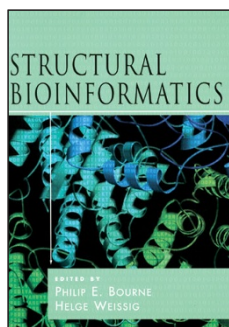


A scoundrel's refuge



Structural Bioinformatics

edited by Philip E Bourne &
Helge Weissig

Wiley, 2003
672 pp. paperback, \$69.95
ISBN 0471201995

Reviewed by
Roland L Dunbrack, Jr

I was once introduced to Sir Sydney Brenner as someone involved in 'bioinformatics.' This is not really how I would describe what I do, but I didn't protest quickly enough: Dr. Brenner laughed heartily and said, "Bioinformatics? The last refuge of scoundrels." I pointed out that I worked on protein structure prediction, and he responded, "Oh, that's different. Much better." This was a great relief to me, and, I'm sure, to Phil Bourne and Helge Weissig who have compiled and edited a timely book entitled *Structural Bioinformatics*.

The Protein Data Bank (PDB) of experimentally determined structures now contains over 21,000 entries that provide more than enough fodder for scoundrels like me and the many authors of chapters in *Structural Bioinformatics* to analyze, classify and predict the structures of proteins. The target audience for the book ranges from undergraduate and graduate students in any area of biology to experts in one area of structural bioinformatics who may need an introduction to other areas. It could easily serve as a sourcebook for courses on structural biology and bioinformatics.

The book begins with a series of introductory chapters on protein and nucleic acid structures, the experimental methods used to determine them and the databases that contain information about them. There is an appropriate emphasis on automated methods developed and used in structural genomics. The book then moves on to analysis of protein structures, including assigning domains and secondary structure, assessments of structure quality, and fold classification from a number of points of view. Several chapters follow discussing interactions of proteins with other molecules including drugs. Some of these have a flavor closer to biophysics than bioinformatics. The last major section includes four chapters on different kinds of protein structure prediction, from comparative modeling and fold recognition where the goal is to use existing structures to produce a three-dimensional model, to *ab initio* structure prediction and local structure prediction (secondary structure, transmembrane segments, coiled-coil regions), where an existing fold is not used directly in modeling. The last section

covers "the future," which consists of one chapter on structural genomics. Frankly, I hope this is not all there is to the future.

Nearly all of the chapters are presented by individuals and groups of people who have made important contributions in each specific area. Some chapters have a limited scope, such that they cover only a specific database or approach that is unique or a specific contribution of the authors. This is true of the chapters on the structure repositories (the PDB and the NDB) as well as structure classification schemes such as CATH and SCOP (the latter is presented by one of the book's editors and not by the SCOP authors themselves). These chapters would be helpful to people unfamiliar with these particular databases.

I found some chapters that compare a number of available databases or programs more useful. Some of these provide detailed comparative tables that examine the many alternatives in each aspect of designing a database or program. This is a very useful exercise, because it sheds light on what combinations of choices have been used and what range of choices is possible. I think it is important for software and databases whose creation is publicly funded to be readily obtainable, and these chapters also provide web addresses for the publicly available tools they describe. The best of these chapters, in my opinion, are those by John Tate on visualization, Roman Laskowski on quality assessment, Adam Godzik on fold recognition and Burkhard Rost on local structure prediction. I found the visualization chapter particularly informative.

The historical perspective in some chapters is well worth noting. I think there are actually relatively few new ideas in bioinformatics and computational biology. Much of what is published nowadays was first tried two or three decades ago, but without enough raw data to reach firm conclusions. We are now in the position of having large numbers of structures and sequences as well as fast computers with large hard disks. But it is still useful to be firmly grounded on the earliest efforts in one's field. The chapters on visualization (mentioned above), on domain parsing by Lorenz Wernisch and Shoshana Wodak, and on electrostatics by Nathan Baker and Andy McCammon are good examples of this. Some chapters also have extensive annotation of the references, which among other things is very useful in sorting out historical contributions of early papers.

Of course even in a book with 29 chapters, there are some other areas that could have been covered. Perhaps the most important of these would be statistical methods and algorithms commonly used in structural bioinformatics. For instance, Bayesian methods have made a strong impact in computational biology because they allow for the development of informative prior distributions that seamlessly handle data-poor and data-rich situations in statistical analysis of complex data. They also allow for hierarchical models that are appropriate for complex data such as protein structures. Some general methods such as clustering, combinatorial search, and Monte Carlo methods might also have been covered in a methods section.

In any case, this book is a useful and timely summary of a rapidly expanding field. "With the advent of structural genomics" (a phrase used in several of these chapters), comparative analyses of protein structures will increasingly yield new insights and new technologies. Scoundrels, indeed! ■

Roland L. Dunbrack, Jr. is at the Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111, USA.
e-mail: RL_Dunbrack@fccc.edu