Preface

Why have I written this book, and in particular as a teaching book?

A common misconception about an experimental method is that it is claimed to be accurate. Such a claim has some justification because every experimentalist strives to eliminate the potential systematic errors in that method. But of course, if we wear the hat of the natural philosopher, as physics was referred to before the 'modern era', we can never determine the truth but only an approximation to it. If, however, we combine one method's results with another method's results we are in a good place because we can assess how well those two sets of results agree. Here we arrive at the domain of accuracy. Where then does that leave the 'single method'? Each method on its own can legitimately be described as precise when the random errors inherent in it have been minimized.

In the teaching of the physics of experimental measurements a common analogy used is that of a dartboard. You can see this if you go to Google Images and search on 'precision and accuracy'. You are inundated with many dartboards as images with each one showing darts in a cluster away from the bullseye centre of the dartboard, thereby being precise but not accurate versus a cluster of poorly placed darts around the centre being imprecise but accurate. It is a striking depiction of precision versus accuracy. However, as a metaphor of the situation of experimental methods the dartboard and its darts gloss over the key fact that in real life we don't know where the bullseye is! The truth is never known.

Where does theory fit into this description? Theory can be a powerful instrument in its own right to discern truth. A classic example is Einstein's posing of the question: what if the speed of light is finite? This question, through his pencil and paper equations, led to consequences and these could be assessed. That theory provides a benchmark, namely the speed of light is indeed finite, and that is the truth. Orthodoxy would require us, however, to state something like: this is a fundamental constant of the universe now but what if its value changed? I have to concede that is an important caveat but nevertheless Einstein's special relativity theory is as close as one can get as an exemplar of truth in physics.

In the modern era a different kind of theory has emerged as we can enjoy the fruits of databases, which in crystallography mean such as the Cambridge Structure Database and the wwPDB (the worldwide Protein Data Bank) and for the latter the highly important amino acid sequences databases complementary to the PDB. A new kind of theoretical analysis has become possible, such as the, by now classic, example of protein three-dimensional (3D) fold prediction. This involves analysing an amino acid sequence against related experimental structures in the PDB, where the precision of each structure is ranked according to the consistent standard of the PDB's Validation Report. Hence phylogenetic changes in a specific sequence suggest conservation of the 3D structure is needed. Thereby a 3D fold for an as yet unknown experimental structure could be derived by prediction algorithms using input from those different databases. The assessment of success, or confidence in the prediction, is the closeness of fit of a prediction to a precise experimental structure where one is available. It is a prediction, though, and thereby a hypothesis of what a 3D structure would look like for an amino acid sequence that has not yet had its structure determined. Experimental data, such as from X-ray crystallography, are needed to know whether it is an appropriate prediction. A major step in crystallography is still always the determination of the crystallography phase of each Bragg reflection. In macromolecular crystallography this was improved enormously, i.e. speeded up, using tuneable synchrotron radiation and the multiple wavelength anomalous dispersion method. Today, however, to use a predicted 3D polypeptide fold is most often a good starting point for phase determination.

The above preamble then provides a context for explaining my motivation for writing this teaching book. It is a book about each experimental method we can use to determine as precisely as possible a 3D structure of a biological macromolecule or portions of it. In this book I also assess how well two or more methods agree in their 3D determined structures. This allows us to dissect our knowledge of the accuracy of what we know. Of course, if we know how accurate our structures are we have a better chance, our best chance, to understand their functions. The dynamics of structures are at the heart of elucidation of function. In the history of our field two competing ideologies existed specific to interaction: the lock and key mechanism of function where rigid entities interacted in 'opening a door' versus the induced fit mechanism as one entity moulded itself into another like a hand into a glove. The latter offers greater structural flexibility. If that is the case, then we must also understand the uncertainties in our determination of atomic positions in our static structures. This is a total void in our PDB entries of atomic coordinates. As data depositors we do

not provide either uncertainty estimates on our coordinates or on our atomic B factors. But methods of analysis do exist for each experimental method that we can use to give estimates of these uncertainties. This teaching book will describe these, also assessing their weaknesses as well as strengths. At the very least we must not quote precisions that convey a false confidence in our results; a classic case is the macromolecular crystallography atomic B factor quoted to two decimal places when its precision is only an integer.

I dedicate this book to Durward Cruickshank who profoundly influenced me in our discussions that led up to his seminal paper on the Diffraction Precision Index that he published in 1999 (Cruickshank, D. W. J. (1999), *Remarks about protein structure precision*, Acta Cryst., D55, 583–601). See also Beagley B. and Helliwell J. R. (2018), *Durward W. J. Cruickshank, 7 March 1924–13 July 2007*. Biogr. Mems Fell. R. Soc., 65, 71–87, http://doi.org/10.1098/rsbm.2018.0018.