International Journal of Chemical Sciences and Applications.
Vol 1, Issue 1, June, 2010, pp 16-24.

APPLICATION OF CHEMOINFORMATICS FOR INNOVATIVE DRUG DISCOVERY

Vishnu J. Gaikwad
Department of Biotechnology, Tatyasaheb Kore Institute of Engineering and Technology, Warananagar,
Tal: Panhala, Dist: Kolhapur, Pin: 4116113, MS, India

ABSTRACT
Chemoinformatics is the mixing of those information resources to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the area of drug lead identification and optimization.
Since then, both spellings have been used, and some have evolved to be established as cheminformatics. Steps in the drug discovery process: disease selection, target hypothesis, lead compound identification (screening), lead optimization, clinical trial and pharmacogenomic optimization. Traditionally, these steps are carried out sequentially. Structure linear notations convert chemical structure connection tables to a string, a sequence of letters, using a set of rules. The earliest structure linear notation was the Wiswesser Line Notation (WLN).
Using all the advanced chemoinformatics system, it enhances the drug discovery rapidly and with low cost and helps to eminent scientists to synthesize the drugs.

Keywords: Chemoinformatics, lead optimization, clinical trial and pharmacogenomic optimization.

INTRODUCTION:
The term Chemoinformatics was defined by F.K. Brown [1][2] in 1998: Chemoinformatics is the mixing of those information resources to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the area of drug lead identification and optimization.
Since then, both spellings have been used, and some have evolved to be established as Cheminformatics, [3] while European Academia settled in 2006 for Chemoinformatics. [4] The recent establishment of the Journal of Cheminformatics is a strong push towards the shorter variant.
Cheminformatics combines the scientific working fields of chemistry and computer science for example in the area of chemical graph theory and mining the chemical space.[5][6] Cheminformatics can also be applied to data analysis for various industries like paper and pulp, dyes and such allied industries. Virtual libraries of classes of compounds (drugs, natural products, diversity-oriented synthetic products) were recently generated using the FOG (fragment optimized growth) algorithm. [7]

Small molecules with at most a few dozen atoms play a fundamental role in organic chemistry and biology. They can be used as combinatorial building blocks for chemical synthesis [8, 13, 14], as molecular probes for perturbing and analyzing biological systems in chemical genomics and systems biology [9, 14, 15],

and for the screening, design, and discovery of useful compounds. These include of course new drugs [11, 12], the majority of which are small molecules. Furthermore, huge arrays of new small molecules can be produced in a relatively short period of time [10, 13].

Data:

large and well-annotated datasets are essential for developing statistical machine learning methods in chemoinformatics, whether supervised or unsupervised, including predictive classification, regression, and clustering of small molecules and their properties [16, 17]. Several parallel support

have emerged recently to start to address the data bottleneck, including PubChem (http://pubchem.ncbi.nlm.nih.gov), the Harvard Chembank [18], UCSF's ZINC [19], and the UCI ChemDB [20]. The UCI ChemDB is a public database containing over 4M compounds as well as

a repository of annotated datasets that can be used to develop statistical machine learning methods. Together, these datasets already pose important challenges for both supervised and unsupervised machine learning methods, from clustering to kernel methods [17, 21].

Drug Discovery

There are seven steps in the drug discovery process: disease selection, target hypothesis, lead compound identification (screening), lead optimization, pre-clinical trail, and clinical trial and pharmacogenomic optimization. Traditionally, these steps are carried out sequentially [22]. The average cost of creating a NCE in a major pharmaceutical company was estimated at around $7,500/compound [23]. To reduce costs, pharmaceutical companies have had to find new technologies to replace the old "hand-crafted" synthesis and testing NCE approaches. Since 1980, with the advent of high throughput screening (HTS), automated techniques have made possible robotized screening. Through this process, hundreds of thousands of individual compounds can be screened per drug target per year [23, 24]. In response to the increased demand for new compounds by biologists, chemists started using combinatorial chemical technologies to produce more new compounds in shorter periods. Combinatorial chemistry (CC) systematically and repetitively yields a large array of compounds from sets of different types of reagents, called "building blocks". By 2000, many solution- and solid-phase CC strategies were well-developed [25].

Drug and Lead properties.

Technologies have been developed to recognize drug-like compounds from a diverse compound library [26-32]. These drug-like measuring and filtering technologies have partly solved the screening problems. However, they have not been good enough to completely solve these problems. It has been observed that many drug-like compounds, which should be potential candidates; do not come up as hits when they are screened against biological targets. It is believed that further refinement of the filtering technologies should be made in order to recognize *lead-like* compounds [33-34] instead of *drug-like* compounds. Intrinsically, lead-likeness and drug-likeness are the descriptors of potency, selectivity, absorption, distribution, metabolism, toxicity, and scalability

Drug Discovery Process and Early ADMET Prediction.

One now finds too many hits when searching for lead candidates, thus lead optimization is stymied. To get more target structural information, high-throughput protein crystallization has been explored [35-36]. lead optimization remains the most serious bottleneck. In

addition, we know that, about forty percent of all development candidates fail due to absorption, distribution, metabolism, excretion and toxicity ("ADMET") problems [37]. HTS for pharmaceutical discovery was used as a filter in order to identify the few potentially promising hits in a corporation's synthetic archive. Therefore, HTS data analyses were focused on hits, and the bulk of the non-hit data was ignored [38]. Cheminformatics methods must be applied while generating data using high throughput techniques in order to assure that good ADMET properties are achieved while making and screening compounds, This approach is called a multi-parametric optimization strategy [39].

Chemical structure database.

Since structure and substructure searches are typical NP problems, they were computationally costly [40]. In order to make structure and sub-structure searching feasible on slow computer systems, many methods were attempted in order to find concise structural representations, such as, linear notations. These convert structural graphs to strings that can easily be searched by a computer. The data screening strategies filtered out the compounds were not the main structural features (search keys) in a given query. Then, an atom-by-atom search algorithm was applied (this was usually time consuming) to a smaller number of compounds. Subsequently, screening approaches have been used in most of chemical database management systems.

Linear notations.

Structure linear notations convert chemical structure connection tables to a string, a sequence of letters, using a set of rules. The earliest structure linear notation was the Wiswesser Line Notation (WLN). ISI® adopted WLN to be used in some of their products in 1968 and, it is still use today. It was also adopted in the mid 1960s for internal use by many pharmaceutical companies. At that time (mid 60s to 80s) [41], it was considered the best tool to represent, retrieve and print chemical structures. In WLN, letters represents structural fragments and a complete structure is represented as a string. This system efficiently compressed structural data and, was very useful to storing and searching chemical structures in low performance computer systems. However, the WLN is difficult for nonexperts to understand. Later, David Weininger suggested a new linear notation designated as SMILESTM [42-43]. SMILESTM is widely accepted and used in many chemical database systems.

Visualizing structures from graphed data points

Chemical structure graphs are chemists' natural language. Since a compound library is mapped to points on a two-dimensional graph, a reasonable requirement is for one to have an easy way to see the structure by pointing to the corresponding dot. This problem has been well resolved by Spotfire® software [44][45].

The criteria used for selecting descriptors should be: (1) the selected descriptors should be bioactivity related (requiring correlation analysis), (2) the selected descriptors should be informative (should have diversified value distributions), (3) the selected descriptors should be independent of each other (if two descriptors are correlated to each other, related property will be unfairly biased), (4) the selected descriptors should be simple to extract, easy to explain to a chemist, invariant to irrelevant transformations, insensitive to noise, and efficient to discriminate patterns in different categories (specificity). After comparing performance and predictability

in high throughput data mining, researchers from multiple groups have consistently concluded that 2D descriptors perform significantly better than 3D descriptors [46-49].

Clustering and Partitioning.

The term cluster analysis (CA) was first used by Tryon, in 1939. Actually CA encompasses a number of different classification algorithms. A general question in many areas of an inquiry is how to organize the observed data into meaningful structures, that is, how to develop taxonomies [50]. Hierarchical clustering rearranges objects in a tree-structure (see Table 1). Javis-Patrick (also known as nearest neighbor cluster algorithm) is commonly used to cluster chemical structures [51].

One of the most popular decision tree techniques is recursive partitioning (RP). It has been reported that RP algorithms can partition on data sets with over 100,000 compounds and 2,000,000 descriptors, in less than an hour [52-53]. RP algorithms can also be used to build multivariable regression models.

Virtual library generation and Virtual screening.

As equipment is being automated and miniaturized, HTS capacity keeps expanding [54].

But, increased HTS efforts have not significantly increased drug discovery successes [55]. Considering total lead-like molecular space, the total percentage of compounds that current technologies have made and screened, is still small. This has made way for the birth of *in silico* or virtual screening (VS) technology [56-58].

In conjunction with high-throughput screening technology, virtual screening has become a main tool for identifying leads [56]. Virtual screening is actually one of the computational tools used to filter out unwanted compounds from physical libraries or *in silico* libraries. In order to reduce drug discovery costs, one needs to remove undesired compounds as early as possible. Filters have been built based upon oral bioavailability, aqueous solubility, and metabolic clearance and, chemically reactivity or toxic chemical groups [59-60]. A virtual screening method for identification of “frequent hitters” in compound libraries has been reported [61]. If the target structure is known, one of the structure-based virtual screening methods that can be used is high throughput docking [62-63]. If the target structure is unknown, but the ligands from the literature or, competitors are known, then, similarity approaches can be applied [64-66]. If neither target structure nor ligand structure is known, then SAR patterns can be derived from experimental screening data by statistical approaches [67-69]. Also, virtual screening is a great tool for the design of a combinatorial library with a given target. For example, Hopfinger and coworks have constructed a combinatorial library of glucose inhibitors of glycogen phosphorylase *b* using virtual screening technology and 4D-QSAR analyses [70]. Using the 4D-QSAR model developed for a training set of 47 glucose analogue inhibitors of glycogen phosphorylase, the investigators have developed a virtual approach to screen a focused combinatorial virtual library of 225 inhibitors. Analysis of the binding predictions across the virtual library reveals patterns of structureactivity information. The patterns are then used to design new focused libraries. A recent review has indicated that HTS and VS are moving toward integration [71].

In silico ADMET.

Under multi-parametric optimization drug discovery strategies, there is no excuse for failing to know the relative solubility and permeability rankings of collections of chemical compounds for lead identification [72]. The method used (VolSurf) transforms 3D fields into descriptors and correlates them to the experimental permeation by a discriminant partial least squares procedure [73]. Human serum albumin (HSA) protein is the major transporter of non-esterified fatty acids, as well as of different drugs and metabolites, to different tissues. HSA allows solubilization of hydrophobic compounds, contributes to a more homogeneous distribution of drugs in the body, and increases their biological lifetime. The binding strength of any drug to serum albumin is the main factor for availability of that drug to diffuse from the circulatory system to target tissues. All these factors cause the pharmacokinetics of almost any drug to be influenced and controlled by its binding to serum albumin [74]. Binding to HSA turns out to be determined by a combination of hydrophobic forces together with some modulating shape factors [75]. This agrees with X-ray structures of HSA alone or, bound to ligands, where the binding pockets of both sites and II are composed mainly of hydrophobic residues [76]. HTS has been used for metabolism and pharmacokinetics [77-78]. *In vitro* approaches determine metabolic stability, screening for inhibitors of specific cytochrome P450 isozymes and, identifying the most important metabolites. QSAR and pharmacophore models, protein models, and expert systems. QSAR and pharmacophore models predict substrates and inhibitors of a specific cytochrome P450 isozyme [79-80]. Protein models rationalize metabolite formations and identify possible substrates, potential metabolites or, inhibitors by means of docking algorithms [81-82]. Stereoelectronic factors involved in metabolic transformations can be taken into account using quantum chemical calculations. Expert systems are predictive databases that attempt to identify potential metabolites of a compound as determined by knowledge based rules defining the most likely products [83-84]. The mechanistic approach involves human experts who make a considered assessment of the mechanism of interaction with a biological system, taking the molecular properties, biological data, and chemical structures into account [85]. The correlative approach uses an unbiased assessment of the data to generate relationships and predict toxicity. It is capable of discovering potentially new SARs [86]

Conclusion and future scope:

Using all the advanced chemoinformatics system, it enhances the drug discovery rapidly and with low cost and helps to eminent scientists to synthesize the chemical molecules which laeds to helps the society.

References:

1. F.K. Brown (1998). "Chapter 35. Chemoinformatics: What is it and How does it Impact Drug Discovery". Annual Reports in Med. Chem. 33: 375.doi:10.1016/S0065-7743(08)61100-8.
2. Brown, Frank (2005). "Editorial Opinion: Chemoinformatics – a ten year update". Current Opinion in Drug Discovery & Development 8 (3): 296–302.
3. Chemoinformatics (http://www.molinspiration.com/chemoinformatics.html)
4. Obernai Declaration (http://infochim.ustrasbg.fr/chemoinformatics/Obernai%20Declaration.pdf)
5. Gasteiger J. (Editor), Engel T.(Editor): Chemoinformatics : A Textbook. John Wiley & Sons, 2004, ISBN 3-527-30681-1

6. A.R. Leach, V.J. Gillet: An Introduction to Chemoinformatics. Springer, 2003, ISBN 1-4020-1347-7
7. Kutchukian, Peter; Lou, David; Shakhnovich, Eugene (2009). "FOG: Fragment Optimized Growth Algorithm for the de Novo Generation of Molecules occupying Druglike Chemical". Journal of Chemical Information and Modeling 49: 1630–1642.
8. Agra¯otis, D.K., Lobanov, V.S., and Salemme, F.R., Combinatorial informatics in the post-genomics era, *Nature Reviews Drug Discovery*, 1:337{346, 2002.
9. Dobson, C.M., Chemical space and biology, *Nature*, 432:824{828, 2004.
10. Houghten, R.A., Parallel array and mixture-based synthetic combinatorial chemistry: tools for the next millenium. *Annual Review of Pharmacology and Toxicology*, 40:273{282, 2000.
11. Jonsdottir, S.O., Jorgensen, F.S., and Brunak, S., Prediction methods and databases within chemoinformatics: Emphasis on drugs and drug candidates, *Bioinformatics*, 21:2145{2160, 2005.
12. Lipinski, C. and Hopkins, A., Navigating chemical space for biology and medicine, *Nature*, 432:855{861, 2004.
13. Schreiber, S.L., Target-oriented and diversity-oriented organic synthesis in drug discovery, *Science*, 287:1964{1969, 2000.
14. Schreiber, S.L., The small-molecule approach to biology: chemical genetics and diversity-oriented organic synthesis make possible the systematic exploration of biology, *Chemical and Engineering News*, 81:51{61, 2003.
15. Stockwell, B.R., Exploring biology with small organic molecules, *Nature*, 432:846{854, 2004.
16. Micheli, A., Sperduti, A., Starita, A., and Biancucci, A.M., A novel approach to QSPR/QSAR based on neural networks for structures, *Soft Computing Approaches in Chemistry*, Cartwright, H. and Sztandera, L.M. (ed.), *Springer Verlag* 265{296, 2003.
17. Ralaivola, L., Swamidass, S.J., Saigo, H., and Baldi, P., Graph kernels for chemical informatics.Neural Networks, 2005. Special issue on Neural Networks and Kernel Methods for Structured Domains, 2005, In press.
18. Strauseberg, R.L. and Schreiber, S.L., From knowing to controlling: a path from genomics to drugs using small molecule probes, *Science*, 300(5617):294{295, 2003.
19. Irwin, J.J. and Shoichet, B.K., ZINC{A free database of commercially available compounds for virtual screening, *Journal of Chemical Information and Computer Sciences*, 45:177{182, 2005.
20. Chen, J., Swamidass, S.J., Dou, Y., Bru, J., and Baldi, P., ChemDB: A public database of small molecules and related chemoinformatics resources, *Bioinformatics*, 2005, In Press.
21. Swamidass, S.J., Chen, J., Bruand, J., Phung, P., Ralaivola, L., and Baldi, P., Kernels for small molecules and the prediction of mutagenicity, toxicity, and anti-cancer activity, Proceedings of the 2005 ISMB Conference, *Bioinformatics*, 21(Supplement 1):i359{368, 2005.
22. Augen, J. “The evolving role of information technology in the drug discovery process”, *Drug Discov. Today*, **2002**, *7*, 315-323.
23. Gallop, M. A.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gordon, E. M. “Applications of Combinatorial Technologies to Drug Discovery. 1. Background and Peptide Combinatorial Libraries”, *J. Med. Chem.*, **1994**, *37*, 1233-1251.
24. Hecht, P. “High-throughput screening: beating the odds with informatics-driven chemistry”,*Curr. Drug Discov.*, January 2002, 21-24. *Molecules* **2002**, 7589
25. Hall, D. G.; Manku, S.; Wang, F. **“**Solution- and Solid-Phase Strategies for the Design, Synthesis, and Screening of Libraries Based on Natural Product Templates: A Comprehensive Survey”, *J. Comb. Chem.*, **2001**, *3*, 125-150
26. (a) Bemis, G. W.; Murcko, M. A. “The properties of known drugs. 1. Molecular Frameworks”, *J.Med. Chem.*, **1996**, *39*, 2887-2893; (b) Bemis, G. W.; Murcko, M. A. “The properties of knowndrugs. 2. Side Chains”, *J. Med. Chem.*, **1999**, *42*, 5095-5099.
27. Ajay; Walters, W. P.; Murcko, M. A. “Can we learn to distinguish between “drug-like” and “nondrug-like” molecules?” *J. Med. Chem.*, **1998**, *41*, 3314-3324.
28. Sadowski, J.; Kubinyi, H. “A scoring scheme for discriminating between drugs and non-drugs”,*J. Med. Chem.*, **1998**, *41*, 3325-3329.
29. Xu, J.; Stevenson, J. “Drug-like Index: A New Approach To Measure Drug-like Compounds and Their Diversity” *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 1177 –1187.
30. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. “Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings”, *Adv. Drug Deliv. Rev.*, **1997**, *23*, 3-25.

31. Clark, D. E. and Pickett, S. D., "Computational methods for the prediction of 'drug-likeness'",*Drug Discov. Today*, **2000**, *5*, 49-58.
32. Matter, H.; Baringhaus, K.-H.; Naumann, T.; Klabunde, T.; Pirard, B. "Computational approaches towards the rational design of drug-like compound libraries", *Comb. Chem. High T.Scr*., **2001**, *4*, 453-475.
33. Oprea, T. I., Davis, A. M., Teague, S. J., and Leeson, P. D. "Is There a Difference between Leads and Drugs? A Historical Perspective", *J. Chem. Inf. Comput. Sci*., **2001**, *41*, 1308 -1315.
34. Proudfoot, J. R. "Drugs, Leads, and Drug-Likeness: An Analysis of Some Recently Launched Drugs", *Bioorg. Med. Chem. Lett*., **2002** (in press).
35. Stewart, L.; Clark, R.; Behnke, C. "High-throughput crystallization and structure determination in drug discovery", *Drug Discov. Today*, **2002**, *7*, 187-196.
36. Luft, J. R.; Wolfley, J.; Collins, R.; Bianc, M.; Weeks, D.; Jurisica, I.; Rogers P.; Glasgow, J.; Fortier, S.; DeTitta, G. T. "High Throughput Protein Crystallization: Keeping up with the Genomics", 2002,www.imca.aps.anl.gov/~ahoward/luft_ab.html
37. (a). Kennedy, T. *Drug Discov. Today*, **1997**, *2*, 436-444.(b). Start-Up: Windhover's Review of Emerging Medical Ventures, July 2000, page 34,www.windhoverinfo.com/contents/monthly/exex/e_2000900126.htm
38. Manly, C. J.; Louise-May, S.; Hammer, J. D. "The impact of informatics and computationalchemistry on synthesis and screening", *Drug Discov. Today*, **2001**, *6*, 1101-1110.
39. Baxter, A. D. and Lockey, P. M., " 'Hit' to 'lead' and 'lead' to 'candidate' optimization using multi-parametric principles", *Drug Discov. World*, **2001**, *2*, 9-15.
40. Xu, J. "GMA: A Generic Match Algorithm for structural Homomorphism, Isomorphism, Maximal Common Substructure Match and Its Applications", *J. Chem. Inf. Comput. Sci*., **1996**,*36*, 25-34.
41. http://www.asis.org/Features/Pioneers/wiswess.htm
42. Weininger, D. **"**SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules", *J. Chem. Inf. Comput. Sci.* , **1988**, *28*, 31-6.
43. http://esc.syrres.com/interkow/docsmile.htm
44. http://www.spotfire.com/
45. Xu, J. "SCA: New Cluster Algorithm for Structural Diversity Analysis and Applications", The First Spotfire Users Conference, Philadelphia, May 30, 2001.
46. 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.
47. Matter, H.; Pötter, T. "Comparing 3D Pharmacophore Triplets and 2D Fingerprints for Selecting Diverse Compound Subsets", *J. Chem. Inf. Comput. Sci*., **1999**, *39*, 1211 -1225.
48. Estrada, E.; Molina, E.; and Perdomo-Lopez, I. "Can 3D Structural Parameters Be Predicted from 2D (Topological) Molecular Descriptors?", *J. Chem. Inf. Comput. Sci*., **2001**, *41*, 1015 -1021.
49. Xue, L.; Stahura, F. L.; Godden, J. W.; Bajorath, J. "Mini-fingerprints Detect Similar Activity of Receptor Ligands Previously Recognized Only by Three-Dimensional Pharmacophore-BasedMethods", *J. Chem. Inf. Comput. Sci*., **2001**, *41*, 394 -401.
50. (a). Tryon, R. C. *J. Chronic Dis*., **1939**, *20*, 511-524.(b). http://www.statsoftinc.com/textbook/stcluan.html
51. Willett, P. "Similarity and Clustering in Chemical Information Systems", Research Studies Press, Wiley: New York, 1987.
52. Rusinko, A., III; Farmen, M. W.; Lambert, C. G.; Brown, P. L.; Young, S. S. "Analysis of a Large Structure/Biological Activity Data Set Using Recursive Partitioning", *J. Chem. Inf. Comput. Sci*., **1999**, *39*, 1017-1026.
53. Rusinko, A., III; Young, S. S.; Drewry, D. H.; Gerritz, S. W. "Optimization of Focused Chemical Libraries Using Recursive Partitioning", *Comb. Chem. High T. Scr*., **2002**, *5*, 125-133.
54. Mander, T. "Beyond uHTS: ridiculously HTS?", *Drug Discov. Today*, **2000**, *5*, 223-225.
55. Valler, M. J.; Green, D. "Diversity screening versus focused screening in drug discovery", *Drug Discov. Today*, **2000**, *5*, 286-293.
56. Walters, W. P.; Stahl, M. T.; Murcko, M. A. "Virtual screening – an overview", *Drug Discov. Today*, **1998**, *3*, 160-178.
57. Joseph-McCarthy, D. "An overview of *in silico* design and screening: Toward efficient drug discovery", *Curr. Drug Discov*., March 2002, 20-23. *Molecules* **2002**, 7595

58. Bajorath, J. "Virtual screening in drug discovery: Methods, expectations and reality", *Curr. Drug Discov.*, March 2002, 24-27.
59. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings", *Adv. Drug Deliver. Rev.*, **1997**, *23*, 3-25.
60. Huuskonen, J.; Rantanen, J.; Livingstone, D. "Prediction of aqueous solubility for a diverse set of organic compounds based on atom-type electrotopological state indices", *Eur. J. Med. Chem*., **2000**, *35*, 1081-1088.
61. Zuegge, J.; Schneider, G.; Coassolo, P.; Lave, T. "Prediction of hepatic metabolic clearancecomparison and assessment of prediction models", *Clin. Pharmacokinet*., **2001**, *40*, 553-563.
62. Roche, O.; Schneider, P.; Zuegge, J.; Guba, W.; Kansy, M.; Alanine, A.; Bleicher, K.; Danel, F.; Gutknecht, E. M.; Rogers-Evans, M.; Neidhart, W.; Stalder, H.; Dillon, M.; Sjogren, E.; Fotouhi, N.; Gillespie, P.; Goodnow, R.; Harris, W.; Jones, P.; Taniguchi, M.; Tsujii, S.; von der Saal, W.; Zimmermann, G.; Schneider, G. "Development of a Virtual Screening Method for Identification of 'Frequent Hitters' in Compound Libraries", *J. Med. Chem.*, **2002**, *45*, 137 -142.
63. Abagyan, R.; Totrov, M. "High-throughput docking for lead generation", *Curr. Opin. Chem. Biol*., **2001**, *5*, 375-382. Diller, D. J.; Merz, Jr., K. M. "High throughput docking for library design and library prioritization", *Proteins*, **2001**, *43*, 113-124.
64. Willett, P. "Chemoinformatics – similarity and diversity in chemical libraries", *Curr. Opin. Biotech*., **2000**, *11*, 85-88.
65. Hopfinger, A. J.; Duca, J. S. "Estimation of molecular similarity based on 4D-QSAR analysis: formalism and validation", *J. Chem. Inf. Comput. Sci*., **2001**, *41*, 1367-1387.
66. Makara, G. M. "Measuring molecular similarity and diversity: total pharmacophore diversity", *J. Med. Chem.*, **2001**, *44*, 3563-3571.
67. Hopfinger, A. J.; Duca, J. "Extraction of pharmacophore information from high-throughput screens", *Curr. Opin. Biotech.*, **2000**, *11*, 97-103.
68. Roberts, G.; Myatt, G. J.; Johnson, W. P.; Cross, K. P.; Blower, Jr., P. E. "Lead Scope: Software for Exploring Large Sets of Screening Data", *J. Chem. Inf. Comput. Sci*., **2000**, *40*, 1302 -1314.
69. Willet, P.; Gedeck, P. "Visual and computational analysis of structure-activity relationships in high-throughput screening data", *Curr. Opin. Chem. Biol*, **2001**, *5*, 389-395.
70. Hopfinger, A. J.; Reaka, A.; Venkatarangan, P.; Duca, J. S.; Wang, S. "Construction of a Virtual High Throughput Screen by 4D-QSAR Analysis: Application to a Combinatorial Library of Glucose Inhibitors of Glycogen Phosphorylase b", *J. Chem. Inf. Comput. Sci*., **1999**, *39*, 1151 -1160.
71. Good, A. C.; Krystek, S. R.; Mason, J. S. "High-througput and virtual screening: core lead discovery technologies move towards integration", *Drug Discov. Today*, **2001**, *5*, (suppl.).
72. Lipinski, C. A. "Poor aqueous solubility – an industry wide problem in ADME screening", **2002**Spotfire Users Europe Conference.http://www.spotfire.com/images/pdf/presentations2002/Chris_Lipinski_Lead_Identification_Europe.pdf
73. Crivori, P.; Cruciani, G.; Carrupt, P. -A.; Testa, B. "Predicting Blood-Brain Barrier Permeation from Three-Dimensional Molecular Structure", *J. Med. Chem.*, **2000**, *43*, 11, 2204 -2216.
74. Herve, F.; Urien, S.; Albengres, E.; Duche, J.-C.; Tillement, J. "Drug Binding in Plasma. A Summary of Recent Trends in the Study of Drug and Hormone Binding", *Clin. Pharmacokinet*., **1994**, *26*, 44-58
75. Colmenarejo, G.; Alvarez-Pedraglio, A.; Lavandera, J. –L. "Cheminformatic Models To Predict Binding Affinities to Human Serum Albumin", *J. Med. Chem.*, **2001**, *44*, 4370 -4378.
76. Carter, D. C.; He, X.-M. "Structure of Human Serum Albumin", *Science*, **1990**, *249*, 302-303.
77. Roberts, S. A. "High-throughput screening approaches for investigating drug metabolism and pharmacokinetics", *Xenobiotica*, **2001**, *31*, 557-589.
78. Watt, A. P.; Morrison, D.; Evans, D. C. "Approaches to higher-throughput pharmacokinetics (HTPK) in drug discovery", *Drug Discov. Today*, **2001**, *5*, 17-24.
79. Ekins, S.; Bravi, G.; Blinkley, S.; Gillespie, J. S.; Ring, B. J.; Wikel, J. H.; Wrighton, S. A. "Three- and four-dimensional quantitative structure activity relationship analyses of cytochrome P-450 3A4 inhibitors", *J. Pharm. Exp. Ther*., **1999**, *290*, 429-438.

80. Ekins, S.; Bravi, G.; Blinkley, S.; Gillespie, J. S.; Ring, B. J.; Wikel, J. H.; Wrighton, S. A. "Three and four dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2D6 inhibitors", *Pharmacogenetics*, **1999**, *9*, 477-489.
81. De Groot, M. J.; Vermeulen, N. P. "Modeling the active sites of cytochrome P450s and glutathione S-transferases, two of the most important biotransformation enzymes", *Drug Metab. Rev.*, **1997**, *29*, 747-799.
82. Keseru, G. M. A. "Virtual high throughput screen for high affinity cytochrome P450cam substrates. Implication for *in silico* prediction of drug metabolism", *J. Comput.-Aided Mol. Des.*, **2001**, *15*, 649-657.
83. Darvas, F.; Marokhazi, S.; Kormos, P.; Kulkarni, P.; Kalasz, H.; Papp Á. In "*Drug Metabolism, Databases and High Throughput Testing During Drug Design and Development*"; Erhardt, P. W., Ed.; Blackwell Science: Cambridge, MA, 1999, pp. 237-270.
84. Klopman, G.; Tu, M. In "*Drug Metabolism, Databases and High Throughput Testing During Drug Design and Development*"; Erhardt, P. W., Ed.; Blackwell Science: Cambridge, MA, 1999, pp 271-276
85. Greene, N. "Computer Software for Risk Assessment", *J. Chem. Inf. Comput. Sci.*, **1997**, *37*, 148-150.
86. Richard, A. M. "Application of SAR methods to non-congeneric databases associated with carcinogenicity and mutagenicity: issues and approaches", *Mutation Res.*, **1994**, *305*, 73-97.