Visualizing Phosporilation Experiments Data In The Context Of Known Protein Interactions

Radu Jianu* David H. Laidlaw† Arthur Salomon‡

Brown University

1 Introduction

Phosphorilation experiments indicate which proteins respond to a specific stimulus and can be the first step in protein pathway analysis. From the large number of proteins that result from phosphorilation, only a few can be investigated with other types of experiments that are more time and resource consuming. Since the phosphorilation experiments reveal little about which activated proteins also play key roles in the pathway, scientists have to rely on experience and previous knowledge to select proteins that are likely to reveal interesting aspects.. *We propose a visual analysis system that integrates data from phosphorylation experiments with protein interaction information contained in open source databases.*

2 Motivation

Basic protein visualization systems have already been created [][][]. They are mainly lightweight components associated to protein-interaction databases and only intended to aid the user in browsing the data. We think that more complex visualization tools geared towards advanced proteomic analysis, that can deal with large amounts of data and allow proteomicists to integrate their own experimental data would help proteomics make the most of high-throughput technology and advance at a more rapid pace.

3 Methods

We have prototyped a preliminary system that allowed us to establish a collaboration with proteomics researchers and gain valuable feedback. We then started working on a second, more advanced version, building on the accumulated experience.
Our first attempt consisted in projecting proteins indicated by phosporylation experiments onto interaction data derived from the Human Protein Reference Database (HPRD). Our system used a traditional node-edge representation powered by a simulated annealing algorithm. It also included features such as filtering, highlighting or hiding information to make the representation more readable [Figure 1].
Our proteomics collaborators appreciated it as a useful way of looking at their experimental data in the context of already known interactions. However, they also identified some drawbacks: the amount of information was still overwhelming and the placement of proteins, generated exclusively by the the graph-drawing algorithm, was not consistent with the proteomicists' mental representation of pathways (shaped by conventions and anatomical realities). This caused the user to feel disoriented and added some extra adapting and learning time.

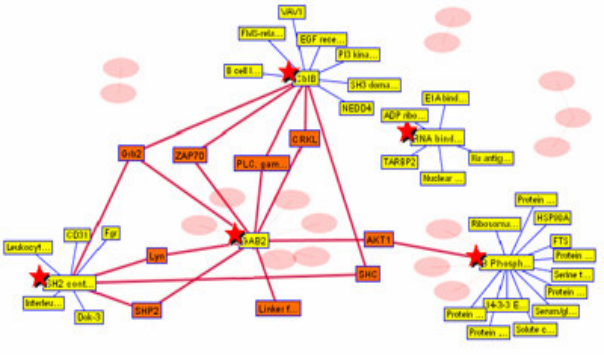

Figure 1: Representation of HPRD interaction data with overlaid phosphorilation information (proteins marked with star). Important proteins are highlighted while irrelevant proteins are hidden.

We have tried to solve these issues in our second system. We found that the STRING [] database could potentially solve the cluttering problem since it provides confidence levels to protein interactions. We use these to filter out information that is not relevant or might be considered not reliable enough.

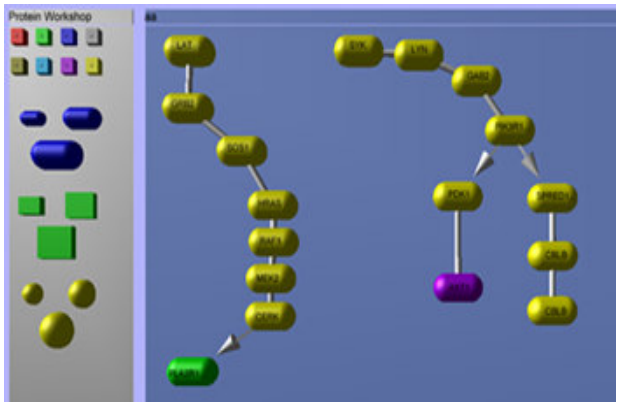

Figure 2: The pathway skeleton. Users assemble a pathway consisting of well-established proteins and interactions using a simple graphical interface.

* e-mail: jr@cs.brown.edu
† e-mail: dhl@cs.brown.edu
‡ e-mail: as@brown.edu

We have also completely changed the interaction visualization approach. Users first specify manually a pathway skeleton containing well-established proteins and connections. Since proteomicists only use a few pathways over a long period of time this seems feasible enough. Starting from this now familiar overview representation, the researcher can zoom in on specific proteins and crawl through the protein-interaction graph locally, viewing a single level at once.

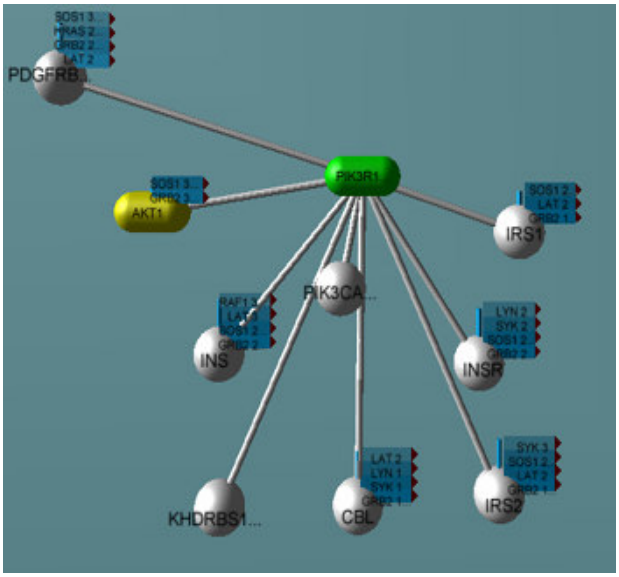

Figure 3: Local view of the pathway exploration tool. Signposts guide the user to proteins of interest.

We have used several techniques to ensure that the user does not get disoriented in the local view. Positions of proteins that are not in the pathway skeleton (don't have a position specified by the user) are computed by interpolating between the position of the ones present in the skeleton depending on how far (in terms of number of edges) each of them is. This creates a sense of Euclidian space allowing for better orientation.
We also place sign-posts indicating what proteins the user might reach if he chooses a certain path and how far away they are, thus helping the user to move in a goal driven way.

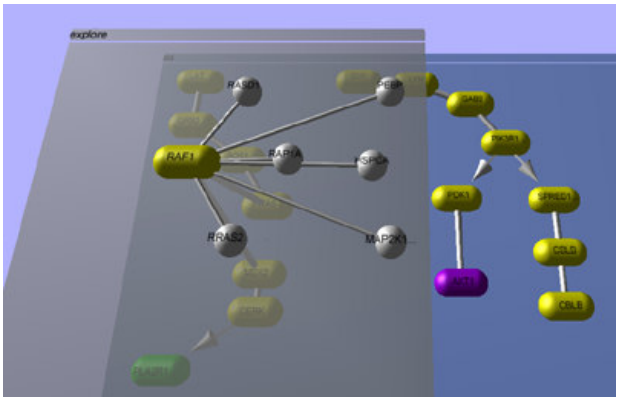

Figure 4: Using multiple 3D layers to explore the protein pathway.

We are currently experimenting with representations that would allow the user to perform the local exploration while maintaining the "big-picture" view in a secondary plane. We plan to make use of two or more 3D layers to organize the data in a more efficient way. Instead of switching to a completely different view-mode a new 3D layer is placed on top of the pathway skeleton and thus the local exploration takes place while keeping the user tightly linked to the overview image of the pathway [Figure 4].

4 Conclusion

We have introduced a new approach to visual analysis of proteomic experimental data by putting it in context with what is already known in the field. Further work is needed to improve the visualization and exploration paradigm, deal with time information that can also be derived from phosphorilation experiments and take into consideration other protein related dimensions that may provide insight to proteomics researchers.

Referernce

[1] Harel, D. and Davidson, R. Drawing graphs nicely using simulated annealing.

[2] Ju, BH. 2003. Visualization and analysis of protein interactions. Bioinformatics 2003, 317-318.

[3] Xenarios, I. 2000. DIP: the Database of Interacting Proteins. Nucleic Acids Research 2000, 289-291

[4] Mrowka, R. 2001. A Java applet for visualizing protein-protein interaction. Bioinformatics 2001, 669-670.

[5] C.v. Mering, L. J. Jensen, B. Snel, S. D. Hooper, M. Krupp, M. Foglierini, N. Jouffre, M. A. Huynen and P. Bork 2005. STRING: known and predicted protein protein associations, integrated and transferred across organisms. Nucleic Acids Research, 2005, Vol. 33, Database issue D433-D437.