Journal of Biotechnology Letters, ISSN: 0976–7045 & E-ISSN: 0976–7053 , Vol. 1, Issue 1, 2010, PP-10-14

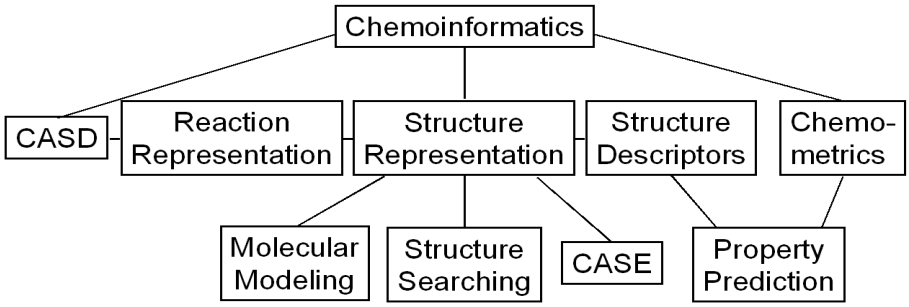
Chemoinformatics-An emerging field for Computer Aided Drug Design

Saurabh Shukla[1], Sanjay Kumar Choubey[1], Prashant Srivastava[1] and Gomase V.S.[2]
[1]Yeshwant College of Information Technology (Bioinformatics & Biotechnology), Parbhani, MS
[2]School of Technology, S.R.T.M. University, Sub-Centre, Latur, 413512, India

Abstract- In the last decade, thousands of drug molecules are newly developed. So, its very important to manage these useful chemical data with the help of Chemical Databases. Thus, there is an emergence of novel field named as "*Chemoinformatics*". One can easily solve the chemical based problems with the help of Informatics approaches. The use of information technology and organic chemistry playing an extremely crucial role in the drug discovery process. Thus, Chemoinformatics is a generic term that encompasses the design, creation, organization, management, retrieval, analysis, dissemination, visualization, use of chemistry based information and gives the promising results in the ADME properties prediction and in the Medicinal Chemistry.

Introduction

Chemoinformatics is one of the emerging and applied branch of Drug Discovery by which one can easily solve the chemical problem with the help of computational approach. Now-a-days, most of the chemical compounds and compositions are associated with the biological targets, such as proteins, GPCR's etc. Thus, every chemical characterized by a focus on the computer processing of databases of chemical structures, (either 2d and 3d); thus, this section will hence highlight specifically the development of techniques for searching chemical structures and their properties. Recent years have seen a tremendous increase in the technologies available for the discovery of new drugs.[1] Functional genomics research has led to the identification of an unprecedented number of potential therapeutic protein targets; *combinatorial libraries* has expanded the size of compound collections; and high-throughput screening (HTS) has enabled the screening of million-compound libraries. Thus, day by day, massive amounts of physical and chemical property data are generated each year for new and existing chemical substances. Thus, the management of this useful information is very important for any researcher. Fortunately, those trained in chemical informatics can provide tools to acquire, organize, and evaluate data tools that yield new insights for further chemical research.[2] Chemical informatics companies combine molecular simulation and data analysis techniques with high quality graphical visualization to obtain stunning results. Chemical informatics thus helps chemists investigate new problems and organize and analyze scientific data to develop novel compounds, materials, and processes through the application of information technology. The following is the current schema of entire Chemoinformatics [3].



Combinatorial Libraries in Chemoinformatics……?

Finding of novel drug is a complex process. Historically, the main source of biologically active compounds used in drug discovery programs has been natural products, isolated from plant, animal or fermentation sources.[4] Hence, *Combinatorial library* is one of the important new methodology developed by researchers in the pharmaceutical industry to reduce the time and costs associated with producing effective and competitive new drugs. By accelerating the process of chemical synthesis, this method is having a profound effect on all branches of chemistry, but especially on drug discovery [5]. Through the rapidly evolving technology of combi-chemistry, it is now possible to produce compound libraries to screen for novel bioactivities. This powerful new technology has begun to help pharmaceutical companies to find new drug candidates quickly, save significant money in preclinical development costs and ultimately change their fundamental approach to drug discovery[6].

Chemical Data→Storage in Databases→Data Information→Data Retrieval→Analysis

The current schema of Chemoinformatics in drug designing is given below:

Analysis Of Predesigned drug Structure→Structural Property Prediction (QSAR)→Property Prediction by Smiles format→ Perform some modification in prior drug→Again predict the drug property→If variation occurs in novel structure→ Save that structure and design a Fragment library.

WHAT ARE FRAGMENTS LIBRARIES…..?

The fragment libraries are collection of a large number of organic compounds (including aromatic, bicyclic, amines, carboxylic acid, nucleic acid, amino acid), various drugs, inorganic compound, organometallic complexes, transition metal complexes and commonly used ions and functional groups. The genomics revolution has provided a deluge of new targets for drug discovery.To fascilitate the drug discovery process, many researchers are turning to fragments based approaches to find lead molecules more efficiently.[6] One such method is, allows for the identification of small molecule fragments that binds to specific regions of a protein target.These fragments can then be elaborated combined with other molecules or combined with one another to provide high affinity drug leads. Fragments offer the prospect of a more efficient approach to drug discovery resulting in the generation of high-quality leads with a better chance of success in clinical development.

PROPERTIES OF FRAGMENTS LIBRARY :

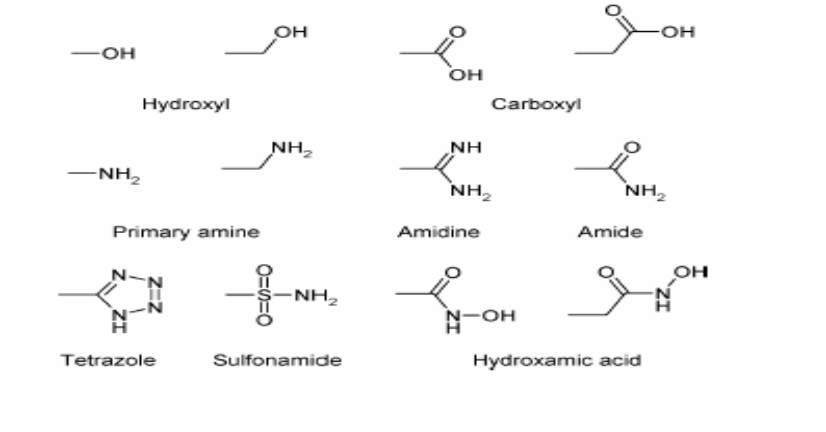
1- Low molecular weight (<300).
2- Bind readily to drug targets but show low potency (.05---- 1.0 mM)
3- Occupy a small region of chemical space.
4- Generally more soluble, which is ideal for structural studies.
5- Rapid optimization of potency, can be achieved by X-ray crystallography.

HOW DO WE USE FRAGMENT'S LIBRARY.

Two fragments sets were used in this work, fragments were small ,used as probe, and small no of interactions within the protein.[12]Generally, fragments have molecular weight between 100 to 250 and are relatively simple with few functional groups, suitable for rapid optimization .Compounds that were unlikely to be soluble in experimental conditions used for crystallographic screening were not included in the fragments sets.

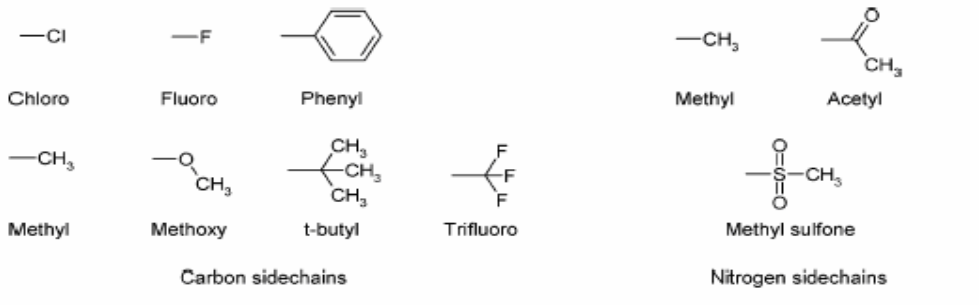
Applications of Fragment's library:

Regarding the fragment's library design, one can design various drug-lead compounds with the help of Functional groups modification at any side chain positions. Thus, some of the preferred side chains are given below



Preferred side chains involved in lead designing.

LIPOPHILIC/SECONDARY SIDE CHAINS:- The properties of the fragments are further modulated by allowing substitution with a set of secondary side chains. Most of these are lipophilic and are intended to pick up hydrophobic interactions in a protein binding site.[11] The secondary side chains for carbon atoms are shown in given Figure-



Secondary side chains

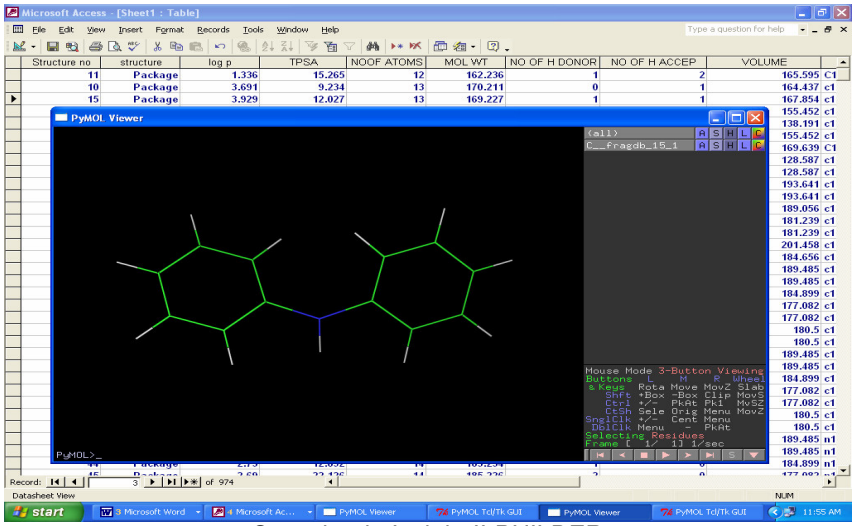
LEAD MOLECULE BUILDING WITH INSIGHTII

InsightII is a sophisticated molecular modeling environment that provides a powerful graphical interface to best of breed algorithms for molecular dynamics, homology modeling, de novo design, and electrostatics—making it the perfect solution for protein modelers, computational chemists, and structural biologists.[9] Insight II is a 3D graphical environment for molecular modeling. Its powerful user interface enables the seamless flow of data between a wide range of scientific applications. The insight II environment integrates builder molecules, development tools, force fields, simulation and visualization tools with tools specifically developed for applications in the life and material sciences. Insight II runs on Silicon Graphics workstations and servers. Several researchers receive application and system administration under a research license in parallel with the administration of this teaching license through the research computing group.[10] Insight II runs on the public SGI workstations located in several buildings on campus.

Methodology of Fragment's library creation : In the terms of methodology, first we designed the molecules in the Insight II software, with the help of builder. The path to design the molecules is given below- The Drug fragment sets were built using the BUILDER module of InsightII. Molecular Builder ----→ Draw structure-----→ Molecule----→ Put----→PDB---→Execute--→ Sybyl--→ Execute---→ Optimize---→ Save---→.mol2 ---→ .pdb

To see the result ----→ Get ----→PDB

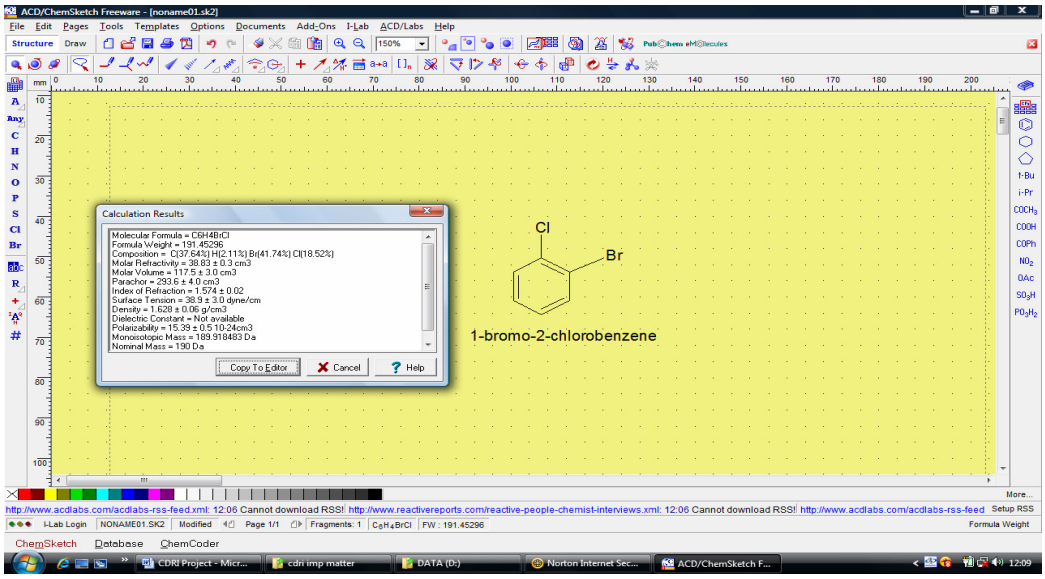
---→ Sybyl



Snapshot in Insight II-BUILDER

Some Useful outputs:

One of the important tool is also applicable for the novel lead designing-*"Chemsketch"*



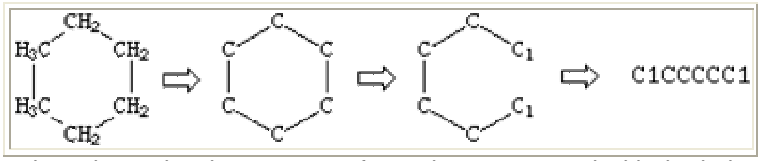
With the help of organic Builder which is embedded in this software, user can design any new drug molecule, and calculates the properties of that drug.[7]

- User can also predict the smiles format of any newly designed drug.

Smiles format……?

SMILES is simply defined as a comprehensive chemical nomenclature, and the meaning is (**S**implified **M**olecular **I**nput **L**ine **E**ntry **S**pecification). SMILES is widely used as a general-purpose chemical nomenclature and data exchange format. However, SMILES differs in several fundamental ways from most chemical nomenclatures and other chemical formats. ***In the present time a lot of databases are available which focuses the drug molecular property with the help of Smiles format.*** Cyclic structures are represented by breaking one bond in each ring. The bonds are numbered in any order, designating ring opening (or ring closure) bonds by a digit immediately following the atomic symbol at each ring closure.[10] This leaves a connected non-cyclic graph which is written as a non-cyclic structure using the three rules described above.

Cyclohexane and its smiles format:

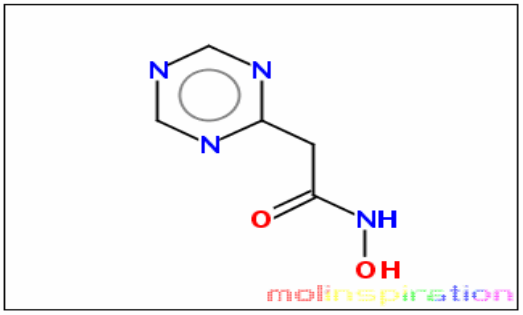


Thus, one can analyze the molecular property of any drug compound with the help of *Molinspiration database*, [8]because there is a need of input of only smiles format.
Molinspiration →calculate chemical properties → then one can clearly gain the overall properties like in the following manner,

- Logp
- Molecular wt
- Atomic volume
- Tpsa (Total polar surface area)
- No of H acceptor
- No of H donor
- No of atoms

molinspiration

SMILES n1cncnc1CC(=O)NO
miSMILES ONC(=O)Cc1ncncn1



miLogP	-1.624	NEW version2.2
TPSA	88.002	
natoms	11	
MW	154.129	
nON	6	
nOHNH	2	
nviolations	0	
nrotb	2	
volume	127.78	

Get data as text (for copy / paste).

This was request 4 out of 3000 available this month for your site 203.190.146.194.
With technology from Molinspiration you can easily setup similar service also directly on your intranet.
Comments or questions ? See FAQ and do not hesitate to contact us by email !

The above snapshot showing the observed drug properties on the basis of smiles format

Conclusion

The above novel aspects showing some promising features in Chemoinformatics. Thus, one can predict the Structure Activity Relationship (SAR property), with the help of Reaction databases and some online servers. In the High Throughput Screening, one of the important approach is Lead designing on the basis of prior drugs, so after any side chain modification there is a variation occurs in the molecular property of any novel lead compound, hence with the Informatics approaches one can easily pointsout the property based prediction and gives the promising results in the the ADME (Administration, Distribution, Metabolism & Excretion) property prediction and Computer Assisted Drug Design (CADD). Hence, Chemoinformatics is one of the current stage of development in the de-novo design of a novel drug.

References

[1] J. Gasteiger, Editor, (2003).*Handbook of Chemoinformatics – From Data to Knowledge*, Wiley-VCH, Weinheim
[2] R.V. Williams, M.E. Bowden, Chronology of Chemical Information Science
[3] R.T. Bottle, J.F.B. Rowland, Information Sources in Chemistry, 4th edition.
[4] R.E. Maizell, How to find Chemical Information, 3rd edition, Wiley, New York.
[5] Oprea, T. I.; Davis, A. M.; Teague, 2001 S. J.; Leeson, P. D. A historical perspective. *J. Chem. Inf. Comput. Sci.*, *41*, 1308-1315.
[6] Teague, S. J.; Davis, A. M.; Leeson, 1999 P. D.; Oprea, T. The Design of Leadlike Combinatorial Libraries. *Angew. Chem., Int. Ed.*, *38*, 3743-3748.
[7] Fejzo, J.; Lepre, C. A.; Peng, J. W.; Bemis, 1999 G. W.; Ajay; Murcko, M. A.; Moore, J. M. The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem. Biol.*, *6*, 755-769.
[8] Borman, S. (1990) New QSAR . *Chem. Eng. News*, **68:** 20-23.
[9] Chemoinformatics – A Textbook, *J. Gasteiger,* (2003). *T. Engel, Editors*, Wiley-VCH, Weinheim
[10] Campbell, S. F. Science, . 2000, *99*. art and drug discovery: a personal perpsective. *Clin. Sci.*
[11] K. Funatsu, S. Sasaki, (1999)*Tetrahedron: Comput. Methodol.*, **1988**, 127-137. K Satoh, K. Funatsu, *J. Chem. Inf. Comput. Sci.*, **39**, 316-325.
[12] International Chemometrics Society: http://www.mamics.nysaes.cornell.edu/ chem-society.html