

Underdeveloped territories of bioinformatics

Molecular Bioinformatics: Algorithms and Applications by Steffan Schulze-Kremer. Walter de Gruyter, New York, 1995, \$98.00, (hbk).

John C. Wootton

The phrase "molecular bioinformatics" evokes a hot research area that is both intellectually exciting and also essential in most of today's discovery-oriented biotechnology and pharmaceutical companies. There is certainly a demand for textbooks, guidebooks or overviews that might help a suitably talented experimental molecular biologist or a computer scientist learn to become an effective, marketable computational biologist. Contrary to the implications of its title and cover notes, Schulze-Kremer's thought-provoking book is not really such an overview, and the reader will need significant background knowledge of both computer methods and molecular biology to be able to fully appreciate the author's perspective and vision, and to follow the detailed expositions.

Schulze-Kremer deliberately restricts his definition of "bioinformatics" to relatively underdeveloped territories. Artificial intelligence and machine learning are emphasized heavily, whereas methods based on numerical computation, statistics, probability and graph theory have been largely avoided. Some recent innovations are covered in depth, particularly if they have already made an impact on the classification and prediction of macromolecular structures, but this choice of coverage excludes most of the current, well-tested, widely used algorithms for analyzing large databases of genomic, sequence or three-dimensional structural data or for macromolecular modeling, molecular mechanics, and dynamics, or for research in large-scale combinatorial technologies.

Molecular biology has attracted a number of computer scientists seeking application domains for emerging, experimental computational methodologies, including approaches (e.g., genetic algorithms and

neural networks) that have explicitly borrowed biological models and metaphors. Apart from neural networks, few such methods have yet been systematically evaluated for their potential to generate new biological discoveries. The principal strength of Schulze-Kremer's book is that it describes and assesses many of these rapidly evolving methods and demonstrates the existence of promising biological applications for each of them in a single volume from the perspective of a single expert.

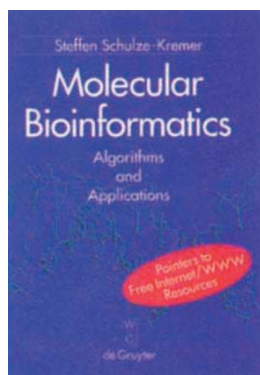
The approaches covered include: Symbolic knowledge processing, especially knowledge representation and expert systems based on Lisp; automatic logic inference based on predicate logic and constraint satisfaction, particularly Prolog applications; inductive machine learning, especially probabilistic classification and use of heuristic decision trees; genetic algorithms and their variants, including simulated annealing; and artificial neural networks, emphasizing perceptrons, back-propagation, and Kohonen networks. Detailed applications to the classification and prediction of protein structure and substructure are used to illustrate almost all of these methodologies. In addition, applications are described for transcriptional regulation systems, experimental design in genetic engineering, process design for protein purification, interpretation of NMR spectra, prediction of signal peptide sequences, and recognition of exon-intron junctions. However, one increasingly important domain for many of these methodologies is not empha-

sized in this book, namely, the association of knowledge of complex biological functions such as metabolism, signal transduction and morphogenesis with correlated data of genomic sequence and molecular structure.

This book does not stand alone. Potential purchasers can view the Web-savvy author's home page, with links to the complete table of contents and preface (<http://mycroft.rz-berlin.mpg.de/~steffen/>). Samples from chapter 5 on evolutionary computation are also available as part of the author's open World Wide Web (WWW) educational resources <http://www.techfak.unibielefeld.de/bcd/Curric/ProtEn/contents.html>). The book also includes a list of Internet/WWW entry points as sources of further information in bioinformatics.

The author has a talent for capturing in a nutshell the essentials of a methodology together with concise views of its domain of applicability and its strengths and limitations. In contrast, the examples and demonstrations presented are far from concise: several programs and outputs are listed in full and discussed in depth, including analysis of their shortcomings. This may allow the assiduous reader to make informed independent judgments of the algorithms' potentialities, but is less accessible to others.

Schulze-Kremer's approach is refreshingly eclectic and avoids promoting any single methodology. Inevitably, not all experts will agree with all his assessments, and some will be irritated by the oversimplified historical perspectives. Considering its high density of content, the book contains relatively few detailed errors. It would benefit from a glossary, however, especially as jargon abounds and some genetic terms have different meanings in biology than they do in genetic algorithms.



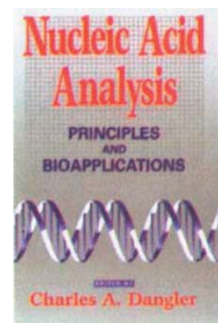
Application better than principles

Nucleic Acid Analysis: Principles and Bioapplications, Charles A. Dangler (Ed.). John Wiley & Sons, New York, 1996, pp. 284, \$89.95, (hbk).

Graham R. Taylor

How can a nonspecialist quickly become familiar with a new area of technology? Scientists and technologists, those who fund

them, and those who hope to profit from their activities are continually striving to keep abreast of new developments. Fortunately, several well-trodden paths to knowledge are at their disposal. These include database searches, literature surveys, review articles, seminars and courses, but perhaps the most widely used approach is to peruse a medium-sized book



John C. Wootton is at the Computational Biology Branch, National Center for Biotechnology Information, Building 38A, Room 8N805, National Library of Medicine, National Institutes of Health, Bethesda, MA 20894 (wootton@ncbi.nlm.nih.gov).

Graham R. Taylor is at the Yorkshire Regional DNA Laboratory, Ashley Wing, St James' Hospital, Leeds LS9 7TF, UK (gtaylor@hgmp.mrc.ac.uk).