Preface

A NATO Advanced Research Workshop (ARW) on *'Structural Biology and Functional Genomics'* was held from May 4th to 8th 1998 at the International Centre for Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy. 65 scientists from 17 countries attended (Bulgaria, Canada, Croatia, Czech Republic, France, Germany, Greece, Hungary, Israel, Italy, Mexico, Russia, Slovenia, Switzerland, Turkey, United Kingdom and the United States).

Biomedical research will be revolutionized by the current efforts to sequence the human genome and the genomes of model organisms. A notable finding from the already completed genome sequences of microbes and lower eukaryotes is that about 50% of all the genes code for proteins of unknown functions. A major challenge to the biomedical community will be to determine the functions of these unknown proteins and the roles they play in the multitude of life processes. This is the emerging field of functional genomics. The difficulty of assigning functions to unknown proteins cannot be overestimated. It will be necessary to identify all the components of conserved protein and protein/nucleic acid complexes. An unknown protein found in these complexes can then be labeled with a probable function. In parallel with functional genomics the objectives of structural genomics will be to determine by sychrotron radiation and multidimensional nuclear magnetic resonance spectroscopy high resolution structures of unknown proteins, of conserved protein complexes and of different classes of proteins e.g. all the proteins in a microbe, all the proteins involved in a biological pathway etc. This protein structure initiative will provide the structural basis for understanding biological function. Because of the very large numbers of unknown proteins high throughput capabilities will have to be developed for protein purification, characterization and crystallization. Fortunately the timely availability of highly collimated very bright synchrotron radiation beams allow the structures of proteins and their complexes even when in small crystals to be determined.

It is estimated that about 5% of sequences in mammalian genomes code for proteins. The remaining sequences have been called 'junk' DNA. This is an unfortunate terminology in that a multitude of non-coding sequences have important biological functions. These include telomeric and centromeric DNA sequences; trinucleotide repeats associated with degenerative diseases and other micro and mini-satellite sequences; sequences that position and organize nucleosomes and chromosomes. To understand the functions of non-coding DNA sequences it will be necessary to; i) solve their structures and determine their effects on the associated functions; ii) understand the physics and theory of DNA flexibility, bendability and supercoiling and iii) elucidate changes in DNA structure with all aspects of DNA processing such as replication, transcription, DNA repair and spermiogenesis. This will involve solving the structures of DNA/protein complexes isolated at different stages of these processes as well as studying the constituent molecules with techniques like combinatorial mutagenesis. The big problem is to put all the above approaches together to understand the organization, structures, and functions of chromosomes.

The articles in this special issue address many of the problems outlined above and in addition provide a 'snapshot' of the transition we are undergoing to genome wide approaches to understand biological functions.

Sándor Pongor
E. Morton Bradbury

Trieste, January 1999

Acknowledgements

The Organizers thank the help and advice of Professors Arturo Falaschi and Francisco E. Baralle as well as the financial and logistical support of the International Centre for Genetic Engineering and Biotechnology (ICGEB), Trieste. The manuscripts of this special issue were prepared for publication by Kristian Vlahovicek, ICGEB.