John von Neumann Institute for Computing

Protein-Protein Interaction Prediction

F. Fink, St. Ederer, W. Gronwald

published in

From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 209-212, 2008.

http://www.fz-juelich.de/nic-series/volume40

Protein-Protein Interaction Prediction

Florian Fink, Stephan Ederer, and Wolfram Gronwald

Institute of Functional Genomics, University of Regensburg, Germany
E-mail: {florian.fink, wolfram.gronwald}@klinik.uni-regensburg.de

Based on amino acid based pair-potentials and intermolecular energies we calculate score distributions for protein-protein complexes that exist in nature and those which do not exist. The distributions of the two groups are then found to be different in maximum and in shape. This opens the possibility to discriminate between complexes that exist and those which do not.

1 Introduction

1.1 From Protein Structure to Complexes

Proteins are an integral component for most of the mechanisms taking part in the cell. One important aspect in the research on proteins is their three-dimensional structure. The most common methods to determine the structure are X-ray cristallography and NMR spectroscopy, and due to them the number of known protein structures is actually rapidly growing. However, cellular functions are rarely carried out by single proteins but by complexes of several interacting proteins. High-throughput methods for detecting protein interactions, like yeast2hybrid, produce a huge number of such expected protein-protein interactions. Unfortunately it is not possible to determine the structures for all of them by experimental methods because there are limitations concerning large or transient complexes. In addition, if possible, the experimental structure determination of complexes is a very time-consuming and challanging process. For that reason computationally approches such as docking algorithms to predict the structure of protein-protein complexes are needed.

1.2 Docking

The hypothesis underlying docking predictions is that the native complex structure is the state with the lowest free energy accessible to the system. There are quite different approches on how to develop docking algorithms but the common, basic idea is to first do a sampling step followed by a scoring step. Scoring means, to analyse the putative complex structures generated in the first step with regard to chemical and physical aspects. Selecting suitable aspects and weighting them in an appropriate way is one of the great challenges in docking. The aim is to rank all putative structures in a way that most of the native-like structures are found in the top part of the ranked output.

2 Motivation

Since protein complexes play a major roll in cellular processes and experimental methods like yeast2hybrid are not always applicable and often contain a considerable number

of false positives[2], there is a need for computational methods predicting protein-protein interactions.

On the other hand, methods providing the three-dimensional structure of known protein complexes (docking algorithms) are already available. In their scoring step a great amount of different possible complex structures of the same two proteins is compared to choose those that are near-native. If this is possible, it must even be possible to do this analysis on complex structures of different protein pairs and by this get information on the probability that two specific proteins do interact at all. That means in other words to do docking with different proteins, even those that do not interact or are not known to do so and finally, after the interpretation of the structures, get as a result whether two proteins are suggested to built complexes in nature or not.

This is actually a computational method to predict protein-protein interaction.

3 Method and Results

3.1 Overview

For becoming able to predict protein-protein interaction what we actually need is a method that discriminates between complexes that exist in nature (native complexes) and those that do not (false complexes). This difference is mesured by (upp to now) three scoring functions (amino acid based pair-potentials, van der Waals energy and electrostatic energy), and becomes apparent in different score distributions for native and false complexes (see figure 1).

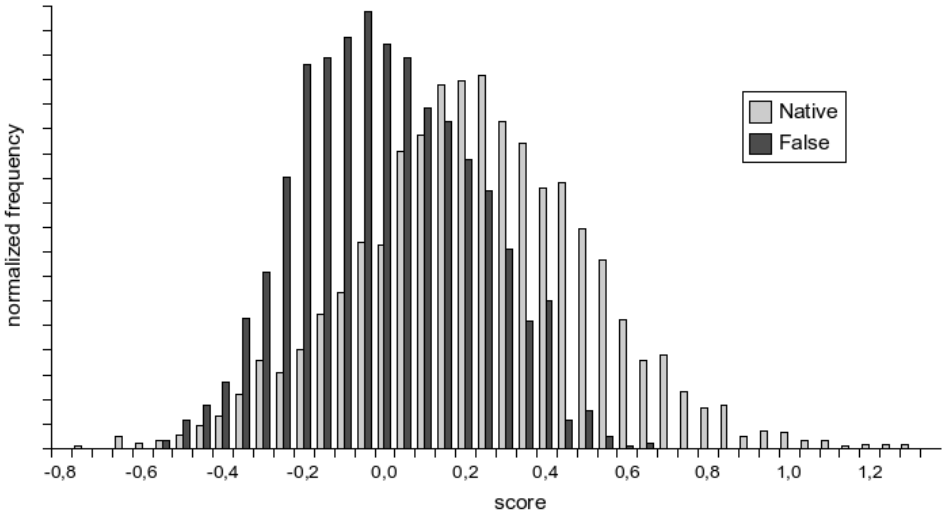

Figure 1. Score distributions from amino acid based pair-potentials.

3.2 Getting Native and False Complexes

Native structures can easily be obtained from the Nussinov database[4] that contains over 2000 non-homologous protein-protein complexes, whereas there does not exist a database for false complexes. For this reason we produced them on our own by somehow abusing the docking program HADDOCK[1]: We docked proteins that are not supposed to build complexes. This ensures that even the false complexes are in the best possible confirmation and hence really comparable to the native ones.

3.3 Scoring Functions

We use **amino acid based pair-potentials** that were obtained by Wolowski et al.[3], and calculated distributions for the native complexes from the Nussinov database and our self-produced false complexes. In figure 1 it can be seen, that scores from the two groups are not totally separated, but that there is an evident difference in the shape of the two curves.

Van der Waals energy and **electrostatic energy** are both calculated between all intermolecular atompairs in the complex and can be combined. Their sum is called interaction energy. We have not yet obtained score distributions for this, but the two examples in table 1 show, that the energies for the native complexes are considerably lower then for the false ones.

Receptor	Ligand	$E_{int} = E_{vdw} + E_{elec}$ [kcal/mol]
Barnase	Barstar	-264.4
Barnase	Soybean trypsin inhibitor	-242.4
Barbase	APPI	-214.0
Barnase	Ovomucoid 3rd domain	-192.0
Barnase	Pancratic secretory trypsin inhibitor	-189.4

Table 1. Intermolecular energies of one nativ complex (shaded in grey) and four false complexes. The energy is always the average of ten complexes that were top ranked from the docking algorithm.

4 Conclusion

We could show that it is possible to find scoring functions that can discriminate between native and false protein-protein complexes. By combination of the three presented scores and maybe even more in future, it will be possible to predict whether a hypothetical complex can be supposed to exist in nature or not.

Acknowledgments

We would like to thank the bavarian genomic network for financial support.

References

1. C. Dominguez, R. Boelens, A. M. Bonvin, *Protein-Protein Docking Approach Based on Biochemical or Biophysical Information*, J. Am. Chem. Soc. **125**, 1731–1737, 2003.
2. C. von Mering, R. Krause, B.Snel, M. Cornell, S. G. Oliver, S. Fields, P. Bork, *Comparative assessment of large-scale data sets of protein-protein interactions*, Nature **417**, 399–403, 2002.
3. V. Wolowski, *Computational analysis of protein-protein complexes related to knowledge-based predictions of interaction*, (Diploma Thesis, University of Hagen, Germany, 2008).
4. `http://bioinfo3d.cs.tau.ac.il/Interfaces/Non-Redundant`.