Journal of Proteomics & Bioinformatics - Open Access

www.omicsonline.com

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jpb.1000147

JPB/Vol.3 Issue 8

Cheminformatics

Nutan Prakash* and Dinta A. Gareja

Department of Biotechnology, Shree M . & N. Virani Science College, India

Abstract

Cheminformatics is the application of computational methods to chemical problems, with particular emphasis on the manipulation of structural information. Cheminformatics is a relatively new field of information technology that focuses on the collection, storage, analysis, and manipulation of chemical data. The chemical data of interest typically includes information on small molecule formulas, structures, properties, spectra, and activities (biological or industrial). Cheminformatics originally emerged as a vehicle to help the drug discovery and development process; however Cheminformatics now plays an increasingly important role in many areas of biology, chemistry, and biochemistry. Cheminformatics can also be applied to data analysis for various industries like paper and pulp, dyes and such applied industries.

Introduction

Cheminformatics is the use of computer and informational techniques, applied to a range of problems in the field of chemistry. Also known as Cheminformatics and chemical informatics, these techniques are used in pharmaceutical companies in the process of Drug Discovery. Cheminformatics combines the scientific working fields of Chemistry and computer science especially in the area of chemical Graph theory and mining the chemical space. It is to be expected that the chemical space contains at least 10^{62} molecules. Cheminformatics is a generic term that encompasses the design, creation, organization, management, retrieval, analysis, dissemination, visualization, and use of chemical information. The transformation of data into information and of information into knowledge is an endeavor needed in any branch of chemistry not only in drug design. Chemistry has produced an enormous amount of data and this data avalanche is rapidly increasing. More than 45 million chemical compounds are known and this number is increasing by several millions each year. Novel techniques such as combinatorial chemistry and high-throughput screening generate huge amounts of data. All this data and information can only be managed and made accessible by storing them in proper databases. That is only possible through Cheminformatics.

Definition

Cheminformatics is the application of informatics methods to solve chemical problems.

Role of Natural Product Chemistry in Cheminformatics

A natural product is an important sector in the area of drug discovery and development. Most encouraging is the continuing emergence of new natural product chemo types with interesting structures and biological activities and potential for sub-library generation of targeted screening. Increasingly available as pure compounds, natural products are highly amenable to the much broader screening opportunities presented by the new targets. Regardless of chemical library input, natural products are uniquely well placed to provide structural information from which virtual compounds can be created by computational chemistry and applied technologies. The structural versatility of natural products is expected to play a major role in modern drug discovery programs (Moore and Nisbet, 1997).

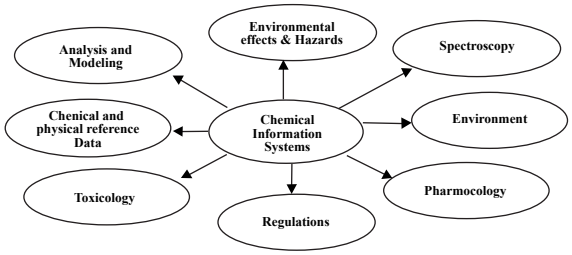

Figure 1:

Cheminformatics and Drug Discovery

Introduction

Competition and cost has changed the drug design paradigm from the hit and trial approach to the drug design approach allowing the tailor-made design of active molecules. This has resulted in both targeted drug discovery and reduced drug development cycle time. The need for introducing newer molecules that are superior using an

Figure 2:

***Corresponding author:** Nutan Prakash, Department of Biotechnology, Shree M . & N. Virani Science College, India, E-mail: nutanp@vsc.edu.in

Received July 31, 2010; **Accepted** August 31, 2010; **Published** August 31, 2010

Citation: Prakash N, Gareja DA (2010) Cheminformatics. J Proteomics Bioinform 3: 249-252. doi:10.4172/jpb.1000147

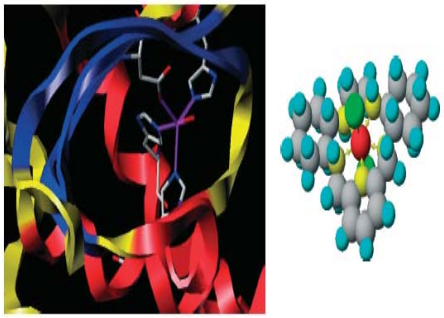

Figure 3:

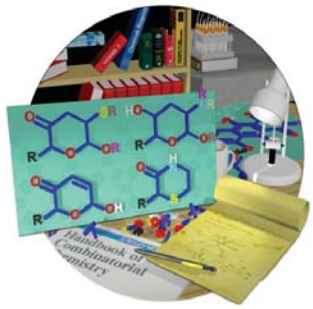

Figure 4:

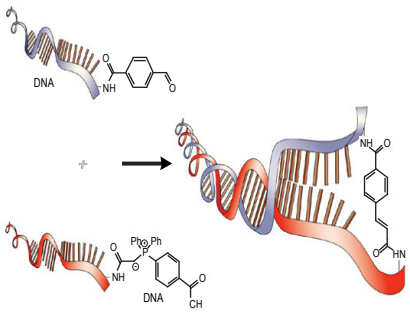

Figure 5:

automated approach will make drug discovery a highly knowledge specific and efficient process.

Traditional drug discovery process

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 (Augen, 2002), and if one of the steps is slow, it slows down the entire process. These slow steps are bottlenecks.

The old bottlenecks and HTS Technologies

Previously, the main bottlenecks in drug discovery were the time and costs of making (or finding) and testing new chemical entities (NCE). The average cost of creating a NCE in a major pharmaceutical company was estimated at around $7,500/compound (Augen, 2002). In order 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 (Barrett et al., 1994; Hecht, 2002). Since biologists can now test thousands of compounds per day, chemists are required to make enough compounds to meet the needs of biologists. But, can chemists make thousands of compounds a day?

Combinatorial chemistry

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 (Hall et al., 2001). Parallel syntheses techniques are nowadays used in all major pharmaceutical companies. By increasi ng the capabilities of making and testing compounds, it was hoped that the drug discovery process could be accelerated dramatically. Unfortunately, this did not turn out to be the case. Seeking the reasons for these disappointing results, it was believed that increasing the chemical diversity of compound libraries would enhance the drug discovery process. Cheminformatics approaches would now be introduced in order to optimize the chemical diversity of libraries.

Chemical diversity and cheminformatics

It was soon realized that millions of compounds could be made by CC technologies. However, this procedure did not yield many drug candidates. In order to avoid wasting CC efforts, it was believed that it would be best to make chemically diverse compound libraries. In order to make a compound library with great chemical diversity, a variety of structural processing technologies for diversity analyses were created and applied. These computational approaches are the components of Cheminformatics. After 1990, many chemical-diversity-related approaches were developed, such as structural descriptor computations, structural similarity algorithms, classification algorithms, diversified compound selections, and library enumerations. However, help from these diversity analyses approaches has been limited. More hits have been found from these chemically diverse libraries, but most of these hits do not result in new drugs. Therefore, the process of making and screening drug-like compounds came under question.

Applications

The use of information technology and management has become a critical part of the drug discovery process as well as to

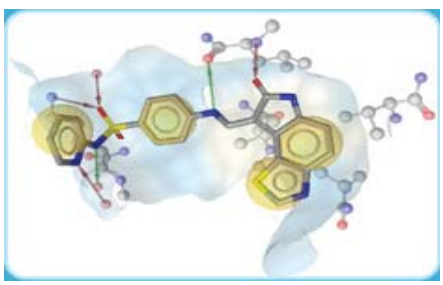

Figure 6:

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jpb.1000147
JPB/Vol.3 Issue 8

solve the chemical problems. So, Cheminformatics 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 organization. Cheminformatics is the use of computer and informational techniques, applied to a range of problems in the field of chemistry. Also known as Cheminformatics and chemical informatics, these techniques are used in pharmaceutical companies in the process of Drug Discovery.

Current Status

Recent advances in virtual screening track computational capability and as the processing power of computers improves, so do screening speed and complexity. Parameters such a structure, function or chemical space allow for a nearly limitless array of screening options. The use of screening data for development decision making is predicated from the management and interpretation of the data. Extraction of information from the data is the vital link between theoretical design and the drug candidate. Finally, it is the integration of iterative results from computation to activity that drives the cycle forward. Library chemistry and high-throughput screening require the greater use of chemo informatics to increase their effectiveness. However to identify types of procedure which yield the best result and address factors such as cost, availability and synthetic feasibility

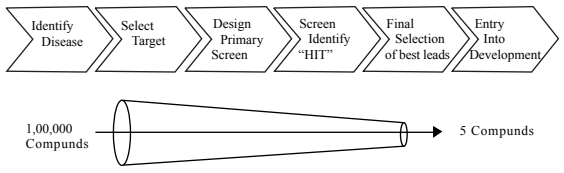

Figure 7: Drug discovery funnel.

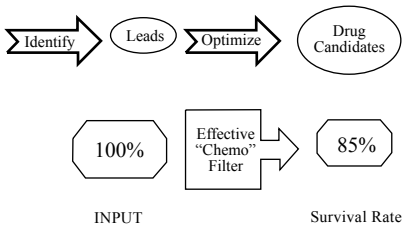

Figure 8: Need for effective chemo informatics filter.

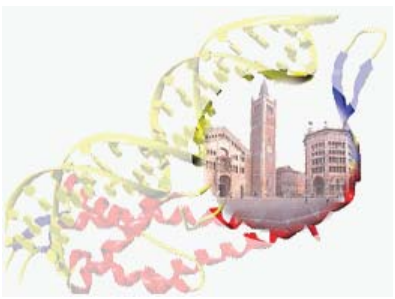

Figure 9:

rest with the user's decision. In parallel, another area that is gaining greater importance is the development of filtering procedures which identifies molecules that exhibit some sort of undesirable characteristic (toxicity, high reactivity etc.)

Without a proper knowledge base, lead optimization is a search in the vast darkness of chemistry space. It may lead to the wrong direction in the drug discovery program. Establishing a proper database with complete test results may lead to organizational success in drug discovery developments (Figure 8).

Combinatorial chemistry has opened up new strategies for a more comprehensive parallel approach to sweeping and searching during lead optimization, which has necessitated the development of suitable and new library design principles.

Recent Development in Cheminformatics

The technological developments in combinatorial synthesis and HTS have brought about an increase, by several orders of magnitude, in the volumes of data that need to be processed in drug discovery programmes. This explosion of both structural and bioactivity data has further hastened the need to integrate two areas of chemical computation that had previously developed, on the whole, separately. Chemical information techniques have been developed for the storage and retrieval of information from databases of chemical articles and chemical structures, both corporate and public. The computer processing required for such system is relatively simple in nature, although extremely impressive in terms of the data volumes involved (hundreds of thousands or millions of molecules). Conversely, molecular modeling techniques have traditionally been used for the detailed analysis of datasets that contain a few tens, or at most a few hundreds, of molecules, with the aim of using knowledge about their conformations and energies, *inter alia*, to predict their biological activities. Extending these methods for analyzing SARs to data volumes typical of those routinely handled in chemical information systems, is a data mining challenge that is now being faced by most drug discovery organizations. Third, there is no doubt that informatics is an idea whose time has come, or is coming, in an increasingly wide range of disciplines. Bioinformatics is, of course, now a widely recognized discipline, the establishment of which has been driven mainly by the data explosion resulting from the Human Genome and related sequencing projects. Medical informatics and health informatics are also well established and references are starting to appear to, for example, educational informatics and neural science informatics.

Grand Challenges for Cheminformatics

There are three "grand challenge" areas. They should be an important focus for cheminformatics.

Overcoming stalld drug discovery

After the impressive successes in drug discovery toward the end of the last century, productivity in the pharmaceutical industry has declined as expenses have gone up. Cheminformatics can help by enabling fast, cheap virtual experiments to prioritize real experiments. As more drug discovery research is carried out in academia, institutes and small companies, and solutions will require pieces from cheminformatics, bioinformatics and other disciplines, cheminformatics knowledge and tools should be made as widely available as possible.

Figure 10:

Green chemistry and global warming

Global warming and preserving the environment will be one of the biggest challenges for mankind this century. Fundamental to this will be finding chemicals which are less polluting or less toxic to the environment, or improving chemical use to minimize environmental impact (e.g. in petrochemicals). Cheminformatics already has much to offer through computational toxicology and predictive modeling.

Uuderstanding life from a chemical perspective

Chemicals are being found to be increasingly important in cellular functions, for example through small molecule modulators and epigenetic. This has led to fields such as chemical biology, and more recently systems chemistry (Ludlow and Otto, 2008) and systems chemical biology (Oprea et al., 2007), which seek to understand biological systems from a chemistry perspective. Integration of Cheminformatics and bioinformatics methods will be key to this.

References

1. Augen J (2002) The evolving role of information technology in the drug discovery process. Drug Discov Today 7: 315-323.
2. Barrett RW, Dower WJ, Fodor SPA, Gallop MA, Gordon EM (1994) Applications of Combinatorial Technologies to Drug Discovery. 1. Background and Peptide Combinatorial Libraries. J Med Chem 37: 1233-1251.
3. Brown FK (1998) Chemo informatics: what is it and how does it impact drug discovery. Annu Rep Med Chem 33: 375-384.
4. Crippen CA, Parks GM, Topliss JG (1998) The measurement of molecular diversity by receptor site interaction simulation. J Comput Aided Mol Des 12: pp441-449.
5. Chemspider (http://www.chemspider.com/).
6. Green R, Hann M (1999) Chemo informatics – a new name for an old problem. Curr Opin Chem Biol 3: pp.379-383.
7. Hagler AT, Xu J (2002) Cheminformatics and Drug Discovery. Molecules 7: 566-600.
8. Hall DG, Manku S, Wang F (2001) Solution- and Solid-Phase Strategies for the Design, Synthesis, and Screening of Libraries Based on Natural Product Templates: A Comprehensive Survey. J Comb Chem 3: 125-150.
9. Hecht P (January 2002) High-throughput screening: beating the odds with informatics-driven chemistry. Curr Drug Discov 21-24.
10. Karthikeyan M, Krishnan S (2002) Cheminformatics: A tool for modern drug discovery. International Journal of Information Technology and Management 1: 69-82.
11. Ludlow RF, Otto S (2008) Systems chemistry. Chem Soc Rev 37: 101-108.
12. Moore M, Nisbet LJ (1997) Will natural products remain an important source of drug research for the future? Curr Opin Biotechnol 8: pp708-712.
13. Oprea TI, Tropsha A, Faulon JL, Rintoul MD (2007) Systems chemical. Nat Chem Biol 3: 447-50.
14. PubChem [http://pubchem.ncbi.nlm.nih.gov/].
15. Wild DJ (2009) Grand Challenges for Cheminformatics. J Cheminform 1: 1.