Chemoinformatics: manipulating chemical information to facilitate decision-making in drug discovery

The rapid identification of higher quality drug candidates is now possible with the aid of chemoinformatics

Timothy J. Ritchie, Laboratory Head, Medicinal Chemistry, Novartis

Profound changes have taken place in the practice of drug discovery over the past few years, particularly in the areas of combinatorial chemistry, HTS, virtual-molecule generation and the concept of 'drug-like' molecule properties. These changes have forced pharmaceutical companies to rethink how chemical information is managed. Thus, the area of 'chemoinformatics' has emerged to encompass these new approaches and to support decision-making in the search for novel drug candidates.

The 're-birth' of chemical information

Chemoinformatics is a relatively new term described, by Frank Brown in 1998 (Ref. 1), as the use of information resources to generate knowledge from data, which allows for faster and more informed decisions to be made in drug lead-identification and optimization activities.

Many people view chemoinformatics as an extension of 'chemical information', which is a well established concept covering many areas that employ chemical structures, data storage and computational methods, such as compound registration databases, on-line chemical literature, SAR analysis and molecule-property calculation.

However, the recent appearance of the term 'chemoinformatics' is perhaps appropriate, and also timely, given the profound changes that have taken place in the practice of drug discovery over the past few years[2].

The emergence of new chemoinformatic initiatives as a consequence of these changes, some of which are described later, have been in some respects driven by technological advances in chemistry, biology and computing, and also by the realization that to maximize the chances of success in developing new medicines, it is crucial to select the most promising chemical leads and drug candidates from all the possible alternatives early in the discovery process. In addition, the more that is known about a drug candidate at an early stage will minimize the likelihood of hidden 'surprises' emerging later in development, and also ensures that non-viable compounds are identified and eliminated quickly.

The chemical data 'big bang'

The routine performance of high-throughput biological screening of combinatorial libraries, containing thousands of chemical compounds, has resulted in an unprecedented chemical and biological data explosion. This has quickly overwhelmed existing database technologies and has forced a rethink of how large data-sets could be registered, stored and analyzed. New ways have had to be invented to enable the mining of databases, visualization of results and recognition of non-obvious patterns and trends to transform raw data into useful information.

Test early, decide early

There has been an increasing desire within the pharmaceutical industry to profile candidate-drug molecules much more extensively in the lead identification and optimization phases, with the eventual goal of producing higher quality development compounds that are equipped with appropriate physicochemical properties, thus increasing their chances of becoming efficacious and profitable marketed drugs.

The recent advances in high-throughput *in vitro* assay technologies, which allow the measurement of important parameters, such as aqueous solubility, membrane permeability, metabolic stability and potential drug–drug interactions, have helped to guide the selection of appropriate chemical-lead classes for further optimization and to flag

1359-6446/01/$ – see front matter PII: S1359-6446(01)01895-5

potential weaknesses in compounds early in the optimization process. The analysis of results from these screens, together with data from pharmacokinetic studies *in vivo*, has also highlighted the optimal physicochemical properties, or combination of properties, that are required for good oral bioavailability, brain penetration, adequate plasma half-life, and so on[3].

A universe of virtual molecules

It has also become clear that even the most efficient combinatorial chemistry approaches can generate only a minute fraction of the 1×10^{40} virtual drug molecules that could be prepared. Therefore, attention has turned to the possibility of profiling candidate molecules for desired properties entirely by computational methods, so that only the compounds that fulfil the required criteria are actually synthesized and tested in the laboratory. To this end, a considerable amount of work has been undertaken to realise fast, high-capacity computational methods that are able to calculate relevant physicochemical molecular properties accurately *in silico*, and predict the behaviour of virtual molecules in relevant situations. Although these activities should perhaps be more correctly called 'chemometrics', they are closely associated with 'chemoinformatics' as a whole.

Outlook

As described earlier, chemoinformatics encompasses the numerous chemical information sources that, together, should ensure that the right decisions are made at the right time during the course of a hit-to-lead and lead-to-drug candidate process. However, for this to occur smoothly and efficiently, the software applications and their associated end-user interfaces, through which the information is searched, sorted and analyzed, must be reliable and simple to use. However, they should also be flexible and powerful enough to handle the expanding amounts of data that can now be generated. There is also a growing need for the seamless integration with other information sources, such as genomic and proteomic biological databases, reagent vendor-information, electronic laboratory notebooks and on-line scientific literature.

Thus, one of the major challenges that companies now face, is how best to integrate and present this vast and complex array of information, so that the people who actually make the decisions in drug discovery will be able to make the right choices quickly. Those companies that achieve this goal will undoubtedly gain a competitive advantage in the race to discover, develop and market new medicines in the 21st century.

References

1 Brown, F.K. (1998) Chemoinformatics: what is it and how does it impact drug discovery? *Annu. Rep. Med. Chem.* 33, 375–384

2 Hann, M. and Green, R. (1999) Chemoinformatics – a new name for an old problem? *Curr. Opin. Chem. Biol.* 3, 379–383

3 Blake, J.F. (2000) Chemoinformatics – predicting the physicochemical properties of 'drug-like' molecules. *Curr. Opin. Biotech.* 11, 104–107

Timothy J. Ritchie
Novartis Institute for Medical Sciences
5 Gower Place
London, UK WC1E 6BN
tel: +44 (0)20 7333 2168
fax: +44 (0)20 7387 4116
e-mail: tim.ritchie@pharma.novartis.com