Foreword

Computational biology: More than just a set of techniques

The third Indo-French Bioinformatics meeting was held in the National Centre for Biological Sciences, Bangalore, in June 2005. The subject matters covered by the speakers in this meeting were quite diverse and represented the rapidly changing contemporary phase of biocomputing. The idea of devoting a special issue of *Journal of Biosciences* to the evolving field of computational biology was an offshoot of this meeting.

A number of different terms, such as "computational biology", "bioinformatics" and "*in silico* biology" are commonly used to refer to applications of computational and mathematical approaches in biology. While one can debate about the meanings of these terms, we have adopted the most general definition of Computational Biology to include the use of computational techniques, tools, software, databases, data analyses, modelling and simulations to address biological problems. The contents of this Special Issue cover different aspects of complexity in biological systems and consider proteins, genes, interactions, pathways, processes and cells. One could have extended the range to include tissues, organelles, organs, organisms, populations and ecology.

Computational biology is an approach based on quantitative and computational methods. It is not a set of techniques similar to molecular cloning, the polymerase chain reaction (PCR) or spectroscopy. Its importance can be gauged from the fact that the computational point of view is turning out to provide insights in diverse areas such as neurobiology, immunology, infectious disease, genetics, cellular signal transduction, cancer and developmental biology. For example, it is common to use bioinformatics tools to predict antigenic determinant sites or to learn the properties of a critical protein involved in developmental processes using three-dimensional modelling. However, the same approach cannot be used for solving different problems. An understanding of the question is essential for applying the right tools. The ready availability of many computer-based programs and databases in the World-Wide Web is a boon, but the possibility of wrong interpretation must always be kept in mind.

Another facet of computational biology is its ability to generate new hypotheses, theories and fresh understanding of fundamentals – features that support the viewpoint of computational biology as a branch of biology. This is particularly relevant in the post-genomic era. Organism-wide information (pertaining, among other things, to genomic sequences, microarray, mass spectrometry and yeast two-hybrid data, expression profiles and 3-D structures) is getting accumulated at an unprecedented speed. Also, the more conventional experimental approaches in biology continue to contribute an enormous amount of data. Integration of information obviously adds value to the individual datasets; it enables generation of new hypotheses that may be probed using experiments. An organized and systematic integration of published information adds to the generation of new theories and models. Indeed, the difference between publications in biology and biological databases is blurring with time. Simulation is another kind of computational biology that supports the viewpoint that computer-based methods should be thought of as a novel way of doing biology. Simulation of signalling pathways, biological processes, structures of large multi-domain, multi-component molecular machines and molecular dynamics generates hypothesis which may take a long time to validate experimentally.

In this Special Issue, the papers encompass both the technical and the scientific components of computational biology. We have tried to arrange the articles so as to reflect an increasing level of complexity, beginning with genes and genomes and going on to cells. Microsatellites with sequence repeats commonly occur in genomic DNA of many organisms including mycobacteria. Sreenu *et al* note

interesting features of distribution of sequence repeats and occurrence of long and short microsatellites. Based on a phylogenetic analysis, Usha and coworkers analyse regions of recombination in the Rep gene and its implications for Gemini virus infections. Basin and Raghava have combined artificial neural networks and quantitative matrix methods effectively to predict MHC class I restricted T-cell epitopes. Genome-wide analyses of repertoires of proteins provide insights into the evolution of structures and functions. Metpally *et al* present an inventory of nuclear hormone receptors encoded in the genome of *Tetraodon nigroviridis* that are detected using homology-detection methods. The prediction of three-dimensional structures can aid our understanding of how proteins work. While the code that relates amino acid sequence and 3-D structures still remains a Holy Grail, attempts are on to discover it. de Brevern and coworkers and Jayaram and coworkers use simplified viewpoints for predicting folding patterns in proteins. Balaji and Srinivasan argue that the traditional approaches to model the evolution of proteins fail when the evolutionary divergence between homologues is high and they demonstrate the use of 3-D structures, instead, in modelling the evolution. Analysis of interactions between proteins could help obtain insights on the structures of protein assemblies. Computational approaches are particularly valuable since determination of 3-D structures of many biologically important complexes using experimental approaches is notoriously difficult. Sasin *et al* provide a convenient framework to analyse protein surfaces while Pal *et al* report features of the structural roles of segments that make protein-protein interaction interfaces.

Modelling and simulation of biological pathways and processes contribute richly in understanding signal transduction and other cellular events. Miyano and coworkers propose elegant steps in modeling an apoptosis pathway. Stephan *et al* analyse neural interactions between different brain regions – an area of high interest in the field of neuro-imaging as well as electrophysiology. Molina and coworkers use operations involving annotations in adding value to the knowledge of the TCA cycle. Rangamani and Iyengar describe strategies in modelling cellular signalling processes. Finally, Wingender and coworkers outline the steps required for extending computational approaches to study higher levels of biological complexity.

We thank all the authors who contributed articles in this special issue and all reviewers for spending their valuable time to assess the manuscripts.

N Srinivasan[1], R Sowdhamini[2] and Alok Bhattacharya[3]
*[1]Molecular Biophysics Unit, Indian Institute of Science,
Bangalore 560 012, India*

*[2]National Centre for Biological Sciences, Tata Institute of Fundamental Research,
UAS-GKVK Campus, Bellary Road, Bangalore 560 065, India*

*[3]School of Life Sciences, Jawaharlal Nehru University,
New Mehrauli Road, New Delhi 110 067, India*