Combinatorial Chemistry and Molecular Diversity. An Overview†

Wendy A. Warr

Wendy Warr & Associates, 6 Berwick Court, Holmes Chapel, Cheshire, CW4 7HZ, England

Received October 1, 1996⊗

Combinatorial chemistry is no longer viewed as a speculative technique for R&D in the pharmaceutical and agrochemical industries but has become an established important technology. High throughput screening and combinatorial chemistry are altering the character of discovery research. This paper introduces the principles of design and synthesis of small molecule libraries and discusses software being developed in generation of molecular diversity, in molecular design and analysis, in handling the information explosion, in information management, and in decision support to enable drug discovery groups to accelerate the planning and development of libraries, improve work group communication, and facilitate the generation of ideas.

INTRODUCTION

The healthcare industry is going through a period of unprecedented change. Governments worldwide are experiencing spiraling costs in healthcare provision and have put pressure on the pharmaceutical industry to reduce the cost of drugs. This is in addition to the pressure on the industry from regulatory authorities (in some cases time to market is actually increasing rather than decreasing) and from the threat to revenues as key products come off patent.

For many years the industry had high profit margins and special growth, but now it has been estimated that the world pharmaceuticals market may grow at no more than 5% for the rest of the decade. It is getting more expensive to develop drugs (Datamonitor estimated $350 million in March 1996), and in spite of all attempts to accelerate speed to market, the average number of years to bring a product through has actually increased. The hurdles of safety, efficacy, quality, and health economics take longer to surmount.

The pharmaceutical industry has reacted by cutting overheads and jobs. Now companies are being managed much more tightly, and some will not survive. Medium-sized European businesses are suffering particularly because lack of scale prevents them from investing properly in R&D and developing new products. Even the biggest drug companies are reducing R&D spending or at least slowing the increase.

Peter Machin[1] of SmithKline Beecham, using data supplied by Drews,[2] has shown that the industry needs a 2−4-fold increase in its output of new chemical entities (NCEs). In 1990−1994, the top 20 companies produced on average one new chemical entity a year.[1,3] Increasing this figure to three NCEs a year could be achieved if 30 000 new compounds were made each year by each company (leading to 30 development compounds and three NCEs), and genomics research were to produce 75 new biological targets of which 40% were useful.

Using traditional methods of synthesis, a medicinal chemist can produce perhaps 50−100 compounds per year. Once high throughput screening started to make an impact, the demand for substances to test increased dramatically, and medicinal chemists began to seek new sources of chemicals and faster ways of synthesizing compounds in-house. Combinatorial chemistry and multiple parallel synthesis became attractive propositions.

The term "combinatorial chemistry" tends to be used as an umbrella term for a number of technologies, including multiple parallel synthesis. "Molecular diversity" is another term that also has varied definitions. A chemical "library" is a set of "diverse" compounds based on a common structural template: it may contain mixtures of compounds or it may be an array of discrete compounds produced by multiple parallel synthesis. Several review articles on combinatorial chemistry have appeared.[4−14] Lebl maintains a list of literature references on a World Wide Web site,[15] and there is an Internet listserver devoted to molecular diversity.[16]

RATIONAL AND IRRATIONAL DRUG DESIGN

The philosophy of testing as many compounds as possible as fast as possible has led to the term "irrational drug design". This is not to suggest that the "rational" drug design concepts embraced by computational chemists are now going out of favor. Indeed, there is now increased interest in software for detecting dissimilarity amongst compounds: both diverse and focused libraries of compounds need to be designed, and cross-matching of databases has become more important. High throughput screening and combinatorial chemistry might be producing new leads, but they do not directly produce new *drugs*: lead optimization is an essential part of the drug design process. QSAR studies are needed to handle the information coming out of the high throughput screening process. Advances in crystallography, NMR, and genomics are leading to a greater understanding of disease processes, and new targets are being discovered every year. The discipline of bioinformatics is blossoming.

NATURAL DIVERSITY

Opinions vary on the value of natural products as samples for high throughput screening. On the one hand, there is the advantage of the vast diversity of the world of nature (many of today's best-selling drugs have a "natural product" origin in certain meanings of that phrase), but, on the other hand, there is the disadvantage of the length of time needed

† Paper given at the Fourth International Conference on Chemical Structures, Noordwijkerhout, The Netherlands, June 4, 1996, as an introduction to the symposium on Combinatorial Chemistry.

⊗ Abstract published in *Advance ACS Abstracts,* January 1, 1997.

S0095-2338(96)00142-4 CCC: $14.00 © 1997 American Chemical Society

No. of amino acid residues	Peptides	No. of distinct peptides
1	H-X_1-OH	20
2	H- X_1X_2-OH	400
3	H- $X_1X_2X_3$-OH	8000
4	H- $X_1X_2X_3X_4$-OH	160,000
5	H- $X_1X_2X_3X_4X_5$-OH	3,200,000
6	H- $X_1X_2X_3X_4X_5X_6$-OH	64,000,000
7	H- $X_1X_2X_3X_4X_5X_6X_7$-OH	1,280,000,000
8	H- $X_1X_2X_3X_4X_5X_6X_7X_8$-OH	25,600,000,000

Figure 1. Combinatorial explosion: peptides.

to determine the active material in a natural product mixture and to work out ways of isolating and synthesizing a pure chemical. David Gani of St. Andrew's University[17] has quoted a saying "nature does not think, nature tries everything, nature has forever, nature backs the winner [by natural selection]" perhaps proving that "thinking can be dangerous". This is an interesting comment on irrational drug design. Certainly nature has made a remarkable number of molecules using just 20 amino acids.

HISTORICAL BACKGROUND

Combinatorial chemistry has intimate links with solid-phase synthesis. Merrifield's synthesis of peptides on solid phase in the 1960s was a forerunner in this field.[18] Subsequently, Leznoff[19,20] published solid-phase synthesis of small molecules. Geysen[21,22] did solid-phase synthesis on "pins", polymeric rods arranged in a 96-well microtiter plate format, whereas Houghten[23] carried out peptide synthesis on "tea bags" (resin-containing, porous polypropylene bags). Furka[24] devised the concept of split synthesis, later called the "split and mix" or "one bead one compound" approach, once it was applied to the synthesis of libraries of peptides on solid support. Lam and co-workers[25] used the method to make libraries of peptides for screening while immobilized on beads, and Houghten[26] used the split synthesis approach and tea-bag chemistry to make soluble peptide libraries. The Ugi multicomponent reaction[27−29] was a forerunner of current solution phase combinatorial chemistry. It has been employed by Armstrong at the University of California and adapted to solid-phase at Ontogen Corporation. These various technologies are described in more detail below, as are more recent methods for building libraries of small molecules rather than peptides.

THE INFORMATION EXPLOSION

Combinatorial chemistry is often viewed as a "numbers game". The earliest combinatorial libraries were peptides. An indication of the numbers of compounds that could be made is given in Figure 1. Dr. Joseph C. Hogan, of ArQule, has calculated that a complete library of 60-mers, based on 20 amino acids, consists of 20^{60} or 1.15×10^{78} different peptides, but there is not enough mass in the universe to provide one molecule of each peptide.[30]

There are three types of library: combinatorial, permutational, and binary. Pirrung has discussed the parallelism advantages of the three types.[31] Phage display libraries are an example of permutational libraries. Taking one card at a time from a pack gives 52! possibilities; if you put each card back after you have used it, there are 52^{52} possibilities. The number of peptides formed from genetically coded

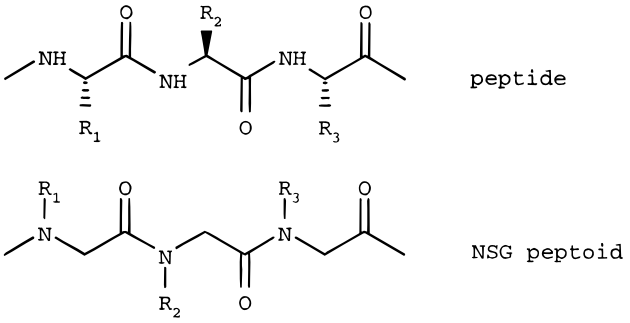

Figure 2. Peptoids.

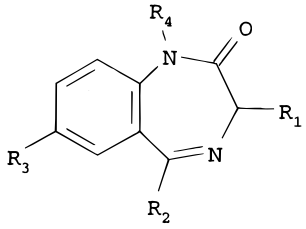

Figure 3. Benzodiazepines.

amino acids is 20^l where l is the number of amino acid residues (thus 20^3 or 8000 tripeptides). The numbers in Figure 1 are derived in a similar manner.

Besides peptides, in the early days of combinatorial chemistry, other oligomeric libraries were made, all on the same principle of adding A to B to make AB, then adding C to make ABC, then adding D to make ABCD, and so on. Peptides, however, do not make good drugs, and variations were thus tried, e.g., Chiron's N-substituted glycine (NSG) "peptoids" (see Figure 2) where the side chains are attached to the amide nitrogen.[32−33] The next step was to extend the idea of libraries to small molecules. Almost simultaneously, the teams of Sheila DeWitt[34] at Parke-Davis and Jonathan Ellman[35−39] at UC Berkeley made *combinatorial* libraries of benzodiazepines (see Figure 3). Such libraries are not of the oligomeric ABCD... type described above, but are based on the concept of a central core, or "scaffold" with substituents chains A, B, C, D, etc. attached at various positions. Ellman's team have also made various other libraries.[40−41]

Pirrung[31] has likened a combinatorial library to combining five pairs of pants and four shirts all in different colors. This gives you 20 (i.e., 4×5) different outfits. If you have 4 hydrophobic amino acids, 5 basic amino acids, and 10 small hydrophobic amino acids,[42] the total library size is $4 \times 5 \times 10$ or 200. Such libraries may be made in "pools" or mixtures of compounds (e.g., 4×50 or 5×40) or (as in the case of DeWitt's benzodiazepines) in arrays of discrete compounds.

LIBRARY GENERATION

Combinatorial chemistry is very often carried out on solid support. The advantages of using solid support lie in product isolation and manipulation and in ease of automation. Also, solid-phase synthesis allows excess of reagents to be used to drive reactions to completion. The advantages may well outweigh the disadvantages, namely limited scale, the functionalization needed for solid-phase attachment, the orthogonal chemistries for linkers and capping, and the specialized protocols for reaction monitoring. Solution phase

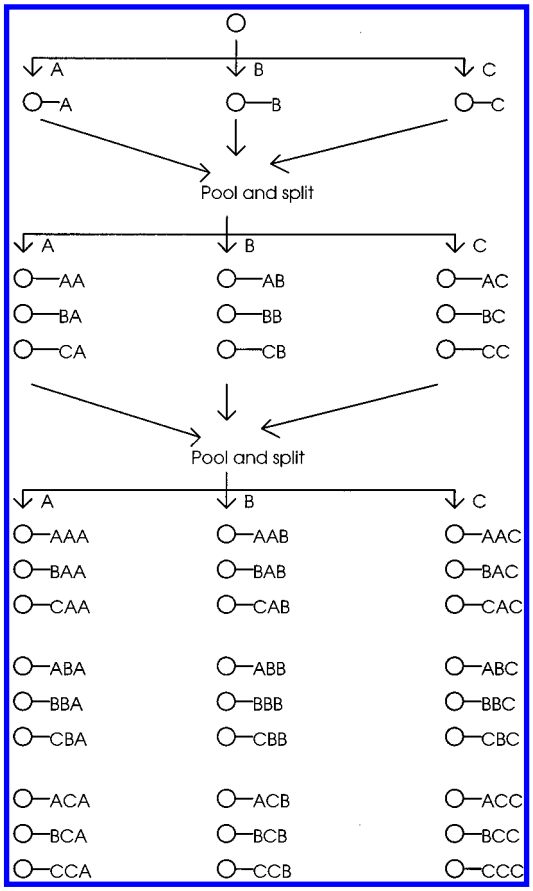

Figure 4. Split synthesis.

synthesis has the advantages of easier sample manipulation and limitless scale without the additional need to develop the reaction for solid phase. Products are directly assayable, and intermediates can be purified. However, sample isolation is a problem.

The concept of "split synthesis", carried out on beads made of polymeric material, is illustrated in Figure 4. The circles represent beads. After each synthesis stage the beads are mixed together and then split into batches for further synthesis. In this particular case 27 (i.e., 3^3) compounds are made in just nine (i.e., $3 + 3 + 3$) reactions.

One problem with mixtures is that, after an active pool has been identified in a screen, it is necessary to determine which compound in the mixture caused the activity. Recursive deconvolution is one technique for finding the active component. Suppose that the central column of nine trimers in one pot in the above diagram were active. All those compounds have a terminal B. The object now is to make three mixtures of three compounds to determine whether the most active mixture has A, B, or C in the middle. Samples of the intermediate on resin will have been retained, so addition of B to the dimers, i.e., resynthesis, is carried out to make three "pots" containing BAA, BAB, and BAC in the first, BBA, BBC, and BBB in the second, and BCA, BCB, and BCC in the third. Suppose now that the active pot is the first, i.e., the best component for the middle is A. Now the three individual compounds BAA, BAB, and BAC are made to determine which is the most active. Other techniques, such as "positional scanning", can be used to determine the active component of a mixture.[43] After deconvolution, some but not all, information needed for QSAR studies can be retained.

Table 1. Advantages and Disadvantages of Synthesis Methods

split/pool	multiple simultaneous synthesis
large number of compounds rapidly generated	smaller number of compounds
mixtures	one structurally known compound/well
deconvolution is time consuming	labor intensive and slower synthesis
information content is questionable	information content intact
screening limitations	screen in available assays

Tagging[44] is an alternative method of detecting actives in a mixture. In this technique an "identifier" is tagged onto each molecule as it is synthesized. After cleavage (in the case of chemical tags) the tag is identified, revealing the identity of the molecule to which it was attached. Peptide tags can be handled by Edman degradation, and oligonucleotides can be amplified by PCR and sequenced. Still and coworkers[45,46] have also developed a method using halogenated organics as "bar codes" that can be read by selective photochemical deprotection of the tags from selected positive beads followed by analysis by EC-GC. More recently, radio frequency tags and related techniques have been used.[47,48]

There are some disadvantages to the use of mixtures. For example, several compounds with modest activity may be mixed in one pool leading to a false positive from the summation of the activities. An alternative to generation of mixtures by combinatorial chemistry is high speed parallel synthesis of very large numbers of individual organic molecules. They are often synthesized and delivered in 96-well plate format, and they are assayed in solution. No encoding or deconvolution is necessary: the structure and synthesis of a hit compound is immediately available from its position (usually from A1 to H12) in the spatial array. Scale-up and analogue programs can be launched immediately. The entire repertoire of synthetic organic chemistry can be accessed. Solid-phase and solution synthesis can be used as appropriate for the chemical reactions selected.

Advantages and disadvantages of the two library generation strategies are summarized in Table 1. Broadly speaking, there is a tendency nowadays toward "discretes" rather than mixtures or toward smaller libraries and smaller mixtures. It is possible to handle large libraries with effective tagging, but people are becoming wary of "playing the numbers game" without good reason. Incidentally, some financial analysts use library sizes and high throughput screening statistics as a way of valuing a company for mergers, acquisitions, and partnerships.

AUTOMATION

Numerous companies are involved in the automation of high throughput synthesis and screening: Advanced ChemTech,[49] Argonaut Technologies,[50] Bohdan Automation,[51] Chiron Mimotopes,[52] CombiChem,[53] CRS Robotics,[54] Diversomer Technologies,[55] Myriad,[56] Ontogen,[57] Sagian (and ORCA),[58] Tecan,[59] Tomtec,[60] and Zymark,[61] to name but a few. MDL,[62] Chemical Design,[63] and other companies have announced alliances with various automation companies. There is debate about the workstation concept, where a robot moves plates from one workstation to another, *versus* the all-in-one synthesizer concept. Closed systems more easily allow inert atmospheres to be used.

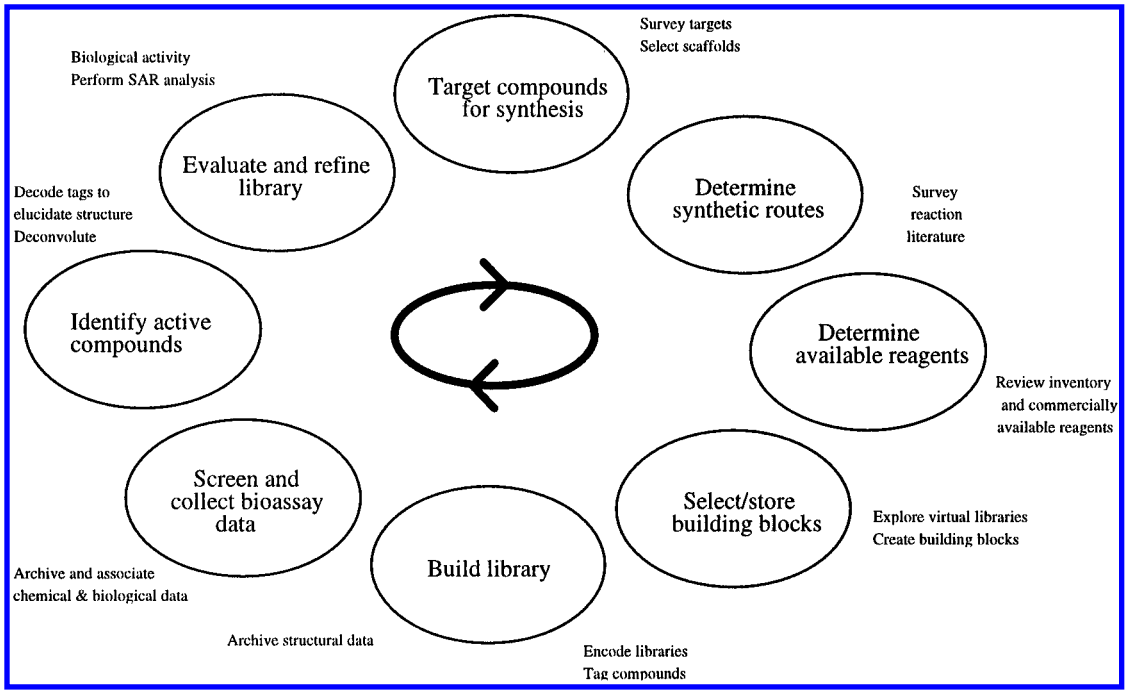

Figure 5. Combinatorial chemistry workflow.

SOLID-PHASE SYNTHESIS DATABASES

The literature on solid-phase synthesis is growing rapidly. Two companies have announced databases covering the solid-phase synthesis literature. MDL and FIZ Chemie Berlin have a collaboration to produce the Solid-Phase Organic REactions (SPORE) database[64] and Synopsys Scientific Systems[65] together with Oxford Diversity is building the SPS database.

THE DRUG DESIGN PROCESS: CHEMISTRY

The stages of the combinatorial chemistry drug design process are illustrated in Figure 5. Companies such as Daylight Chemical Information Systems,[66] MDL Information Systems, Chemical Design, and Tripos[67] have software for reagent selection, building virtual chemical libraries, library registration, and diversity analysis. The handling of generic structures is of fundamental importance in the field of chemical libraries. Barnard and Downs[68] have concluded that combinatorial libraries are a restricted type of Markush structure. Successful techniques have been developed over the last 15 years to handle Markush structures from patents, and these techniques are potentially applicable to storage, searching, and diversity analysis in combinatorial libraries. Markush, or generic, structures are used in chemical patents, as query structures in substructure search systems, in QSAR analyses, and in combinatorial libraries. Four types of variability have been distinguished:

s-variation (substituent variation)
  a list of alternative values for an R-group
h-variation (homology variation)
  a generically described group, e.g., "alkyl"
p-variation (position variation)
  variable point of attachment
f-variation (frequency variation)
  multiple occurrence of groups

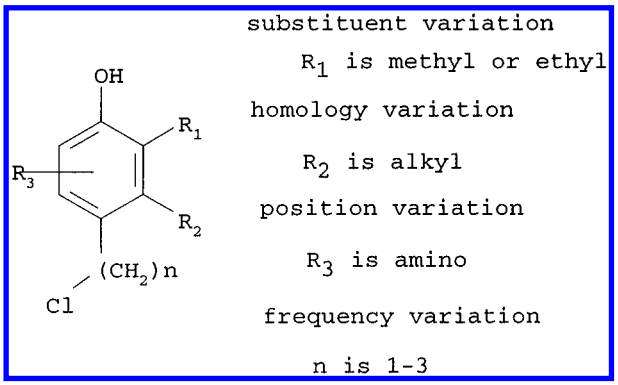

Figure 6. Markush structure.

Table 2. Applicability of Variations in Markush Structures

	substituent	homology	position	frequency
patents	√	√	√	√
queries	√	(√)	(√)	(√)
QSAR	√		√	
libraries	√		(√)	(√)

An example is given in Figure 6. Barnard and Downs concluded originally that not all types of variation are applicable to the four uses of Markush structures, as shown in Table 2. Since their original paper, they are beginning to have second thoughts about Table 2: they think that more variations may actually be applicable to libraries in some senses.

Combinatorial library problems include a combinatorial explosion of coverage and "segmentation" of the library (arbitrary boundaries between monomers). Solutions to both of these problems can be provided by techniques developed for Markush structures in patents. Markush representation is clearly essential since storage of enumerated specific compounds is impractical. Structure and substructure search must be based on Markush representation, not on enumerated specifics. The search must not depend on the original segmentation[68].

Table 3. Comparison of Structural Descriptors

metric	good news	bad news
2D fingerprints (e.g., keys, Tanimoto)	convenient (free with your database!)	highly dimensional, Receptors can not read! non-intuitive clusters non-Euclidean
connectivity	convenient, medium dimensional	receptors (still) can not read (structures)!
log P/molar refractivity	low dimensional, very relevant to pharmacokinetics	no selective binding, estimation can fail
CoMFA	very relevant to binding	need aligned 3D models, highly dimensional
"3D" fingerprints	binding-relevant	need 3D models, "all" conformations, non Euclidean

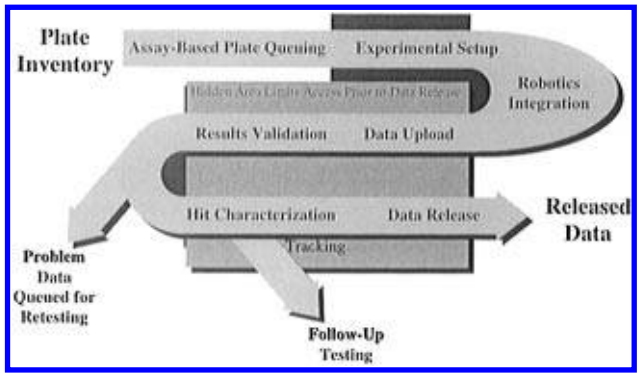

Figure 7. Screening process.

SOFTWARE FOR DATA MANAGEMENT AND WORKFLOW

To date, I am aware of four workflow software packages for high throughput screening: ActivityBase from ID Business Solutions,[69] BioSTAR (Biological Sample Tracking and Registration) from Tripos, CSAP (Chemical Screening Analysis Package) from MLR Automation,[70] and MDL SCREEN from MDL Information Systems (see Figure 7).

MDL SCREEN manages a plate-based sample inventory, the robotics assay process, and results validation and release. It stores 10^9 data points (limited only by hardware). It integrates robotics automation systems and MDL products such as ISIS and Central Library. It is a workflow management system which emphasizes integration and performance. It is based on ORACLE (as opposed to a desktop rdbms), so it is a multiuser, server-based system which can handle very large volumes of data. MDL also emphasizes its data model and its data validation facilities.

This is not to say that other systems lack any or all of the above features: MDL SCREEN has been picked simply as an example. It is far too early to do an objective comparison of the four systems. BioSTAR is a client-server system built using (Macintosh-based) 4D from ACI US. It contains links to ORACLE and can be integrated with systems from Tripos. BioSTAR was developed originally at Onyx Pharmaceuticals where a database expert worked alongside bioscientists to emulate their workflow procedures. ActivityBase is being integrated with Chemical Design's software for combinatorial chemistry. It is based on Access and has direct links to ORACLE and SQL Server. Curve fitting and SAR generation modules have recently been added. CSAP is an Access database application, designed by high throughput screening professionals, that can run as a standalone database on a single PC for small biotech companies or can be connected directly to a powerful client server application such as ORACLE for large high throughput screening programs.

SELECTION AND CLUSTERING

Let us now return to the topic of chemistry and the problem of the enormous number of compounds that could be made (or bought) and tested and perhaps tested unnecessarily if they are similar to previous compounds. Weininger[71] illustrates "selection" with the following statement:

> "There are 10^{180} possible drugs, 10^{18} likely drugs, 10^7 known compounds, 10^6 commercially available compounds, 10^6 compounds in corporate databases, 10^4 compounds in drug databases, 10^3 commercial drugs and 10^2 profitable drugs".

The procedure for generating a diverse library, or comparing libraries or databases, consists of calculating structural descriptors for the molecules, weighting them according to some scheme, and then calculating a similarity coefficient. Cluster-based, dissimilarity-based, or partition-based compound selection is then carried out. Since there are other detailed papers in this publication, and the subject has been reviewed elsewhere,[72−74] the following material concentrates on work that has not yet appeared in the literature.

Teig[75] of CombiChem believes in "good" libraries, that maximize information, as opposed to "diverse" libraries. He believes that pairwise diversity measures (distances between molecules) do not correlate with information. His is not the only team to believe that 2D measures of diversity are inadequate, but many pharmaceutical companies have produced useful results from generation of Daylight 2D fingerprints as chemical descriptors and clustering algorithms for their analysis. Brown and Martin[76] have published a comparison of clustering methods and descriptors for use in compound selection. Workers at Chiron have published a paper on multidimensional scaling and D-optimal design.[77]

A comparison of descriptors presented by Cramer[78] at a meeting in 1995 is given in Table 3. Cramer later reported[79] on validity of descriptors. If a diversity measurement has neighborhood behavior (is "valid") for a particular data set, then a plot of the absolute differences in that measurement *versus* the absolute differences in the biological activity will have a characteristic appearance, an absence of data points in the upper left triangle of the plot. (In this region of the plot, any data points indicate that a small change in the diversity property produced a large change in the biological activity, which completely contradicts the desirable neighborhood behavior.) A χ^2 test can be used to assess statistically whether the density of data points in some Lower Right Trapezoid (LRT) is higher than the overall average density of points. Cramer showed that topomeric steric CoMFA, 2D fingerprints-side chains, and topomeric Hbond spatial fingerprints are valid descriptors; 2D fingerprints-whole molecule, atom pairs (Sheridan), and autocorrelation (Moreau), are somewhat valid; and connectivity indices (HDI, first 10), partition coefficient (CLogP), molar refractivity (CMR), "strain energy" (Tripos force field), and random number are not valid.

In the Optiverse library design process of Tripos and Panlabs, a virtual library of over 100 million compounds is computer screened by 2D, 3D, and ensemble properties and structure. Thus a diverse library of 100 000 compounds for synthesis is designed. The process involves collating reagents, then selecting reagents, then building product combinations, and finally selecting products. The product selection process is a second filter based on Tanimoto coefficients and 2D fingerprints.

Molecular Simulations (MSI) is another company[80] exploring solutions to the productivity and data management challenges associated with compound libraries, assessing library SAR, and integrating library design and analysis with synthesis and high throughput screening informatics. MSI also integrates its combinatorial chemistry products with comprehensive software tools for receptor and analogue-based design, but the company does not intend to compete with products for registering and handling chemical libraries. Rather, it is concentrating on molecular design and analysis tools, using shape-based 3D searching techniques as well as traditional pharmacophore-based 3D searching. Mason and co-workers[81,82] and Martin[74] are also amongst those currently interested in 3D methods more relevant to biological interactions. Mason has successfully used Chemical Design's product, ChemDiverse,[83] for handling pharmacophore diversity.

CONCLUSIONS

In Figure 5 the drug design cycle was illustrated. The goal is not just to go round and round the cycle and then finally break out with a potential drug candidate. If sufficient information is gathered and well used, it should be possible to assemble yet more knowledge of receptors and disease processes which will further the process of bringing better drugs to market faster.

REFERENCES AND NOTES

(1) Machin, P. Combinatorial chemistry. Paper given on April 23, 1996, at Silicon Graphics, Reading, UK.

(2) Drews, J. The impact of cost containment on pharmaceutical research and development. 1995 Center for Medicines Research Annual Lecture.

(3) Figures from *Drug News and Perspectives*, February 1995 quoted by Machin, *loc. cit.*

(4) The March 1996 issue of *Acc. Chem. Res.* is dedicated to combinatorial chemistry.

(5) Hermkens, P.; Ollenheijm, H.; Rees, D. Solid-phase Organic Reactions 1992−1995. *Tetrahedron* **1996**, *52*, 4527−4554.

(6) Thompson, L. A.; Ellman, J. A. Synthesis and Applications of Small Molecule Libraries. *Chem. Rev.* **1996**, *96(1)*, 555−600.

(7) Lowe, G. Combinatorial Chemistry. *Chem. Soc. Rev.* **1995**, *24(5)*, 309−382.

(8) Terrett, N. K.; Gardner, M.; Gordon, D. W.; Kobylecki, R. J.; Steele, J. Combinatorial synthesis—the design of compound libraries and their application to drug discovery. *Tetrahedron* **1995**, *51(30)*, 8135−8173.

(9) Gallop, M. A.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gordon, A. M. Applications of combinatorial technologies to drug discovery 1. *J. Med. Chem.* **1994**, *37*, 1233−1251.

(10) Gordon, E. M.: Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. Applications of combinatorial technologies to drug discovery 2. *J. Med. Chem.* **1994**, *37*, 1385−1401.

(11) Desai, M. C.; Zuckermann, R. N.; Moos, W. H. Recent advances in the generation of chemical diversity libraries. *Drug Dev. Res.* **1994**, *33*, 174−188.

(12) Moos, W. H.; Green, G. D.; Pavia, M. R. Recent advances in the generation of molecular diversity. *Ann. Rep. Med. Chem.* **1993**, *28*, 315−324.

(13) Combinatorial Libraries: Synthesis, Screening and Application Potential; Cortese, R., Ed.; Walter de Gruyter: Berlin, 1996.

(14) Ecker, D. R.; Crooke, S. T. Combinatorial drug discovery: which methods will produce the greatest value? *Biotechnology* **1995**, April 13, 351−360.

(15) ESCOM Science Publishers B. V., *Mol. Diversity* URL http://vesta.pd.com.

(16) To subscribe to the mol-diversity listserver send a message SUBSCRIBE MOL-DIVERSITY ⟨firstname lastname⟩ to LISTSERV@LISTSERV.ARIZONA. EDU.

(17) Gani, D. Challenges in combinatorial chemistry: synthesis, labeling, screening and deconvolution. Paper given at the meeting Synthetic Challenges in Combinatorial Chemistry, on February 27, 1996, organized by the Society of Chemical Industry, London, England.

(18) Merrifield, R. B. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85*, 2149−2154.

(19) Leznoff C. C. The use of insoluble polymer supports in organic chemical synthesis. *Chem. Soc. Rev.* **1974**, *3*, 65−85.

(20) Leznoff C. C. The use of insoluble polymer supports in general organic synthesis. *Acc. Chem. Res.* **1978**, *11*, 327−333.

(21) Geysen, H. M. Meloen, R. H.; Barteling, S. J. Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid.. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 3998−4002.

(22) Geysen, H. M.; Barteling, S. J.; Meloen, R. H. Small peptides induce antibodies with a sequence and structural requirement for binding antigen comparable to antibodies raised against the native protein. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 178−182.

(23) Houghten, R. A. General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5131−5135.

(24) Furka, A; Sebestyen, F.; Asgedom, M.; Dibo, G. General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Peptide Protein Res.* **1991**, *37*, 487−493.

(25) Lam, K. S. *et al.* A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **1991**, *354*, 82−84

(26) Houghten, *et al*. *Nature* **1991**, *354*, 84−86.

(27) Ugi, I. In *Isonitrile Chemistry*. Blomquist, A. T., Ed.; Academic Press: New York, 1971; p 133.

(28) Ugi, I.; Dömling, A.; Hörl, W. *Endeavour* **1994**, *18*, 115.

(29) Dömling, A.; Ugi, I. *Angew. Chem., Int. Ed. Engl.* **1994**, *32*, 563.

(30) Hogan, J. C. Automated high throughput synthesis and analysis of small molecular thematic arrays. Paper given at the Cambridge Healthtech Institute Conference *Exploiting molecular diversity: small molecule libraries for drug discovery* (January 1996). (A version of the proceedings is available from Wendy Warr & Associates.)

(31) Pirrung, M. Indexed combinatorial libraries: nonoligomeric chemical diversity for the discovery of novel enzyme inhibitors. Paper given at the Cambridge Healthtech Institute Conference *Exploiting molecular diversity: small molecule libraries for drug discovery* (January 1996). (A version of the proceedings is available from Wendy Warr & Associates.)

(32) Simon, R. J. *et al.* Peptoids: a modular approach to drug discovery *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9367−9371.

(33) Simon, R. J. *et al.* (1994) In *Techniques in Protein Chemistry*; Crabb, J. W., Ed.; Academic Press: New York; Vol. 5, pp 533−539.

(34) DeWitt, S. *et al*. "Diversomers": an approach to nonpeptide, nonoligomeric chemical diversity. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6909−6913.

(35) Bunin, B. A.; Ellman, J. A. A general and expedient method for the solid-phase synthesis of 1,4-benzodiazepine derivatives. *J. Am. Chem. Soc.* **1992**, *114*, 10997.

(36) Plunkett, M. J.; Ellman, J. A. Solid-phase synthesis of structurally diverse 1,4-benzodiazepine derivatives using the Stille coupling reaction. *J. Am. Chem. Soc.* **1995**, *117*, 3306−3307.

(37) Bunin, B. A.; Ellman, J. A. Increasing the diversity of a 1,4-benzodiazepine library through side-chain functionalisation. *Polymer Preprints* **1994** *35*, 983−984.

(38) Bunin, B. A.; Plunkett, M. J.; Ellman, J. A. The combinatorial synthesis and chemical and biological evaluation of a 1,4-benzodiazepine library. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4708−4712.

(39) Ellman, J. A "Solid phase and combinatorial synthesis of benzodiazepine compounds on a solid support". U.S. Patent No. 5,288,514, issued 1994.

(40) Valerio, R. M.; Bray, A. M.; Stewart, K. M. Multipin solid-phase synthesis of acyl 2,3-diaminopropionic acid oligomers. *Int. J. Peptide Protein Res.* **1994**, *44*, 158−165.

(41) Virgilio, A. A.; Ellman, J. A. Simultaneous solid-phase synthesis of β-turn mimetics incorporating side-chain functionality. *J. Am. Chem. Soc.* **1994**, *114*, 11580−11581.

(42) Zuckermann *et al*. Encoded combinatorial peptide libraries containing non-natural amino acids. *J. Am. Chem. Soc.* **1993**, *115*, 2529−2531.

(43) Dooley, C. T.; Houghten, R. A. The use of positional scanning synthetic peptide combinatorial libraries for the rapid determination of opioid receptor ligands. *Life Sci.* **1993**, *52*, 1509−1517.

(44) Janda, K. D. Tagged versus untagged libraries: methods for the generation and screening of combinatorial chemical libraries. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10779−10785.
(45) Borchardt, A.; Still, W. C. Synthetic receptor binding elucidated with an encoded combinatorial library. *J. Am. Chem. Soc.* **1994**, *116*, 373−4.
(46) Needels, M. C. *et al*. Generation and screening of an oligonucleotide-encoded synthetic peptide library. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10700−10704.
(47) Nicolaou, K. C., Xiao, X-Y.; Nova, M. P. Radio frequency encoded combinatorial chemistry. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2289.
(48) Moran, E. J. *et al.* Radio frequency tag encoded combinatorial library method for the discovery of tripeptide-substituted cinnamic acid inhibitors of the protein tyrosine phosphatase PTP1B. *J. Am. Chem. Soc.* **1995**, *117*, 10787−10788.
(49) Advanced ChemTech, 5609 Fern Valley Road, Louisville, KY 40228−1075, U.S.A., tel. 502 969 0000 or 800 456 1403, fax 502 968 1000.
(50) Martin, J. F. New instruments for solid-phase synthesis. Paper given at the Cambridge Healthtech Institute Conference *Solid-phase synthesis: developing small molecule libraries* (January 1996). (A version of the proceedings is available from Wendy Warr & Associates.) Argonaut Technologies, Inc., 887 Industrial Road, Suite G, San Carlos, CA 94070, U.S.A., tel. 415 598 1350, fax: 415 598 1359.
(51) Harness, J. R. Automation of high throughput synthesis. In *Molecular Diversity and Combinatorial Chemistry: Libraries and Drug Discovery*; Chaiken, I. M., Janda, K. D., Eds.; American Chemical Society: Washington, D.C., 1996; Chapter 18. Bohdan Automation, Inc., 1500 McCormick Boulevard, Mundelein, IL 60060, U.S.A., tel. 708 680 3939, fax. 708 680 1199.
(52) Bray, A. M. Simultaneous multiple synthesis by the multipin method: techniques for multiple handling and reaction optimization on solid phase. Paper given at the Cambridge Healthtech Institute Conference *Solid-phase synthesis: developing small molecule libraries* (January 1996). (A version of the proceedings is available from Wendy Warr & Associates.) Chiron Mimotopes, 11055 Roselle Street, San Diego, CA 92121, U.S.A., tel. 619 558 5800, and 800 644 1866, fax 619 558 5810.
(53) Myers, P. L. A totally integrated chemistry approach to drug discovery. Paper given at the Cambridge Healthtech Institute Conference *Exploiting molecular diversity: small molecule libraries for drug discovery* (January 1996). (A version of the proceedings is available from Wendy Warr & Associates.) CombiChem, Inc., 9050 Camino Santa Fe, San Diego, CA 92121, U.S.A., tel. 619 530 0484, fax 619 530 9998.
(54) CRS Robotics Corporation, 5344 John Lucas Drive, Burlington, Ontario, Canada, L7L 6A6, tel. 905 332 2000, fax 905 332 1114.
(55) DeWitt S. H. *et al*. DIVERSOMER technology: solid-phase synthesis, automation, and integration for the generation of chemical diversity. *Drug Dev. Res.* **1994**, *33*, 116−124. DIVERSOMER Technologies, Inc., 2800 Plymouth Road, Ann Arbor, MI 48105, U.S.A., tel. 313 996 7418, fax 313 998 2782.
(56) The Technology Partnership plc, Melbourn Science Park, Cambridge Road, Melbourn, Royston, Hertfordshire, SG8 8EE, UK, tel. +44 1763 262626, fax +44 1763 261582.
(57) Harris, A. L.; Toyonaga, B. E. Application of polymer-supported chemistry to the discovery and optimization of lead drug candidates. In *Molecular Diversity and Combinatorial Chemistry: Libraries and Drug Discovery*; Chaiken, I. M., Janda, K. D., Eds.; American Chemical Society: Washington, D.C., 1996; Chapter 26. Ontogen Corporation, 2325 Camino Vida Roble, Carlsbad, CA 92009, U.S.A., tel. 619 930 0100, fax: 619 930 0200.
(58) Sagian, PO Box 68536, Indianapolis, IN 46268, U.S.A., tel. 317 328 3588, 800 352 4975, fax 317 328 3589.
(59) Tecan, P. O. Box 13953, Research Triangle Park, NC 27709, U.S.A., tel. 919 361 5200, 800 338 3226, fax 919 361 5201.
(60) Tomtec, 607 Harborview Road, Orange, CT 06477, U.S.A., tel. 203 795 5030, fax 203 248 5724.
(61) Zymark Corporation, Zymark Center, Hopkinton, MA 01748-1668, U.S.A., tel. 508 435 9500, fax 508 435 3439.
(62) MDL Information Systems Inc., 14600 Catalina Street, San Leandro, CA 94577, U.S.A., tel. 510 895 1313, fax 510 352 2870.
(63) Chemical Design Ltd., Roundway House, Cromwell Park, Chipping Norton, Oxon. OX7 5SR, UK, tel. +44 1608 644000, fax +44 1608 642244.
(64) Grethe, G. Solid-Phase Organic REactions Database (SPORE). Paper given at the Cambridge Healthtech Institute Conference *) Solid-phase synthesis: developing small molecule libraries* (January 1996). (A version of the proceedings is available from Wendy Warr & Associates.)
(65) Synopsys Scientific Systems Ltd., 175 Woodhouse Lane, Leeds LS2 3AR, UK, tel. +44 113 245 3339, fax +44 113 243 8733.
(66) Daylight Chemical Information Systems, Inc., 27401 Los Altos, Suite # 370 Mission Viejo, CA 92691, U.S.A., tel. 714 367 9990, fax: 714 367 0990.
(67) Tripos Inc. 1699 South Hanley Road, St. Louis, MO 63144, U.S.A., tel. 314 647 1099, 800 323 2960, fax 314 647 9241.
(68) Barnard, J. M.; Downs, G. M. Applications of Markush structure techniques to handling combinatorial libraries. Paper given at 210th National ACS Meeting, Chicago, August 1995.
(69) Id Business Solutions, Surrey Technology Center, Occam Road, Guildford, Surrey, GU2 5YG, UK, tel. +44 1483 295950, fax +44 1483 295951.
(70) MLR Automation, 204 Littlefield Avenue, South San Francisco, CA 94080, U.S.A., tel. 415 589 1197, fax 415 589 1078.
(71) I am grateful to John Bradshaw of Glaxo Wellcome for drawing this quotation to my attention.
(72) Johnson, M. A.; Maggiora, G. M. *Concepts and Applications of Molecular Similarity*; J. Wiley and Sons: New York, 1990.
(73) Willett, P. Computational tools for the analysis of molecular diversity. In *Combinatorial Chemistry: A Short Course*; Czarnik, A., Hobbs DeWitt, S., Eds.; ACS Books: Washington DC, in press.
(74) Martin, Y.; Brown, R. D.; Bures, M. G. Quantifying Diversity. In *Combinatorial Chemistry and Molecular Diversity*; Kerwin, J. F., Gordon, E. M., Eds.; Wiley: New York, in press.
(75) Teig, S. Diversity, shmiversty: is your library any good? Paper given at the Cambridge Healthtech Institute Conference *Exploiting molecular diversity: small molecule libraries for drug discovery* (January 1996). (A version of the proceedings is available from Wendy Warr & Associates.)
(76) Brown, R. D.; Martin, Y. C. Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572−584.
(77) Martin, E. J.; Blaney, J. M.; Siani, M. A.; Spellmeyer, D. C.; Wong, A. K.; Moos, W. H. *et al.* Measuring diversity: experimental design of combinatorial libraries for drug discovery. *J. Med. Chem.* **1995**, *38*, 1431−1436.
(78) Cramer, R. Informational challenges for high throughput library screening. Paper given at the Cambridge Healthtech Institute Conference *Exploiting molecular diversity: small molecule libraries for drug discovery* (January 1995). (A version of the proceedings is available from Wendy Warr & Associates.)
(79) Cramer, R. Validating new molecular diversity measurements. Paper given at the Cambridge Healthtech Institute Conference *Exploiting molecular diversity: small molecule libraries for drug discovery* (January 1996. (A version of the proceedings is available from Wendy Warr & Associates.)
(80) Molecular Simulations Inc., 9685 Scranton Road, San Diego, CA 92121-3752, U.S.A., tel. 619 458 9990, fax 619 458 0136.
(81) Ashton, M. J.; Jaye, M. C.; Mason, J. S. New Perspectives in Lead Generation. II Evaluating Molecular Diversity. *Drug Discovery Today* **1996**, *2*, 71−78.
(82) Pickett, S. D.; Mason, J. S.; McLay, I. M. Diversity Profiling and Design Using 3D Pharmacophores: Pharmacophore-Derived Queries (PDQ) *J. Chem. Inf. Comput. Sci.*, in press.
(83) Davies, K. Using pharmacophore diversity to select molecules to test from commercial catalogs, including DIVERSet and HTS Chemicals. In *Molecular Diversity and Combinatorial Chemistry: Libraries and Drug Discovery*; Chaiken, I. M., Janda, K. D., Eds.; American Chemical Society: Washington, D.C., 1996; Chapter 27, pp 309−316.

CI9601426