ProCope—protein complex prediction and evaluation

Jan Krumsiek, Caroline C. Friedel* and Ralf Zimmer

Institut für Informatik, Ludwig-Maximilians-Universität München, Amalienstraße 17, 80333 München, Germany

Received on May 27, 2008; revised on July 7, 2008; accepted on July 15, 2008

ABSTRACT

Summary: Recent advances in high-throughput technology have increased the quantity of available data on protein complexes and stimulated the development of many new prediction methods. In this article, we present ProCope, a Java software suite for the prediction and evaluation of protein complexes from affinity purification experiments which integrates the major methods for calculating interaction scores and predicting protein complexes published over the last years. Methods can be accessed via a graphical user interface, command line tools and a Java API. Using ProCope, existing algorithms can be applied quickly and reproducibly on new experimental results, individual steps of the different algorithms can be combined in new and innovative ways and new methods can be implemented and integrated in the existing prediction framework.

Availability: Source code and executables are available at http://www.bio.ifi.lmu.de/Complexes/ProCope/

Contact: Caroline.Friedel@bio.ifi.lmu.de

1 INTRODUCTION

Protein complexes as modular components of the cells carry out many important molecular functions. Thus, the identification and characterization of protein complexes is a key to understanding cellular processes. Since the publication of two genome-wide tandem affinity purification (TAP) assays for yeast (Gavin et al., 2006; Krogan et al., 2006) many efforts have been put into the prediction of the actual protein complexes from these experimental datasets (Collins et al., 2007; Friedel et al., 2008; Gavin et al., 2006; Hart et al., 2007; Krogan et al., 2006; Pu et al., 2007; Zhang et al., 2008). Generally, these methods first calculate scoring networks for protein interactions from the purification results which are then clustered to derive protein complexes.

Here we present ProCope, an extensible software package written in Java for the prediction and evaluation of protein complexes from purification datasets which integrates efficient implementations of the major prediction methods. The ProCope package provides a convenient graphical user interface (GUI), command line tools suitable for batch job processing and a Java application programming interface (API). The Java API makes it possible to use the functionality of ProCope within other software applications or to extend the user interfaces with new prediction algorithms. Thus, ProCope is a useful tool for researchers both for applying published methods on new datasets to obtain reproducible and high-quality predictions and for developing and evaluating new prediction methods.

2 IMPLEMENTATION

2.1 Methods

The general procedure for protein complex prediction and evaluation used within ProCope is outlined in Figure 1. ProCope provides implementations for the following scoring methods: Socio-affinity scores (Gavin et al., 2006), Purification Enrichment scores (Collins et al., 2007), a scoring scheme based on the hypergeometrical distribution (Hart et al., 2007), Dice coefficients (Zhang et al., 2008) and bootstrap confidence scores (Friedel et al., 2008).

For the clustering of PPI networks, ProCope provides access to the Markov Clustering algorithm (van Dongen, 2000) and implements several variants of hierarchical agglomerative clustering (Murtagh, 1984). Furthermore, two methods have been implemented for identifying proteins involved in more than one complex (Friedel et al., 2008; Pu et al., 2007).

The accuracy of scoring networks in predicting interactions within reference complexes can be assessed using receiver operating characteristic (ROC) curves. Predicted complexes can be evaluated by calculating sensitivity and positive predictive value compared to reference complexes (Brohée and van Helden, 2006) or functional similarity and co-localization within protein complexes. Functional similarity is evaluated in terms of semantic similarity (Schlicker et al., 2006) of the GO annotations within the complex. Several alternative definitions of semantic similarity are implemented. Furthermore, ProCope provides two different co-localization measures based on experimental data (Friedel et al., 2008; Pu et al., 2007).

All methods have been tested extensively against the results of the original publications. ProCope has been designed to be easily extensible as well as highly efficient to allow for higher order algorithms which require many repeated calculations such as the bootstrap algorithm we presented recently (Friedel et al., 2008).

2.2 User interfaces and Java API

ProCope provides an intuitive, easy to use GUI (see Figure 2) which can also be used as a Cytoscape plugin (Shannon et al., 2003). The user can quickly load purification data, calculate and cluster scoring networks, load evaluation data, compare complex sets, apply cut-offs on score networks or complex sets and much more. A Java Webstart version of the GUI is also available to start the program directly from within the webbrowser without the need for installing

*To whom correspondence should be addressed.
Assessed without delay. Moreover, new powerful approaches can appropriate classes and the performance of the methods can be predictions methods can be implemented quickly by extending provided by ProCope. New interaction score definitions or complex methods can rely on the infrastructure and evaluation methods extended using the Java API. Researchers developing new prediction documentation).

from the scoring networks and evaluate the results (see the ProCope the package. In addition to the GUI, a set of command line tools are provided for calculating interaction scores, clustering and complex evaluation. With these tools, different calculation processes can be easily integrated in scripts and batch jobs and distributed on computing clusters. All functionalities of ProCope can be accessed via a well-documented Java API and integrated easily into new software programs. Furthermore, the user interfaces of ProCope offer plugin functionalities to extend them by custom score calculation and clustering methods. Detailed examples on using the GUI, command-line tools and the Java API including example source codes are provided with the ProCope release.

3 RESULTS AND DISCUSSION
ProCope is an extensible Java suite for the prediction and evaluation of protein complexes from affinity purification assays which implements state-of-the-art methods published in the field over the last years. Its potentials can be illustrated with two examples. First, using the GUI, biologists can rapidly analyse data from new affinity purification experiments using previously published methods or a combination of those. Thus, time-consuming and error-prone reimplementations are avoided. Only few steps are necessary to load the purification data, calculate reaction scores, predict complexes from the scoring networks and evaluate the results (see the ProCope documentation).

Second, the existing functionalities of ProCope can be easily extended using the Java API. Researchers developing new prediction methods can rely on the infrastructure and evaluation methods provided by ProCope. New interaction score definitions or complex predictions methods can be implemented quickly by extending appropriate classes and the performance of the methods can be assessed without delay. Moreover, new powerful approaches can be made available to the research community immediately using the plugin option of the user interfaces. Using ProCope, we have developed an unsupervised bootstrap approach which does not require additional training data apart from the purification experiments (Friedel et al., 2008). Bootstrap confidence scores have been shown to be more accurate than previously published scoring methods and predicted complexes are of equivalent quality to the best published predictions.

Because of the easy access to its methods and the efficiency and extensibility of its algorithms, the ProCope package will be a valuable tool for researchers seeking to apply existing algorithms to new data or developing new and innovative prediction methods.

Requirements
ProCope requires the Java 5.0 (or higher) Runtime Environment, which is freely available at http://www.java.com.

Conflict of Interest: none declared.

REFERENCES