

REVIEW ARTICLE

Cheminformatics: A New Approach

Neeraj Upmanyu^{1*}, Surya Gupta², B. N. Prakash², Gopal Garg³ and P. Mishra⁴

¹Shri Ram College of Pharmacy, A. B. Road, Banmore, (MP) India.

²Rajiv Gandhi Institute of Pharmacy, Satna (MP) 485001

³Institute of Pharmacy, Pt. Ravishankar Shukla University. Raipur (C.G.)

⁴G.L.A Institute of Pharmacy, Mathura (U.P.)

* Corresponding Author E-mail: neerajupmanyu@rediffmail.com

ABSTRACT

Cheminformatics is a spectacular advancement in drug discovery and development. Cheminformatics is the combination of chemical synthesis, biological screening and data mining approaches used to guide drug discovery and development. Cheminformatics applies information technology to chemical data and includes topics such as chemical databases, combinatorial library design, computational chemistry, structure activity relationship, data sequencing, visualization and structure-based drug design. Thus cheminformatics is a boon to medicinal chemistry specially drug design.

KEY WORDS Cheminformatics, Combinatorial Chemistry, Molecular Modeling, Ligands, Drug Design.

INTRODUCTION:

Cheminformatics is the use of computer software to assist in the acquisition, analysis and management of data and information relating to chemical compounds and their properties. The cheminformatics groups offer programs and databases (mainly for organic and sometimes for inorganic application) related to small molecules, complementing the activities of the bioinformatics groups, who concentrate on biological macromolecules. In case of chemistry, where information system have long played a vital role in pharmaceutical and biotechnology research and where there is need not only to handle information in traditional, textual and numeric data but also databases containing information regarding the two dimensional (2D) or three dimensional (3D) structures of molecules

Recent developments, such as combinatorial chemistry, high-throughput screening and laboratory information management system, are bringing about drastic increases in the volumes of biological, numerical, structural and textual data that is being preceded in research programs. This has produced the discipline of cheminformatics, which involves the creation, retrieval, organization, dissemination and processing of chemical information.

Chemical information system:

The primary purpose for chemical information system is to be able to identify a chemical substance, find compounds similar to target compound and determine the

location of the compound. The hub of the chemical information system is the inventory system.

Emergence of cheminformatics:

The year 1961 first saw the journal of chemical documentation. It is the journal, later on named as Journal of Chemical Information and Computer Science.

The first book on computer handling information relating to chemical structure was published in 1971. The first international conference on this subject was held in 1973 at Noordwijkerhout and after 1987 regularly every three years.¹⁻⁴

Cheminformatics as a whole covers following informations, i.e., chemical information sources, medicinal chemistry, chemical database design and their management, data sequencing, mining and visualization, computational chemistry and modern combinatorial chemistry.

MEDICINAL CHEMISTRY:

In ancient times most of the drugs were obtained from the plants. It is the nineteenth century that chemistry developed as a science, both in terms of experimental procedures and scientific theory. Methods of organic synthesis were developed that helped chemistry to alter structures in a predictable way. Medicines before nineteenth century were derived from natural sources. Chemical reactions on isolated active principles resulted in synthesis of series of analogs, some of which had useful pharmacological activities and could be used in medicines. Such analogs are defined as semi synthetic drugs e.g. diacetylmorphine i.e. Heroin is a semi synthetic drug synthesized from morphine.

Received on 05.03.2008

Modified on 10.03.2008

Accepted on 25.03.2008

© RJPT All right reserved

Research J. Pharm. and Tech. 1(1): Jan.-Mar. 2008; Page 2-5

In Medicinal Chemistry we go through several aspects like prodrugs and soft drugs concept. Prodrug technology is applied to all the drug administration routes (e.g. oral, parental, dermal etc.) and dosage forms (e.g. tablet, solution etc.) and designed to minimize the difficulties faced during pharmaceutical and pharmacokinetic phases.

Another aspect of medicinal chemistry is the study of drug targets i.e. receptors (which are specific areas of certain proteins and glycoproteins that are found either embedded in cellular membranes or in the nuclei of living cells). Any endogenous or exogenous chemical agent that binds to receptor is known as ligands. Drugs that bind to receptor and give a similar response are known as agonists where as drugs that bind to receptor but do not cause a response are termed as antagonists.

Molecular modeling is another aspect of medicinal chemistry. Three main methods are: -

1. Building using standard geometrics – especially bond lengths and bond angles.
2. Building by using fragments which are corrected by optimization.
3. Building using data obtained from physical experiment – usually X-ray crystallography or structure deduced from NMR data.

Molecular modeling also includes stereochemistry in which different aspects like three dimensional arrangement of atoms, isomers, enantiomers, diastereoisomers and epimers (special type of diastereoisomers) and geometric isomers are involved.^{5,6}

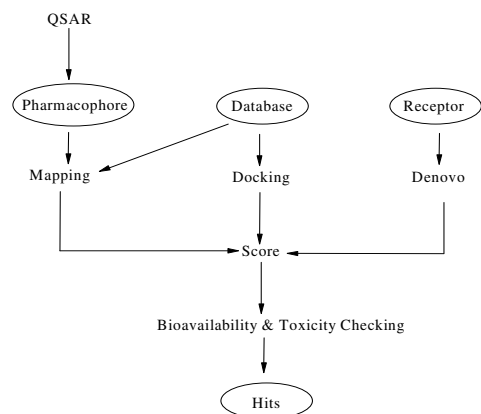


Fig. 1 : Scheme of drug design with help of computer tools

COMBINATORIAL CHEMISTRY:

Combinatorial chemistry is a collection of different methodologies for the synthesis of multiple compounds. A combinatorial chemistry program is designed to support lead discovery with several well-defined characteristics. Combinatorial chemists develop protocols for transferring data from instrument to instrument through a seamless data flow. The development of new processes for the generation of collections of structurally related compounds (libraries) with the introduction of combinatorial approaches has

revitalized random screening as a paradigm for drug discovery and has raised enormous excitement about the possibility of finding new and valuable drugs in short times and at reasonable costs.

Combinatorial chemistry is carried out either on a solid support or in solution to produce library consisting of either individual compounds or mixtures.

Another important study is high throughput screening (HTS) which is the process, by which a large number of compounds can be tested, in an automated fashion for activity as inhibitors (antagonists) of a certain biological function (Fig.2). The technique originated for screening of microbial natural products for the discovery of new antibiotics. In other words HTS is a multidisciplinary process directed towards the identification of leads. HTS helps in selection of hits amidst large collections of compounds. In this the samples and assays are the inputs and hits and lead the outputs.

Combinatorial libraries:

Different sort of libraries like peptide library are made. The number of compounds in a library may vary easily. The major difference between a combinatorial library and collection of compounds is that all components of a combinatorial library share something common (design, synthetic history etc.).

The analysis of these combinatorial libraries involves:

1. Structure identification by NMR (¹H and ¹³C) and MS (including high resolution mass spectrometry) with additional confirmation of structure provided by IR spectroscopy, X-ray etc.
2. Estimation of purity performed by elemental analysis, TLC, HPLC with UV/VIS detection.
3. Determination of amount to synthesized compound, usually by weighing.

The ability to identify new potential ligands against a variety of targets of interests has been drastically improved through the use of random peptide libraries. The great power of biological libraries lies in the possibility to make use of “biopanning” which represents an in-vitro selection process that has been employed to isolate target binding sequences from a large pool of different molecules.

So combinatorial technology is a platform technology that integrates with other functional genomics and proteomics are used to focus new lead generation.⁷⁻¹³

Chemical database design and their management

A database is simply an organized collected of related data, typically stored on disk and accessible by many concurrent users. Databases are generally separated into application areas. Databases are managed by a DBMS (Database management system).

Data and its storage may be considered to be the heart of any information system. For data to be of value it must be presented in a form that supports the various operational, financial, managerial, decision making, administrative and clerical activities within an organization. Different

types of databases are used now-a-days like biocatalysis database, the MOS (methods in organic synthesis) database, failed reaction database, protecting group database, solid phase synthesis database etc.

Chemical Information Sources^{14,15}

Chemical information sources may be different like literature, journals, internet etc. possibilities for chemists are numerous, profound and barely perceived. Currently, over 200 departments of chemistry around the world have some of the www site and are listed at <http://www.Shef.ac.uk/uni/academic/A-C/chem/chemistry-wwww-sites.html>.

COMPUTATIONAL CHEMISTRY:

Computational chemistry is the one which includes fundamentals of chemistry, quantum mechanics or mathematics and chemical computations. The term computational chemistry implies as a well developed mathematical method which can be automated for implementation on a computer. Applications for computational chemistry in chemical processing involves new bioprocesses, catalyst design, improved reaction mechanism, efficient process design, product development (polymers, pharmaceuticals), environmental modeling etc. computational technologies are embodied in every aspect of chemical research and development, design and manufacture. The field of computational chemistry, broadly defined includes computational molecular science, empirical correlations such as linear free energy relationships and quantitative structure property relationships (QSPR) and aspects of process modeling and simulation. Various tools have also been used for a diversity of practical applications involving adhesives, coating, polymers and surfactants for the prediction of toxicity of chemicals.

Current applications of computational chemistry involve atmospheric chemistry, drug design, catalyst/biocatalyst design, material design, adhesive/coating design, lubricant chemistry etc. In pharmaceutical industry computational methods have played vital role in structure based drug design, most recently in development of current generation of HIV protease inhibition.

Chemical Data Sequencing Mining and Visualization:

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies to focus on most important information in their data warehouses. So data mining is about the discovery of knowledge, rules, or patterns in large quantities of data. An alternative term is "Knowledge Discovery in Databases" (KDD). There are several data mining techniques like cluster analysis, decision trees, rule induction, neural networks, on-line analytical processing etc.

Data visualization makes it possible for the analyst to gain a deeper, more intuitive understanding of the data. Data-mining allows the analyst to focus on certain patterns and trends and explore in depth using visualization. Different soft wares for data-base management are used like ORACLE, IBM,

INFORMIX. Other soft wares used for chemical data mixing like AUTOCORR, C@ROL, SURFACE, SONNIA, CLIFF, GAMA (for 3-d features), PAOZA, PETRA, ROTATE, STERGEN, WODCA, AUTONOM, CHEMKEY, CHEMPROTECT, CHIRON etc.

So competition requires timely and sophisticated analysis on an integrated view of the data.

DRUG DESIGN AND DISCOVERY:

The drugs used in medicine are developed from so-called **Lead Compound**. The approach to drug design depends on objectives of design team. The general steps for designing drug are depicted in Fig.1

Development of a drug is performed in five classical steps.

1. Search for active site (biological development phase)
2. Search and optimization of active compound (chemical research phase)
3. Testing of active compound (pre-clinical phase)
4. Clinical trails
5. Approval process

"Drug-design" is the determination of molecular structure of target, which makes sense of the word "rational".

Different chemical parameters involved in drug design are:

- i. stereochemistry of enzymes
- ii. penetrating of membranes
- iii. shapes of molecules
- iv. rigidity and flexibility of molecules

In drug designing several terms like, QSAR, Bio-isosterism etc. are involved.

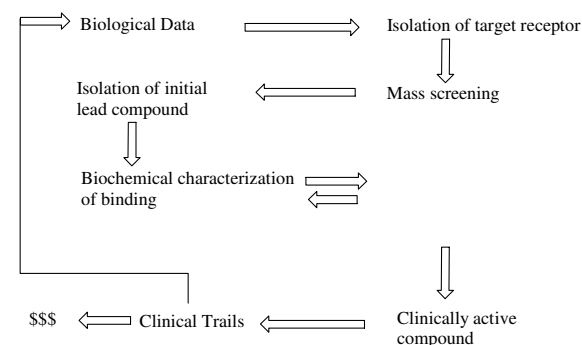


Fig . 2 : Drug discovery pipeline.

QSAR:

Quantitative structure activity relationship derives model which describes the structural dependence of biological activities either by physico-chemical parameter (Hansch analysis) or by three dimensional (3-D) molecular property profiles of the compound [Comparative molecular field analysis (COMFA)].¹⁶⁻¹⁸

Bio-isosterism

Bioisosteres are groups or molecules that are structurally similar and show same type of biological activity and phenomena. These involve true and partial bioisosteres. A true bioisosteres has same type of bioactivity as an analog and has same magnitude of biological response, whereas partial bioisosteres has same type of bioactivity but different magnitude of response.

Different physicochemical factors in drug design are also involved like ionization constants, chelation, partition coefficient, steric factors, electronic factors, etc.

Hansch analysis in drug design is extra thermodynamic approach which correlates biological activity values with physicochemical properties. By linear, non-linear regression analysis.^{18,19}

Drug discovery pipeline: The development of any potential drug begins with years of scientific study to determine the biochemistry behind a medical problem for which pharmaceutical intervention is possible. The result is the determination of specific receptor targets that must be modulated to alter their activity in some way. The modern day drug discovery pipeline is shown in Fig. 2.

So the advancement in the field cheminformatics are playing an essential role in generating, selecting and optimizing qualified drug leading from increasing amount of potential drug targets.

Relevance of Cheminformatics in India:

In Indian cheminformatics is in its infancy with its strength in chemical information source, computing, chemical data-base design and drug design, the country is positioned ideally to emerge as a front runner in cheminformatics.

The Indian government and industry should take advantage of huge potential involve in it. With the initiative of DBT (Department of Biotechnology) Govt. of India, R & D Organizations and Pharma and Software Company, Cheminformatics is making inroad in Indian Software, Pharmaceutical Companies and Other industries.

Several companies like MDZ, Tripos, Pharmacopoeia software company, Chem. Navigator, Anaays Pharmaceuticals (Young Drug Discovery Company), these all are engaged in making use of cheminformatics.

If the industry and the government work together, it can play a vital role in the 50-100 billion market of the pharmaceutical sector.

FUTURE PROSPECTS:

Cheminformatics is the latest emerging branch not only in India but also in European countries like USA, UK, Japan etc. with huge grants spent and thousands of jobs available. The future prospects of cheminformatics professionals are growing day by day. The need of trained manpower in agriculture, pharmaceutical and

biotechnological industry, molecular modeling, computational chemistry and library design and data base and publishing organizations is growing.

So the boundary between cheminformatics and bioinformatics is disappearing and bioinformatics blurs into cheminformatics with the process of target identification, target validation and assay development. The integration of cheminformatics with related disciplines such as genomics, proteomics and application of these sciences are must sought careers across globe and will come with flying colours in near future.²³⁻²⁵

REFERENCES:

1. Nourse, J.G. et al., **J Chem Inf Comput Sci**, 1988, 32-34.
2. Weininger, D., **J Chem Inf Comput Sci**, 28, 1988, 31-37.
3. Weininger, D., **J Chem Inf Comput Sci**, 29, 1989, 97-98.
4. Viswanandhan, V.N., Ghose, A.K., Revankar. G.R., and Robins, R.K., **J. Chem. Inf. Comput. Sci.**, 1989, 29,163-169.
5. Thomas, G. In; Medicinal Chemistry An Introduction 1st Edn., John Wiley and Sons. Ltd New York, 200,49.
6. Foye, W.O., Zemke, Z., Thomas and Willams, A. David, Principles of Medicinal Chemistry, 4th edition, B.I. Waverly Pvt. Ltd., New Delhi, 1996, 1-4.
7. Kundu, B., Kare, S.K, Rastogi, S.K., **Prog Drug Res**, 53, 1999, 89-90
8. Terett, N., Combinatorial Chemistry, 1st edition, 1998, Oxford University Press, New York, 20.
9. Patchett, A.A., **J Med Chem**, 36, 1993, 2051-2055.
10. Gallop, M.A., Barrett, R.W., Dower, W.J., Fodor, S.D.A., Govdon, Z.H., **J Med Chem**, 37, 1994, 1233-1234.
11. Merrifield, R.B., **J Amer Chem Soc**, 83, 1963, 2149-2143.
12. Wilson S. R., Czarnik, A. W., Combinatorial Chemistry Synthesis and Application, First edition, John Wiley and Sons, Singapore, 1997, 130.
13. Lam, K. S., Lebi M., Krchnak, V., **Chem Rev**, 1997, 97, 911.
14. Smith, E. G, Baker, P. A., The Wirwesser Line Formycle Chemical Notation, (WIN), Chemical Information Management Inc., New Jersey, 1973.
15. Daylight Theory Manual, Chapt. 7-10, "Daylight Chemical Information System Inc.", 18500, Von Kothas Ave # 450, Irvine CA, USA.
16. Thomas, G In; Medicinal Chemistry An Introduction 1st Edn., John Wiley and Sons, New York, 200,32.
17. Kubiniji, H., "QSAR: Hansch analysis and Related approaches", VCH, Weinheins, New York, 1993, 58.
18. Kubinyi, H., **J. Med. Chem.** 1976, 19, 587-600.
19. Crammer III, R & D, "Pettersen D. E. and Bunce, J. D. "**J. Am. Chem. Soc.**", 1988, 110, 5959-5967.
20. Hansch C, Lc. A, Unger S. H., Kin, K. H., Nikaitami, D. and Lein E. Z., "**J. Med. Chem**", 1973, 16, 1207-1216.
21. Singh, R., "Synthetic Drugs" First edition, Mittal Publications, Delhi, 2002.
22. Ariens, E. J., "Drug design", First edition, Academic Press, New York, 42, 1571.
23. Delgado, J. N., and Remers, W.A., In Text Book of Organic, Medicinal and Pharmaceutical Chemistry, 9th edition, J. B. Lippincott Company, Philadelphia-1, 1998, 17
24. Gordon, E. and Vorapagel, E.R., **J. Med. Chem.**, 1994, 39, 1385-1386.
25. Gallop, m.a., Barrett, R.W., Dower, W.J., Fodor S.P.A and Gordon, E.M., **J. Med. Chem.**, 1994, 37, 1233-1236

