

Preface

Nucleotide sequences of nucleic acids (DNA or RNA) and amino acid sequences of proteins are the most widely used type of biological information. Nucleic acid sequences are relatively easy to determine experimentally, and can be used to predict the sequences of proteins they encode. Thanks to advances in high-throughput DNA sequencing, the number of sequences in public databases grows exponentially. Thereby, sequence databases became one of the largest data resources in biology, providing researchers with a wealth of information, and challenging computer scientists to develop tools that would allow for rapid discrimination between relevant and irrelevant information for various types of analyses. While DNA sequence is mostly a carrier of genetic information, genes typically exert their function through the RNA and protein molecules they encode. A growing body of evidence points toward the importance of various RNA molecules that do not encode proteins, but fulfill various regulatory and/or catalytic roles. However, proteins that are produced by translation of coding RNAs remain the most important type of macromolecules required for the vital functions of the cell. Therefore, this book focuses on computational analyses of proteins, although many of these analyses are relevant also to nucleic acids (e.g. sequence alignment and prediction of protein-nucleic acid complexes).

Proteins perform most essential structural, enzymatic, transport, and regulatory functions in the cell. Protein functions are strictly determined by their structures, which can be organized into four levels of hierarchies with increasing complexity. These levels are: primary, secondary, tertiary, and quaternary structure. Above the structure of individual proteins and complexes there is another level of complexity, namely networks of interactions. The description of this hierarchy extends beyond the area of traditional structural bioinformatics and enters the realm of systems biology. The chapters in this book are meant to provide summaries of the concepts underlying protein sequence-structure-function relationships at these different levels of organization, and to serve as a comprehensive resource of state-of-the-art methods (as of 2008) for the corresponding bioinformatics analyses. The book is aimed at advanced graduate students, postdocs and faculty members. Although individual chapters have been written by several different authors and each can stand on its own, the book has been designed as a cohesive whole rather than a collection of independent reviews.

The primary structure corresponds to a linear sequence of amino acid residues linked together by peptide bonds into a polypeptide chain. The C- α atoms and peptide bond atoms form the main chain backbone, and other atoms protrude away as side chains. Functionally and evolutionarily relevant elements of primary structure are sequence motifs, domains and modules. In this volume, the bioinformatic tools for protein sequence analysis, including detection of motifs and domains, are reviewed in Chapter 1 by Kaminska *et al.*, while

tools for protein function prediction from sequence are reviewed in Chapter 3 by Chitale *et al.* The level above primary structure is the secondary structure, defined as the local conformation of a polypeptide chain. It is characterized by the presence of recurring elements stabilized by hydrogen bonds between the main chain carboxyl and amino groups of different residues. Two very common types of secondary structures are corkscrew-like helices and extended strands; these regular elements are connected by loops that exhibit a plethora of different conformations. Further, some regions of polypeptide chain may exhibit no stable conformation and appear intrinsically disordered. Bioinformatics methods for predicting secondary structure and regions of disorder are reviewed in Chapter 2 by Majorek *et al.* The next level up is tertiary structure, which describes the three-dimensional arrangement of secondary structural elements and connecting regions within individual protein domains. Two different approaches for tertiary structure prediction, by 'template-based' and 'template-free' methods, have been reviewed in Chapter 4 by Kosinski *et al.* and in Chapter 5 by Gront and Kolinski, respectively. Methods for quality assessment of protein models (i.e. MQAPs) have been reviewed by Wallner and Elofsson in Chapter 6. Methods for predicting possible sites of interactions with other molecules and potential enzymatic functions from protein structure are reviewed in Chapter 7 by Kinoshita and coworkers, and Chapter 8 by Torrance and Thornton, respectively. Beyond the tertiary structure is the quaternary structure, which usually refers to the non-covalent association of two or more polypeptide chains into a protein complex. Here, this term will be also used to describe non-covalent interactions between domains within a single polypeptide. The so-called docking methods for prediction of the structure of protein-protein and protein-nucleic acid complexes from the structure of the isolated subunits are reviewed in Chapter 9 by de Vries and coworkers. The relationship of protein networks to protein function is reviewed in Chapter 10 by Nabieva and Singh, while Chapter 11 by Tress and coworkers integrates the whole spectrum of studies, from sequences to structures and networks, and places them in the context of systems biology.

The successful publication of this book would not have been possible without much appreciated support of many people. First, I owe an enormous debt to the various authors who contributed to this volume. Their hard work, dedication, expertise have made this book something we can all be very proud of. Thank you! Sincere thanks are extended to Paul Deards and Richard Davies, along with everyone else at Wiley who helped to manage this publication project. Thanks are also owed to members of my laboratory, both in Warsaw and in Poznań, for tolerating my prolonged absence during the major writing and editing phase. I must thank Ichizo Kobayashi and the University of Tokyo for hosting and supporting me while this work has been carried out, and all members of the Kobayashi laboratory for their hospitality (and nourishing me with sushi, sashimi, tempura, okonomiyaki, fugu, soba, udon, dried octopus, and yes of course, sake and shochu).

Janusz M. Bujnicki
Warsaw-Poznań-Tokyo
2007-08