# STRING and STITCH: known and predicted interactions between proteins and chemicals

Lars Juhl Jensen[1,2*], Michael Kuhn[1*], Manuel Stark[3], Samuel Chaffron[3], Christian von Mering[3†] and Peer Bork[1,4†]

[1] European Molecular Biology Laboratory, Heidelberg, Germany [2] Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark
[3] University of Zurich, Zurich, Switzerland [4] Max-Delbrück-Center for Molecular Medicine, Berlin, Germany * These authors contributed equally †Joint PIs

Nature Precedings : doi:10.1038/npre.2008.2265.1 : Posted 6 Sep 2008

## Motivation

A full description of a protein's function requires knowledge of all molecules with which it specifically associates. From a functional perspective, 'association' can mean direct physical binding, but can also mean indirect interaction such as participation in the same metabolic pathway or cellular process.

## STRING: protein-protein interactions

Currently, information about protein association is scattered over a wide variety of resources and model organisms. The STRING web resource (von Mering et al., 2007) aims to simplify access to this information by providing a comprehensive, quality-controlled collection of protein–protein associations for a large number of organisms. We achieve this by combining curated knowledge on protein complexes and pathways, information extracted through automatic literature mining (Saric et al., 2006), gene coexpression and interaction data from high-throughput experiments, and associations predicted by genomic context methods (Korbel et al., 2004). To integrate and rank the associations from such heterogeneous sources, we benchmark all associations against a gold standard, use orthology assignments to extend the interactions beyond the organisms in which they were originally described, and finally combine the resulting confidence scores using a probabilistic framework. The resulting functional association network covers over 1.5 million genes from 373 genomes and can be accessed via an intuitive web interface at http://string.embl.de/

## STITCH: protein-chemical interactions

Knowledge about interactions between proteins and small molecules is also essential for the understanding of molecular and cellular functions. For example, enzymes interact with their substrates, products and allosteric inhibitors, and drugs interact with their target proteins. However, information on such interactions is as widely dispersed as information on protein–protein interactions. To facilitate access to this data, the STITCH resource (Kuhn et al., 2008) integrates information about interactions from metabolic pathways, crystal structures, binding experiments and drug–target relationships. Inferred information from phenotypic effects, text mining and chemical structure similarity is used to predict relations between chemicals. The resource further allows exploring the network of chemical relations, also in the context of associated binding proteins. Our database contains interactions for over 68,000 different chemicals, including 2200 drugs, and connects them to the same gene set used in STRING. The STITCH web resource is available at http://stitch.embl.de/

## References

C. von Mering, L.J. Jensen, M. Kuhn, S. Chaffron, T. Doerks, B. Krüger, B. Snel and P. Bork, "STRING 7 – recent developments in the integration and prediction of protein interactions", Nucleic Acids Research, 35, D358–D362, 2007.

J. Saric, L.J. Jensen, R. Ouzounova, R. Rojas and P. Bork, "Extraction of regulatory gene/protein networks from Medline", Bioinformatics, 22, 645–650, 2006.

J.O. Korbel, L.J. Jensen, C. von Mering and P. Bork, "Analysis of genomic context: Prediction of functional associations from bidirectionally transcribed gene pairs", Nature Biotechnology, 22, 911–917, 2004.

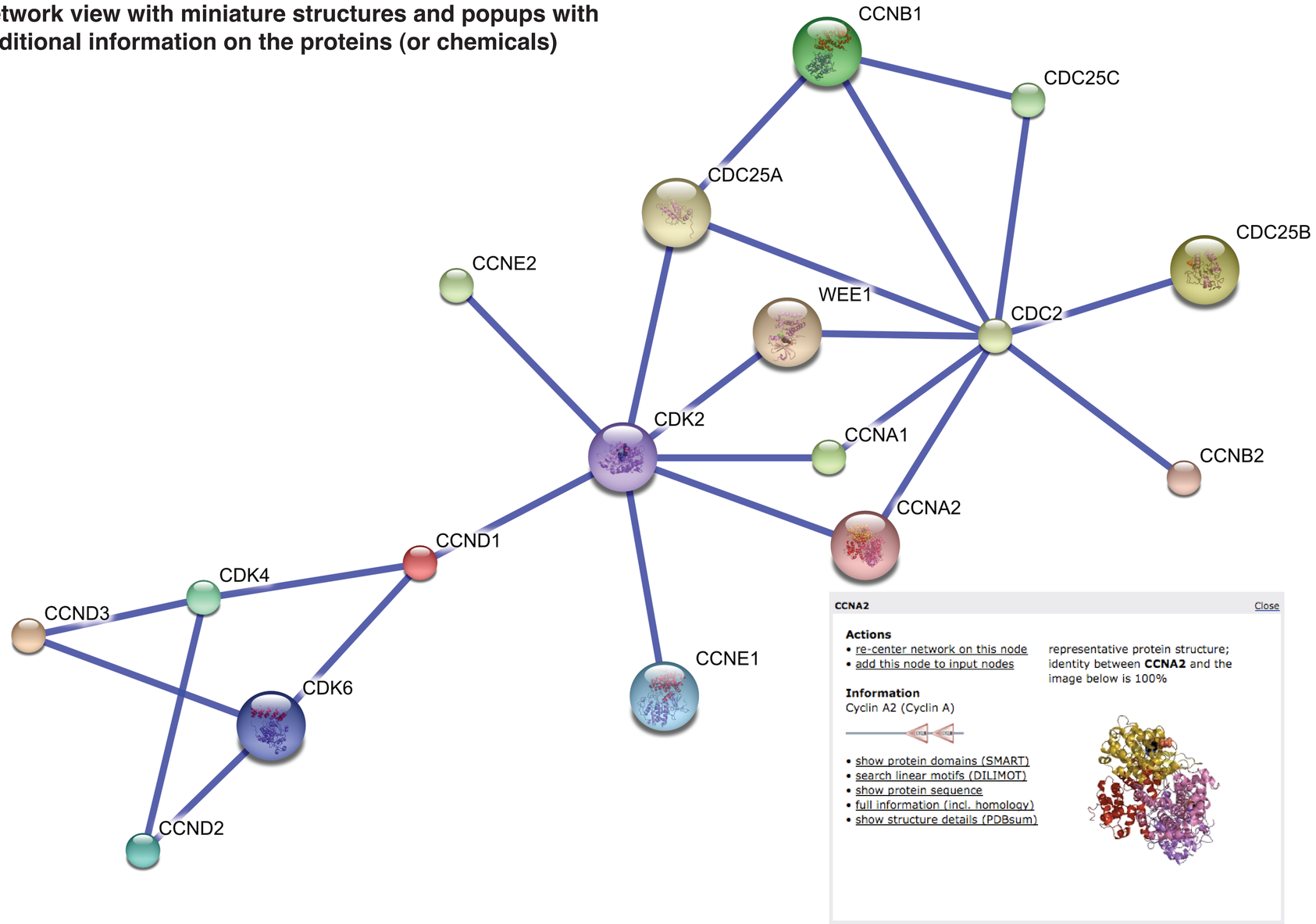
M. Kuhn, C. von Mering, M. Campillos, L.J. Jensen and P. Bork, "STITCH: interaction networks of chemicals and proteins", Nucleic Acids Research, 36, D684–D688, 2008.

European Molecular Biology Laboratory

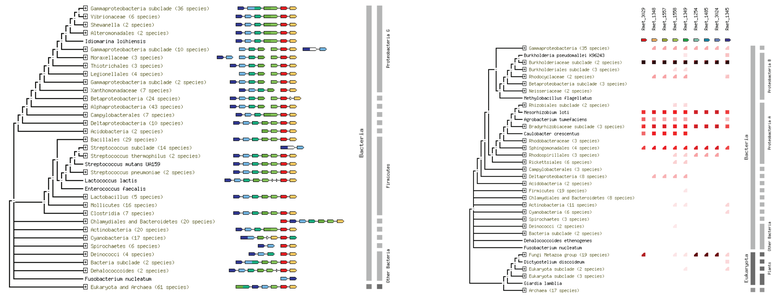
Swiss Institute of Bioinformatics

University of Zurich

## Network view with miniature structures and popups with additional information on the proteins (or chemicals)

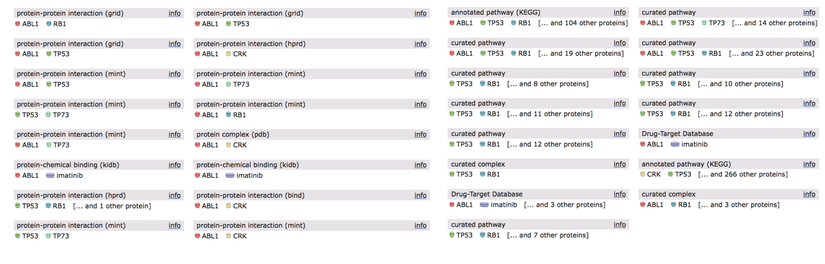



## Prediction of interactions from genomic context



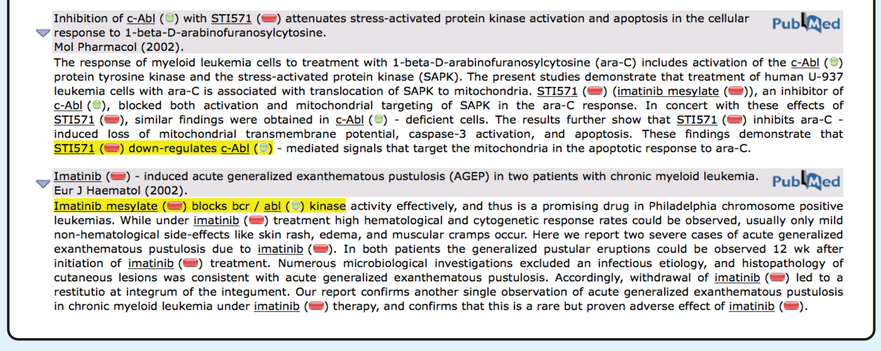
Gene neighbourhood Cooccurrence

## Integration of interaction and pathway databases



Primary experimental data Pathway and drug-target databases

## Literature mining based on cooccurrence and NLP




## The web resources string.embl.de and stitch.embl.de