# Chemical space navigation in lead discovery

Tudor I Oprea

The number of new chemical entities has remained rather constant (averaging 37 per year) in the past decade, despite the multiple-fold increase in the number of compounds that are being made and tested. Chemical space requires novel methods that can handle the increasing number of potentially accessible molecules. Neighborhood behavior, as an approach to similarity, and chemical property space navigation are some of the recent advances that are discussed, in the context of lead discovery and appropriate pharmacokinetic properties.

**Addresses**
EST Lead Informatics, AstraZeneca R&D Mölndal, S-43183 Mölndal, Sweden; e-mail: tudor.oprea@astrazeneca.com

**Current Opinion in Chemical Biology** 2002, **6**:384–389

1367-5931/02/$ – see front matter


***Published online 15 April 2002***

**Abbreviations**

| | |
|---|---|
| **FBPF** | fuzzy bipolar pharmacophore fingerprint |
| **HTS** | high-throughput screening |
| **NB** | neighborhood behavior |
| **NCE** | new chemical entity |
| **PCA** | principal component analysis |

## Introduction

The vastness of chemical space cannot be better illustrated than by using David Weininger's example [1]: considering all the derivatives of *n*-hexane, starting from a list of 150 substituents, and enumerating all the possibilities, from mono- to 14-substituted hexanes, one can easily reach over $10^{29}$ structures. While most of them might be, to date, synthetically inaccessible, a small number of building blocks can still give an infinite number of possibilities, as seen in living systems: 20 natural amino acids and five nucleotides form the basis of most proteins and nucleic acids, respectively. As chemical space is limited at the low-molecular-weight end only, methods to navigate it are becoming increasingly important. Not only for molecular biology and material science, but also for medicinal chemistry — and in particular for lead discovery.

The pharmaceutical industry has benefited, in recent years, from a technological breakthrough in terms of rapid access to a large number of novel compounds and related biological data, through combinatorial chemistry [2••], and high-throughput screening (HTS) [3]. However, this plethora of chemical and biological data has yet to translate into clinical success [4•], despite the logical progression of HTS actives into lead compounds and drug candidates to the launched status [5•]. In fact, the number of new chemical entities (NCEs) has remained rather constant in the past decade [6•], even though the number of molecules being synthesized and screened increased by several orders of magnitude during the same period. Because poor pharmacokinetic behavior is, besides poor efficacy, the main reason for failure in the clinical trials [7], this review focuses on chemical space navigation tools, in particular on methods that increase the awareness for pharmacokinetic properties, besides molecular diversity.

**Box 1**

According to the *Webster's New World Dictionary & Thesaurus* (Macmillan Publishers), **sim·i·lar·i·ty** (sim′ ler′ ē), **n.,** *pl.* **-ties**, is **1.** the state or quality of being similar; resemblance or likeness; **2.** a point, feature, or instance in which things are similar. Synonyms for similarity include: likeness, connection, comparison, parallel, correspondence, match, and relationship – all of which indicate that a point of reference is required.

## Chemical similarity validation

One of the misleading concepts in chemical space navigation relates to chemical, or molecular similarity [8,9,10••–12••]. Similarity is, by definition (Box 1), related to a particular framework: that of a descriptor system (we need a metric by which to judge similarity), as well as that of an object, or class of objects (we need a reference point to which objects can be compared). Similarity depends on the choice of molecular descriptors [13•], the choice of the weighting scheme(s) and the similarity coefficient itself [10••]. Its essence, thus, remains in the eye of the beholder: their nature will classify oranges and tomatoes in a different category (i.e. fruit vs. vegetable), although they appear to be quite similar from a geometrical perspective (spheroidal, perhaps 7–15 cm in diameter). In fact, even in the macro world, similarity perception is not as trivial as one might think, as illustrated by the fruit vs. vegetable debate — see Box 2. By the same token, 3D-based pharmacophore [14] fingerprints [15] are expected to afford different similarity results than, for example, the Daylight fingerprints [16], as recently shown by Dixon and Merz [17•], but not necessarily because one metric is better than the other. Rather, one is expected to use different similarity or, indeed, diversity metrics for different purposes.

For example, 'supine' is dissimilar when compared to 'ovine', 'lupine' and 'vulpine' — which, in turn, can be clustered (i.e. similar) when the choice is from this list alone, but all of the above would stand out as outliers (i.e. dissimilar) in a comprehensive list of cities in Sweden. Thus, the misleading character of 'molecular (dis)similarity' relates to our (in)ability to suitably compare the proverbial apples and oranges, while choosing a metric that is most fitting for our requirements. If one is tempted to estimate the validity of any molecular (dis)similarity metric by comparing its output to the binding affinities of ligands to a promiscuous target like albumin or the 3A4 isozyme of cytochrome P450, one should not neglect the fact that such

targets have a built-in ability to bind an impressive variety of endogenous and exogenous structures. On the other hand, biological targets that are much more precise in terms of the molecular recognition features that are likely to result in high-affinity, high-specificity ligands, are more likely to assist us in validating and refining chemical similarity methods that, in turn, can be used to rapidly scan large databases for ligands of interest. The higher the specificity and selectivity of our target, the more likely we are to find that molecular similarity tools become relevant. By the same token, the term 'maximum diversity' can not, or should not, be taken out of context: maximum diversity can only refer to the number of combinations available in a limited pool of reagents [18–20] but should be regarded as an inefficient strategy for chemical space navigation [21••]. Chemical diversity is infinite — our understanding of it is only limited by time and resources.

## Chemical space neighborhood and lead-like space

When compounds need to be prioritized for synthesis, or even for HTS assays [22•], one needs to ask the question: serendipity [23••] notwithstanding, how similar (or how diverse) do I want my compound library to be, based on prior target-related knowledge? Cheminformatics [24,25•] has an increasingly important role in the decision-making process [26], namely in constantly re-evaluating biologically relevant information, both at the chemical (2D, 3D) and structural (3D) level, in the lead and drug discovery process (see Figure 1). Reagent selection is more than ever dependent on chemical space navigation tools, and should be guided by structural chemistry and biological testing: as the information content increases, one should use targeted diversity methods [27,28] in order to maximize the chances of finding active compounds.

The task of navigating in chemical space is more challenging in the absence of 3D-structural target information. An approach that goes beyond 'maximum diversity' starts from a pre-defined activity space for over 584 drugs and drug-like compounds, tested on 42 different targets (Horvath D, Jeandenans C, personal communication), activity space that allows an unambiguous mapping in chemical and biological space. Similarity is then validated against two criteria: 'completeness' (i.e. the number of compound pairs with similar activity profiles that have high similarity scores), and 'consistency' (i.e. the number of compound pairs with different *in vitro* profiles that have low similarity values) (Horvath D, Jeandenans C, personal communication). These two criteria are used to validate a novel similarity metric, based on fuzzy bipolar pharmacophore fingerprints (FBPFs) [29], that relies on the concept of neighborhood behavior (NB) [30]. An extension of the similarity postulate, which states that structurally similar molecules are expected to display similar biological activity, the NB concept states that more active compounds than random are expected to be found in a subset of the structurally nearest neighbors of an active compound, given two sets of the same size (one random, one NB-selected). The FBPF-based autocorrelograms [29] have been further benchmarked against other similarity metrics (Horvath D, Jeandenans C, personal communication), and found to be significantly more appropriate in capturing neighborhood behavior in a more general manner.

**Box 2**

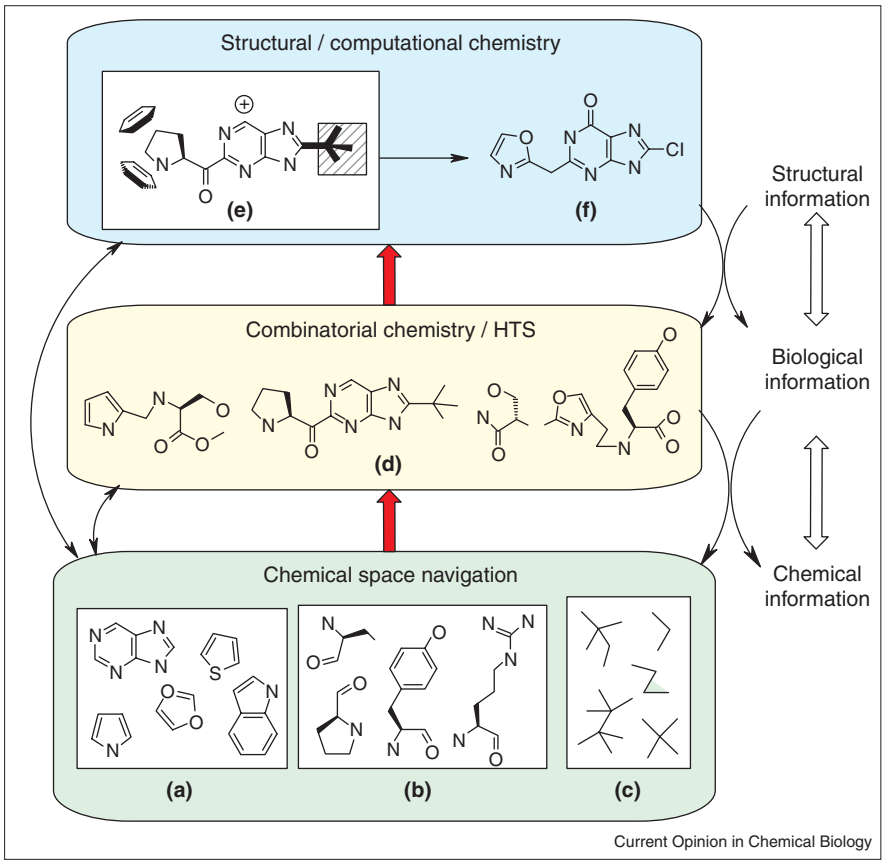
"Botanically speaking, the tomato is a fruit, but legally speaking it is a vegetable. The Supreme Court of the United States said so in 1893. An importer had argued that tomatoes were fruit and therefore not subject to a duty in effect at that time. The Court held that the tomato is a vegetable because it was usually served at dinner in, with, or after the soup, or with fish or meats that constitute the main part of the meal. This is less true now than it was then, for today a much larger part of our tomato crop is made into juice, but the tomato remains, legally, a vegetable."

Extracted from http://aggie-horticulture.tamu.edu/plantanswers/vegetables/vegancestors.html

In a different study, the autocorrelogram-based ALMOND descriptors [31] (ALMOND is available from Multivariate Infometric Analysis srl, Perugia, Italy, http://www.miasrl.com) have been shown to provide reliable quantitative structure–activity models for a range of targets: glycogen phosphorylase b, corticosteroid-binding globulin and serotoninergic (5-$HT_{2A}$) receptors, thus providing a novel class of alignment-independent, chemically interpretable, pharmacophore-related descriptors. One can expect autocorrelogram-based pharmacophore tools to emerge as a method of choice in exploring neighborhood behavior, and probing chemical space.

The NB concept is a similarity-rooted extension of the related medicinal chemistry tenet of 'lead series': structurally related chemicals (lead series) are expected to exhibit a good range of biological activities (preferably three orders of magnitude or more), to provide the basis for (quantitative) structure–activity relationship analyses. Besides the good pharmacokinetic properties required for high-quality leads (discussed in the next section), other features of promising lead series include favorable intellectual property position and relatively simple structures, suitable for further chemical alterations of chosen molecules ('leads') in order to optimize the target activity. The ability to yield such lead series from combinatorial libraries has been put to question [32•], based on Lipinski's earlier work [33] that analyzed high-molecular-weight, low-polarity trends at Pfizer. This led to in-depth analyses regarding the exact nature of chemical leads [34,35] that established the presence of a lead-like space. Thus, on average, leads are expected to have lower molecular complexity [34] when compared with drugs (fewer rings, fewer rotatable bonds), to be smaller (lower molecular weight), and more polar [35]. Furthermore, the intellectual property position in a lead series could be strengthened by theoretical or experimental means. On the theoretical side, lead hopping [36] can be used by jumping across different series [37] using the 3D-based topomer similarity technique [38]

**Figure 1**




The chemical information flow in the drug-discovery process: chemical space navigation allows one to select reagents (e.g. **(a)** heterocyclic scaffolds, **(b)** amino acids and **(c)** spacers) that yield, via multiple parallel syntheses, **(d)** products, which are biologically tested by HTS. Structural information (from X-ray, NMR or computational chemistry models) could then provide an in-depth understanding of the intermolecular forces that are relevant for the biological target in question (e.g. **(e)** steric, electrostatic and π–π interactions), which may lead to **(f)** optimized ligands. Red arrows indicate a logical progression, whereas white arrows indicate the interdependence between chemical, biological and structural information. Line arrows indicate feedback loops.

(Topomer-based similarity techniques are available from Tripos Inc., St Louis, Missouri, http://www.tripos.com), or by surfing across the 'scaffold space' with SORT&gen [39] (SORT&gen is available from SPECS and BioSPECS, Rijswijk, The Netherlands, http://www.specs.net). On the experimental side, one could resort to high-energy gamma-ray radiations (Kessler U, Pilger BD, Zerbe O, Scapozza L, Folkers G, personal communication) or to high-speed microwave chemistry [40].

## Chemical property space navigation

Using the analogy of an intercity distances table, in contrast to a geographical map, Martin and Critchlow [41•] pointed out the advantage of having a chemical space map, rather than mere distance-based 'diversity' in combinatorial library design. Having the right inter-object distances is clearly not enough, as one is likely to be successful in finding a list of, for example, five cities in Western Europe that have identical (or close) distances to cities on the East Coast of the United States. In the absence of a proper map, a sixth city in Eastern Europe could, in the wrong context, be placed somewhere in the Atlantic Ocean. In other words, context-sensitive information is required while evaluating chemical spaces, even though appropriate measures may have been taken with respect to distance-based (dis)similarity. Thus, an effective chemical space mapping effort requires a property-based system, besides an efficient similarity/diversity metric.

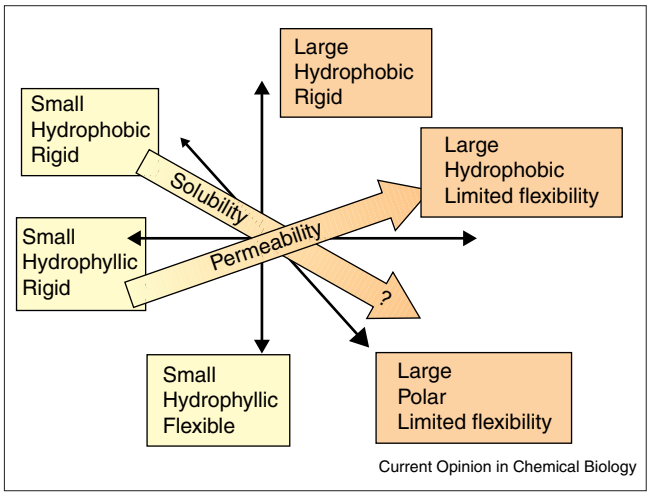
Early attempts to map physical properties described the two-dimensional **BC(DEF)**, 'bulk' and 'cohesiveness' parameters [42], derived from six physical properties (aqueous solvation energy, partition coefficient, boiling point, molecular refractivity, volume and vaporization enthalpy) for a set of 114 pure liquids. This scheme was shown to work quite well for a set of 139 diverse structures [43]. By analogy to the Mercator convention in geography, we recently suggested *chemography*, a combination of rules (not unlike the **BC(DEF)** dimensions) and objects (chemical structures), to provide a consistent, global chemical space map [44,45]. Chemographic rules included, initially, general properties such as size, lipophilicity, and hydrogen bond capacity, while objects include 'satellites', intentionally placed outside the drug-like space, as well as 'core' objects, selected mostly from a list of orally available drugs. ChemGPS, the chemical global positioning system, comprises both the 'core' and 'satellite' molecules. Chemographic map coordinates are extracted, in ChemGPS, by principal component analysis (PCA) [46], from a (fixed) list of molecular descriptors that evaluate the above-mentioned rules on a single set of molecules. PCA-score prediction is used, then, to project new molecules on the

pre-established map — providing thus a consistent and systematic method to map the chemical property space. A map derived with the same choice of rules (dimensions) and the same set of compounds (objects) can then be used to compare PCA scores across a large number of chemicals, because it does not change with chemistry and time. ChemGPS is therefore well suited to become a reference system for comparing multiple combinatorial libraries, and for keeping track of previously explored regions of the chemical space.

The ability of ChemGPS to provide a global property space map was evaluated using atom-count, atom-type count, topological, electrostatic and physico-chemical property descriptors [44] derived from 2D-structures, for a set of 45 heteroaromatic compounds [47] previously described with GRID [48], for a set of 87 α-amino acids with known z-scores (principal properties) [49], as well as for a set of 8600 monocarboxylic acids, for which extensive comparisons between local and global PCA models were performed. Based on the above descriptors, a nine-dimensional map was derived that could be interpreted in terms of, for example, size, hydrophobicity and flexibility (see Figure 2).

Pharmacokinetic property prediction is currently being considered early on in the process of drug discovery — to the extent that combinatorial chemistry libraries are being scrutinized for ADME (absorption, distribution, metabolism and excretion) properties [50,51]. This has recently been extended to virtual screening efforts [52,53]. GPSVS, the combination of ChemGPS molecules (objects) and VolSurf descriptors (Volsurf is available from Molecular Discovery Ltd, London, UK, http://www.moldiscovery.com) [54] (rules) was used to map the chemical space with respect to passive permeability and solubility — key properties according to the US Food and Drug Administration [55]. VolSurf has been extensively validated in both oral absorption [56,57] and blood-brain-barrier permeation [58•] models. We have recently shown [59] that, in the absence of any biological input, GPSVS (a model based on PCA prediction), correlates well with passive transcellular permeability, as illustrated for the Caco-2, ghost erythrocyte and blood-brain barrier datasets — for the first dimension, and with solubility, as investigated using the octanol–water partition, and intrinsic solubility datasets — for the second dimension, respectively. The two GPSVS dimensions are no longer orthogonal, forming an angle of approximately 43° — see also Figure 2. The global property map of chemical space, starting from 2D-based [44] and 3D-based [59] descriptors is depicted in Figure 2: a general sense of property location can be derived, based on the fact that the first dimension in the two ChemGPS models has a weak ($r^2 = 0.61$) correlation. The question mark in the solubility arrow is a reminder that the exact relationship between chemical properties, in a general sense, and aqueous solubility, needs further investigation.

**Figure 2**




Chemical space map interpretation based on the combined ChemGPS models derived from 2D (black arrows) and 3D (colored arrows) descriptors. Small structures are located on the left side, whereas large structures are on the right side. Hydrophobic compounds occur more often on the upper side, and hydrophilic compounds are more frequent on the lower side. Rigid structures are more frequent below the paper plane, whereas flexible structures occur more often above. Permeability has some degree of correlation with size and hydrophobicity, whereas solubility cannot be rigorously placed.

## Conclusions

Recent advances in the area of neighborhood behavior and chemical property space mapping enable us to complement 'chemical similarity' in ways that are meaningful for combining chemical, biological and structural information. Pharmacokinetic and chemical property space can be simultaneously queried, together with biological activity space and chemical fingerprints or pharmacophores. This is likely to enable chemistry projects to reach an informed decision in the early stages of lead identification, thus allowing the progression of higher-quality leads through the drug-discovery pipeline.

## Acknowledgements

I would like to thank Drs Katherine Andrews Cramer and Richard D Cramer (both from Tripos, Inc.) and Dr Dragos Horvath (from CEREP) for sharing their work prior to publication.

## References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Weininger D: **Combinatorics of small molecular structures.** In *Encyclopedia of Computational Chemistry*, Vol 1. Edited by Von Ragué Schleyer P, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer HF III. New York: Wiley; 1998:425-530.

2. •• Lebl M: **Parallel personal comments on 'classical' papers in combinatorial chemistry.** *J Comb Chem* 1999, **1**:3-24.
Insightful, personal comments and key references from authors involved early on in the field of combinatorial chemistry.

3. Fox S, Farr-Jones S, Yund MA: **High throughput screening for drug discovery: continually transitioning into new technology.** *J Biomol Screening* 1999, **4**:183-186.

4. • Horrobin DF: **Innovation in the pharmaceutical industry.** *J Royal Soc Med* 2000, **93**:341-345.
This paper articulates some of the issues related to marketing-driven preclinical research in the pharmaceutical industry.

5. • Drews J: **Drug discovery: a historical perspective.** *Science* 2000, **287**:1960-1964.
A synopsis of the drug-discovery process is provided.

6. • Gaudillière B, Bernardelli P, Berna P: **To market, to market — 2000.** *Annu Rep Med Chem* 2000, **36**:293-318.
Each year, *Annual Reports in Medicinal Chemistry* summarizes the NCEs introduced into the world market for the first time in the previous year. 35 NCEs were introduced into their first markets in 2000. Previous years are as follows, with the number of NCEs first introduced in the market that year given in brackets: 1991 (36), 1992 (36), 1993 (43), 1994 (a record 44 NCEs), 1995 (35), 1996 (38), 1997 (39), 1998 (27), and 1999 (35). The average for this decade is 37 NCEs.

7. Kennedy T: **Managing the drug discovery/development interface.** *Drug Discov Today* 1997, **2**:436-444.

8. Willett P: *Similarity and Clustering Techniques in Chemical Information Systems*. Letchworth: Research Studies Press; 1987.

9. Johnson MA, Maggiora GM: *Concepts and Applications of Molecular Similarity*. New York: Wiley; 1990.

10. •• Willett P: **Chemoinformatics — similarity and diversity in chemical libraries.** *Curr Opin Biotechnol* 2000, **11**:85-88.
Good overview of cluster-based, partition-based, dissimilarity-based and optimization-based selection techniques for combinatorial synthesis planning.

11. •• Lewis RA, Pickett SD, Clark DE: **Computer-aided molecular diversity analysis and combinatorial library design.** *Rev Comput Chem* 2000, **16**:1-51.
This is an exhaustive review related to molecular similarity and diversity. Methods to probe chemical diversity, including molecular descriptor validation, are discussed in the context of combinatorial library design.

12. •• Martin YC: **Diverse viewpoints on computational aspects of molecular diversity.** *J Comb Chem* 2001, **3**:231-250.
This review provides first-hand accounts and key references from authors involved early on in the field of molecular similarity and diversity.

13. • Todeschini R, Consonni V: *Handbook of Molecular Descriptors*. Weinheim: Wiley-VCH; 2000.
This book covers ca. 3000 molecular descriptors.

14. Gund P: **Three-dimensional pharmacophoric pattern searching.** *Prog Mol Subcell Biol* 1977, **11**:117-143.

15. Olender R, Rosenfeld R: **A fast algorithm for searching for molecules containing a pharmacophore in very large virtual combinatorial libraries.** *J Chem Inf Comput Sci* 2001, **41**:731-738.

16. Daylight Chemical Information System, Santa Fe, New Mexico. http://www.daylight.com

17. • Dixon SL, Merz KM: **One-dimensional molecular representations and similarity calculations: methodology and validation.** *J Med Chem* 2001, **44**:3795-3809.
This paper provides a simplified, one-dimensional projection of molecular structures, that can be generated from 2D or 3D structures.

18. Leach AR, Bradshaw J, Green DVS, Hann MM: **Implementation of a system for reagent selection and library enumeration, profiling, and design.** *J Chem Inf Comput Sci* 1999, **39**:1161-1172.

19. Shi S, Peng Z, Kostrowicki J, Paderes G, Kuki A: **Efficient combinatorial filtering for desired molecular properties of reaction products.** *J Mol Graph Model* 2000, **18**:478-496.

20. Lobanov VS, Agrafiotis DK: **Stochastic similarity selections from large combinatorial libraries.** *J Chem Inf Comput Sci* 2000, **40**:460-470.

21. •• Lipinski CA: **Drug-like properties and the causes of poor solubility and poor permeability.** *J Pharmacol Toxicol Methods* 2000, **44**:235-249.
Analysis of clinical candidates at Merck indicates that the 'rational design approach' leads to poorer permeability, whereas Pfizer's 'HTS approach' appears to result in clinical candidates with poorer solubility.

22. • Agrafiotis DK, Rassokhin DN: **Design and prioritization of plates for high-throughput screening.** *J Chem Inf Comput Sci* 2001, **41**:798-805.
The issue of plate selection for HTS is discussed in view of plate-based molecular diversity and similarity to known lead(s). User-defined selection criteria (e.g. purity), can also be included in these experimental design schemes.

23. •• Kubinyi H: **Chance favors the prepared mind — from serendipity to rational drug design.** *J Rec Signal Transduction Res* 1999, **19**:15-39.
Fast-paced review of serendipitous drug discoveries, emphasizing structure-based design aspects. "Screening, especially…HTS, can be considered as a systematic approach to benefit from mere chance".

24. Brown F: **Chemoinformatics: what is it and how does it impact drug discovery.** *Annu Rep Med Chem* 1998, **33**:375-384.

25. • Hahn MM, Green R: **Cheminformatics — a new name for an old problem?** *Curr Opin Chem Biol* 1999, **3**:379-383.
Both software and hardware are discussed from a practitioner's perspective.

26. Olsson T, Oprea TI: **Cheminformatics: a tool for decision-makers in drug discovery.** *Curr Opin Drug Discov Dev* 2001, **4**:308-313.

27. Li J, Murray CW, Waszkowycz B, Young SC: **Targeted molecular diversity in drug discovery — integration of structure-based design and combinatorial chemistry.** *Drug Discov Today* 1998, **3**:105-112.

28. Lewis RA: **The design of small- and medium-sized focused combinatorial libraries.** In *Molecular Diversity in Drug Design*. Edited by Dean PM, Lewis RA. Dordrecht: Kluwer Academic Publishers; 1999:221-248.

29. Horvath D: **High throughput conformational sampling and fuzzy similarity metrics: a novel approach to similarity searching and focused combinatorial library design and its role in the drug discovery laboratory.** In *Combinatorial Library Design and Evaluation for Drug Design*. Edited by Ghose AK, Viswanadhan VN. New York: Marcel Dekker Inc.; 2001:429-472.

30. Patterson DE, Cramer RD, Ferguson AM, Clark RD, Weinberger LE: **Neighborhood behavior: a useful concept for validation of 'molecular diversity' descriptors.** *J Med Chem* 1996, **39**:3049-3059.

31. Pastor M, Cruciani G, McLay I, Pickett S, Clementi S: **GRid-INdependent Descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors.** *J Med Chem* 2000, **43**:3233-3243.

32. • Teague SJ, Davis AM, Leeson PD, Oprea TI: **The design of leadlike combinatorial libraries.** *Angew Chem Int Ed Engl* 1999, **38**:3743-3748. German version: *Angew Chem* 1999, **111**:3962-3967.
Low-molecular weight and low-hydrophobicity are discussed as key properties when designing combinatorial libraries.

33. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ: **Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.** *Adv Drug Deliv Rev* 1997, **23**:3-25.

34. Hann MM, Leach AR, Harper G: **Molecular complexity and its impact on the probability of finding leads for drug discovery.** *J Chem Inf Comput Sci* 2001, **41**:856-864.

35. Oprea TI, Davis AM, Teague SJ, Leeson PD: **Is there a difference between leads and drugs? A historical perspective.** *J Chem Inf Comput Sci* 2001, **41**:1308-1315.

36. Cramer RD, Poss MA, Hermsmeier MA, Caulfield TJ, Kowala MC, Valentine MT: **Prospective identification of biologically active structures by topomer shape similarity searching.** *J Med Chem* 1999, **42**:3919-3933.

37. Cramer RD, Jilek RJ, Andrews KM: **Dbtop: topomer similarity searching of conventional structure databases.** *J Mol Graph Model* 2002, in press.

38. Andrews KM, Cramer RD: **Toward general methods of targeted library design: topomer shape similarity searching with diverse structures as queries.** *J Med Chem* 2000, **43**:1723-1740.

39. De Laet A, Hehenkamp JJJ, Wife RL: **Finding drug candidates in virtual and lost/emerging chemistry.** *J Heterocyclic Chem* 2000, **37**:669-674.

40. Kappe CO: **High-speed combinatorial synthesis utilizing microwave irradiation.** *Curr Opin Chem Biol* 2002, **6**:this issue.

41. • Martin EJ, Critchlow RE: **Beyond mere diversity: tailoring combinatorial libraries for drug discovery.** *J Comb Chem* 1999, **1**:32-45.
Chemical property binning, via multi-dimensional scaling, could facilitate chemical diversity void identification.

42. Cramer RD: **BC(DEF) parameters: 1. The intrinsic dimensionality of intermolecular interactions in the liquid state.** *J Am Chem Soc* 1980, **102**:1837-1849.

43. Cramer RD: **BC(DEF) parameters: 2. An empirical structure-based scheme for the prediction of some physical properties.** *J Am Chem Soc* 1980, **102**:1849-1859.

44. Oprea TI, Gottfries J: **Chemography: the art of chemical space navigation.** *J Comb Chem* 2001, **3**:157-166.

45. Oprea TI, Gottfries J: **ChemGPS: a chemical space navigation tool.** In *Rational Approaches to Drug Design.* Edited by Höltje HD, Sippl W. Barcelona: Prous Science Press; 2001:437-446.

46. Jackson JE: *A Users Guide to Principal Components*. New York: Wiley; 1991.

47. Clementi S, Cruciani G, Fifi P, Riganelli D, Valigi R, Musumarra G: **A new set of principal properties for heteroaromatics obtained by GRID.** *Quant Struct Act Relat* 1996, **15**:108-120.

48. Goodford PJ: **Computational procedure for determining energetically favourable binding sites on biologically important macromolecules.** *J Med Chem* 1985, **28**:849-857.

49. Sandberg M, Eriksson L, Jonsson J, Sjöström M, Wold S: **New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids.** *J Med Chem* 1998, **41**:2481-2491.

50. Darvas F, Dorman G: **Early integration of ADME/Tox parameters into the design process of combinatorial libraries.** *Chim Oggi* 1999, **17**:10-13.

51. Pickett SD, McLay IM, Clark DE: **Enhancing the hit-to-lead properties of lead optimization libraries.** *J Chem Inf Comput Sci* 2000, **40**:263-272.

52. Oprea TI, Zamora I, Svensson P: **Quo vadis, scoring functions? Toward an integrated pharmacokinetic and binding affinity prediction framework.** In *Combinatorial Library Design and Evaluation for Drug Design.* Edited by Ghose AK, Viswanadhan VN. New York: Marcel Dekker Inc; 2001:233-266.

53. Oprea TI: **Virtual screening in lead discovery: a viewpoint.** *Molecules* 2002, **7**:51-62.

54. Cruciani G, Crivori P, Carrupt PA, Testa B: **Molecular fields in quantitative structure-permeation relationships: the VolSurf approach.** *J Mol Struct (Theochem)* 2000, **503**:17-30.

55. Anonymous: **Waiver of *in vivo* bioavailability and bioequivalence studies for immediate-release solid oral dosage forms based on a biopharmaceutics classification system.** 2000. Available from http://www.fda.gov/cder/OPS/ BCS_guidance.htm.

56. Guba W, Cruciani G: **Molecular field-derived descriptors for the multivariate modeling of pharmacokinetic data.** In *Molecular Modeling and Prediction of Bioactivity.* Edited by Gundertofte K, Jørgensen FS. New York: Kluwer Academic/Plenum Publishers; 2000:89-94.

57. Zamora I, Oprea TI, Ungell AL: **Prediction of oral drug permeability.** In *Rational Approaches to Drug Design*, Edited by Höltje HD, Sippl W. Barcelona: Prous Science Press; 2001:271-280.

58. Crivori P, Cruciani G, Carrupt PA, Testa B: **Predicting blood-brain barrier permeation from three-dimensional molecular structure.** *J Med Chem* 2000, **43**:2204-2216.
• A comprehensive list of drugs that penetrate (or not) the blood-brain-barrier.

59. Oprea TI, Zamora I, Ungell AL: **A pharmacokinetically based mapping device for chemical space navigation.** *J Comb Chem* 2002, in press.