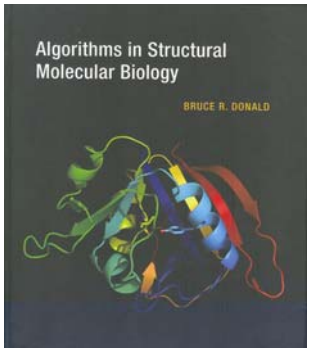


Algorithms in Structural Molecular Biology. Bruce R. Donald. The MIT Press, 55 Hayward Street, Cambridge, Massachusetts, 02142, USA. 2011. xxxi + 429 pp. Price: US$ 65.00/£48.95.

The first protein structures of myoglobin and haemoglobin were determined more than 50 years ago. The study of biology at the molecular level has become possible due to the availability of three-dimensional structures of macromolecules like proteins and nucleic acids. The unique structure adopted by a protein enables it to perform its biological function. As of July 2012, more than 83,000 structures are available in the Protein Data Bank (PDB). The explosive growth in the PDB is mainly due to advancement in technology, software development and also due to the structural genomics initiatives in the post-genomic era. Both NMR and X-ray crystallography have played a major role in biomolecular structure determination. Over the years, sophisticated software tools have been developed that are used at various levels of structure determination in both X-ray and NMR methods. While completely automated structure determination has helped accelerate the entire process to quickly determine the structures, the users most often treat the user-friendly software tools as a black box and they seldom have knowledge about the algorithms behind the software tools. Unless the user knows the algorithm and its limitations, he/she has the least chance to improve and develop a completely new procedure. In this context, the book under review assumes great significance and is a comprehensive collection of algorithms used in structural biology. Majority of these algorithms have focused on structure determination using NMR.

The book under review contains 50 short chapters presumably to facilitate developing modular courses around a specific theme from specific selected chapters. The first eight chapters briefly explain protein structure, basics of NMR, different kinds of data collected from NMR experiments, use of residue dipolar couplings (RDCs) in solving the structures and the JIGSAW algorithm which is used for NMR peak assignment. Chapter 9 mainly deals with the design of peptides. This chapter discusses algorithms to construct an all-atom representation of the protein backbone from Cα atoms and to design membrane-binding α-helical peptide. Building foldamers from new compounds that can produce stable conformation with desired biological function is also presented in this chapter. Chapter 10 devotes approaches to enzyme redesign. Studies on several prominent enzymes such as subtilisin are cited as examples. The main focus of chapters 11 and 12 is automated protein design. While the algorithm developed by Bassil Dahiyat and Stephen Mayo is discussed in detail in chapter 11, an ensemble scoring method K* used to model the protein–ligand binding in active site mutants is the main topic in chapter 12. Nonribosomal peptide synthetase enzymes are used as examples in chapter 12.

Chapter 13 sees the return of RDCs in NMR and associated computational algorithms related to assignment of RDCs, structure determination, estimation of alignment tensor and detecting homology for an unknown protein for which experimentally determined RDCs are available. Nuclear vector replacement algorithm is presented in detail in chapter 14. Its applications include automated resonance assignment from chemical shifts, RDCs, amide exchange data and unassigned peaks in NOSEY spectra. It helps to detect structural homology of a target protein from known models. A revised version of the review article published in 2009 by Bruce Donald and Jeffrey Martin in *Progress in NMR Spectroscopy* is presented in chapters 15–18. Advances in NMR assignment and protein structure determination based on sparse RDCs are described. The algorithms RDC-EXACT and SYMBRANE are discussed in detail and the results from representative proteins are presented. Readers are walked through the analogies of these methods with molecular replacement in X-ray crystallography. The application of mass spectrometry is briefly illustrated through mass spectrometry classification algorithms (MSCAs) in the next chapter. Chapter 20 has taken up the topic of an important problem of loop closures in protein structures. Two algorithms, Probik and ChainTweak, are described in this context and compared.

Chapter 21 and the next two chapters deal with normal mode analysis, an important technique to explore the dynamics of biological molecules. The algorithms FIRST, ROCK and FRODA are presented. Applications of these algorithms include sampling protein conformations, protein–ligand flexible docking and exploring the internal mobility of proteins. In chapter 24, the fragment ensemble method is introduced that can be used to generate an ensemble of protein conformations. Application of this method is illustrated for protein G and ubiquitin. The concept of free energy is introduced in chapter 25. Graphical models such as Bayesian network, Markov random fields and factor graphs are presented. Belief propagation algorithm and its application to estimate the free energy of a protein structure are explained. The problem of protein design is presented in the same chapter and the use of graphical models and belief propagation in protein design has been discussed. Chapter 26 presents different entropy components and their implications in ligand binding and ligand design.

Chapters 27 and 28 describe proteins in polyketide synthetase and nonribosomal peptide synthetase systems. The algorithm 'NOE matching' that can help compute the protein–ligand binding pose using minimal NMR data is described in chapter 29. The advantage of this method is that no protein resonance assignment is required. A protein structure prediction algorithm incorporating side-chain and backbone flexibility, called SoftROC, is discussed in chapter 30. This method combines genetic algorithm and Monte Carlo procedures and its application on 434 cro and T4 lysozyme is illustrated.

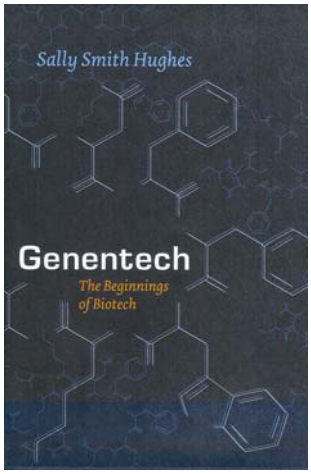
Chapters 31 and 32 provide introduction to 'distance geometry' and J. B. Saxe's reduction of two problems 'Partition' and '3SAT' to the problem of distance geometry. Different steps of 'AutoStructure', a software used in NOE assignment, and its evaluation using three different human protein NMR test datasets are presented in chapter 33. Another algorithm, MARS, a method for backbone resonance assignment, is presented in chapter 34. It is also used in

backbone assignment with known protein structure using RDCs. Dynein light chain and PDZ domain of PTP-Bas are two examples discussed in chapter 35 in which structures are incorrectly determined due to wrong assignment of NMR data. Details of Semidefinite programming and its applications in side-chain positioning problem and the sensor network localization problem are explained in chapter 36. Graph-embedding problem with local angle information is compared with the protein structure determination using RDCs in the same chapter. Graph cuts with broader applications in structural biology are discussed in chapters 37 and 38. Studying the structure of intrinsically disordered and unstructured proteins is an important and challenging area in structural biology. Using ensemble-averaged RDCs and small-angle X-ray scattering data to derive the structure of an unfolded protein is presented in chapter 39. Chapters 40 and 41 revisit the protein–ligand binding and present methods for flexible protein–ligand docking using continuum solvent model and metadynamics. The problem of comparing an ensemble of structures with a reference structure is addressed in chapter 42. Two different experimental approaches, mass spectroscopy and NMR, can complement each other when the systems are difficult to study. This is the major focus in chapter 43 in which mass spectroscopy-assisted NMR assignment is discussed. Another algorithm, Autolink, an automated method for NMR resonance assignment is presented in detail in chapter 44. A version of the Rosetta structure prediction program, called CS-Rosetta, for predicting the structure from experimentally determined NMR chemical shifts is described in chapter 45. Its accuracy in predicting 16 protein structures and its limitations are discussed in the same chapter. A new support vector machine-based approach for predicting substrate specificity of enzymes is discussed in chapter 46. CRANS (cross-rotation analysis), an algorithm to predict and analyse non-crystallographic symmetry in X-ray diffraction is dealt with in chapter 47. Phase problem, molecular replacement method in X-ray crystallography and application of normal mode analysis are some of the highlights in chapter 48. Another application of genetic algorithm in structural biology, namely optimizing the charge–charge interactions on the surface of protein structures, is presented in chapter 49 and the example protein used is Fyn SH3 domain. The final chapter talks about computational topology, triangulation, alpha-shapes and applications in the contexts of protein structure.

It was almost 50 years ago, that G. N. Ramachandran and his colleagues made the fundamental contribution to structural biology in the form of the Ramachandran map. Today, it is not uncommon to encounter a student of biochemistry, biophysics or related discipline who cannot fully describe the significance of this map or the method used to calculate the φ–ψ plot. It is in this context that this book is an important step in introducing and familiarizing the algorithms used in the software tools for biomolecular structure determination. This book presents a nice collection of several popular algorithms used in structure determination and analysis of biomolecules. Perhaps, this is probably a first book of this kind, although several books describing the algorithms used in sequence analysis are available. With major emphasis on NMR data, the title *Algorithms in NMR Structural Biology* would have been more appropriate for the book. Computer science students with exposure to biochemistry and NMR would find the book helpful. However, the readers will have to go back to the original published papers to understand the algorithmic details in depth. As far as the biology/biochemistry students are concerned, they need a solid mathematics background to understand the different topics discussed in the book. The link between the chapters could have been better and the transition between one topic and another sometimes seems to be abrupt (for example, mass spectroscopy is suddenly introduced in chapter 19). More examples in detail to illustrate the application and significance of the algorithms would help the structural biologists with biology background to take an active interest in the book. Overall, this book will be a useful reference for advanced researchers working in the area of NMR structural biology.

R. SANKARARAMAKRISHNAN

Department of Biological Sciences and Bioengineering,
Indian Institute of Technology Kanpur,
Kanpur 208 016, India
e-mail: rsankar@iitk.ac.in



Genentech – The Beginnings of Biotech. Sally Smith Hughes. The University of Chicago Press, Chicago 60637, USA. 2011. xv + 213 pp. Price US$ 25.00/ £16.00.

The biotechnology company Genentech was co-founded by a researcher, Herbert Boyer and a venture capitalist, Robert A. Swanson about 36 years ago. While Swanson (who is no more) might be largely unheard of in the scientific community, Boyer happens to be a well-known figure, particularly in life sciences. The research Boyer carried out in collaboration with Stanley Cohen on recombinant DNA (rDNA) technology won them both the prestigious Lasker award in 1980 (with two other scientists).

The popularity of the duo and their work is apparent from the number of hits that Google returns when one types 'Boyer and Cohen'. Following their collaboration on rDNA technology, both independently continued their scientific pursuits. Boyer went on to shape a new company (Genentech) with Swanson, whereas Cohen served as the scientific advisor to another company (Cetus). If one could compare, Cohen's corporate involvement was much less than Boyer's. A brief biographical account of the three key figures in Genentech's history – Boyer, Cohen and Swanson – is provided in the book.

Boyer and Swanson were the ideal co-founders of the first-of-its-kind venture into the biotech industry, with Boyer at the forefront of rDNA technology and Swanson eager to market its applications. Even if the book is spun around the development of a business enterprise, what would interest a broad scientific