Preface

Structural genomics is a technology- and methodology-driven field with the simple aim to determine many protein structures at a lower cost per structure than by using traditional approaches and to populate protein fold space to as high a degree of completeness as possible. The field emerged in the mid to late 1990s from distinct pilot projects primarily in Canada, Japan, and the U.S., and its inception was spurred by the abundance of genomic information and the methodological/technological advances made in protein expression, computational methods, biomolecular NMR and X-ray crystallography. The various pilot projects had shared technological goals, but the field struggled to find a common scientific objective in large part because the various pilot projects were launched with different objectives in mind and to meet the goals of different funding agencies.

Almost ten years later, structural genomics still has, at its core, technology development but has also "crystallized" around two coherent scientific themes — discovery and applied structural genomics. The technological development effort (Chapter 2, Chapter 3, Chapter 4, and Chapter 5) is viewed with great enthusiasm by the general scientific community because of the interest in adapting structural genomics-derived technology for their own scientific projects. Since the late 1990s perhaps the most significant improvements have been made in the analysis of X-ray diffraction data; the integrated packages that identify heavy atom sites, phase X-ray diffraction data and build the protein chain are truly remarkable.

The scientific objectives of discovery structural genomics are akin to those of traditional structural biology, which seek to achieve greater understanding of well-studied biological systems. In fact, discovery structural genomics is targeted to study genes/proteins of unknown structure in order to discover new information about folding, specific structural features, or function. In this guise, discovery structural genomics gives structural biologists a chance, for the first time, to generate hypotheses about new proteins, rather than to work on proteins preordained to be interesting.

The relative merits of discovery structural genomics and structural biology can be, and have been, debated. In some respects, the debate is slightly clouded by the fact that there is no commonly accepted definition of structural genomics or a consensus of its objectives. In this way, pundits have the opportunity to both define structural genomics and the reasons for its existence or demise. Taking our definition of discovery structural genomics, with tens of thousands of proteins of unknown structure and function to be analyzed, it is not unreasonable to suggest that in the long haul, discovery structural genomics may have a greater impact on our understanding of biology than will traditional structural biology. The parallel strategy adopted by structural genomics is already generating more structural information per unit time per person than other approaches. In addition, discovery structural genomics is less subject to the mercurial trends of biomedical research, which tend

to focus researchers into narrowly defined areas largely defined by funding possibilities. With these different views in mind, we have devoted five chapters (Chapter 1, Chapter 6, Chapter 7, Chapter 8, and Chapter 9) to review the premises and the promises of structural genomics from the discovery view and how the generated structural data is used in computational approaches.

Applied structural genomics (Chapter 12) describes the focused application of the structural genomics technologies to proteins of known function, but of interest to the biotechnology and pharmaceutical industry. This effort often targets families of proteins of known relevance to drug discovery (e.g., protein kinases) and also applies the structural genomics suite of technologies to generate structures of a given protein in complex with a number of different ligands such as chemical inhibitors.

Finally, with any field devoted to generating experimental information comes the problem of data visualization and analysis. The computational problems raised by structural genomics are not unique, and it could be argued that they have yet to emerge. Structural genomics efforts are contributing only 10% of the new structures to the PDB, hardly an unmanageable deluge. However, in the future, we can expect that this rate will increase and that an increasing number of structures will be deposited without companion biological experimentation. The analysis and interpretation of these new structures must be driven computationally; we have devoted two chapters (Chapter 10 and Chapter 11) to these impending problems.

Finally, it is always difficult to compile an up-to-date book in a technology-driven field. What we have attempted to do is to provide a snapshot of the present and some view to the future of the field. The contributors to this book are each at the forefront of their fields, and it will be fun to read this book again in five years to see how well they each prognosticated.

Aled Edwards

Martin Norin

Michael Sundström