

# Prediction of Protein Functions Based on K-Cores of Protein-Protein Interaction Networks and Amino Acid Sequences

|                                                                      |                                                                      |                                                                  |
|----------------------------------------------------------------------|----------------------------------------------------------------------|------------------------------------------------------------------|
| <b>Md. Altaf-Ul-Amin</b> <sup>1,2</sup><br>amin-m@is.aist-nara.ac.jp | <b>Kensaku Nishikata</b> <sup>1</sup><br>k-nishik@is.aist-nara.ac.jp | <b>Toshihiro Koma</b> <sup>1</sup><br>t-koma@is.aist-nara.ac.jp  |
| <b>Teppei Miyasato</b> <sup>1</sup><br>teppe-m@is.aist-nara.ac.jp    | <b>Yoko Shinbo</b> <sup>1</sup><br>y-shinbo@is.aist-nara.ac.jp       | <b>Md. Arifuzzaman</b> <sup>2</sup><br>arif-z@bs.aist-nara.ac.jp |
| <b>Chieko Wada</b> <sup>3</sup><br>cwada@lif.kyoto-u.ac.jp           | <b>Maki Maeda</b> <sup>3</sup><br>mmaeda@gtc.aist-nara.ac.jp         | <b>Taku Oshima</b> <sup>1</sup><br>taku@bs.aist-nara.ac.jp       |
| <b>Hirotsada Mori</b> <sup>1,4</sup><br>hmori@gtc.aist-nara.ac.jp    | <b>Shigehiko Kanaya</b> <sup>1</sup><br>skanaya@gtc.aist-nara.ac.jp  |                                                                  |

<sup>1</sup> Nara Institute of Science and Technology(NAIST), 8916-5 Takayama, Ikoma, Nara 630-0101, Japan

<sup>2</sup> NEDO(New Energy and Industrial Technology Development Organization), Toshima, Tokyo 170-6028, Japan

<sup>3</sup> CREST JST(Japan Science and Technology Corp.), Kawaguchi, Saitama 332-0012, Japan

<sup>4</sup> Inst. Adv. Bioscience, Keio University, Baba-machi, Tsuruoka, Yamagata 997-0035, Japan

**Keywords:** protein-protein interaction, k-core subgraph, protein function prediction

## 1 Introduction

Comprehensive analysis of protein-protein interactions and amino acid sequences plays important roles to understand protein functions in molecular level. A k-core of a network or a graph is a subgraph in which all nodes are connected to at least k other nodes in the subgraph. In a protein-protein interaction network, a node represents a protein and an edge represents an interaction between proteins. A k-core of a protein-protein interaction network usually contains cohesive groups of proteins. On the other hand, phylogenetic analysis classifies proteins into groups based on the similarity of amino acid sequences. In this work, we use k-cores of protein-protein interaction networks and phylogenetic analysis to predict the functions of some function-unknown proteins of *E.coli*.

## 2 Method

Let, a graph  $G = (V, E)$  consists of a finite set of nodes  $V$  and a finite set of edges  $E$ . A subgraph  $S = (V', E')$  where  $V' \in V$  and  $E' \in E$  is a k-core or a core of order k of  $G$  if and only if  $\forall v \in V' : deg(v) \geq k$  within  $S$  and  $S$  is the maximal subgraph of this property [1]. The core of maximum order is referred to as the main core or the highest k-core of the graph. A k-core subgraph of a graph can be generated by recursively deleting the vertices from the graph whose degree is less than k. In this work, we apply the notion of k-core subgraph to predict protein functions.

