

POSTER PRESENTATION

Open Access

Prediction of highly-connected 'hub'-proteins in protein interaction networks using QSAR

M Hsing^{1*}, K Byler², A Cherkasov¹

From 5th German Conference on Cheminformatics: 23. CIC-Workshop
Goslar, Germany. 8-10 November 2009

Proteins that are most essential for functioning and viability of bacterial cell have been shown to exhibit larger number of interactions with other cell components. Thus, by identifying the most connected proteins (or hubs) in protein interaction networks (PINs), one may discover prospective drug targets that can be utilized to combat emergent and drug-resistant pathogens such as Methicillin-Resistant *Staphylococcus aureus* 252 (MRSA). The advantage of using such hub proteins as drug targets lies in their essentiality, non-replaceable position in the PIN and lower rate of mutation, which can help to counter bacterial resistance.

However, finding or predicting such hub proteins remains a challenging task as the corresponding experiments are very costly, while traditional bioinformatics approaches generally fail in forecasting PIN data due to the general lack of agreement between the existing datasets [1].

Thus, we have decided to utilize various structural and physicochemical features of proteins, related to traditional QSAR properties for predicting highly connected proteins. Using our own in-house generated PIN for the MRSA cell we have trained a boosting tree-based classifier that uses 75 physical and chemical QSAR descriptors computed for all proteins in the interaction network [2]. The utilized parameters included molecular weight, net charge, isoelectric point, hydrophobicity, surface area, solvent accessibilities, electronegativity, secondary structure composition, surface coils and flexibility among other QSAR descriptors.

The developed QSAR model has yielded a high prediction accuracy of 80% for the validation set and was used to predict additional hubs in the rest of the MRSA proteome. The predicted hubs have then been evaluated experimentally and 55% of them were confirmed as high

interactors what corresponds to >5 fold dataset enrichment for potential hub-proteins provided by the developed QSAR model.

Thus, the successful development of accurate hub classifiers demonstrated that highly-connected proteins tend to share certain structural and physicochemical features that can be characterized and quantified by conventional QSAR descriptors.

It is anticipated that the developed hub classifiers will represent a useful tool for the prediction of highly-interacting proteins and can find broad application for planning and executing large-scale proteomic experiments and for identification of novel and prospective anti-bacterial drug targets – even in those organisms that currently lack protein interaction data.

Author details

¹University of British Columbia, 2329 West Mall, Vancouver, BC Canada.

²Rochester, USA.

Published: 4 May 2010

References

1. Huang H, Jedynek BM, Bader JS: *PLoS Comput Biol* 2007, **3**:e214.
2. Byler K, Hsing M, Cherkasov A: *QSAR & Combinatorial Science* 2009, **28**:509-519.

doi:10.1186/1758-2946-2-S1-P35

Cite this article as: Hsing et al.: Prediction of highly-connected 'hub'-proteins in protein interaction networks using QSAR. *Journal of Cheminformatics* 2010 **2**(Suppl 1):P35.

¹University of British Columbia, 2329 West Mall, Vancouver, BC Canada

