Foreword to the First English Edition

In June 1967, Oleg Ptitsyn became the Head of the Laboratory of Protein Physics at the new Institute of Protein Research at Pushchino. Three months later, he was joined by Alexei Finkelstein; first as a research student and then as a colleague. Their approach to the study of proteins was different from that common in the West, being strongly influenced by the Russian school of polymer physics. One of its most distinguished members, Michael Volkenstein, had been Ptitsyn's PhD supervisor. Together Ptitsyn and Finkelstein created, at Pushchino, one of the world's outstanding centers for the study of the physics and chemistry of proteins.

Certain areas of their work, particularly that on protein folding, have become well known in Europe, India and America, either directly through their papers or indirectly through the elaboration of their work by two of their former students: Eugene Shakhnovich, in America, and Alexei Murzin, in England. However, it became obvious when Finkelstein or Ptitsyn talked with colleagues, that the range of their work went far beyond what was commonly known in the West. We would find that fundamental questions that were our current concerns had already been considered by them and they had some elegant calculation that provided an answer. Usually, they had not got round to publishing this work. Now, to make their overall achievements available to more than just their friends and the students of Moscow University; Finkelstein has written this book on the basis of lectures he and Ptitsyn gave to these students.

In the breadth of its range, the rigor of its analysis and its intellectual coherence, this book is a *tour de force*. Of those concerned with the physics and chemistry of proteins, I doubt if there can be any, be they students or senior research workers, who will not find herein ideas, explanations and information that are new, useful and important.

Cyrus Chothia, FRC
MRC Laboratory of Molecular Biology
Cambridge, UK
2002

Preface

This book is devoted to protein *physics*, that is, to the overall topics of structure, self-organization and function of protein molecules. It is written as a course of lectures, which include a more-or-less conventional introduction to protein science (specifically, in the first and last parts of the book), as well as the story of in-depth protein science studies (in the middle of the book). Thus, the book is a kind of monograph embedded in the course of lectures, and it bridges the gap between introductory biophysical chemistry courses and research literature.

The course is based on lectures given by us (earlier by O.B.P. and later by A.V.F.), first at Moscow PhysTech Institute, then at Pushchino State University, and now at the Pushchino Branch of Lomonosov Moscow State University and at Biology and Bioengineering & Bioinformatics Departments of Lomonosov Moscow State University. Initially, our students were physicists, then mainly biologists with some chemists. That is why, by now, the lectures have not only been considerably up-dated (as research never stops) but also thoroughly revised to meet the requirements of the new audience.

Since what you will be reading has a form of a course of lectures, repetition is hardly avoidable (specifically, we have repeated some figures). Indeed, when delivering a lecture one cannot refer to "Figure 2 and Equation 3 of the previous lecture"; however, we have done our best to minimize repetition.

The following comments will help you to understand our approach and the way in which the material is presented:

On the "lecturer". All lectures are presented here in the way they are delivered, that is, from the first person—the lecturer.

On the "inner voice". The personality of the "lecturer" comprises both authors, O.B.P. and A.V.F., although this does not mean that no disagreement has ever occurred between us concerning the material presented! Moreover, sometimes each of us felt that the problem in question is disputable and subject to further studies. We made no attempt to smooth these disputes and contradictions over, so the "lecturer's" narration is sometimes interrupted or questioned by the "inner voice." Or the "inner voice" may simply articulate the frequently asked questions or elaborate the discussion.

Why "protein physics"? Because we are amazed to see how strongly biological evolution enhances, secures and makes evident the consequences of the physical principles underlying molecular interactions in the protein. It is

also striking how much our understanding of biological systems, proteins in particular, has developed through the application of physical methods. We see this from mass spectrometry and single-molecule techniques to electron or atomic force microscopy, X-ray crystallography and NMR studies of proteins. There is hardly any other area of contemporary science in which the traditional boundaries between disciplines and philosophies have been so clearly breached to such great profit.

On the physics and biology presented in these lectures. In the course of lecturing, we shall take the opportunity to present physical ideas, such as some elements of statistical physics and quantum mechanics. These ideas, to our minds, are absolutely necessary not only for an understanding of protein structure and function but also for general scientific culture, but usually a "normal" biology graduate has either completely forgotten or never known them. On the other hand, among a myriad of protein functions, we will discuss only those absolutely necessary to demonstrate the role of spatial structures of proteins in their biological, or rather biochemical, activities.

On "in vivo" and "in vitro" experiments. The terms "in vivo" and "in vitro", as applied to experiments, are often understood differently by physicists and biologists. Strictly speaking, between the pure "in vivo" and the pure "in vitro" there are a number of ambiguous intermediates. For example, protein folding in a cell-free system (with all its ribosomes, initiation factors, chaperones, crowding, etc.) is unequivocally an "in vivo" experiment in the physicist's view (for a physicist, "in vitro" would be mostly a separate protein in solution; even the cell-free system contains too many biological details). But for a biologist, this is undoubtedly an "in vitro" experiment (since "in vivo" is referred to a living and preferably intact organism). However, structural studies of a separate protein in an organism are hardly possible. Therefore, reasonable people compromise by making biologically significant "in vivo" events accessible for experimental "in vitro" studies.

On experiment, physical theory and calculations. Experiment provides the basic facts underlying all our ideas of phenomena, as well as a lot of refining details. However, the experimental methods used in protein physics are only briefly described in this book, with references to excellent textbooks that describe them in detail. The same, and to an even greater extent, concern chemical and biochemical methods (such as those used in purification of proteins or genetic engineering), which are just barely mentioned in the book.

Theory allows us to understand the essence and interrelation of the phenomena and is helpful in planning informative experiments. Some basic physical theories—in a simplified form, naturally—are included in our lectures, not only because they permit us to put in order and comprehend from a common standpoint the vast experimental material, but also because they are elegant. Besides, we believe that knowledge of basic physical theories and models is essential to human culture.

Calculations connect theory with experiment and verify key points of the theory. However, not everything that can be calculated must be calculated; for example, it is easier just to measure water (or protein) density than to calculate it from first principles. And not only is this easier but it yields a more accurate result, since a detailed calculation requires many parameters that can hardly be accurately estimated.

On physical models, rough estimates and computer (in silico) experiments. In these lectures, we will often discuss simple models, that is, those drastically simplified compared with reality, and use rough estimates. And we want you, after reading these lectures, to be able to make such estimates and use simple models of the events in question. The use of simple models and rough estimates may seem to be quite old-fashioned. Indeed, it is often believed that with the powerful computers available now, one can enter "all as it is in reality": water molecules, salt, coordinates of protein atoms, DNA, etc. fix the temperature, and obtain "the precise result." As a matter of fact, this is a Utopian picture. The calculation—we mean a detailed one (often made using so-called molecular dynamics)—will take days and cover only some nano- or microseconds of the protein's life, because you will have to follow the thermal motions of many thousands of interacting atoms. In any case, this calculation will not be absolutely accurate either, since all elementary interactions can be estimated only approximately. And the more detailed the description of a system is, the more elementary interactions are to be taken into account, and the more minor errors will find their way into the calculation (not to mention the increased computer time required). Eventually, you will obtain only a more-or-less precise estimate of the event instead of the desired absolutely accurate description of it—with days of a supercomputer's time spent. Meanwhile, what really interests you may be a simple quick estimate such as whether it is possible to introduce a charge into the protein at a particular site without a risk of protein structure explosion. That is why one of our goals is to teach you how to make such estimates. However, this does not mean that we will simply ignore computer experiments. Such experiments yield a lot of useful information. But the computer experiment is a real experiment (although in silico, not in vitro or in vivo). It involves highly complex systems and yields facts requiring further interpretation, which, in turn, demands simplified but clear models and theories.

On equations. We are aware that mathematical equations can be a difficulty for biologists, so we have done our best to refrain from using them, and only those really unavoidable have survived (some very useful but more complicated equations are used in the Appendices and Problem sections). Our advice is: when reading these lectures, "test words by equations." It is certainly easier for biologists to read words only, but they are often ambiguous, so word-verifying by equations and vice versa will help your understanding. To avoid going into insignificant (and unhelpful) detail, we shall often use

approximate calculations; therefore you will often see the symbols "≈" (approximately equal to) and "∼" (of the same order of magnitude as).
On references and figures. The references are included in the text, legends to figures and tables; the lists of references are positioned at the end of each Lecture or Appendix.

Protein structures are drawn using the programs MOLSCRIPT (Kraulis, P.J.,J. Appl. Cryst. 24, 946–950); WHAT IF (Vriend, G., J. Mol. Graphics 8, 52–56); RASMOL (Sayle, R., Milnerwhite, E.J., Trends Biochem. Sci. 24, 374–376); Insight II (Molecular Simulations Inc., 1998); ViewerLite (Accelrys Inc., 2001); and coordinates taken from the Protein Data Bank (initially described in: Bernstein, F.C., Koetzle, T.F., Meyer, E.F., Jr., Brice, M.D., Kennard, O., Shimanouchi, T., Tasumi, T., J. Mol. Biol. 112, 535–542). Many figures are adapted from the literature, with appropriate references and permissions. Most of other figures are purposely schematic.

On tastes. These lectures, beyond doubt, reflect our personal tastes and predilections and are focused on the essence of things and events rather than on a thorough description of their details. In the main, they contain physical problems and theories, while only the necessary minimum of experimental facts are given and experimental techniques are barely mentioned. (Specifically, almost nothing will be said in these lectures about the techniques of X-ray crystallography and NMR spectroscopy that have provided the bulk of our knowledge of protein structure; we tried to compensate this by reference to excellent textbooks on experimental methods.)

Therefore, these lectures are by no means a substitute for regular fact-rich biophysical and biochemical courses on proteins. When referring to specific proteins, we merely give the most important (from our viewpoint) instances; only absolutely necessary data are tabulated; all values are approximate, etc.

On small print. This is used for helpful but not essential excursus, additions and explanations.
On the personal note in these lectures. We shall take the opportunity of noting our own contributions to protein science and those by our co-workers and colleagues from the Institute of Protein Research, Russian Academy of Sciences. This will certainly introduce a "personal note" into the lectures and perhaps make them a bit more vivid.
What is the difference between 1st and 2nd English editions of Protein Physics? To my (A.V.F.) surprise, the current revision of the book showed no need of any correction of the initial text. I have made only a few cuts there. But of course I have updated the book to include new fascinating material. Specifically, this concerns novel information on intrinsically disordered proteins, amyloid aggregation, protein folding in vivo, protein motors, misfolding, chameleon proteins, advances in protein engineering and design and advances in modelling of protein folding. Also, the revised version includes a Problems section (with solutions).