

Preface

This book developed out of a review we wrote for the journal *Reports of Progress in Physics* (Taylor *et al* 2001). That work was aimed at non-(protein) specialists and in particular towards those with a more physical background. This is an aspect that is retained in the current work but we hope the work is, nonetheless, still easily accessible to readers without any (or much) training in physics or mathematics. Some large parts of the review remain unaltered in the current work, specifically, the introductory sections on basic protein structure (just because it is tedious to rewrite the same material). Other parts have reduced in size, largely as the result of comments from less specialized readers who thought they were a bit boring (and were brave enough to tell us!). The main change, however, is the inclusion of much new material which I hope should appeal to those who take a physical view of the world. These include a new section on distance geometry, an 'Ising' model for protein domain definition, an expanded section describing a 'periodic table' of protein structure, and a Fourier analysis of symmetry in structure plus lots more on topology and knots. The quotes on some of these terms indicate that they should be taken to indicate more a similarity to the physical theory rather than an exact application. This is a common feature in the methods (some would say pseudomethods) described later and is true also of many of the descriptions of protein topology. Rather than irritate, I hope that these loose approaches may help inspire someone to re-examine the problem themselves and maybe succeed in dragging the theory of protein structure a little closer to an exact science.

The work does not pretend to be a comprehensive source for the analysis of proteins and there are many methods, even some we have worked on ourselves, that are not covered. In particular, there is virtually no mention of protein sequence analysis, even when this relates to structure (such as in structure prediction and threading). There is also no mention of statistical methods, such as artificial neural nets or associative memory Hamiltonians, even though these (and other statistical methods) have been widely applied in the analysis and prediction of structure. Our decision in this has been to concentrate on simple geometric methods that are deterministic (as distinct from probabilistic). Even with this restriction, there is a large body of work on the mechanical properties of proteins that we have not ventured into. These include analytical methods,

such as normal-mode analysis and the calculation of atomic interactions using molecular dynamics. Although probably of considerable interest to those who work in physics, this is a large field in itself that has been covered widely in many books. In contrast to molecular dynamics, our aim in this work has been to concentrate on the more abstract simplifications of protein structure that allow a better understanding of the overall fold of the structure. This view also corresponds with another neglected topic: that analyses the properties of simple force fields on simplified proteins that are represented on a regular lattice. There is some mention of these studies when we deal with protein folds on a secondary structure lattice, but nothing to justify the large body of interesting work on these models.

On the original review we had a few co-authors and are grateful to them for letting us use what remains of their contributions in the current work. They were: Nigel Brown (structure comparison) and Alex May (structure classification). In addition, we were aided in the original Fourier work by Tom Flores, Jaap Heringa and Franck Baud. Linus Johannassen worked with me on the 'fold-trees' while David Jones and Kjell Petersen are thanked for providing some figures. Much of the other material is derived from our own papers in *Nature* ('periodic table' and protein knots), the *Journal of Molecular Biology* (stick models) and *Protein Engineering* (symmetry and domains) plus various sources (including those mentioned) for the distance geometry material. (Specific acknowledgements can be found throughout the text.) We are grateful to those journals for allowing us to reuse some figures and text (with most of the boring bits left out). This material has been heavily edited so any errors in the current work are almost certainly the result of our own recent activity.

Most of the methods described in this book have corresponding computer programs and these can be found (as C source code) at the ftp site of the Division of Mathematical Biology at the National Institute for Medical Research: <http://mathbio.nimr.mrc.ac.uk/ftp/wtaylor>. (Other activities in the Division can be seen by omitting the part following 'uk'). Whilst happy to hear of any 'bugs' in these programs, we cannot provide any help in getting them installed or in running them. The more commonly used geometric routines have been collected in the Web site associated with the book, along with some explanation (which is sadly lacking in most of my other programs). These constitute a little library of utility functions that are widely used in our programs (on the Web site) and an understanding of them may help anyone who is brave enough to venture into these larger bodies of code.

In the course of putting this work together, we have sometimes wondered what sort of a book it is. It is not a textbook, although it does contain some pedagogic material, neither is it a research work even though it does contain some new ideas and results. It sometimes comes close to a review but as such it is particularly biased towards our own work to the extent that, if we were older, it might be mistaken for a retrospective. In the end, we like to think of it as a collection of methods and ideas. Often the two are linked with the method

implementing the idea but some of the methods have given rise to new unexplored ideas and some of the ideas need new methods to be developed. Rather than a weakness, we hope this open-ended aspect is what our readers will value most.

Willie Taylor and András Aszódi

1 May 2004