biology. For Wilkins, it started a journey along the path to unravelling the complexity of living processes.

Harvesting data

In 1950, after being a lecturer in Physics at St Andrews University (UK), Wilkins followed Randall to King's College, London (UK), where he eventually became Assistant Director of the Medical Research Council Biophysics Unit and, in 1955, its Deputy Director. In his early days at King's, Wilkins had the freedom to switch from trying (unsuccessfully) to cause mutations in fruit flies with ultrasonic radiation to 'mucking about with tobacco-mosaic virus', a reference to structural work on this popular study organism.

'Wilkins was a polite, intensely private and self-effacing man.'

Eventually, Wilkins arrived at studying DNA using X-ray diffraction techniques. DNA purification techniques had been greatly improved by the Swiss scientist Rudolf Signer. The breakthrough came with the use of calf thymus cells as the source of DNA, which could be extracted as long fragile threads suitable for structural examination. Working with Ray Gosling, who is currently Professor Emeritus at the University of London, Wilkins found that the addition of water to DNA obtained using this method enabled the production of a gel from which fibres could be extracted. Gosling wound 35 such fibres around a paperclip, examined them by X-ray diffraction and obtained '…these wonderful spots', which demonstrated that DNA had been crystallized. Amazingly, the work was not immediately pursued because of a lack of instrumentation in the laboratory.

Once a new X-ray tube with a beam of sufficient intensity became available, the project had become the responsibility of Wilkins' new research assistant, Rosalind Franklin. The intricacies of the relationship that developed between Wilkins and Franklin have been well-documented. Through Wilkins, the information that DNA was a 'monoclinic, face-centred unit cell' eventually reached the fertile mind of Francis Harry Crompton Crick. The rest is history and the double-helical structure of DNA soon followed in 1953.

Public opinion, social responsibility

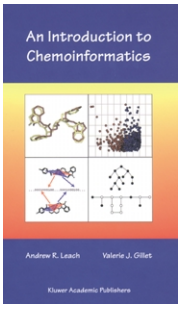
Wilkins was a polite, intensely private and self-effacing man. He had a strong dislike of pompous ideas or actions and was particularly proud of his involvement in the Campaign for Nuclear Disarmament. For 22 years, Wilkins was president of the British Society for Social Responsibility in Science. He was also an active undergraduate lecturer on the social impact of the biological sciences. These activities grew out of his lifetime experiences and the strength of conviction he felt towards the need for scientific undertakings to reflect social responsibility. He will be remembered for having taken part in perhaps the greatest biomedical discovery of the 20th century.

Maurice Wilkins died on 5th October 2004. He was 87.

References

1 Schrödinger, E. (1944) *What is Life?*, Cambridge University Press

Barry Whyte
Consultant, Rochat & Partners
Rue Du-Roveray 12
Geneva, Switzerland
e-mail: barry@rochat-pr.ch



Chemo-informatics: Concepts, Methods and Tools for Drug Discovery

Methods in Molecular Biology, Volume 275, Edited by Jürgen Bajorath, Humana Press, US$125.00, 524 pages, ISBN 1-58829-261-4

Chemoinformatics: Concepts, Methods and Tools for Drug Discovery is a timely review of current chemoinformatics trends within both academia and industry. Jürgen Bajorath, renowned for his work in this field, has edited this latest addition to the *Methods in Molecular Biology Series* and has drawn upon the vision and practices of over 40 leaders within the field to yield a weighty tome dedicated to this rapidly evolving discipline.

In almost every chapter, the contributors discuss the raison d'etre for chemoinformatics. The emergence of both combinatorial chemistry and HTS has led to compound library profiling at a hitherto unknown rate. As a consequence, the bottleneck of drug discovery was shifted from the practicalities of screening, to the sifting of many millions of datapoints. Many

  PII: S1359-6446(04)03263-3

believe that chemoinformatics can help manage this process more effectively and efficiently.

Two years ago, Claus and Underwood's seminal review [1] offered an insight into the challenges and opportunities provided by chemoinformatics as part of the drug discovery process. In addition to offering a timely update, this collection of essays also focuses on the practical aspects of chemoinformatics and assumes that the reader has some prior knowledge. After an informative introduction by the Editor, the initial chapters launch straight into current concerns with the drug discovery process, including the rational *in silico* design of compound libraries with a particular focus on similarity (or diversity) measurements. Maggiora and Shanmugasundaram point out that assessment of chemical similarity can be subjective and cite an assorted armamentarium of recent approaches including an in depth treatise on the mathematics underlying molecular similarity measurements. They summarize how diverse chemical space can be compared and broken down into mathematical form enabling rational comparators of diversity to be formulated. A raft of alternative similarity measurements is also provided, including that of Tamimoto, now often cited as the method of choice for candidate selection in a growing number of commercially available chemical libraries.

Intelligent data management is a second important theme recurring throughout the book. It helps maximize efficiency and minimize the costs of drug discovery. Data reduction methods are an essential tool for managing everexpanding datasets in lead discovery. The effectiveness of clustering, nearest neighbour searching and subsequent scoring functions are described, as are partitioning methods designed to rank the potential effectiveness of drugs. In addition, there is an increasing number of algorithms that effectively measure molecular connectivity and account for overlapping physical parameters, thereby enabling structurally diverse compounds to be linked according to particular functionality.

One notable approach to data management when grouping potential drug candidates is Robert S. Pearlman's well-cited BCUT methodology [2]. Here, in addition to structural similarities, molecular connectivity is employed to link numerous physical parameters such as polarizability, hydrogen bond donating and accepting ability, and charge distribution. These characteristics are all related to the ability of a drug candidate to interact with a receptor and form the basis of numerous commercially available chemoinformatics software packages. Many of these packages are not designed to reduce the fallibility of the medicinal chemist but to provide an intelligent informatics platform for rapid decision-making. Apart from the ability to cluster drug candidates according to physicochemical similarity, such dedicated software can also score these qualities and continually reprioritize the drug desiderata without compound elimination. The medicinal chemist is far from redundant in such processes. Beyond library synthesis, careful examination the soundness of the drug candidate is imperative, and important parameters such as the drug ADME–Tox profiles have to be considered in great detail. To some extent, one might view such multi-parametric methodologies as a complex extrapolation of Lipinski's rule of five [3].

With Moore's law [4] – computational power doubles every 24 months – still being maintained more than four decades later, *in silico* systems for drug discovery can now be seen as essential tools that have the ability to process the data being generated by pharma R&D at a colossal rate. This continual increase in processing power will have an enormous impact on chemoinformatics in the 21st Century. However, to further progress, data heterogeneity needs to be reduced. This means that informatics systems are required that allow data mining and intelligent decision-making processes to be based upon a complete information dataset for every potential drug candidate and at any point during the discovery process. As such, several contributors also discuss the practical considerations involved in hyphenating chemoinformatics and HTS to provide systems that effectively overcome data jams and streamline the decision-making process. The chapter by Bembenek *et al.* describes a web-based informatics system, designed to manage and data-mine both commercially available compound libraries and disparate data sources using user-friendly web-based tools and flexible implementation. However, there are still many challenges involved and current constraints include the incredible quantity of data points involved, logistical issues such as compound availability, compound storage and also the additional statistical issues that assure the quality of the data produced. It is opportune that a recent paper in *Drug Discovery Today: BIOSILICO* [5] also supports the need for simplicity and user-friendliness of such web-based informatics tools.

From an aesthetic perspective, Bajorath has edited this text effectively and many of the chapters described in this volume overlap, enabling the pertinent themes to flow. It is apparent that, particularly for dataset selection and directed library design, many of the approaches are variations on a theme and therefore the reader is offered a well-rounded view of this subject matter. Numerous chapters are reliant on 'screen dumps', which are a useful feature pictorially and help demonstrate the software packages under discussion. In addition, much of the text is augmented with numerous chemical structures, molecular models, histograms

and tabulated data that assist the reader's understanding. All chapters are well referenced with the majority of peer-reviewed citations having been published within the last four years offering the reader a comprehensive up-to-date treatise.

There is a mix of contributors that illustrates the wide-reaching role chemoinformatics in the pharmaceutical industry yet there is also a hint of self-justification, as can be expected from any relatively new and radical approach in an industry under continual pressure to deliver new drugs targets. Chemoinformatics is seen to contribute, regardless of whether one is defining library components, predicting log P, studying ADME–Tox and pin-pointing potential substrates for P450s, to name but a few examples. Hence, Bajorath offers a snapshot of current practices and a vision of the future within the pharmaceutical industry. Ultimately, the key factor in drug discovery is ensuring that the desired candidate is fit for purpose in the final therapeutic indication. This text neatly demonstrates how chemoinformatics can provide a streamlined platform that dovetails all the salient information pertaining to a drug candidate prior to reaching a Go/No Go decision point.

There are already numerous postgraduate courses available for the study of chemoinformatics and this is testament to the future potential of this rising star for drug discovery. This text offers a host of ideas and methodologies that will supplement the knowledge of both the student and practiced cheminformatician alike.

References

1 Claus, B.L and Underwood, D.J (2002) Discovery informatics: its evolving role in drug discovery. *Drug Discov. Today* 7, 957–966
2 Pearlman, R.S. and Smith, K.M. (1998) Novel software tools for chemical diversity. *Perspect. Drug Discov. Design* 9–11, 339–353
3 Lipinski, C. *et al*. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* 23, 3–25
4 Moore, G.E. (1965) 38 (8), 19 April 1965
5 Ertl, P. *et al*. (2004) Web-based cheminformatics tools deployed via corporate Intranets. *Drug Discov. Today: BIOSILICO* 2, 201–207

Michael E. Cooper
G.E. Healthcare
Biosciences
Discovery Systems
The Maynard Centre
Cardiff
UK CF14 7YT
e-michael.e.cooper@ge.com