ELSEVIER

Journal of Chromatography B, 725 (1999) 3–15

JOURNAL OF CHROMATOGRAPHY B

Review

Combinatorial chemistry, automation and molecular diversity: new trends in the pharmaceutical industry

Luc Van Hijfte*, Gilbert Marciniak, Nicolas Froloff

Synthélabo Biomoléculaire, Department of Combinatorial Chemistry, 16 Rue d'Ankara, 67080 Strasbourg Cedex, France

Abstract

Combinatorial chemistry has emerged as a set of novel strategies for the synthesis of large sets of compounds (combinatorial libraries) for biological evaluation. Within a few years combinatorial chemistry has undergone a series of changes in trends, which are closely related to two important factors in libraries: numbers and quality. While the number of compounds in a library may be easily expressed, it is a lot more difficult to indicate the degree of quality of a library. This degree of quality can be split into two aspects : purity and diversity. The changing trends in combinatorial chemistry with respect to the strategies, the technologies, the libraries themselves (numbers and purity aspects) and the molecular diversity are outlined in this paper.

Keywords: Reviews; Combinatorial libraries; Molecular diversity

Contents

*Corresponding author. Tel.: +33-3-8860-8700, fax: +33-3-8845-9070.

0378-4347/99/$ – see front matter
PII: S0378-4347(99)00007-9

1. Introduction

The search for cures of diseases has been intriguing mankind for several thousands of years. When looking back over the past of drug discovery research, an almost continuous change in ''research trends'' may be observed.

Several decades ago drug research was almost entirely relying on a purely empirical approach, utilising ''molecular diversity'' primarily originated from mother nature. People were relying on plant extracts as medications. Towards the end of the last century, organic synthesis made its entry, and chemists started to introduce new chemical diversity synthetically, but the drug discovery remained entirely trial and error.

Since the 1960s, the understanding of enzymatic reactions and drug–receptor interactions improved considerably what may be considered as the childhood of rational drug design. Chemists were charged with the task of synthesizing compounds based on these new understandings, as exemplified by the discovery of numerous ''transition state'' analogs or the well-known suicide inhibitors [1].

With the development of computer science and better spectroscopic tools such as nuclear magnetic resonance (NMR) and X-ray crystallography during the 70s and 80s, researchers immediately recognized the power of these new technologies, and great efforts were put in unraveling and/or modeling of three-dimensional (3D) structures of target proteins. For numbers of years, researchers were hoping to design ''the drug'' aided by computers, but were rapidly realizing that the phenomenon of drug–protein interactions was more subtle than initially expected [2].

This brings us towards the end of the 80s, and gears were changed once more bringing us back to the good old empirical approach, a new trend now accentuated by the explosion of molecular biology, gene research and not to forget, the fast evolution towards better and more powerful informatics tools. Rapidly, it became clear that a vast number of new enzymes and receptors were to be discovered, and high-throughput screening made its entrance as a new drug discovery tool kit. Medicinal chemists realized that their classical way of carrying out synthesis could not cope with the numbers of compounds needed for HTS, leading to a new issue: how to produce rapidly a large molecular diversity? Slowly but steadily, people started reporting their answers to this important new question, setting the basis for the latest trend in pharmaceutical industry: combinatorial chemistry.

This brief introduction is intended to give a short overview of the different strategic and technical options which do exist in this very rapidly emerging and changing field, to discuss briefly the libraries, and the corresponding ''numbers'' and ''quality'' (purity and diversity) aspects. This will bring the readers of this special issue in the context of combinatorial chemistry and allow them a better understanding of some of the ''quality control'' issues treated in the following articles. For more detailed reports on combinatorial chemistry, the readers are referred to a number of excellent books and reviews which have appeared over the last years [3–16].

2. Strategies in combinatorial chemistry

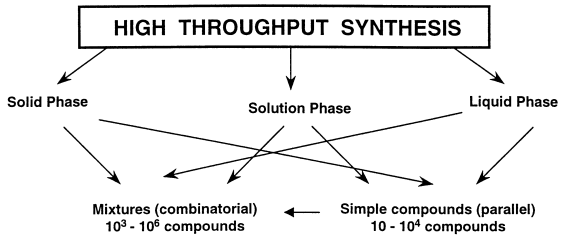
The strategy at the basis of combinatorial chemistry is extremely simple, namely the combination of chemical building blocks, potentially in all combinations and permutations. Thus the challenge of the chemists is not anymore to find conditions to convert A and B to C, but to find conditions optimal for the conversion of a series of As and Bs to a large number of Cs. As can be seen in Table 1, with the peptides formed from the 20 natural amino acids, the number of compounds which virtually may be accessed based on this principle rapidly approaches tens of thousands, even millions.

Although the idea by itself may be considered as extremely simple, it does not bring an answer to the remaining question, which is how to deal with an increasing number of reactions. In fact, two different strategies may pop up in our minds rapidly: either carry out reactions as mixtures, or do reactions in parallel and get away from the time consuming parts in the classical synthesis, which is mainly the separation/purification process. A very nice retrospective view on this issue has recently been published by Curran, who analyses this problem based on phase labeling and phase separation [17]. Based

Table 1
Number of potential compounds in ''natural'' peptides

Oligomer size	Number of building blocks	Total number of potential compounds in library
Dimer	20	$20^2=400$
Trimer	20	$20^3=8000$
Tetramer	20	$20^4=160\,000$
Pentamer	20	$20^5=3\,200\,000$

on the two above mentioned ideas, different strategies enabling the chemists to speed up their syntheses have been developed: solid-, solution- or liquid-phase, either as mixtures or as discrete compounds (multiple parallel synthesis), as depicted in Scheme 1. We will give a short description of the different methodologies, and discuss briefly their advantages and inconveniences.



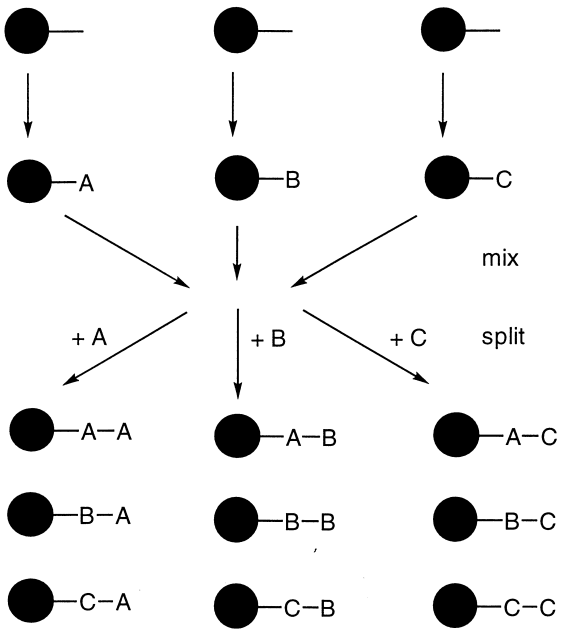
Scheme 1. Strategic options for the synthesis of combinatorial libraries.

2.1. Solid-phase synthesis

The first combinatorial approaches found their origin in the famous solid-phase peptide synthesis, first described in 1963 by Merrifield [18]. A solid phase, such as polystyrene–divinylbenzene or polyacrylamide, is utilized to immobilize a first amino acid, and in an automated manner the peptide is constructed via repetition of coupling-deprotection sequences. Notwithstanding the early remark by Merrifield concerning solid-phase chemistry ''A gold mine awaiting organic chemists'' and a number of early reports in the seventies, the application of a solid-phase strategy for the generation of large number of derivatives only started its breakthrough in the eighties, thanks to the pioneering work of researchers as Furka [19,20], Geysen et al. [21], Houghten [22], and others, which had rapidly realized the immense power of solid-phase synthesis: the possibility of utilising excess of reagent to drive the reactions to completeness, and simply removing the reagents at the end of the reaction by filtration and washing of the solid-phase.

The solid-phase approach was initially applied to the combinatorial synthesis of mixtures, utilising either the tea-bag method [22] or the mix–split [19,23] approach. The tea-bag method consists of a resin loaded into small polypropylene bags (like tea-bags), which are subsequently immersed in solutions containing mixtures of amino acids with appropriate coupling agents. This approach may be considered as the fastest approach for the rapid generation of peptidic libraries. As the compounds are screened in mixtures of several thousands of molecules, some ingenious ways of deconvolution were to be developed, such as iterative deconvolution and positional scanning [24], in order to trace back the active components in the mixture. Finally, the presumed active substances are resynthesized in order to confirm the biological activity.

A second approach which was introduced in the 80s for the synthesis of mixtures is the famous ''one bead, one compound'' or mix–split method, as exemplified in Scheme 2. In contrast to the tea-bag method, one given bead is always reacted with only one monomer in each diversity step, with the consequence that one bead contains only multiple copies of the same molecular entity. Since the partitioning of the beads is done in an entirely statistic manner, it is crucial to work with redundancies of three or more, in order to ascertain that at least 95% of the compounds are synthesized [25]. Also, as one bead contains one compound, it is possible to screen the library both ''on-bead'' or after previous cleavage. The elucidation of the structures may be accomplished on either residual compound on the active



	A	B	C
A	AA	AB	AC
B	BA	BB	BC
C	CA	CB	CC

Scheme 2. The mix/split (one bead–one compound) method versus parallel (matrix) synthesis.

beads, or by determining the structure of the code when a coding strategy is applied [19,58]. Because of the low quantity of derivative available on one bead, structural elucidation is only possible by mass spectrometry.

The advantages of the synthesis of mixtures is the fast production of the library. The big drawback of the method is the lack of analytical control during the synthesis, lack of accurate structure elucidation methods, and the fact that the compounds are screened as mixtures and/or on the beads, which inevitably leads to numerous false positives and negatives. It may thus be concluded that this type of methodology gives very high numbers of compounds, but the quality of the library is rather low, due to a complete absence of purity control.

Although the method may still be considered as useful to generate generic peptide libraries where the chemistry has been thoroughly developed, the method was rapidly abandoned in the context of small molecule libraries. When changing from amide bond formations towards other types of chemistries, such as nucleophilic displacements, C–C bond formations, etc., it was rapidly observed that it was far from evident to develop conditions enabling to react highly diverse species, and thus more analytical control was needed in order to ensure that the compounds were indeed formed. These considerations redirected the solid-phase approach to the parallel or matrix synthesis. A new challenge was now the development of methodologies to synthesize libraries as distinct molecules in a timely manner. Rapidly, Geysen et al. [21] reported the pin-technology and Dewitt and co-workers [26,27] and researchers from Parke Davis described the diversomer technology. At a later stage, robotic systems, automatic synthesizers, etc. were introduced, each with their advantages and inconvenients, as explained further in the synthesis technology chapter. It is clear that the parallel synthesis method is much slower than the mixture method for the preparation of large libraries. On the other hand, the quality of the library may be better controlled, and no coding or deconvolution methods are required, allowing a faster follow up at the stage of confirmation of biological activity. Finally, the chance for false positives and negatives is greatly reduced. Still today, one of the main drawbacks of the solid-phase synthesis is the inadequate synthetic repertoire in this newly developed field, especially if compared with the traditional solution-phase chemistry, making chemical development work tedious and time-consuming. Considerable efforts are nowadays directed towards the development of novel solid supports, new linkers and to the adaptation of a wide variety of known organic reactions to the solid-phase conditions [28–30].

From an analytical point of view, the solid-phase chemistry has necessitated the rediscovery of a certain number of analytical techniques. First of all, since the molecules of interest are covalently bound

to a solid-phase, it is impossible for the chemist to follow the reactions using thin-layer chromatography (TLC) or high-performance liquid chromatography (HPLC), and prior cleavage is necessary to observe the reaction products. Often, the cleavage conditions are quite aggressive, what may lead to undesired side-reactions during this step. For this reason, analytical techniques for detection of derivatives bound to the resin such as Fourier transform infrared (FT-IR) and the Magic Angle Spinning NMR technology [31,32] have gained in interest.

2.2. Solution phase

In the early days of combinatorial chemistry, the solution-phase approach was less commonly employed, not withstanding more than 50 years of prior experience in this field. This lack of interest was mainly due to problems to separate reagents from the required molecule by time consuming extractive work-ups and purifications.

Much of the initial excitement about combinatorial chemistry revolved about the ability to screen large mixtures of compounds in solution against biological targets rapidly and simultaneously, and in this context a number of groups have reported solution-phase mixture synthesis. As with solid-phase mixture synthesis, it is very important to elucidate which components in a mixture are at the origin of the biological activity. A number of deconvolution strategies from orthogonal mixture synthesis [33] to simple HPLC fractioning [34] have been and are still being employed.

Based on recent developments in the field of resin scavengers, polymer-supported reagents and/or purification by solid-supported liquid extraction [35–40], the approach has regained a great interest for the parallel synthesis of libraries, and may nowadays be considered as a very powerful and complementary tool with respect to the previously described solid-phase approach.

An important advantage of the solution-phase approach is the ease of chemistry development, based on previous experience, and also the possibility to apply traditional analytical control. Furthermore, there is no need for a functional group for attachment to the solid support, and the technology may be considered as less expensive, avoiding extensive washing steps and the use of expensive linkers and resins. Conversely, for the preparation of large solution-phase libraries, the purification options are still limited. Solution-phase chemistry may thus lead to libraries with lower numbers and with a higher ''diversity'' quality. A number of high-throughput purification systems has now been developed to increase the purity [41].

2.3. Liquid phase

The liquid-phase approach, a special case of solution-phase, maybe considered as the most recently developed but also least applied methodology for high-throughput synthesis purposes. The technique is based on the linking of the molecule to a so-called ''liquid polymer'' [42–46] or, more generally, to a phase label [17]. In the case of the liquid polymers, depending on the solvent, the support will either be soluble or will precipitate. This approach enables chemists to carry out the synthesis in solution, then precipitate the polymer with a suitable solvent, in order to separate the desired precipitated molecule linked to the polymer from the excess reagent. Another example is the phase labeling using polyfluorinated chains on the molecule, rendering them highly soluble in polyfluorocarbons, which in turn are insoluble in water and organic solvents. The fluorolabeled compounds may thus be separated from organic and inorganic side products [17]. Notwithstanding the very appealing combination of advantages derived from the solution and solid-phase approach, the liquid phase method has not experienced a real breakthrough, and would necessitate the design of special robotics equipment.

2.4. Other approaches

For the sake of completeness, it should be mentioned that, in addition to the above described purely synthetic approaches, a number of ''biological combinatorial synthesis'' methodologies have been developed [47–49]. These techniques vary from the use of enzymes as catalysts for ''random'' conversion of organic substances to the very recently reported ingenious but sophisticated methods of combinatorial biosynthesis.

Combinatorial biosynthesis is a set of techniques

in which DNA is manipulated and transferred from parent organisms into host microorganisms. The expressed biosynthetic genes may generate libraries of ''unnatural natural products''.

The outcome of biological combinatorial synthesis is mixtures, thus elucidation of the structures of active compounds may be a difficult task. Furthermore, an often encountered problem is the synthetic feasability and, in some cases, preparation of gram quantities of a hit by classical chemistry has taken years.

3. Synthesis technologies

3.1. Parallel synthesis

With the development of very powerful technologies for high-throughput screening programs, it became rapidly clear that the synthesis of new chemical entities was for the first time the rate limiting step of the discovery process. Combinatorial chemistry has thus been developed over the past few years as the methodology to help chemists to generate hundred to few thousand compound libraries. In the early stage, combinatorial libraries were focused mainly on the preparation of peptides or oligonucleotides using manual methods and semi- or automated systems [50]. Many efforts were also dedicated in the past to the automation of synthetic chemistry [51,52] but none of these systems have shown to be adapted to the synthesis of large numbers of compounds. In today's pharmaceutical industry emphasis has been put in the synthesis of libraries of discrete small molecules and in order to provide a major increase in production capacity per unit of time, new semi- or automated tools have been introduced in the synthesis processes of small molecules libraries [53–56]. These new technologies, as well as combinatorial chemistry, although not always well accepted by traditional synthetic or medicinal chemists, have now been recognized as a core technology in pharmaceutical industry.

As previously cited, the pioneers in the field were Geysen et al. with the Multipin synthesis approach [21]. However, high-throughput parallel synthesis took its real dimension with the Diversomer approach described by Dewitt et al. [26,57] and Czarnick et al. [58]. Since then automated systems have been developed with more or less success by many biotechnological and pharmaceutical companies and numerous manual, semi-automated and automated systems for both solution- and solid-phase parallel synthesis are now commercially available [54,55] (see also the manufacturers page of the Laboratory Robotics Interest Group homepage at http://lab-robotics.org).

The systems differ in many aspects: degree of automation (manual, semi-automated, automated or robotics) ; type of chemistry (solid- and/or solution-phase); type of reaction vessels (closed reactors or reaction block in glass, polypropylene, PTFE); number of reaction vessels (12 to up to 96); reaction block format (proprietary, 96-well format); solvent/reagent delivery system [manual, valve blocks, liquid handling robots (1 to 96 needles)]; solvent/reagent removal method (bottom filtration or aspiration); mixing (bubbling, magnetic stirring, orbital shaking, vortex mixing); heating and cooling capabilities; cleavage on-line, etc.

Dedicated workstations complete the arsenal of tools for the production of libraries: compound weighing and dissolution, resin dispensing, off-line incubation, high-throughput vacuum evaporation, solid-phase extraction and high-throughput purification systems.

As shown briefly, there is no single answer in terms of automation and technology but rather there exists a series of choices which should be guided mainly by the chemistry to be performed.

As combinatorial technologies become widely used for the synthesis of new compounds, there seems to be a regain of interest in the workstation approach and in semi-automated systems.

This new trend could be explained by the fact that automation of organic chemistry is still extremely challenging, especially in the context of combinatorial library synthesis, where in first instance a high degree of chemical diversity is sought. Furthermore, automated (and robotic) systems are still very expensive and have, in general, low throughput capacity. They also require maintenance and laboratory modifications which increase the cost. Finally, they are often complicated and thereby necessitate highly skilled people for their functioning which limits their accessibility to all the chemists. The

software and hardware are in general not prone to modifications. If robotic systems allow much more flexibility than automated synthesizers, they suffer mainly from long development times with the danger of becoming obsolete at the time of their use in library production.

Despite these limitations, and as technologies and combinatorial chemistry evolve rapidly, it is clear that automation will play a major role in the production of our future libraries and will liberate the chemists from repetitive and laborious tasks associated with the traditional manual synthetic processes.

3.2. The split and pool method

Among the different techniques to encode compound libraries [58], IRORI has developed a combinatorial synthesis system based on the association of radiofrequency memory tags with two different types of microreactors [59,60]. The MicroKans consist of small porous cans containing resin beads and the MicroTubes are inert polymer tubes with a grafted polystyrene surface. By tagging each reaction microreactor with a radiofrequency memory tag, it is possible to follow exactly which compound is synthesized into each microreactor at any stage of the synthesis. During the library synthesis, the microreactors are pooled after each synthetic step and then sorted electronically. For the synthesis of large libraries (ca. few thousands), IRORI has developed an automated sorting machine with the capacity of sorting 10 000 microreactors per night. The methodology may be considered as an improved mix/split approach, since the statistical character of the mixing process has been completely removed, thus allowing to work with a redundancy of exactly 1 (ratio No. reactors/No. compounds). Another improvement compared to the original mix/split method is the amount of each compound prepared (up to mg quantity). At the end of the synthesis, the cleavage may be carried out on each microreactor separated in an array format, thus providing one discrete compound per well. This technique offers a combination of advantages of a mix/split approach (numbers of reaction steps greatly reduced) and parallel synthesis (absence of time consuming deconvolution and possibility to analyze rigorously every discrete compound). Notwithstanding its very appealing advantages, it should be mentioned that the technique still requires some practical improvements, mainly with respect to the washing and cleavage steps. A very similar approach has been developed in parallel by Chiron Technologies using the TranSort, TranStems and SynPhase crowns technology, as described recently by Giger [61].

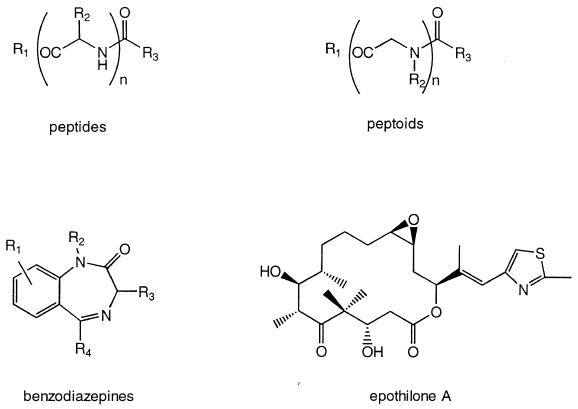
3.3. Miniaturization

With the tendency towards decreasing quantities of substances for screening and increasing dramatically the number of compounds to be synthesized, combinatorial chemistry necessitates miniaturization. Thus microdevices are developed for application in ultra high-throughput screening and for high-throughput synthesis. Among these new technologies is the microfluidic chip based system that is developed by Orchid Biocomputer [62]. The chip incorporates microfabricated components for valving and pumping of fluids with no moving parts integrated within a 3D fluidic network. The transfer of fluids is achieved by electrodynamic or electroosmotic pumping. The syntheses are performed on commercially available solid-phase synthesis beads. The chip in its final version will have the capability to perform up to 10 000 parallel syntheses per run.

Although these technologies look very attractive and promising, they are at a too earlier stage of development to assess their real potential. However, it is clear that they will be invaluable tools for very large library synthesis in terms of speed, performance, throughput, cost and automation.

4. Combinatorial libraries

Historically, solid-phase combinatorial chemistry finds its origin in the solid-phase peptide chemistry, introduced in the sixties by Merrifield [18], and for about ten years the combinatorial approach remained in the context of the preparation of peptide and oligonucleotide libraries. These libraries, often containing millions of compounds, were either prepared utilizing the teabag or mix–split method. One ingenious parallel synthesis approach for peptides has been developed by researchers in Affymax, based on



Scheme 3. Examples of combinatorial chemistry libraries.

spatially addressed deprotection using light and light sensitive protective groups (VLSIPS-approach) [63].

The first non-peptidic libraries made their entry in the late 80s, early 90s, and were still copies of the ''peptide technology'', in the sense that they were the result of a linear construction of different monomeric building blocks. The famous peptoids first described by Zuckermann et al. at Chiron [64] (see Scheme 3) are a very nice example of these types of peptidomimetic libraries. The advantage of these libraries compared to peptidic libraries is their stability to metabolic degradation.

In 1992, Ellman [65] published his first benzodiazepine library synthesis, introducing the small ''drug-like'' molecule libraries. Ellman has clearly demonstrated the ''combinatorial power'', by synthesizing in parallel an 11 000-member library of benzodiazepines [65], which has led to a number of interesting hits on a number of new therapeutic targets. These first reports by Ellman's group have undoubtedly contributed to the enormous explosion combinatorial chemistry has known since then in the pharmaceutical industry.

The types of libraries have recently been extended to the field of natural (or ''natural product like'') product synthesis. A number of outstanding syntheses have recently been reported by very prestigious groups, for example the epothilone A and B analog library synthesis by Nicolaou and co-workers [66,67], and an over two million ''natural product like'' library by Schreiber [68]. As can be seen in these syntheses, great progress is being made in the synthetic repertoire available for solid-phase synthesis, which undoubtedly may give a new boost to this more recently developed strategy.

Depending on the purpose of a library, the numbers and quality aspects will be different. For lead generation libraries, it is commonly accepted that the numbers of compounds should be higher (10^3–10^6), the purity is of somewhat less importance, and finally the diversity descriptors used should be general ones. For lead optimization libraries, the numbers tend to be lower (10^2–10^3), the purity of the derivatives should be excellent, in order to accomplish SAR studies, and finally the diversity should be pharmacophore driven. Researchers have realized that the combinatorial chemistry should not only be a ''numbers game'', and it should be noted that over the last years, the ''smarter, not bigger'' concept has made its entry.

5. Molecular diversity

Molecular diversity is a set of (computational) techniques for describing molecules and classifying them into useful groups (for detailed recent reviews see Refs. [69–72]). In practice, a set of descriptors is calculated for each molecule in a database. Each molecule projects as a point into a multidimensional space. The coordinates of the molecule are the values taken by the descriptors for that particular molecule. Clustering techniques are then used to identify groups of molecules in descriptor space. Descriptors which can be used for diversity profiling can be derived from the molecular formula (one-dimensional, 1D, descriptors, e.g., the molecular mass), from the two-dimensional (2D) graph (e.g., topological descriptors such as structural fingerprints or substructural keys, or physicochemical descriptors such as the log P), or from the 3D atomic coordinates (e.g., shape indices or pharmacophoric keys).

In pharmaceutical research, the ultimate goal of molecular diversity is to build a fast and efficient expert system based on the knowledge of the medicinal chemists. This system would correlate changes in chemical structure and functionality to changes in mechanism of action on biological targets, and hence to biological activity. In other words, the challenge is thus to be able to superimpose a computed descriptor

space onto the biological activity space, where neighborhood relationships between molecules are specified in terms of affinity and specificity towards macromolecular targets.

The purity of a combinatorial library is a very important issue in combinatorial chemistry. However, the molecular diversity of a library is also becoming an increasingly important issue. We can foresee that in a near future there will be a set of commonly accepted criteria for assessing both the purity and the diversity of a library. These criteria will provide a quantitative measure of the added-value of a particular library. A first example of such criteria that have been already widely used by the medicinal chemists is the Lipinski rule of 5 (see below).

Before the establishment of such criteria, the molecular diversity techniques have to be refined in order to prove truly useful. The most important issue to solve in molecular diversity is descriptor validation. The aim is to tailor molecular descriptors that are useful for the purpose of drug design. Such molecular descriptors should show *neighborhood behavior* [73]: the proximity of any two molecules in descriptor space should correlate with similarity in physico-chemical properties and/or similarity in mechanism of action against macromolecular targets. With rational drug design and QSAR tools, it is now easy to identify a descriptor (e.g., the score against a pharmacophore hypothesis) or a relationship between descriptors (e.g., a QSAR equation) that correlate with the activity for a particular target. However, a set of descriptors that would show neighborhood behavior for any potential biological target has not yet been found. Obtaining such descriptors is nevertheless crucial because proximity of molecules in descriptor space would then mean similarity in biological activity, whereas distance in descriptor space would mean optimal diversity in biological activity.

There is now a plethora of molecular descriptors that have been published. The challenge may still be to design new and better descriptors. But there is already much challenge in identifying the best combination of existing descriptors for the general purpose of quantifying the diversity in biological activity of organic compounds.

2D fingerprints were recently shown to have superior neighborhood behavior compared to other 2D and 3D molecular descriptors [74,75]. A similarity radius for 2D fingerprints could be estimated, and compounds within this similarity radius were shown to have comparable biological properties. 2D fingerprints are widely used as 2D descriptors, and have been implemented in most commercially available software packages. A fingerprint encodes the 2D structure of a molecule as a pattern of bits set within a bit string and is calculated by splitting the 2D structure up into fragments. If a particular fragment is present, then a corresponding bit is set in the bit string. Initially used for increasing the efficiency of 2D searches in chemical databases, 2D fingerprints are now also used for similarity/diversity. The similarity between two molecules is measured by counting the number of common bits in the two corresponding fingerprints. Flower [76] showed that measures of similarity based on fingerprints become meaningful only when a large amount of bits are common, namely when molecules are often virtually identical. Fingerprints may thus prove useful in hit explosion, where it is important to finely tune the degree of 2D similarity of the analogs that are synthesized. They are also useful for identifying molecules in a commercial database that are chemically different from those in a proprietary database. However, a 3D description is needed when one wants to get away from a chemical series while retaining biological activity.

A promising 3D description of molecules are pharmacophore keys [77,78]. A pharmacophore key is a bit-string that is calculated by sampling all energetically-accessible conformations of a molecule, and by identifying all pharmacophores that are displayed by the molecule. A particular bit is set in the bit-string if the corresponding pharmacophore is found on the molecule. Intuitively, pharmacophore keys are more closely related to biological activity than 2D descriptors which only describe atom bonding patterns. They are also true 3D descriptors since they account nicely for conformational flexibility. They allow fast database mining with a pharmacophore query. They also allow iterative pharmacophore mapping from active molecules. When a first set of active molecules is identified in primary screening, the intersection of their pharmacophore keys should include the pharmacophore of interest.

This intersection key may thus help to retrieve new molecules from databases for secondary screening, or to design new combinatorial libraries that are more focused. Intersecting again the pharmacophore keys of the newly identified hits may hence eventually lead to a unique potential pharmacophore. However, pharmacophore keys are still very time consuming for combinatorial library design, because of the need to enumerate all molecules and perform at least a limited conformational search of each molecule. Furthermore, such descriptors have not yet shown superior neighborhood behavior to 2D fingerprints [74,75]. More generally, much development efforts still need to be invested in 3D descriptors. The accurate calculation of the conformational space of various molecules is indeed still a difficult problem. Moreover, efficient pattern recognition techniques of 3D objects are still in their infancy.

Molecular diversity techniques have recently proven useful in characterizing the properties of ''drug-like'' molecules. Lipinski et al. [79] noted that obtaining oral activity from compounds that already have adequate potency in vitro is now a rate limiting step in drug development. They also observed that new HTS techniques tend to select leads that are larger, more lipophilic, less soluble than leads in the pre-HTS era. They suggest to use five simple criteria to guide chemistry synthesis towards more favorable physicochemical properties to minimize the time required to achieve oral activity. This ''rule of 5'' predicts that poor absorption or permeation is more likely when there are more than 5 H-bond donors, more than 10 H-bond acceptors, more than 10 heteroatoms such as N and O, the molecular mass is over 500 and the calculated log *P* is over 5. The rule of 5 does not apply to substrates for transporters or to natural products. Lipinski et al. [79] suggest to include a computational ''absorption alert'' based on the rule of 5 in the compound registration system used by the chemists in the early stages of drug discovery.

Can one distinguish between drugs and nondrugs? Stated differently, does ''drug-likeness'' correspond to a descriptor that can be calculated, or is there a set of appropriate descriptors which can help identify drug-like molecules? Two groups recently showed that it is possible to solve computationally this question with the current molecular diversity techniques [80,81]. Both used two large chemical databases as starting material: a database of nondrug molecules (intermediates and reactive compounds from commercial sources), and a database of drug-like molecules (compounds in clinical trials). Both groups then computed 2D descriptors and trained a neural network system to distinguish between drugs and nondrugs on subsets of both databases. After training, the systems were able to fairly discriminate between drugs and nondrugs. Both groups provided convincing demonstrations that their learning systems did not capture the characteristics of existing drugs, but some general structural features that differentiate potential drugs from the vast majority of organic molecules. Hence using their systems as filters should be quite useful and should not be detrimental to exploring new structural motifs.

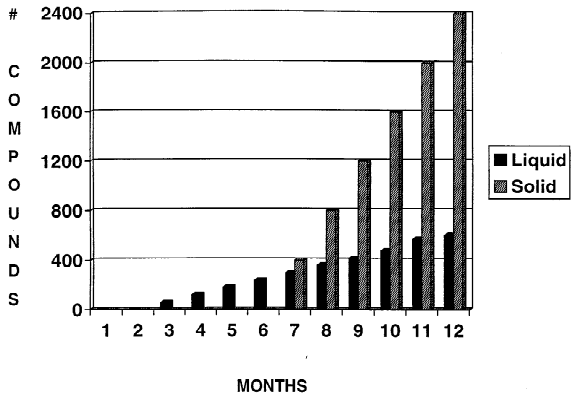
Other interesting work which should and will be pursued in that area is the prediction of toxicity, permeation across the blood–brain barrier or other tissue-specific membranes, synthetic accessibility, metabolic stability.

To summarize, it is reasonable at present to use the following steps for designing prospective combinatorial libraries for primary screening: (i) eliminate non-desired compounds (potential toxicity, ''rule of 5'' not satisfied, nondrug-like molecules, etc.); (ii) calculate 2D descriptors which accurately encode the chemical nature of the molecules, and 3D descriptors which better relate to drug–receptor interactions; (iii) keep molecules that are under a threshold of 2D similarity, in order to reduce chemical redundancy (this process is fast since 2D descriptors can be rapidly calculated); (iv) select molecules that are most diverse according to 3D descriptors, for example those that display the most unrelated pharmacophores.

Is combinatorial chemistry compatible with maximal molecular diversity? Practically, the answer may be no. First, the commercially available reagents and building blocks are often very limited, and starting material that is too diverse may not be amenable to parallel synthesis because of different reaction kinetics. Second, because of practical reasons, combinatorial chemistry will always be planned as matrix synthesis, giving a reagent-driven diversity and not a compound-driven diversity. Finally, the range of available chemistry for high-throughput synthesis is still limited.

With combinatorial chemistry, medicinal chemists

can now make thousands of new molecules in a systematic way. Such power has its drawbacks. It is very easy to make collections of molecules that are useless for drug development, namely molecules that are not ''drug-like''. It is also easy to make collections of very similar molecules since the products result from the combination of a small set of reagents. Therefore, chemists are increasingly asking for efficient molecular diversity techniques to help them identify which reagents to combine and what molecules to make, in order to optimize the resulting libraries. In the case of general libraries for primary screening, the objective is to minimize the chances of missing a hit by covering the maximal biological space with a minimum number of products. In the case of focused libraries, the aim is to rapidly refine a pharmacophore hypothesis through iterative combinatorial synthesis and biological screening runs.



Scheme 4. Simulation of production potential for a given project using either solid or solution.

6. Future perspectives

Combinatorial chemistry may be considered as one of the fastest integrated new methodologies in pharmaceutical industry. It is based on a certain number of different strategies for the high-throughput synthesis of large sets of compounds. Today, it is very difficult to answer the question if one approach is superior to the others. All approaches have shown success stories (but also failure stories), and one should closely evaluate the advantages and inconveniences of the different approaches before initiating a new synthesis. First, the choice will depend on the chemistry involved and the numbers of steps in a synthesis sequence. Some molecules may be excellent scaffolds for a solid-phase approach, while others will be ideal for solution-phase. Another criterion for the choice of a strategy may be the number of compounds in the final library. Automated solid-phase production may, in a given project, be considerably faster then the alternative solution-phase production. Conversely, the chemistry development may be much faster in the latter. This idea is illustrated by a simple simulation (Scheme 4), showing clearly in this case that in order to produce 200 compounds the solution-phase method will be the method of choice. If the production objective in this particular example exceeds 10^3 substances, a solid-phase approach will be the preferred strategy.

Today, a generally observed trend is that the interest in pure ''combinatorial chemistry'', in the strictest sense of the word, is slowly decreasing, making place for high-throughput parallel synthesis, or even high-throughput medicinal chemistry. Combinatorial chemistry has made its entry in the medicinal chemistry lab, where chemists have started to employ some technologies out of the combichem repertoire, in order to speed up their analoging. They are less intrigued by the numbers, but their priority remains the quality of the compounds, in order to get unambiguous biological results for SAR studies.

Where to go from now? The future direction(s) in combinatorial chemistry will undoubtedly depend on new technology development, such as miniaturization and nanotechnologies. Also, combinatorial chemistry now has considerable feedback from the high-throughput screening campaigns, what may direct our future choices: increasing the numbers or increasing the quality of the libraries produced? Finally, one may even wonder if combinatorial chemistry not only represents ''just'' a new trend, which will be rapidly forgotten and replaced by other, more powerful methodologies. It is now possible for the traditional medicinal chemist to take advantage of all technologies developed today to carry out the synthesis at a higher throughput than in the past, what will facilitate meeting the different challenges of quality, quantity and diversity to keep pace with todays increasing flow of new medicinal targets.

References

[1] P. Bey, Enzyme-Activated Irreversible Inhibitors, in: N. Seiler, M.J. Jung, J. Koch-Weser (Eds.), Elsevier, Amsterdam, 1978, p. 27.
[2] Y. Connolly Martin, E. Kutter, V. Austel, in: G.L. Grunewald (Ed.), Modern Drug Research – Paths to Better and Safer Drugs (Medicinal Research Series, Vol. 12), Marcel Dekker Inc., New York and Basel, 1989.
[3] M. Lebl, Z. Leblova, Dynamic Database of References in Molecular Diversity, Internet: http/www.5z.com
[4] F. Balkenhohl, C. von dem Bussche-Hünnefeld, A. Lansky, C. Zechel, Angew. Chem. Int. Ed. Engl. 35 (1996) 2288.
[5] D. Hilvert, S.V. Ley, Curr. Opin. Chem. Biol. 2 (3) (1998).
[6] L.A. Thompson, J.A. Ellman, Chem. Res. 96 (1996) 555.
[7] N.K. Terrett, M. Gardner, D.W. Gordon, R.J. Kobylecki, J. Steele, Tetrahedron 51 (30) (1995) 8135.
[8] P.M. Cowley, D.C. Rees, Curr. Med. Chem. 4 (1997) 211.
[9] R. Storer, Drug Discovery Today 1 (1996) 248.
[10] D.P. Curran, Chemtracts: Org. Chem. 9 (1996) 75.
[11] B.A. Bunin, The Combinatorial Index, Academic Press, San Diego, CA, 1998.
[12] R. Cortese, Combinatorial Libraries – Synthesis, Screening and Application Potential, Walter de Gruyter, Berlin, 1996.
[13] R.A. Houghten, Combinatorial Libraries – the Theory and Practice of Combinatorial Chemistry, ESCOM, Leiden, 1995.
[14] G. Jung, Combinatorial Peptide and Non Peptide Libraries – A Handbook, VCH, Weinheim, 1996.
[15] I. Chaiken, K. Janda, Molecular Diversity and Combinatorial Chemistry – Libraries and Drug Discovery, American Chemical Society, Washington, DC, 1996.
[16] D. Obrecht, J.M. Villalgordo, Solid-Supported Combinatorial and Parallel Synthesis of Small-Molecular-Weight Compound Libraries, Tetrahedron Organic Chemistry Series, Vol. 17, Elsevier, Amsterdam, 1998.
[17] D.P. Curran, Angew. Chem. Int. Ed. 37 (1998) 1174.
[18] R.B. Merrifield, J. Am. Chem. Soc. 85 (1963) 2149.
[19] A. Furka, 2nd International Electronic Conference on Synthetic Organic Chemistry, September 1998, at http//ecsoc2.hcc.ru/posters
[20] J.W. Jacobs, Z.-J. Ni, Combinatorial and Molecular Diversity in Drug Discovery, in: E.M. Gordon, J.F. Kerwin Jr. (Eds.), Wiley–Liss, New York, 1998, p. 271.
[21] H.M. Geysen, R.H. Meloen, S.J. Barteling, Proc. Natl. Acad. Sci. USA 81 (1984) 3998.
[22] R.A. Houghten, Proc. Natl. Acad. Sci. USA 82 (1985) 5131.
[23] M. Lebl, V. Krchnak, N.F. Sepetov, B. Seligmann, P. Strop, S. Felder, K.S. Lam, Biopolymers (Peptide Sci. 37 (1995) 177.
[24] R.A. Houghten, R.C. Pinilla, S.E. Blondelle, J.R. Appel, C.T. Dooley, J.H. Cuervo, Nature 354 (1991) 84.
[25] J.A. Boutin, A.L. Fauchère, TIPS 17 (1996) 8.
[26] S.H. Dewit, J.S. Kiely, C.J. Stankovic, M.C. Schroeder, D.M.R. Cody, M.R. Pavia, Tetrahedron Lett. 35 (1994) 9177.
[27] S.H. Dewitt, A.W. Czarnik, Acc. Chem. Res. 29 (3) (1996) 114.
[28] C. Blackburn, F. Albericio, S.A. Kates, Drugs Future 22 (9) (1997) 1007.
[29] P.H.H. Hermkens, H.C.J. Ottenheijm, D.C. Rees, Tetrahedron 53 (16) (1997) 5643.
[30] P.H.H. Hermkens, H.C.J. Ottenheijm, D.C. Rees, Tetrahedron 52 (13) (1996) 4527.
[31] A.W. Czarnik, Annu. Rep. Combichem. Mol. Divers. 1 (1997) 59.
[32] S.K. Kaldor, M.G. Siegel, Combinatorial and Molecular Diversity in Drug Discovery, in: E.M. Gordon, J.F. Kerwin Jr. (Eds.), Wiley–Liss, New York, 1998, p. 307.
[33] B. Deprez, X. Williard, L. Bourel, H. Coste, F. Hyafil, A. Tartar, J. Am. Chem. Soc. 117 (1995) 5405.
[34] R.H. Griffey, H. An, L.L. Cummins, H.J. Gaus, B. Haly, R. Hermann, P.D. Cook, Tetrahedron 54 (1998) 4067.
[35] C.R. Johnson, B. Zhang, P. Fantauzzi, M. Hocker, K.M. Yager, Tetrahedron 54 (1998) 4097.
[36] S.D. Brown, R.W. Armstrong, J. Am. Chem. Soc. 118 (1996) 6331.
[37] S.W. Kaldor, M.G. Siegel, J.E. Fritz, B.A. Dressman, P.J. Hahn, Tetrahedron Lett. 37 (1996) 7193.
[38] A.A. Virgilio, S.C. Schurer, J.A. Ellman, Tetrahedron Lett. 37 (1996) 6961.
[39] R.J. Booth, J.C. Hodges, J. Am. Chem. Soc. 119 (1997) 4882.
[40] D.L. Flynn, J.Z. Crich, R.V. Devraj, S.L. Hockerman, J.J. Parlow, M.S. South, S. Woodard, J. Am. Chem. Soc. 119 (1997) 4874.
[41] M.G. Siegel, A.J. Shuker, C.A. Droste, P.J. Hahn, C.D. Jesudason, J.H. McDonald, D.P. Matthews, C.J. Rito, A.J. Thorpe, Mol. Divers. 3 (1998) 113.
[42] D.J. Gravert, K.D. Janda, Chem. Res. 97 (1997) 489.
[43] H. Han, K.D. Janda, J. Am. Chem. Soc. 118 (1996) 7632.
[44] H.S. Han, K.D. Janda, Tetrahedron Lett. 38 (1997) 1527.
[45] C. Bolm, A. Gerlach, Angew. Chem. 109 (1997) 773.
[46] C. Bolm, A. Gerlach, Angew. Chem. Int. Ed. Engl. 36 (1997) 741.
[47] S. Borman, C&EN, (1998) 29
[48] V.V. Mozhaev, C.L. Budde, J.O. Rich, A.Y. Usyatinsky, P.C. Michels, Y.L. Khmelnitsky, D.S. Clark, J.S. Dordick, Tetrahedron 54 (1998) 3971.
[49] P.F. Leadley, Curr. Opin. Chem. Biol. 1 (1997) 162.
[50] V. Krchnak, D. Cabel, M. Lebl, Peptide Res. 9 (1996) 45.
[51] J.S. Lindsey, Lab. Info. Manage. 17 (1992) 15.
[52] M. Needels, J. Sugarman, Combinatorial and Molecular Diversity in Drug Discovery, in: E.M. Gordon, J.F. Kerwin Jr. (Eds.), Wiley–Liss, New York, 1998, p. 339.
[53] S.H. Dewitt, A.W. Czarnik, Cur. Biol. Biotechnol. 6 (1995) 640.
[54] J.F. Cargill, M. Lebl, Curr. Opin. Chem. Biol. 1 (1) (1997) 67.
[55] J.H. Hardin, F.R. Smietana, Mol. Divers. 1 (1995) 270.
[56] J.H. Hardin, F.R. Smietana, High Throughput Screening – The Discovery of Bioactive Substances, in: J.P. Devlin (Ed.), Marcel Dekker, New York, 1997, pp. 251–261.

[57] S.H. Dewitt, J.S. Kiely, C.J. Stankovic, M.C. Schroeder, D.M.R. Cody, M.R. Pavia, Proc. Natl. Acad. Sci. USA 90 (1993) 6909.
[58] A.W. Czarnik, Curr. Opin. Chem. Biol. 1 (1) (1997) 60.
[59] K.C. Nicolaou, X.Y. Xiao, Z. Parandoosh, A. Senyei, M. Nova, Angew. Chem. Int. Ed. 34 (1995) 2289.
[60] A.W. Czarnik, in: ISLAR'97 Proceedings, Boston, MA, October 1997, p. 166.
[61] R. Giger, in: Cambridge Healthtech Institute's First European Conference on Strategies and Techniques for Identification of Novel Bioactive Compounds, Barcelona, December 1997.
[62] S. Dewitt, Synthesis on a Chip in Strategic Use of Combinatorial Chemistry, IBC, Frankfurt, June 1998.
[63] M.A. Gallop, R.W. Barrett, W.J. Dower, S.P.A. Fodor, E.M. Gordon, J. Med. Chem. 37 (9) (1994) 1233.
[64] R.N. Zuckermann, E.J. Martin, D.C. Spellmeyer, G.B. Stauber, K.R. Shoemaker, J.M. Kerr, G.M. Fugliozzi, D.A. Goff, M.A. Siani, R.J. Simon, S.C. Banville, E.G. Brown, L. Wang, L.S. Richter, W.H. Moos, J. Med. Chem. 37 (1994) 2678.
[65] J. Ellman, Acc. Chem. Res. 29 (1996) 132.
[66] K.C. Nicolaou, N. Winssinfze, J. Pastor, S. Ninkovic, F. Sarabia, Y. He, D. Vourloumis, Z. Yang, T. Li, P. Giannakakou, E. Hamel, Nature 387 (1997) 268.
[67] K.C. Nicolaou, F. Roschanger, D. Vourloumis, Angew. Chem. Int. Ed. 37 (1998) 2014.
[68] D.S. Tan, M.A. Foley, M.D. Shair, S.L. Schreiber, J. Am. Chem. Soc. 120 (1998) 8565.
[69] W.A. Warr, J. Chem. Info. Comput. Sci. 37 (1997) 134.
[70] J.M. Blaney, E.J. Martin, Curr. Opin. Chem. Biol. 1 (1997) 54.
[71] E.J. Martin, D.C. Spellmeyer, R.E. Critchlow Jr., J.M. Blaney, Reviews in Computational Chemistry, in: K.B. Lipkowitz, D.B. Boyd (Eds.), VCH, New York, 1997, pp. 75–100.
[72] M.G. Bures, Y.C. Martin, Curr. Opin. Chem. Biol. 2 (1998) 376.
[73] D.E. Patterson, R.D. Cramer, A.M. Ferguson, R.D. Clark, L.E. Weinberger, J. Med. Chem. 39 (1996) 3049.
[74] R.D. Brown, Y.C. Martin, J. Chem. Info. Comput. Sci. 36 (1996) 572.
[75] R.D. Brown, Y.C. Martin, J. Chem. Info. Comput. Sci. 37 (1997) 1.
[76] D.R. Flower, J. Chem. Info. Comput. Sci. 38 (1998) 379.
[77] K. Davies, C. Briant, Network Sci., (1995) http://www.netsci.com/Science/Combichem/feature05.html
[78] S.D. Pickett, J.S. Mason, I.M. McLay, J. Chem. Info. Comput. Sci. 36 (1996) 1214.
[79] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Adv. Drug Deliv. Rev. 23 (1997) 3.
[80] A.P. Walters, M.A. Murcko, J. Med. Chem. 41 (1998) 3314.
[81] J. Sadowski, H. Kubinyi, J. Med. Chem. 41 (1998) 3325.