The Meaning of Life

Bioinformatics—From Genome to Drugs. Vol. I + II. Edited by *Thomas Lengauer.* Wiley-VCH, Weinheim 2002. 650 pp., hardcover € 299.00.—ISBN 3-527-29988-2

This latest volume in the series *Methods and Principles in Medicinal Chemistry* deals with the very topical subject of bioinformatics. Bioinformatics is the interface between information science and biology, and involves molecular biology, biochemistry, and genetics on the one hand and computer science on the other. With the first draft of the human genome having been published recently (February 2001), the already burgeoning field of bioinformatics is about to undergo a developmental explosion, likened in impact to the introduction of the Periodic Table 130 years ago. That defined the vocabulary of chemistry and heralded its transition from alchemy to a quantitative discipline. At this defining moment in bioinformatics it seems appropriate to describe and summarize the current status of the field. This book is intended to provide that description, and to be a background and introduction into the field of bioinformatics for a wide general readership, including biologists, pharmacists, medicinal chemists, and medical doctors, and also to be a snapshot of the field as it exists today.

The editor has classified bioinformatics under its background science and basic technologies on one hand, and the application of these technologies towards the ultimate design of new drugs on the other hand. This natural classification has resulted in Volume 14 of the series being published as two volumes. Both volumes follow a similar logical sequence, as described in the subtitle—"From Genome to Drugs".

Book 1 begins with sequence alignment methods, followed by Chapter 3 describing the structure of the genome and the performance of current programs used in the automatic identification of open reading frames, promoter regions, introns, and exons. Chapter 4 describes the even more difficult problem of identifying regulatory regions in the genome. Current tools are based on the identification of sequence transcription factor binding site motifs, either singly or in multiple occurrences considering their relative order. It is emphasized that even such a complex sequence analysis is a highly simplistic view. Transcription initiation may require the complex three-dimensional assembly of many proximal and distal transcription factor binding sites, with accessory proteins in a complex molecular jigsaw puzzle. Once the gene has been identified, the next step on the path from genome to drug is the prediction of the protein structure for which the gene serves as the code. Chapters 5 and 6 describe the current status of protein-structure prediction, beginning with homology modeling, followed by de novo protein-structure prediction. Whereas homology modeling can be quite successful, ab initio structure prediction direct from the sequence is still an unsolved problem. The research community working on protein-structure prediction is continually measuring its success through the CASP (comparative assessment of structure prediction) method of blind prediction experiments, as described here. Once a protein structure has been obtained, the book infers that the last step on the path towards a drug is the docking of small molecule drug candidates into the protein structure and the prediction/scoring of potency. This process is described in Chapter 7. The final chapter of Book 1 describes the prediction of protein–protein and protein–DNA interactions.

Book 2 begins with a review of currently available biological data resources and their integration. Chapter 2 describes bioinformatics support for genome sequencing projects. Interestingly, although the chapter goes to great lengths to describe strategies in genome sequencing projects, the minimization of errors in sequence reading, and assembly, it does not mention the origin of the DNA being sequenced. This is a fascinating question sociologically, but scientifically almost irrelevant, as individual human genomes are 99.9% identical. It appears that both Celera and the Human Genome Project used multiple sources of DNA from both males and females of different ethnic origins to generate a consensus genome. Chapter 3 discusses the importance of sequence polymorphisms and genotype technologies, not only in identifying genetic diseases, but also increasingly in understanding drug efficacy and safety. Chapters 4 and 5 describe proteome analysis and the use of differential gene expression and protein expression in identifying suitable targets for drug-hunting campaigns. Chapter 5 describes current methodologies for the screening of drug databases, as well as recent validation studies of pharmacophore searching and structure-based design methods for drug-hunting programs.

Each subject is covered in sufficient detail to obtain an overview, and the comprehensive list of references for each chapter allows the reader to directly access the original research papers if more detailed description is required. The Internet has had a pivotal role in the

  1433-7851/02/4114-2609 $ 20.00+.50/0

global development of bioinformatics. Many of the biological tools that form the key informatics infrastructure are freely available on the World Wide Web, and URLs are given throughout the text and citation lists. Each subject is critically reviewed, and while most chapters give an acceptable summary, several chapters are truly excellent in their even-handed treatment of their subject matter. These include, for example, the reviews of drug database screening by Stahl, Rarey, and Klebe, and of modeling of protein–protein and protein–DNA docking by Sternberg and Moont, and the truly eye-opening chapter by Thomas Werner describing the analysis of regulatory regions of genomes.

In contrast, some chapters are quite impenetrable, and although they will probably make useful reference sources for practising bioinformaticians, they may cause some headaches for the intended readership. Not enough attention has been given to the base-level knowledge required of the intended reader. For example, abbreviations and acronyms, (ORFs, SDP, *z*-score, SDS) and concepts (pseudogenes, frozen approximations, NP-hard and NP-complete, dynamic programming) are often introduced with little or no explanation. Indeed, the contributions of informatics and statistics to bioinformatics are given only a very cursory treatment. For example, hidden Markov models (HMMs), neural nets, and dynamic programming are introduced without explanation in several chapters, and although the appendix to Book 1 does attempt definitions, these are so superficial as to be of little instructive use.

What is the final impression that the two volumes have left with this reviewer? (who might be described as a practising cheminformatician). We now have the galleys of the book of life. We will soon have the first edition. The book will have little punctuation, and while we have some knowledge of where the punctuation may be, this knowledge is not complete. So the automated identification of genes within the genome is still uncertain. Even with the gene, the next jump from genome to protein is still very large. With structural genomics just beginning on its task of identifying all the possible protein folds, the prediction of protein structure from sequence is a very difficult task. While homology modeling can be successful if one can find good sequence homology between a protein of unknown structure and one that has been previously determined, a change of a single amino acid between two otherwise homologous proteins can cause major changes in binding affinity for common ligands. Homology models may never have sufficient resolution to critically guide drug design. Even with an experimentally determined protein structure (from X-ray crystallography or NMR spectroscopy), the jump from protein to potent ligand is also very large. The docking and scoring of ligand potency, either for small ligands or by the docking of a second protein to the target protein, is still a very inaccurate science. Even where a protein structure has been determined experimentally with high resolution, uncertainties in the structure itself (in atomic positions, presence of water molecules, the protonation state of ionizable amino acid side chains, the thermodynamics of drug receptor interactions within a particular active site, and the extent of induced fit) still make docking and scoring a very significant challenge. Finally, the leap from ligand to drug is also very large. This area of medicinal chemistry is hardly mentioned, and our understanding of the physicochemical properties governing absorption, distribution, metabolism, and elimination, not to mention toxicology, is still developing. We have a long way to go, but we are making significant progress in all these areas.

The aims of these two books are laudible, and I fully support the concept of the books, as a primer for scientists who may benefit from our current bioinformatics tools. They are definitely worth a read, although for several of the chapters two or three readings may be necessary. But, with the text of the book of life now in our hands, the disciplines of bioinformatics, cheminformatics, chemistry, and biology need to come together much more closely than this book suggests, or even appreciates, if we are to stand any chance of mapping a direct path from genome to drug.

Andrew M. Davis
Department of Physical
and Metabolic Science
AstraZeneca R & D Charnwood
Loughborough (UK)

Organobismuth Chemistry. By *Hitomi Suzuki* and *Yoshihiro Matano*. Elsevier Science, Amsterdam 2001. 639 pp., hardcover $ 131.50.—ISBN 0-85404-637-2

Bismuth is the heaviest stable (namely nonradioactive) element of the Periodic System. This may be the reason why it remained in the background of awareness for many chemists. That is certainly not justified, as was already well known long before the appearance of the book *Organobismuth Chemistry*. The introduction to this book is alone sufficient evidence of that. It contains knowledge about the element bismuth in a very concentrated but nevertheless thoroughly informative presentation. This introduction could be used without hesitation as a basis for a lecture on the element bismuth. Everything important is presented briefly, starting with historical aspects, and continuing on to metallurgy and finally pharmaceutical applications.

The main theme of the book is, as the title states, the organic chemistry of bismuth. This is interpreted very broadly: it is taken to include $Bi(OR)_3$ and related compounds, as well as $BiCl_3$ arene complexes, thereby increasing the number of literature citations to more than 1600. There is the impression that the authors have tried to include every relevant original publication. I could not find any omissions, even in those chapters on topics with which I am especially familiar. On the other hand, I also detected gaps of my own knowledge there.

So the book is comprehensive, perhaps even to the extent of the Gmelin publications that have now ceased. It contains many preparative procedures, very much in the style of the Houben–Weyl works. They number about 150, so the book also provides an introduction for experiments. Physical data, including IR, MS, NMR, and UV spectral data, are collected together in tables, and here again the impression of completeness and accuracy of the numerical data is striking. There is clearly a demand for such comprehensive efforts, and if, as here, the result is a book that is easy to read in spite of the wealth of data, thus providing more than a pure reference book, then it is especially welcome.

  1433-7851/02/4114-2610 $ 20.00+.50/0