

Chapter 16

Domain-Based Approaches to Prediction and Analysis of Protein-Protein Interactions

Morihiro Hayashida
Kyoto University, Japan

Tatsuya Akutsu
Kyoto University, Japan

ABSTRACT

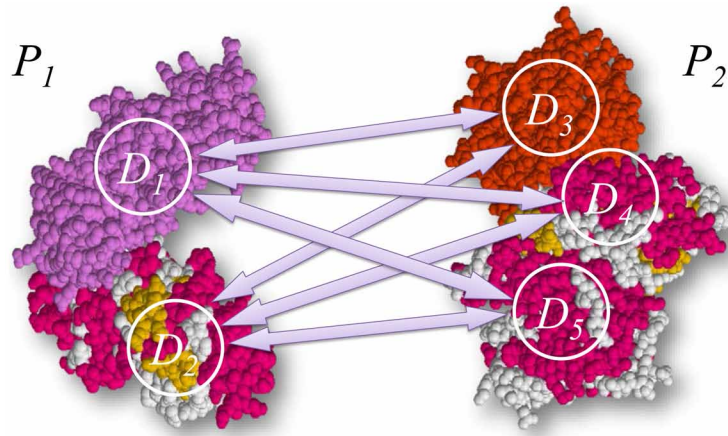
Protein-protein interactions play various essential roles in cellular systems. Many methods have been developed for inference of protein-protein interactions from protein sequence data. In this paper, the authors focus on methods based on domain-domain interactions, where a domain is defined as a region within a protein that either performs a specific function or constitutes a stable structural unit. In these methods, the probabilities of domain-domain interactions are inferred from known protein-protein interaction data and protein domain data, and then prediction of interactions is performed based on these probabilities and contents of domains of given proteins. This paper overviews several fundamental methods, which include association method, expectation maximization-based method, support vector machine-based method, linear programming-based method, and conditional random field-based method. This paper also reviews a simple evolutionary model of protein domains, which yields a scale-free distribution of protein domains. By combining with a domain-based protein interaction model, a scale-free distribution of protein-protein interaction networks is also derived.

INTRODUCTION

Understanding of functions of genes and proteins is important in post-genomic era. Information on protein-protein interactions is useful for understanding protein functions because protein-protein interactions play a key role in many cellular processes. Since the end of the last century, some experimental techniques have been developed for comprehensive analysis of protein-protein interactions, which include two-hybrid systems and proteomics methods. Though these experimental methods revealed many

DOI: 10.4018/978-1-5225-8903-7.ch016

Figure 1. Example of protein domains. Protein P_1 consists of domains D_1 and D_2 , whereas protein P_2 consists of domains D_3 , D_4 and D_5 . In domain-based models, it is assumed that P_1 and P_2 interact with each other if at least one domain pair interacts.



unknown interactions, there were large gaps between results done by different groups (Ito et al., 2001; Uetz et al., 2000). Therefore, computational methods should be developed for inference of protein-protein interactions. For that purpose, various approaches have been proposed. In this paper, we focus on computational and mathematical aspects of domain-based approaches.

A protein consists of one or multiple domains, where a *domain* is defined as a region within a protein that either performs a specific function or constitutes a stable structural unit. Examples of structural domains are illustrated in Figure 1 though domains are sometimes defined not based on structures but based on sequence/functional similarities. In a word, domains are considered as parts of a protein. Though there is no exact or mathematical definition of protein domains, several hundreds of protein domains are currently known. In order to classify domains, several database systems have been constructed, which include Pfam (Punta et al., 2012), InterPro (Hunter et al., 2012) and ProDom (Bru et al., 2005). Furthermore, most of these databases provide facilities to identify protein domains from a given protein sequence. In Pfam, each domain is represented by an HMM (Hidden-Markov Model) and protein domains contained in a given protein sequence are identified by using these HMMs.

Utilizing information of domain organizations of proteins, several methods have been proposed for prediction of protein-protein interactions. In these methods, scores or probabilities of *domain-domain interactions* are first derived from known protein-protein interactions and then these are utilized for calculating the score or probability of protein-protein interaction for given protein sequences. Sprinzak and Margalit (2001) proposed the *association method* for computing the score of each domain pair. Kim et al. (2002) proposed similar scores and applied the scores to inference of protein-protein interactions. Deng et al. (2002) proposed an *EM (Expectation-Maximization) algorithm* for estimating the probability of interaction for each domain pair.

We proposed a *conditional random field* approach using mutual information between domain sequences (Hayashida et al., 2011). In the study, mutual information between positions of amino acid residues in domains is calculated from multiple sequence alignments because it is known that interacting residue pairs tend to have higher mutual information than non-interacting ones (Fraser et al., 2002). In addition,

20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/domain-based-approaches-to-prediction-and-analysis-of-protein-protein-interactions/228632

Related Content

An Eco-Friendly Approach for the Eradication of Heavy Metal Contaminants by Nano-Bioremediation

Chandana Mohanty, Sneha Shriparna Satpathy and Sweta Mohanty (2021). *Recent Advancements in Bioremediation of Metal Contaminants* (pp. 220-236).

www.irma-international.org/chapter/an-eco-friendly-approach-for-the-eradication-of-heavy-metal-contaminants-by-nano-bioremediation/259574

Internet of Things and Nature-Inspired Intelligent Techniques for the Future of Biomedical Engineering

Gur Emre Guraksin (2019). *Biotechnology: Concepts, Methodologies, Tools, and Applications* (pp. 543-561).

www.irma-international.org/chapter/internet-of-things-and-nature-inspired-intelligent-techniques-for-the-future-of-biomedical-engineering/228638

Bacterial Remediation of Phenolic Compounds

Veena Gayathri Krishnaswamy (2019). *Biotechnology: Concepts, Methodologies, Tools, and Applications* (pp. 1910-1943).

www.irma-international.org/chapter/bacterial-remediation-of-phenolic-compounds/228698

Nature: The Design Mentor

(2021). *Inspiration and Design for Bio-Inspired Surfaces in Tribology: Emerging Research and Opportunities* (pp. 188-222).

www.irma-international.org/chapter/nature/257601

Atmospheric Pressure Plasma Systems and Applications

Necdet Aslan (2022). *Emerging Developments and Applications of Low Temperature Plasma* (pp. 49-62).

www.irma-international.org/chapter/atmospheric-pressure-plasma-systems-and-applications/294710