

# Reconstruction and analysis of biochemical networks

Chris Hyland, Liz Gaskell, John Pinney, Glenn McConkey and David Westhead

## Introduction

The rapid proliferation of genome sequencing projects over the last ten years has resulted in an exponential growth in the amount of genomic DNA available to biologists. The focus of genomics research is now moving towards the development of fast, accurate methods of extracting new knowledge from these data. One important target is the elucidation of an organism's metabolic pathway complement from its genome sequence, known as *metabolic reconstruction*.

Knowledge of the presence or absence of specific pathways in a given organism can help improve the quality of genome annotation by highlighting false positives and negatives in assigned gene function. If only one enzyme-encoding gene out of a pathway of several steps is found in an annotated genome, it is likely to be a mistaken assignment. Conversely, if all but one or two enzymes in a pathway have corresponding genes in the annotated genome, the missing steps are likely to be present amongst the unidentified genes and are worth hunting down. Studying the metabolism of disease-causing organisms can also be an excellent means of identifying new drug targets. Many pathogenic bacteria and parasitic eukarya are the subjects of ongoing genome sequencing projects. If metabolic pathways can be identified which are essential in the pathogen but absent in the host, new drugs targeting the enzymes in these pathways are likely to be very effective.

## The metaSHARK project

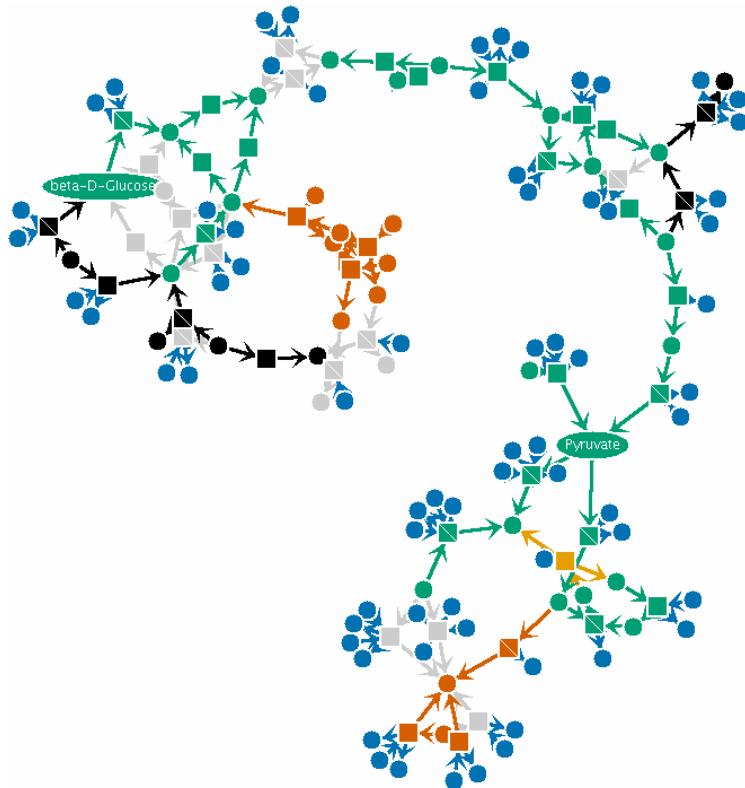
We have developed a comprehensive suite of programs for the representation and analysis of metabolic networks. The **metabolic SearCH And Reconstruction Kit (metaSHARK)** includes an object-oriented database to store knowledge about networks of chemicals and reactions, as well as an automated system to search an unannotated genome for genes with significant similarity to known enzymes from other organisms. These genes are assigned a confidence score based on the strength of their similarity to the test sequences.

The structure of such a metabolic network is very effectively modelled by a type of graph called a *Petri net* (Fig. 1). This is a graph with two types of nodes, called *places* and *transitions*, linked by directed arcs. Metabolites are represented by place nodes, and reactions by transition nodes. Petri nets have been used extensively in computer science to represent complex systems, and have proven to be useful in studies of many kinds of biological network.

Analysis of the Petri net structure can reveal sets of reactions called *elementary modes* - pathways which are stoichiometrically and thermodynamically feasible. The evidence for the presence of each of these modes in an organism's metabolism can then be assessed using the scores of the mode's component reactions. The confidence of gene predictions can be improved by the incorporation of other forms of genomic data, such as gene expression data, to show whether a predicted gene is expressed under a particular condition or at a particular time point in an organism's life cycle. Pathways in the network and elementary modes can then be ranked according to their biological relevance based on the combined expression levels of each gene in the pathway. This data can also be used to produce a list of candidate genes for enzymatic functions that appear to be missing from a particular pathway.

This sort of information is vital for the purposes of identifying good drug targets in pathogens such as *Plasmodium falciparum*, the parasite causing the most virulent type of malaria. Analysis of the recently-released complete parasite genome is revealing novel genes and pathways which are being verified by RNAi experiments in Dr McConkey's group. We hope

that this work will ultimately lead to the identification of new drug targets for this killer disease.



**Fig. 1:** Part of an automated metabolic reconstruction of the avian intracellular parasite *Eimeria tenella*, visualised as a Petri net. Metabolites are represented as circles; reactions as squares. Directed arcs between nodes show the effect of a reaction as the consumption and production of different metabolites. Blue circles represent ubiquitous (“pool”) metabolites, such as water, ATP, NADH etc.. Nodes in green show reactions catalysed by enzymes for which good evidence has been found in the *E. tenella* genome. Nodes in red show reactions catalysed by enzymes for which only tentative evidence has been found. Grey nodes show that no evidence for the catalysing enzymes was found, whilst black nodes indicate that no model data was available in our database, hence no gene search could be performed. The reconstruction clearly shows the presence of a glycolytic pathway in this organism as a route in green from beta-D-glucose to pyruvate. The Petri net structure of the reconstructed network will aid in the development of automated pathway detection algorithms using elementary mode theory, as well as other forms of network analysis.

## Publications

Pinney, J.W., Shirley, M.W., McConkey, G.A. and Westhead, D.R. (2005) metaSHARK: software for automated metabolic network prediction from DNA sequence and its application to the genomes of *Plasmodium falciparum* and *Eimeria tenella*. *Nucleic Acids Res.* (in press).

McConkey, G.A., Pinney, J.W., Westhead, D.R., Plueckhahn, K., Fitzpatrick, T.B., Macheroux, P. and Kappes, B. (2004) Annotating the *Plasmodium* genome and the enigma of the shikimate pathway. *Trends in Parasitology* **20**:60-65

Pinney, J.W., Westhead, D.R. and McConkey, G.A. (2003) Petri Net representations in systems biology. *Biochem. Soc. Trans.* **31**:1513-1515

## Funding

We thank the MRC and BBSRC for funding this project.