Preface

Bioinformatics involve the creation and advancement of algorithms using techniques including computational intelligence, applied mathematics and statistics, informatics, and biochemistry to solve biological problems usually on the molecular level. Major research efforts in the field include sequence analysis, gene finding, genome annotation, protein structure alignment analysis and prediction, prediction of gene expression, protein-protein docking/interactions, and the modeling of evolution.

Computational intelligence is a well-established paradigm, where new theories with a sound biological understanding have been evolving. Defining computational intelligence is not an easy task. In a nutshell, which becomes quite apparent in light of the current research pursuits, the area is heterogeneous with a combination of such technologies as neural networks, fuzzy systems, rough set, evolutionary computation, swarm intelligence, probabilistic reasoning, multi-agent systems etc. The recent trend is to integrate different components to take advantage of complementary features and to develop a synergistic system.

This book deals with the application of computational intelligence in bioinformatics. Addressing the various issues of bioinforatics using different computational intelligence approaches is the novelty of this edited volume. This volume comprises of 13 chapters including some introductory chapters giving the fundamental definitions and some important research challenges. Chapters were selected on the basis of fundamental ideas/concepts rather than the thoroughness of techniques deployed. The thirteen chapters are organized as follows.

In the introductory Chapter, *Tasoulis et al.* present neural networks, evolutionary algorithms and clustering algorithms and their application to DNA microarray experimental data analysis. Authors also discus different dimension reduction techniques.

Chapter 2 by *Kaderali and Radde* provide an overview of the reconstruction of gene regulatory networks from gene expression measurements. Authors present several different approaches to gene regulatory network inference, discuss their strengths and weaknesses, and provide guidelines on which models are appropriate under what circumstances.

Chapter 3 by *Donkers and Tuyls* introduce Bayesian belief networks and describe their current use within bioinformatics. The goal of the chapter is to help the reader to understand and apply belief networks in the domain of bioinformatics. Authors present the current state-of-the-art by discussing several real-world applications in bioinformatics, and also discuss some available software tools.

Das et al. in Chapter 4 explore the role of swarm intelligence algorithms in certain bioinformatics tasks like micro-array data clustering, multiple sequence alignment, protein structure prediction and molecular docking. This chapter begins with an overview of the basic concepts of bioinformatics along with their biological basis and then provides a detailed survey of the state of the art research centered around the applications of swarm intelligence algorithms in bioinformatics.

Liang and Kelemen in the fifth Chapter propose a time lagged recurrent neural network with trajectory learning for identifying and classifying the gene functional patterns from the heterogeneous nonlinear time series microarray experiments. Optimal network architectures with different memory structures were selected based on Akaike and Bayesian information criteria using two-way factorial design. The optimal model performance was compared to other popular gene classification algorithms, such as nearest neighbor, support vector machine, and self-organized map.

In Chapter 6, *Busa-Fekete et al.* suggest two algorithms for protein sequence classification that are based on a weighted binary tree representation of protein similarity data. *TreeInsert* assigns the class label to the query by determining a minimum cost necessary to insert the query in the (precomputed) trees representing the various classes. Then *TreNN* assigns the label to the query based on an analysis of the query's neighborhood within a binary tree containing members of the known classes. The algorithms were tested in combination with various sequence similarity scoring methods using a large number of classification tasks representing various degrees of difficulty.

In Chapter 7, *Smith* compares the traditional dynamic programming RNA gene finding methodolgy with an alternative evolutionary computation approach. Experiment results indicate that dynamic programming returns an exact score at the cost of very large computational resource usage, while the evolutionary computing approach allows for faster approximate search, but uses the RNA secondary structure information in the covariance model from the start.

Schaefer et al. in Chapter 8, illustrate how fuzzy rule-based classification can be applied successfully to analyze gene expression data. The generated classifier consists of an ensemble of fuzzy *if-then* rules, which together provide a reliable and accurate classification of the underlying data.

In Chapter 9, *Huang* and *Chow* overview the existing gene selection approaches and summarize the main challenges of gene selection. Using a typical gene selection model, authors further illustrate the implementation of these strategies and evaluate their contributions.

Cao et al. in Chapter 10, propose a fuzzy logic based novel gene regulatory network. The key motivation for this algorithm is that genes with regulatory relationships may be modeled via fuzzy logic, and the strength of regulations may be represented as the length of accumulated distance during a period of time intervals.

One unique feature of this algorithm is that it makes very limited a priori assumptions concerning the modeling.

In Chapter 11, *Haavisto and Hyötyniemi* apply linear multivariate regression tools for microarray gene expression data. Two examples comprising of yeast cell response to environmental changes and expression during the cell cycle, are used to demonstrate the presented subspace identification method for data-based modeling of genome dynamics.

Navas-Delgado et al. in Chapter 12 present an architecture for the development of Semantic Web applications, and the way it is applied to build an application for systems biology. The architecture is based on an ontology-based system with connected biomodules that could be globally analyzed as far as possible.

In the last Chapter *Han and Zhu* illustrate the various methods of encoding information in DNA strands and present the corresponding DNA algorithms, which will benefit the further research on DNA computing.

We are very much grateful to the authors of this volume and to the reviewers for their tremendous service by critically reviewing the chapters. The editors would like to thank Dr. Thomas Ditzinger (Springer Engineering Inhouse Editor) and Professor Janusz Kacprzyk (Editor-in-Chief, Springer Studies in Computational Intelligence Series) and Ms. Heather King (Springer Verlag, Heidelberg) for the editorial assistance and excellent cooperative collaboration to produce this important scientific work. We hope that the reader will share our excitement to present this volume on '*Computational Intelligence in Bioinformatics*' and will find it useful.

Arpad Kelemen, Ajith Abraham and Yuehui Chen (Editors)
September 2007