Section III

Cheminformatics (in Drug Discovery)

Tudor I. Oprea

There has been a strong impetus to minimize research and development (R&D) costs and the time lapse between idea (e.g., target or lead identification) and drug, and to maximize the rate of success of preclinical R&D projects. Hence, there is a constant drive to re-evaluate the preclinical research process, and to place a stronger emphasis on science, rather than serendipity. This led to a plethora of "revolutionary" methods, each promising to reduce significantly the time from idea to marketed drug, in spite of the increased burden imposed by regulatory agencies and by competition. Methods such as computer-aided drug design, combinatorial chemistry, anti-sense technology and high-throughput screening (HTS) had their share of enthusiasts; while useful in some cases, most of the new techniques did not prove to be more than enabling technologies – useful, but short of miraculous[1].

One of the paramount efforts to revise the input/output (or the signal noise) ratio with regards to chemical aspects of drug discovery has been in the area of cheminformatics. In the strictest sense, chemical informatics integrates data via computer-assisted manipulation of chemical structures[2]. Chemical inventory and compound registration are vital to cheminformatics, but it is their combination with other theoretical tools, linked to physical (organic) chemistry, pharmacodynamics, and pharmacokinetics that bring unique capabilities in the area of lead and drug discovery. In recent years, cheminformatics has emerged as the informatics-driven technological push in preclinical research, since it successfully link all the involved scientific partners, from virtual screening to animal toxicology via one central element: *chemical structure*. The following computational chemistry protocols are cheminformatics activities associated to the drug discovery process:

- Chemical library (virtual or physical) analysis (reagent- and product-based); this includes tools for molecular similarity and diversity;
- Virtual screening (including docking and scoring); this requires 3D structures of the receptor, as well as a model of the binding site;
- Structure-activity relationships (SAR) for lead series identification; if pharmacophore models are available, they can also be used in virtual screening;
- Virtual or high-throughput screening post-processing and analysis; this includes SAR for primary HTS data;
- *De-novo* ligand design for lead generation and lead optimization, when 3D structures of the ligand-receptor complex(es) are available; if the receptor-bound conformation for high-affinity ligands is determined experimentally, the chance for successful lead optimization is increased;
- Quantitative structure-activity relationships (QSAR) for lead hopping and lead optimization, in the absence of 3D structures for the ligand-receptor complex; pharmacophoric patterns can be combined with QSAR or 3D-QSAR techniques in order to increase model quality;

- *In-silico* screening for ADME (Absorption, Distribution, Metabolism, and Excretion), plasma protein binding, metabolism prediction, and toxicology, in order to increase the quality of the compounds that reach late-phase preclinical development;
- Data mining and visualization tools with chemistry awareness: Spotfire™ with the ISIS plug-in[3], Leadscope[4], and VIDA[5] are examples of tools that enhance the information available in the decision making process by providing multiple visualizations, data-dependent filters, substructure searching, and 2D and 3D molecular structures, in order to facilitate understanding of the molecules of interest, and their neighbors, in multi-dimensional spaces, often in a receptor-relevant context.

While the goal of transforming data into information, and information into knowledge[6], is cheminformatics, we consider access to biological (and indeed chemical) data, but *not* the data themselves, as part of cheminformatics[7]. By the same token, several number-crunching activities traditionally associated with computational chemistry, e.g., physical and chemical property calculations, generate more numbers, rather than information. In other words, traditional computational chemistry may not necessarily be part of cheminformatics unless it assists decision-makers to interpret or to clarify the initial physicochemical, structural, and biological data. For example, virtual screening and QSARs transform numbers (data) into models (information, even knowledge), and should therefore be considered an integral part of cheminformatics.

Similar to the role envisioned for medicinal chemistry in the new millennium by Erhardt[8], a major role will be played by cheminformatics in drug discovery in the first decades of the 21st century, as the central instrument for chemistry-related information. This information, however, is unlikely to be transformed into knowledge by medicinal chemists alone[8], as the interpretation basis for such decisions requires in-depth knowledge in several, unrelated, fields: physical organic chemistry (for solubility and passive transcellular permeability), medicinal and combinatorial chemistry (for synthetic prioritization and retrosynthetic analyses), process chemistry (reducing the number of synthetic steps), pharmacology and physiology (for pharmacodynamics and efficacy), and pharmacokinetics and toxicology (for ADME, metabolic stability, acute and chronic toxicity), to name but a few. Although not enabling such competence *per se*, cheminformatics is a computer-based tool primed to assist decision-makers in the drug discovery processes[7]. While it is apparently objective, its computational techniques are ultimately dependent on the (subjective) concepts and thinking that go into the models and assumptions onto which they were built. Thus, cheminformatics provides a forum for the necessary dialogue between theoretical sciences, and experimental chemistry and biology.

---

[1] Horrobin, D.F. *J. Royal Soc. Med.* **2000**, *93*, 341-345.

[2] Brown, F. *Annu. Rep. Med. Chem.*, **1998**, *33*, 375-384.

[3] Potfire and the ISIS Substructure Visualizer (SSV) are available from Spotfire Inc., Göteborg, Sweden; http://spotfire.tibco.com/

[4] Roberts, G.; Myatt, G.J.; Johnson, K.P.; Cross, K.P.; Blower, P.E., Jr. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1302-1314; Leadscope is available from Leadscope Inc., http://www.leadscope.com/

[5] VIDA is available from OpenEye Scientific Software, Santa Fe, New Mexico; http://www.eyesopen.com

[6] Hahn, M.M.; Green, R. *Curr. Opin. Drug. Discov. Dev.* **1999**, *3*, 379-383.

[7] Olsson, T.; Oprea, T.I. *Curr. Opin. Drug. Discov. Dev.* **2001**, *4*, 308-313.

[8] Erhardt, W.P. *Pure Appl. Chem.* **2002**, *74*, 703-785.