SUMMARY OF Ph.D. DISSERTATION

School School of Fundamental Science and Technology	Student Identification Number	SURNAME, First name NAGAMINE, Nobuyoshi

Title

Genome-wide statistical prediction of interactions between biomolecules

Abstract

The fact that the genomes of more than 800 species have been completely decoded demonstrates that many data have been recently produced to elucidate the life. Along with the genome or the `blueprint' of life, the molecular biology has accumulated the information of `parts' constituting the life. However, even though with abundant information on `blueprint' and `parts', integration of these to design the life are still hard. Therefore, this study focused on prediction of interactions between biomolecules in attempt to contribute to overcoming these difficulties.

The biomolecules, including proteins and metabolites, constitute networks and systems to realize biological functions. Predicting interactions, which can be regarded as one of the minimal units of these systems, between them can contribute to the elucidation of biological mechanisms. In particular, the prediction of interaction between proteins relevant to diseases and small molecules can be of help in searching lead compounds in drug discovery and identifying unknown effects and side effects of known drugs, and can be of economic and industrial significance.

In Part 1 of this thesis, the significance of comprehensively predicting interactions between biomolecules in elucidating the biological system was explained.

In Part 2, the protein-protein interaction (PPI) network was utilized to identify cooperative elements in the transcription regulation network. In the computational experiment based on Chromatin Immuno-Precipitation (ChIP) and PPI data of yeast (*S. cerevisiae*), it was discovered that the transcription factors regulating proteins that were located close to one another in the PPI network tended to work cooperatively. This study also revealed meaningful relations between two different biological networks.

In Part 3, the prediction of interactions between proteins and chemical compounds was studied. In the computational experiment using the statistical learning method Support Vector Machine (SVM), relatively accurate prediction of interactions between approved drugs and their target proteins was achieved only by using easily available chemical structure and mass spectrometry data, and amino acid sequence data,. Moreover, in the comprehensive binding ligand prediction and the experimental verification with human androgen receptor, integration of these enabled more effective and efficient interaction predictions.

In Part 4, this study was summarized and future works were discussed.