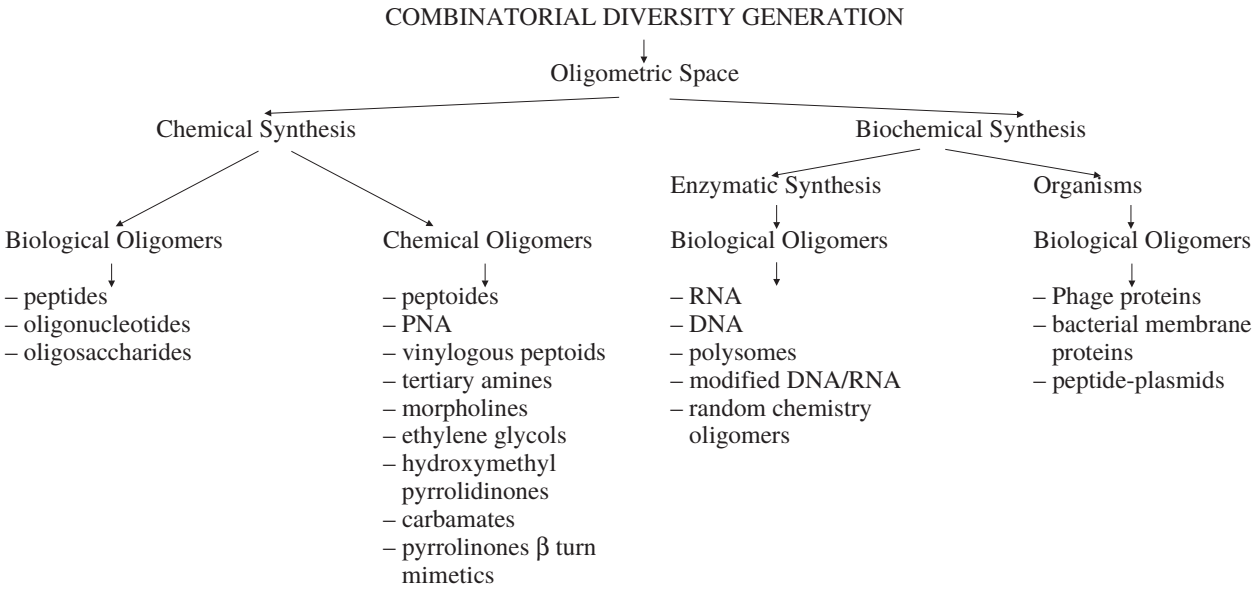


Fig. 3. **Diversity of compounds generated by combinatorial approach**

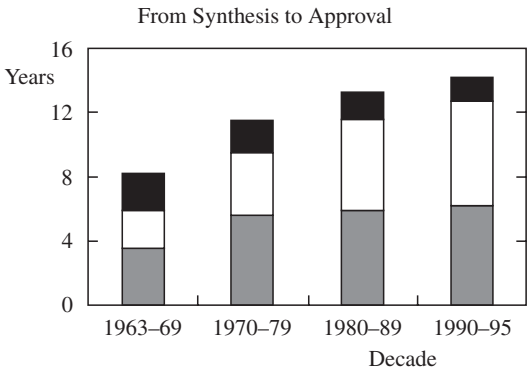


Fig. 5. **Time needed for new drugs development in the last decades**; ■ FDA, □ clinical, ▒ preclinical

FACTORS AFFECTING STRATEGY CHANGES IN DRUG DISCOVERY

1. Biotechnology (genomics):
   provides molecular targets of therapeutic relevance (receptors, hormones, proteins).
2. Combinatorial Technology:
   provide the possibility of generating huge collections of molecules which are simultaneously produced with a built – in decoding capability.
3. High Throughput Screening (HTS):
   provides the possibility of handling many assays at the same time.

Fig. 6. **Key factors affecting drug discovery**

This approach has resulted in many important drugs, however the ratio of novel to previously discovered compounds has diminished with time. In addition, this process is very time consuming and expensive. A limiting factor is the restricted number of molecules available or extract samples to be screened, since the success rate in obtaining useful lead candidates depends directly upon the number of samples tested. Chemical synthesis of new chemical entities often is a very laborious task, and additional time is required for purification and chemical characterization. The average cost of creating a new molecular entity in a pharmaceutical company is around 7500 USD/compound[1]. Generation of natural extracts, while very often providing interesting new molecular structures endowed with biological properties leads to mixtures of different compounds at different concentrations, making thus activity comparisons very difficult. In addition, once activity is found in a specific assay, the extract needs to be fractionated in order to identify the active component. Quite often, the chemical synthesis of natural compounds is extremely difficult, making thus the lead development to in a new drug a very complex task. The time and cost needed for the development of new drugs have been increasing steadily during the past three decades (Figure 5).

Estimated costs for introducing a new drug in the market now reach around 300–400 millions USD, and this process takes around 12–14 years after their original discovery. This increase in time and cost is due mainly to the extensive clinical studies of new chemical entities required by competent regulatory agencies, such as the Food and Drug Administration, and to a lesser extent to the increased costs associated to research. The time and cost required for clinical and preclinical evaluation of new drugs is not likely to decrease in the near future, and as a consequence, a key issue for pharmaceutical companies to stay in the market has been to increase the number of new drugs in their development pipeline. While the pharmaceutical industry was demanding more rapid and cost effective approaches to lead discovery, the advent of new methodologies in molecular biology, biochemistry, and genetics, led to the identification and production of an ever increasing number enzymes, proteins, receptors, involved in biological processes of pharmacological relevance, but also to good candidates for the development of screening assay, complicated even more this scenario. The introduction of combinatorial technologies provided an unlimited source of new compounds, capable to satisfy all these needs (Figure 6).

## Applicability of Combinatorial Technologies

Up to now many active compounds have been selected to date following combinatorial methodologies, and a considerable number of those have progresses to in clinical trials. However, combinatorial chemistry and related technologies for producing and screening large number of molecules find useful applications also in other industrial sectors not necessarily related to pharmaceutical industry. Emerging fields of application of combinatorial technologies are the diagnostic, the down-stream processing, the catalysis, and the new material sectors. In the first case, CC can be successfully applied to the identification of previously unknown epitopes recognized by antibodies in biological fluids associated to pathological conditions.

Combinatorial Technologies have been applied also to the identification of new macromolecules endowed with catalytic activity for reactions where natural enzymes are inactive. This application even if still at the early stage, is calling considerable attention from the industrial sector, since the availability of new enzymes may reduce the production costs of many chemicals.

## Combinatorial Tools

A broad variety of new synthesis and screening methods are currently grouped under the term combinatorial. These methods include parallel chemical synthesis and testing of multiple individual compounds or compounds mixtures in solution, synthesis, and testing of compounds on solid supports, and biochemical or organism-based synthesis of biological oligomers coupled to selection and amplification strategies. Fully automated instruments for the synthesis and for the screening of libraries of compounds are integrated tools in combinatorial technologies, as well as computer assisted approaches for library design. A very important class of molecular libraries is represented by peptide libraries. Peptides are particularly suitable for the construction of libraries since a high degree of structural diversification can be easily achieved simply by varying the peptide sequence length or by the introduction of different amino acids other than those naturally

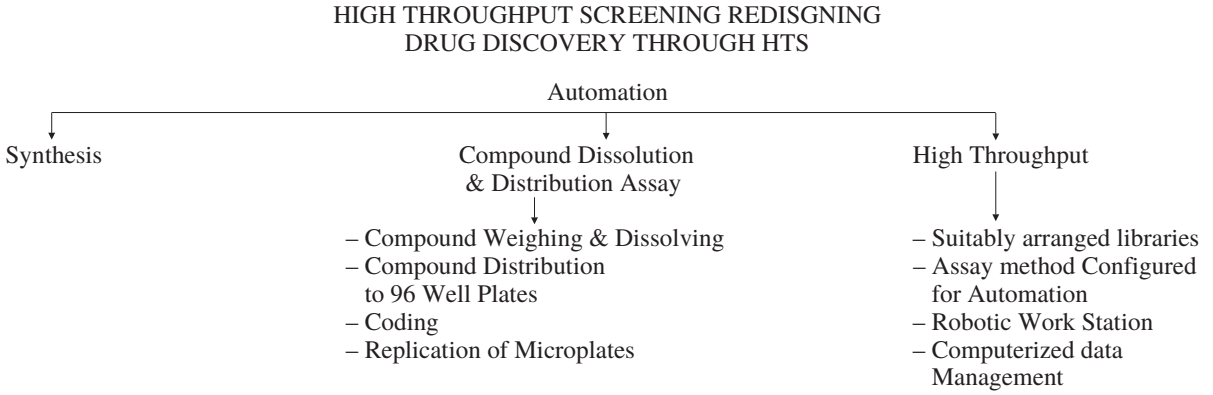


Fig. 7. **Role of automation in CC/CT**

occurring. The number of different peptides obtainable by a combinatorial approach is governed by the simple formula:

$$N = b^x$$

where $N$ is the total number of molecules obtainable, $b$ is the number of residues used in the construction of the library and $x$ is the sequence length. Generation of synthetic peptide libraries generally follow the divide:couple:recombine process[2,5] (DCR), where different aliquots of resin for solid phase synthesis are treated separately with solutions containing different activated amino acids, which are after coupling, recombined, mixed, and then divided again in different aliquots. The process is then repeated several times until the desired length of the library is accomplished.

Peptide libraries could be synthesized even manually in the laboratories where Combinatorial chemistry should be applied without investments.

Another important aspect of combinatorial chemistry is the analytical characterization of molecular libraries. Since a considerable number of different molecules are tested separately or in combination, analytical data should indicate that all the expected components are occurring with a comparable degree of purity. Amino acid analysis by TOF-MALDI mass spectrometry for peptide libraries quality control is often used. The amino acid analysis is useful mainly for the characterization of amino acid-based libraries (peptides, benzodiazepines, hydantoins).

Different techniques such as Electrospray (ES), Matrix Assisted Laser Desorption Ionization (MALDI), Fast Atom Bombardment (FAB) and tandem mass spectrometry have been successfully used to evaluate the composition and purity of synthetic peptide mixtures, but there are no limitations for their use with purely organic libraries. When interfaced to HPLC or capillary electrophoresis the ES becomes the most powerful method for the characterization of even very complex mixtures, since the combination of the two techniques allows the identification of compounds having chemical properties very similar. MALDI is a very a sensitive method and can be used when very small amounts of sample are available.

In combinatorial chemistry, due to the high number of chemical manipulations required to synthesize libraries of compounds and to the high number of screening steps, automation is unavoidable (Figure 7). Many research groups, both in academia and industrial settings are developing automated instruments specifically tailored to these needs, and this technology field is acquiring an extremely important role for the development of combinatorial technologies for the next millennium. However, semi-automated instruments requiring little investment may be constructed in research lab with a low budget.

COMBINATORIAL CHEMISTRY

| ON SOLID PHASE | IN SOLUTION |
|---|---|
| – large excess of reagents allowed | – all organic reactions can be used |
| – multistep synthesis allowed | – no chemistry assessment |
| – easy workup-isolation | – no linker/cleavage chemistry |
| – mix and split possible | – unlimited product quantities |

Fig. 8. C**haracteristics of solid phase and solution phase combinatorial chemistry**

The screening steps required to decipher the active sequence from a molecular library are strictly related to the type of library used, to the synthesis or preparation cycles needed, and to the kind of activity wanted. Molecular libraries can be prepared following chemical or biological approaches. For the first case, libraries can be prepared free in solution or anchored to solid supports, and for these two different situations different screening procedures are required. Resin-released libraries can be conveniently used for the search of molecules able to interfere in solution with a specific biochemical recognition event, such as in the case of hormone-receptor, antigen-antibody, or inhibitor-enzyme interactions. Screening can be conveniently performed evaluating the inhibitory activity of sublibraries, where the nature of at least one functional group of the library is known in a predetermined position, on the assay under consideration. In combinatorial chemistry many different types of libraries can be produced, by using solid phase or solution phase methods (Figure 8).

## Computer-Assisted Combinatorial Chemistry and Molecular Design

The different technologies and strategies used in the production of combinatorial libraries are now so well developed

that it is easy to plan synthetic schemes for the generation of a huge number of compounds. Since the rate at which compounds can be screened does constitute a limitation to the use of combinatorial technologies, it is important to be selective about the compounds which are synthesized (Figure 9).

Computational methods are very valuable from this point of view to assist in the design of combinatorial libraries. The main requirement for lead generation is often to maximize the range of structural types within the library with the expectation that a broad range of activities will result. As a consequence, diversity analysis is an important aspect of library design. The diversity of libraries may be measured by the use of similarity or dissimilarity indexes which make intermolecular comparisons possible. Measures of chemical similarity have been developed for similarity searching in chemical databases. The calculation of the similarity between two molecules involves the characterization of the molecules by using chemical/structural descriptors, and then the application of similarity coefficients to quantify the similarity.

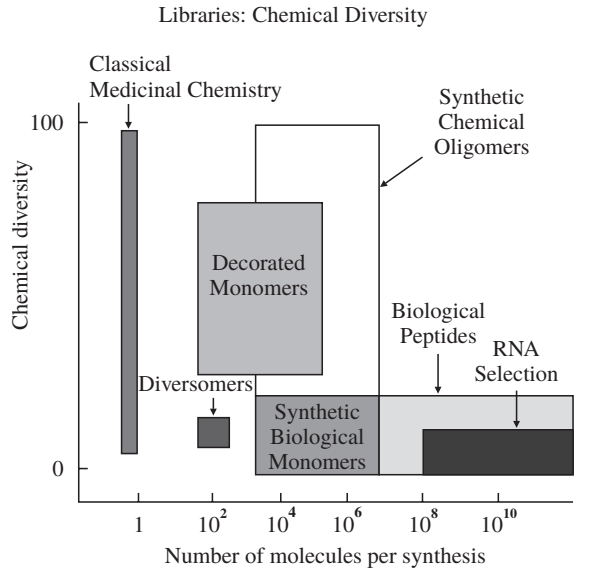


Fig. 9. **Chemical diversity and number of molecules produced by various concepts of synthesis**

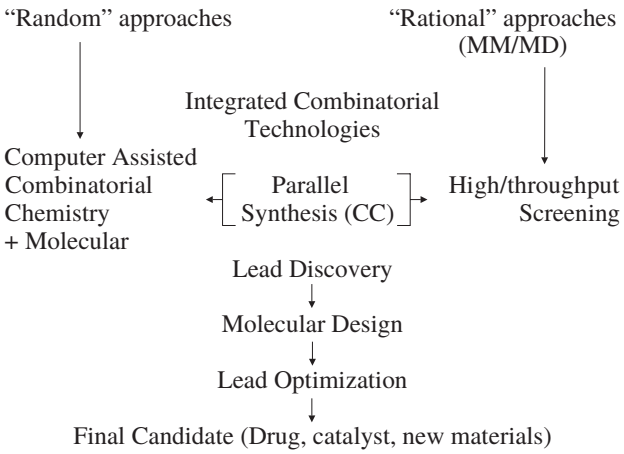


Fig. 10. **Integration of combinatorial technology and molecular design**

Moreover, Molecular Design uses various sophisticated techniques of computer chemistry/molecular mechanics and dynamics (QSAR, neural networking, etc.) for rational search of candidate structures with desired properties and activity with applications in pharmaceutical and chemical companies, and its high tech. and reasonable cost requirements render it suitable for use in SMEs in developing and in transition countries. A clear tendency is observed in industrial companies developing new compounds (drugs, agrochemicals, catalysts, new specialty chemicals, new polymer materials and other) to develop both the strategies (i.e. combinatorial technology and molecular design) in synergy (Figure 10).

## Biological Methods

Biological methods for library preparation are mainly limited to peptide or oligonucleotide libraries. For peptide libraries, methods are based on the construction of a pool of clones each one expressing a different peptide on its surface (Figure 11).

The peptides are fused to proteins normally expressed on the surface of the microorganism used. Phage display libraries are the most commonly used. Screening is accomplished by incubation of the target molecule, adsorbed to a solid support, with the phage population. Active phages will bind the target even after extensive washing steps. Target-bound phages are isolated and propagated by infection of *E. coli* and subjected to an additional round of adsorption to the immobilized target. This procedure increases both the number of active phages and the stringency of selection, since harsher condition may be employed in the washing steps to reduce the number of non--specifically bound phages. As for the case of synthetic libraries, iterative cycles of adsorption, washing, elution and propagation in *E. coli* are performed to enrich the phage population in the active or in few active sequences. Active phages may then be subjected to DNA sequencing in order to decode the active peptide sequence.

The use of biological display libraries for the isolation of peptide ligands is an interesting alternative to chemical libraries. Since 1985 (Ref.[18]), when this technique was first published, many fields of research have benefited from its use.

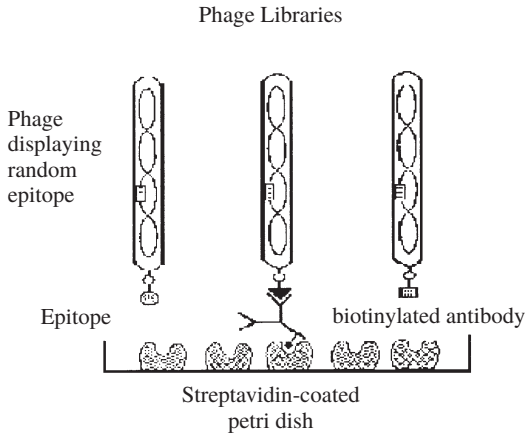


Fig. 11. **Principle of phage libraries preparation**

Web resources relevant to Combinatorial Chemistry and Combinatorial Technologies

There is a sort of information explosion accompanying the development of Combinatorial Chemistry and Combinatorial Technologies and an almost exponential growth of publications and patents in the field. At the same time, several web sites have been established providing updated information. The most important are the following:
http://www.5z.com/divinfo/spos.html

This site is probably the best collection of papers/patents/presentations/abstracts regarding combinatorial technologies in all its aspects and applications.

Theoretically this link may bring the experienced scientist almost everywhere in the field of Combinatorial Technologies, but we will highlight some other relevant web addresses:
http://www.combinatorial.com/

This site is based on a well known book titled "The Combinatorial Index", written by B. Bunin and published by Academic Press. The TOC is electronically available and the new articles appearing are grouped consequently as for their area. The site is devoted to synthetic organic libraries.
http://www.netsci.org/Resources/CCYP/top.html

This is the site of Network Sciences, which hosts an electronic publication related to Combinatorial Technologies and High Throughput Screening among other topics.
http://www.stemcorp.com/organic.htm

Website of STEM corporation, providing many links such as: Discussion group for Organic Synthesis, Discussion group for Drug Discovery and Synthesis, a Lab-Robotics Interest Group and others.
http://www.dl.ac.uk/CDS/sps.html

Database of Solid Phase Synthesis from Synopsis.
http://pubs.acs.org/journals/jcchff/index.html
http://www2.interscience.wiley.com/issn/0006-3592/
http://www.bscipubl.demon.co.uk/cchts/
http://www.wkap.nl./journalhome.html/1381-1991

Websites of of the major scientific magazines in Combinatorial Technologies: respectively Journal of Combinatorial Chemistry (to start in Jan 1999, ACS), Biotechnology and Bioengineering (special issues on Combinatorial Technologies, John Wiley and Sons), Combinatorial Chemistry (Bentham Science Publishers) and High Throughput Synthesis and Molecular Diversity (Kluwer).

Information on the ICS-UNIDO programme on Combinatorial Chemistry and Combinatorial Technologies, together with a database on CC/CT can be found on
http://www.ics.trieste.it

Last but not least, several books and review papers dealing with the topics of CC/CT were recently published[20-25].

## Conclusions

Combinatorial approaches have been introduced from the beginning in the drug discovery field, given their tremendous impact of the identification of new leads. Many active compounds have been selected to-date, following combinatorial methodologies, and a considerable number of those have progressed into clinical trials. However, combinatorial chemistry and related technologies for producing and screening large numbers of molecules also find useful applications in other industrial sectors not necessarily related to the pharmaceutical industry. Emerging fields of application of combinatorial technologies are diagnostics, the down-stream processing, catalysis and the new material sectors.

Many biotechnology/combinatorial-technology companies have been founded in the last few years, with the primary goal to design and produce highly diversified molecular libraries to be screened on selected targets, and the vast majority have definitely caught the attention of pharmaceutical companies.

At the same time, rapidly growing sectors of catalyst design and new material design are going to influence chemical industries as well.

BIBLIOGRAPHY


1. Chabala J. C.: *Drug Discovery Management Subsection, September 19–21, 1993*. Pharmaceutical Manufacturers Association, Philadelphia 1993.
2. Furka A., Sebestyen F., Asgedom M., Dibo G.: *14th Int. Congr. Biochem., Prague, Czechoslovakia*, Abstr. *5*, 47 (1988).
3. Houghten R. A.: Proc. Natl. Acad. Sci. U.S.A. *82*, 5131 (1985).
4. Fassina G., Lebl M., Chaiken I. M.: Collect. Czech Chem. Commun. *53*, 2627 (1988).
5. Furka A., Sebestyen M., Asgedom M., Dibo G.: Int. J. Pept. Protein. Res. *37*, 487 (1991).
6. Hougthen R. A., Pinilla C., Blondelle S. E., Appel J. R., Dooley C. T., Cuervo J. H.: Nature *354*, 84 (1991).
7. Cwirla S., Peters E. A., Barret R. W., Dower W. J.: Proc. Natl. Acad. Sci. U.S.A. *87*, 6378 (1990).
8. Scott J. K., Smith G. P.: Science *249*, 386 (1990).
9. Tuerk C., Gold L.: Science *249*, 505 (1990).
10. Ellington A. D., Szostak J. W.: Nature *346*, 818 (1990).
11. Fassina G., Verdoliva A., Odierna M. R., Ruvo M., Cassani G.: J. Mol. Recogn. *9*, 564 (1996).
12. Lam K. S., Salmon S. E., Hersh E. M., Hruby V. J., Kazmierski W. M., Knapp R. J.: Nature *354*, 82 (1991).
13. Lam K. S., Hruby V. J., Lebl M., Knapp R. J., Kazmierski W. M., Hersh E. M., Salmon S. E.: Bioorg. Med. Chem. Lett. *3*, 419 (1993).
14. Holm A., Meldal M.: *Peptides 1988*, p. 208. ESCOM, Leiden 1988.
15. Gausepohl H., Kraft M., Boulin C., Frank R. W.: *Peptides 1990*, p. 206. ESCOM, Leiden 1991.
16. Knapp D. R., Oatis J. E., Papac D. I.: Int. J. Pept. Protein. Res. *42*, 259 (1993).
17. Ruvo M., Scardino P., Cassani G., Fassina G.: Prot. Pept. Lett. *1*,187 (1994).
18. Smith G. P.: Science *228*, 1315 (1985).
19. Smith G. P., Scott J. K.: Methods Enzymol. *217*, 228 (1993).
20. Miertus S., Fassina G.: *Combinatorial Chemistry and Technology*. M. Dekker, New York 1999.
21. Seneci P.: Chim. Ind. *80*, 1183 (1998).
22. Fassina G., Miertus S.: *Introduction to Combinatorial Chemistry and Technologies*. Special Series on Emerging Technologies. UNIDO, Trieste 1999.
23. Gennari F., Seneci P., Miertus S.: Catal. Rev. – Sci. Eng., in press.
24. Seneci P.: *Solid-Phase Synthesis and Combinatorial Technologies*. Wiley, New York 2000, in press.

25. Miertus S., Fassina G.: *Protocols of Combinatorial Chemistry and Technology*. UNIDO-ICS publication, in press.

**S. Miertus[a], G. Fassina[b], and P. F. Seneci[c]** *([a]International Centre for Science and High Technology of the U.N. Industrial Development Organization, [b]Biopharmaceuticals, Tecnogen S.C.p.A, Piana di Monte Verna, [c]Lead Discovery Department, GlaxoWellcome Medicines Research Centre, Verona, Italy):* **Concepts of Combinatorial Chemistry and Combinatorial Technologies**

A survey of basic concepts of combinatorial chemistry and combinatorial technologies and a great impact of this new approach on the traditional chemistry is presented. The main fields of application of CC/CT are reviewed and the reasons why CC/CT is so strongly needed and demanded are given. Besides obvious utilization of CC/CT in drug discovery, agrochemical research and research and development of new materials and catalysts also gain from this approach. The paper describes the origins and development of the technique, formed on the basis of probabilistic justifications. The applicability of combinatorial technologies and main combinatorial tools are described together with computer-assisted combinatorial chemistry, molecular design and biological methods of CC/CT. A list of important Web resources relevant to the topic is also presented.