

## Sequence Analysis Primer, Michael Gribskov and John Devereux, editors, W. H. Freeman, New York, 1991, 279 pp., \$39.95

Winona C. Barker

National Biomedical Research Foundation, Georgetown University Medical Center, Washington, DC 20007 USA

This volume is the first in the Biotechnical Resource series to be produced by the University of Wisconsin Biotechnology Center. The series is to cover both current and emerging technologies. Those that cover current technologies are directed toward the beginning or intermediate user. These monographs aim to provide an understanding of the principles underlying the methods, to discuss common problems that beginners may face, to describe relevant applications, and to be a useful reference for laboratory workers.

Chapter 1 discusses the management of a DNA sequencing project and the identification of simple sites, transcriptional signals, coding regions, and prediction of conformational features in RNA and DNA. The second chapter is devoted to the physical and structural properties of proteins that can be predicted from the sequence. Chapter 3 discusses methods of detecting similarity and homology, with examples from both nucleic acid and protein sequence analysis. The final chapter serves as a review by discussing the sequencing and computer analysis of the Notch sequence from *Drosophila*. The appendices include a number of useful charts, sources for software and databases, and sections on hardware, electronic communication (including servers and bulletin boards), and submission of data to the major data centers. A reference list provides citations to over 300 of the important original articles in this field. Over 50 terms used in the book are defined in a glossary. The index is very brief and most terms that one would want to look up are subheadings under general terms and can be found only by reading through the entire index. For example, "hydrophobic moment" is found under "amphiphilicity," "DNA sequence assembly" under "consensus," "searching" under "databases," "BLAST" under "dynamic programming," "promoters" under "regula-

tory element types," and "Monte Carlo analysis" under "scoring systems."

Overall, this book is informative, interesting, readable, amply illustrated, and generally attractive. I particularly liked the use of a rather extensive discussion on the dot-matrix method to introduce concepts that also underlie many of the other sequence analysis methods, such as sliding window, local similarity, filtering output, estimating statistical significance, and using Monte Carlo analysis. Chapters 3 and 4 should be very helpful in guiding the novice away from basing conclusions on simplistic assumptions and models; it is made clear that biological significance cannot be judged from measures of statistical significance. Chapter 2 ends with a nice summary; however, the discussion of sequence motifs is too brief and has too few examples. All sections illustrate a variety of sequence analysis programs, although this occasionally leads to inappropriate comparisons, as when a BLAST sequence search in which the epidermal growth factor-like regions have been omitted from Notch is compared with a FASTA search that includes those regions. As might be expected when the chapters are written by different authors, there are some overlaps and occasional gaps. For example, in Chapter 4 a "PEST search program" is mentioned with no citation. The reader is, however, directed to "see Chapter 2" in which I could find no mention of PEST regions or their detection.

Unfortunately, Chapter 3, perhaps the most valuable section, and several of the appendix sections are very poorly edited and contain misspellings, missing words, duplicated words, and even a section out of order (pages 241–242). The other chapters do not suffer from such flaws. It is a shame that the otherwise high quality of the presentation is marred in this way.