Book Reviews

In Silico Technologies in Drug Target Identification and Validation. Drug Discovery Series. Volume 6. Edited by Darryl León and Scott Markel. CRC Press, Boca Raton, FL. 2006. xi + 490 pp. 16 × 24.5 cm. ISBN-13 978-1-57444-478-0. $169.95.

One nightmare of medicinal chemists is being unable to optimize activity and bioavailability of a lead compound. A more dreaded outcome, because of the additional expenditure of time and resources, is to develop an exquisite ligand for a target only to have the candidate fail in the clinic because the target proves to be irrelevant for the disease; i.e., the target is not validated. The more chemists know about how biologists identify and validate drug targets, the more they can help ensure a successful project outcome. The first 300 pages of this volume are dedicated to just these issues, reviewed from a practicing biologist's perspective but described completely enough to provide useful insight for the medicinal chemist and pointing to online databases and tools that enterprising chemists might try using themselves.

Part I (Target Identification) has chapters on pattern matching for sequences, computational protein annotation, impact of genetic variation on drug discovery and development, and mining of gene-expression data. Part II (Target Validation) covers text mining, pathways and networks, molecular interactions in protein complexes, intelligent systems for predicting protein subcellular localization, and use of 3D structures in target discovery and validation. The remainder of the book covers some interesting emerging technologies, as well as computational infrastructure issues. Part III (Recent Trends) covers comparative genomics, pharmacogenomics, use of human simulation models for target identification and validation, and drug discovery using 3D structures of protein targets. Part IV (Computational Infrastructure) reviews database management, BioIT hardware configuration and software architecture, workflows and data pipelines, and biological ontologies.

Sequence pattern matching for proteins, nucleotides, and genes is covered in Chapter 2. An impressive number of databases are described, covering protein patterns (motifs, domains, secondary structure) and nucleotide patterns (cleavage sites, repeating sequences, transcription factors, etc.). A large number of software tools for finding new patterns are described and cited. In Chapter 3, sequence-based methods for functional and structural assignments are reviewed. The authors caution that many sequence annotations found in databases are assigned by sequence similarity and have not been confirmed experimentally, which undoubtedly has contaminated the databases through error propagation. Indirect methods of functional gene assignment are reviewed, including gene fusion data, domain co-occurrence, protein−protein interactions, microarray expression profiles, and others. Methods of building protein 3D structures from sequence information are briefly considered. In addition to sequence similarity methods, ab initio structure prediction methods are reported to have become sufficiently reliable to predict the structure of proteins in the absence of any detectible homology to known protein 3D structures! The authors note that the Protein Data Bank now contains an unprecedented number of protein structures of "unknown function" (from structural genomics centers), and these may be attractive starting points for drug discovery projects if the protein function can be inferred. They conclude with a valuable list of links to almost 200 annotation and function assignment tools mentioned in the text.

In Chapter 4, a strong case is made for the importance of genetic variation for drug action variability in humans. The author notes that an estimated 59% of human genes may undergo alternative splicing, accounting for the complexity of humans despite our number of genes (25 000 or so) being smaller than expected. The types of variation are listed, and human genetic variation databases and Web resources are described, including OMIM (Online Mendelian Inheritance in Man list of human genes and their clinically significant mutations and polymorphisms), HapMap, and many others. Finally in Part I, Chapter 5 covers mining of gene-expression data from microarray experiments to measure gene expression, analyze pathways, and extract other types of information.

Part II (Target Validation) starts with an excellent review of text mining (Chapter 6). Literature resources for text mining are listed, and the limitations of simple Google-type searches are made clear. Examples are given of using text categorization/clustering, ontology-based, and other mining approaches. Applications demonstrated include predicting potential toxicity of various drug targets, performing disease-to-gene linkage analysis, selecting biomarkers, simulating systems biology pathways, and others. However, the authors of the chapter warn that "text mining is not a task to be undertaken lightly." At this point it requires significant computational informatics support and specialized knowledge.

Chapter 7 discusses metabolic, signaling, and regulatory pathways and pathway data acquisition including transcriptomic, proteomic, and metabolomic data. A nice list of available databases is given, followed by discussion of various pathway analysis methods. While the discipline is still developing, it seems clear that pathway analysis should eventually be able to identify the best targets in a pathway for therapeutic intervention.

Chapter 8 makes the important point that we have a "genecentric" view of drug discovery targets at present; however, proteins often act as parts of larger complexes, so we should be considering interacting protein functional units when we design strategies to intervene. The experimental, computational, and hybrid approaches to determining protein−protein and other interactions are reviewed, and the major available interaction databases are listed. This approach will undoubtedly become increasingly important as our knowledge of protein structural complexes increases.

Chapter 9 reviews the important new technique of designing siRNA probes to target mRNA expression levels and thus validate targets, knockout specific proteins, elucidate pathways, and develop new therapeutics. Databases of already available siRNA sequences and effects are listed, as well as public and commercial siRNA design tools.

Chapter 10 considers the issue of protein subcellular localization. Interacting proteins must be in the same subcellular compartment to work together, but proteins are generally synthesized on cytoplasmic ribosomes and then transported from one cellular compartment to another by means of targeting signals on the proteins. Progress has been made in identifying

those translocation signals and thus in predicting protein localization. A number of interesting prediction methods are available as online services and are described and cited in this chapter.

Chapter 11 updates use of 3D structures in target discovery and validation. With some structural information able to be inferred for nearly half of all know protein sequences, valuable insight into function can be gleaned. The authors review how to find structures related to your sequence of interest, how to build and view a structural model, how to determine a ligand binding site, and how to design ligands for it. For each process they cite Web-accessible tools that can be used by nonspecialists.

Part III (Recent Trends) starts with Chapter 12 on comparative genomics, briefly discussing its application for infectious disease, for human disorders, for tracing evolution of organisms, for understanding regulatory pathways, and for agricultural uses. The authors describe some recent tools to perform comparative genomics studies using available sequence databases.

The important discipline of pharmacogenomics is reviewed in Chapter 13. The author begins with case studies of correlating gene variation with therapy variation for several diseases and then discusses the techniques of SNP genotyping, haplotyping, linkage disequilibrium, gene expression, methylation (epigenetics), and proteomics. A number of potential uses of the technique are recounted, as well as informatics challenges and ethical issues. In contrast to some of the more glowing treatments of this subject, the author here attempts to balance the very real potential with the significant difficulties that must be overcome.

Chapter 14 describes very interesting, but preliminary, work to build complex logical and/or quantitative human simulation models (so-called "virtual patients") for candidate drug selection and clinical trial design. Chapter 15 briefly updates methods for performing structure-based drug discovery, namely, QSAR, structure searching, and protein−ligand docking.

Part IV (Computational Infrastructure) begins (Chapter 16) with a review of biological databases, data integration issues (centralized database vs data federation) and software, and major data manipulation software packages. Chapter 17 briefly considers BioIT hardware issues including systems architecture (scalable compute clusters are favored by the author), networking, and security issues. Chapter 18 describes the requirements of a BioIT software architecture and recommends a high-performance compute cluster with a service-oriented architecture (SOA) to meet the needs of a bioinformatics laboratory.

One of the most interesting advances is the appearance of workflow and data pipeline applications that can make routine some of the repetitive calculations that biologists and other scientist must perform. (The author describes computational workflows as working on data sets, while a data pipeline may work iteratively on individual data elements to build up a results data set.) Chapter 19 reviews the background, theory, and available tools (commercial and open source) for building up workflows and data pipelines.

Finally, Chapter 20 deals with the troublesome but important issue of biological ontologies. According to the authors, "an ontology attempts to capture a community's understanding of a domain as a structured collection of vocabulary terms and definitions." They describe the Web ontology language OWL and resulting computer-readable representations of classes from the gene ontology (GO) biological process ontology. Ontologies can facilitate the representation and integration of scientific data and information, can enable process workflow and information sharing, and can increase the power of text mining. The authors paint a bright future for biological ontologies.

As indicated, the book is wide-ranging and yet practical in its review of in silico technologies, and the medicinal chemist will pick up useful background information in important target identification and validation techniques. The authors are clearly knowledgeable about their fields, and the editors have done an excellent job of melding this multiauthored book into a cohesive whole. I found only a few scattered typos that did not interfere with comprehension, and four color plates nicely illustrate some of the techniques. Most chapters are well referenced, mostly through 2004 with a smattering of 2005 citations. Some of the chapters were acronym-heavy and could have benefited from a glossary of terms.

Peter Gund

IBM Federal Healthcare Practice
20463 Sunbright Lane
Germantown, Maryland 20874

JM078007F

10.1021/jm078007f

March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure. Sixth Edition. By Michael B. Smith and Jerry March. John Wiley & Sons, Inc., Hoboken, NJ. 2007. xx + 2354 pp. 16 × 24 cm. ISBN 978-0-471-72091-1. $99.95.

Michael Smith of the University of Connecticut has again done an outstanding job updating the already excellent 5th edition to an even more impressive 6th edition of Jerry March's book. The five new reaction sections along with 5000 new references make this a formidable reference source, with more than 25 000 references in all. If one realizes the breadth of organic chemistry information available, it can be concluded that even in 2354 pages it is impossible to cover all subjects in too great a detail, and inevitably, a lot of information can only be briefly addressed.

In keeping with the March style, this text is divided into two parts. Part 1 (Chapters 1−9) focuses on basic organic chemistry, providing the reader with a comprehensive coverage of the subject matter along with a detailed mechanistic evaluation of chemical transformations. The feature that has made this book a gold standard and a mainstay with chemists is its delineation of reaction type linked with excellent indices, which are cross-linked to synthetic utility. This is illustrated in Part 2 of the text (Chapters 10−19) where a vast array of synthetic transformations is explored. Each chapter offers a solid mechanistic treatment followed by grouping and numbering of the reactions by type. The text offers supporting literature citations for the original work as well as reference to numerous applications. For many of the reactions there is included reference to major reviews, although many of these are somewhat dated. Overall, literature updating has been well done, but in places the most recent reviews could have been included. A comprehensive section on utilization of the scientific literature (i.e., Appendix A) corrects any shortcoming(s) with respect to recent reviews and monographs.

Easy cross-referencing of synthetic reactions is a value that cannot be overstated. The revised Appendix B has all of the reactions from Chapters 10−19 divided into groups representing

a specific compound class. Each of these groups is then subdivided into specific transformations focused on locating the cited compound. This indexing offers the reader a more rapid correlation of each chapter section with specific synthetic transformations. This feature, coupled with the 252-page author index and the 167-page subject index, offers the optimal example of how to provide easy access to a complete text.

The book is an exceptional value for what it contains, so it remains a favorite general organic chemistry text and an easy-to-use one-volume reference. We are confident that this book will remain a dominant reference and that it will reside on many chemists' personal bookshelves.

John W. Ullrich and Richard E. Mewshaw*

Wyeth Research
Department of Chemical and Screening Sciences
500 Arcola Road
Collegeville, Pennsylvania 19426

JM078006N

10.1021/jm078006n