been successfully used not only for finding the sub-optimal alignments but also for locating biologically important alignments.

When we use HMMs for biological sequences, it is a difficult problem to determine the network shapes of the HMMs. Yada et al. used a Genetic Algorithm (GA) on connection matrices and initial parameters for building the optimal network shapes of HMMs, which model and predict the promoters and other signals in DNA. Two related works are presented in the poster session. Tanaka used the Successive State Splitting (SSS) algorithm for the modeling of the amino acid sequence of helices. Fujiwara et al., used modified Iterative Duplication (ID) method and used it for extracting motifs in proteins.

Motif discovery from amino acid sequences is a classical yet important problem in genome informatics in that it provides biologically important knowledge. The paper by Tateishi and Miyano presents a greedy algorithm for finding such motifs with ambiguity just from positive and negative examples, whose basic idea is based on the probabilistic argument for designing approximation algorithms for the maximum satisfiability problem.

Prediction of secondary structure of RNA has been a frequent target of stochastic models. Among them, Stochastic Context Free Grammars (SCFG) have been successfully used. However, the structures which are called psuedoknots cannot be modeled by SCFG because of the features of the nesting of the structures in psuedoknots. Brown and Wilson used intersections of SCFG to approximate the full parsing scores of psuedoknots. More complicated models can express the psuedoknots exactly, but the associated computational cost of parsing is high. In order to predict the structures and the functions of DNA, RNA and protein, the evolutionary information is very important. Gulko and Haussler utilized the evolutionary information for predicting the secondary structure of RNA by considering multiple alignments and phylogenetic trees.

Recently a very interesting computation paradigm has taken place under the name of DNA computing which was initiated by Adleman's remarkable laboratory solution of a small NP-complete problem using DNA molecules in vitro (1994). In parallel to this rapid movement, intensive study on 'theoretical DNA computing' has emerged, based on Head's work (1987) on splicing systems (H systems). This track contains two papers, one by Ferretti and Kobayashi and one by Csuhaj-Varju et al., that belong to the latter category and are devoted to the theoretical study of the universal computability of the extended H systems. These two papers proceed from different models of computation. One is based on Post Normal Systems, while the other is based on the Turing machine/Type-0 grammar approach to computation. The former presents a more simplified construction for the universal computability result, while the latter also discusses variants of the original extended H systems.

Stochastic Models, Formal Systems and Algorithmic Discovery for Genome Informatics

Kiyoshi Asai
Electrotechnical Laboratory (ETL)
1-1-4 Umezono, Tsukuba, Ibaraki 305, JAPAN
asai@etl.go.jp

Tom Head
Dept. of Mathematical Sciences, Binghamto University,
Binghamton, New York 13902-6000, USA
tom@math.binghamton.edu

Katsumi Nitta
Electrotechnical Laboratory (ETL)
1-1-4 Umezono, Tsukuba, Ibaraki 305, JAPAN
nitta@etl.go.jp

Takashi Yokomori
Dept. of Computer Science and Information Mathematics,
The University of Electro-Communications
1-5-1, Chofugaoka, Chofu, Tokyo 182, JAPAN
yokomori@cs.uec.ac.jp

One of the most important targets of computational biology, and of molecular biology itself, is to understand the meaning of the genetic sequences. Because the genetic sequences are the result of evolution, it is quite natural to use stochastic methods for this purpose. At the same time, because genetic sequences have certain rules, typically the base pairing in DNA and RNA, linguistic methods and formal systems are useful. This track, "Stochastic Models, Formal Systems and Algorithmic Discovery," puts its focus on these methods and other algorithmic approaches to the discovery of the meaning of the genetic sequences.

Building the maps of the chromosome from experimental data is one of the very important areas of computational biology. Bhandarkar and Chirravuri used the parallel algorithm of simulated annealing for clone ordering.

Multiple sequence alignment is a very important technique and has been heavily used all over the field of genetic sequence analysis. However, direct calculation of multiple sequence alignment is computationally expensive. Horton's paper provides a branch and bound algorithm which is useful for local multiple alignment and is globally optimal.

Hidden Markov Models (HMMs), a subclass of the stochastic models, have