An optimisation model for metabolic pathways
F.J. Planes\textsuperscript{1} and J.E. Beasley\textsuperscript{2}

\textsuperscript{1}CEIT and TECNUN, University of Navarra, Manuel de Lardizabal 15, 20018 San Sebastian, Spain
\textsuperscript{2}Mathematical Sciences, Brunel University, Uxbridge, UB8 3PH, UK

Abstract
In this paper we detail a novel optimisation model, based upon integer linear programming, to determine metabolic pathways. Our model links reaction stoichiometry with path finding in a single approach. We test the ability of our model to determine 40 annotated \textit{E.coli} metabolic pathways. We show that our model is able to determine 36 of these 40 pathways in a computationally effective manner.

Detailed description
Our model starts from the idea that metabolic pathways do contain at least one directed (metabolic) path from the source compound to the target compound. Our model allows pathways to contain more than one directed path from the source compound to the target compound, as opposed to typical path finding approaches. This makes it suitable to determine (if necessary) metabolic pathways that are branched. In addition, our model directly addresses pathway stoichiometry, namely by requiring that directed paths respect specified stoichiometric constraints. Stoichiometric approaches typically assume a steady state condition by defining a subset of compounds (internal compounds) constrained to be balanced for every metabolic pathway.

We require that intermediate compounds in metabolic paths must be balanced, but do not explicitly state which compounds must be balanced. This “relaxed” stoichiometric constraint (steady state condition) allows the same compound to be balanced in some pathways, whilst being consumed or produced (i.e. unbalanced) in other pathways.

In essence, our model views a metabolic pathway as being made up of a finite set of metabolic paths from the source compound to the target compound with the intermediate compounds on the metabolic paths being balanced. Interestingly, our mathematical model provides a link between path finding approaches and stoichiometric approaches. To the best of our knowledge, no approach in the literature has to date combined both perspectives.

In our optimisation model the minimisation objective adopted gives primary weight to a function related to the connectivity of intermediate compounds in the metabolic paths. Here we aim to avoid highly connected compounds appearing in metabolic paths. We also include in our minimisation objective a secondary factor relating to unbalanced compounds.

Structural modelling of signal transduction in enzyme cascades with the concept of elementary flux modes

Jörn Behre and Stefan Schuster

School of Biology and Pharmaceutics, Section of Bioinformatics, Friedrich Schiller University Jena, Ernst-Abbe-Platz 2, D-07743 Jena, Germany

Signal transduction is essential for all biological functions in living cells. The modelling of signalling networks is attracting more and more interest [ZS05, KSL+06]. The concepts of elementary flux modes (EFMs) [SH94] and extreme pathways [SLP00] are important tools for detecting non-decomposable pathways in metabolic networks. These methods are based on the steady state condition. In signalling networks, mass balance is of minor importance because there the flow of information matters. Nevertheless, it is of interest to adopt pathway detection
methods also for signalling systems. Here we present a formalism that allows the application of elementary flux modes in the case of enzyme cascades operating, for example, by phosphorylation and dephosphorylation. Our approach [BS09] is based on the ideas that the signal may not be diminished along each cascade and that the system has to return to its initial state after each signalling event, which implies that a steady state can be assumed in the sense of an average over time spans including the signalling event and the regeneration period. We illustrate our method by several prototypic single-phosphorylation and double-phosphorylation cascades including convergent and divergent branching. Moreover we apply our formalism to a part of the insulin signalling network. For computing EFMs we use the program Metatool, developed in our group [vKS06].

References

Stoichiometric Network Analysis and (sub)network dynamics
Carsten Conradi

Stoichiometric Network Analysis decomposes large and complex reaction networks in a natural way: each "elementary flux mode" (efm) defines a reaction network in its own right that is at the same time a subnetwork of the overall network. Already in Clark’s classical work (where these subnetworks are called “extreme currents”) the connection between dynamical properties of a subnetwork and the overall network has been studied.

Here we present new results concerning (multiple) steady states and their local stability:

(A) Existence of (multiple) steady states in subnetworks: for subnetworks defined by efms the existence of multiple steady states can be established by analysis of linear inequality systems only. Our results allow the computation of (pairs of) steady states and rate constants for subnetworks defined by efms.

(B) Extension of (multiple) steady states from the subnetwork to the overall network: we propose a computationally simple algorithm that allows to decide whether or not (multiple) steady states can be extended from the subnetwork to the overall network. The only information the algorithm needs are the subnetwork and the overall network together with (multiple) steady states and corresponding rate constants of the subnetwork only. If this extension is possible, then one can (i) compute rate constants for those reactions of the overall network not contained in the subnetwork and (ii) determine the local stability of the extended steady states in the overall network based on the stability properties of the subnetwork steady states (and the newly generated rate constants of the overall network).
Chemical Organization Theory - Advances and Implications

Peter Dittrich

Complex dynamical reaction networks consisting of many components that interact and produce each other are difficult to understand, especially, when new components may appear over time. In this talk, I outline a theory, which has been inspired by artificial chemistry research, to deal with such systems. The theory has been successfully applied to regulated metabolic networks, virus-immune system dynamics, chemical information processing, chemical evolution, and planetary atmosphere photo-chemistries. Here I will focus on theoretical issues by presenting first the basic concepts of the theory (organization = closed + self-maintaining sets of species, union and intersection of organizations, fixed point theorem) together with some very recent results (spatial organizations, limit set theorem). Then I will discuss an interesting theoretical implication, namely that in living systems, the set of organizations always forms a lattice. Finally I will discuss the question concerning the origin of organizations. Are organizational structures generated and formed evolution or by other processes? This question is in particular interesting, because random reaction networks do not posses "interesting" organizations.

The presented results are obtained from joint work with:
Florian Centler (UFZ Leipzig),
Christoph Kaleta (FSU Jena), Thorsten Lenser (FSU Jena),
Naoki Matsumaru (FSU Jena), and Pietro Speroni di Fenizio (Univ. Coimbra).

Determining minimal nutrient sets of metabolic networks

Thomas Handorf et. al.

A key function of metabolism is to chemically convert available nutrients into products which are required by other cellular processes like protein synthesis, DNA replication, energy or cofactor production.

In the last decades numerous approaches have been presented which model these metabolic conversions in various ways. In Handorf et al. (2005) we have proposed the concept of scopes which allows to predict producible metabolites from a given set of substrates if only the topology of the network is known.

Here we study the inverse problem: The proposed algorithm predicts possible nutrient sets which enable the metabolic network to produce a given set of precursor metabolites, which are required by other cellular processes (Handorf et. al (2008)).

We employ a greedy algorithm which tests different nutrient sets by checking the producibility of the precursor set using the concept of scopes. We start with the whole set of network metabolites as nutrients. The metabolites within this set are subsequently removed if they were not essential for the production of the precursor set, resulting in a locally minimal nutrient set. By analyzing a large number of different minimal sets, groups of alternative nutrients can be identified. Further, the essentiality of such a group can be determined, enabling us to conveniently assess the nutritional options of an organism.

In Handorf et. al (2008) we applied this methodology to a large set organism specific networks obtained from the KEGG database. As a tendency, the predicted nutrient sets coarsely correspond to the metabolic live style such as intracellular parasitism or universalism. The specific results for single organisms showed organism specific auxotrophies, for example for homo sapiens various vitamins and the essential amino acids were predicted.

The proposed method is very useful for identifying possible nutrient medias for yet not cultivable organisms. Further, it can be used to assess the completeness of the utilized metabolic network data. In particular, a wrongly predicted essential nutrient points at a missing link in the synthesis pathway of this metabolite.
The increasing number and complexity of bio-models makes automatic procedures for checking the models' properties and quality necessary. Approaches like elementary mode analysis [1], flux balance analysis [2], deficiency analysis [3], and chemical organization theory [4] require only the stoichiometric structure of the reaction network for derivation of valuable information. In formalisms like systems biology markup language (SBML), however, information about the stoichiometric coefficients required for an analysis of chemical organizations can be hidden in kinetic laws. We introduced an algorithm that uncovers this information [5]. This allows us to apply chemical organization theory to SBML models containing kinetic laws. Furthermore, using the new algorithm, we performed a large-scale analysis of the 185 models contained in the manually curated BioModels Database [6]. We found that for 42 models (23%) the set of organizations changes when kinetic laws are considered correctly. We discuss one of these models in detail (BIOMD149, a combined model of the ERK- and Wnt-signalling pathways), whose set of organizations drastically changes when kinetic laws are considered. Third, we found inconsistencies in 5 models (3%) and identified their characteristics. Compared to flux-based methods, chemical organization theory is able to identify those species and reactions more accurately (in 26 cases (14%)) that can be present in a long-term simulation of the model. We conclude that our approach is a valuable tool that helps to improve the consistency of bio-models and their repositories.

The advent of new technologies in molecular biology has lead to high-throughput genome sequencing techniques. Different strategies have been applied to analyze and compare the enormous amount of data generated. The present work introduces a new system for comparing metabolic pathways. SAMPA (System for comparative Analysis of Metabolic Pathways) consists of a database based on the KEGG database (Kanehisa et al., 2000) and of five tools that allow for the metabolic pathway comparison of a group of organisms that can then be clustered according to entirely or partially conserved metabolic paths. The tools are Comparative Analysis (CA), Comparative Analysis with Compounds (CAC), Comparative Analysis with Threshold (CAT), Comparative Analysis with Enzymes (CAE), Comparative Analysis with Intermediates Enzymes (CAIE).

The metabolic pathways of an organism are represented by directed graphs, $D(V,E,L)$ and labeled arcs. Each element of the set $V$ of vertices stands for an enzyme which is represented by its EC number, and each element of the set $L$ of labels stands for a compound which is represented by its KEGG compound ID. The set $E$ of labeled arcs is composed of triplets $(eci, ecj, coij)$, where $eci, ecj \in V$, and $coij \in L$ is a compound which links $eci$ and $ecj$; moreover $coij$ is the product of the chemical reaction catalyzed by $eci$ and the substrate of the chemical reaction catalyzed by $ecj$. The paths are built using a depth first search (DFS) approach, taking into account the direction of the enzymatic reactions (reversible or irreversible) and substrate-product association based on the KEGG database information. To develop the system the PERL language was used for writing the scripts, AJAX for more interactive web interface, Apache as web server and MySQL as database.

As case study to validate the system, metabolic pathways of twelve mycoplasms of which the genomes have been completely sequenced were analyzed to detect a set of similar paths among them and also paths that are exclusive to at least two of them. Mycoplasmas are phenotypically distinct from other bacteria by its reduced genome that ranges from 0.58 to 1.35 Mb and the lack of a cell wall (Razin et al., 1998). More than 190 species of mycoplasms have already been described. Many of them are human and domestic animal pathogens justifying their medical or veterinary relevance.

SAMPA results consist of the clusters identified by their ID (Cluster 1, Cluster 2, etc), the set of metabolic paths and the set of organisms which posses those paths. Additionally, links to the KEGG database website enable to get information on the enzymes from the resulting paths as well as on the organisms of the clusters. Graphical visualization of the clustered paths is available through the KEGG pathway map where enzymes from the selected paths are colored. The tools CA, CAC and CAT generate a circular diagram based on the comparative analysis.

The aim of SAMPA is to fulfill the gaps left by other metabolic pathways comparative analysis programs (Chen et al., 2004; Oehm et al., 2008; Chou et al., 2009). It provides simple and intuitive interfaces to visualize a set of accurate queries. The algorithms were carefully implemented giving satisfactory performance considering the computational complexity of graph searches. Therefore, SAMPA is an effective and robust system for the reconstruction, comparative analysis and visualization of metabolic pathways that may give additional clues for genome annotation.
References

Optimization and metabolic pathway analysis

Francisco J. Planes¹, Alberto Rezola¹ and Luis F. de Figueiredo²,³

¹CEIT and TECNUN (University of Navarra), San Sebastián, Spain; ²Friedrich-Schiller-University Jena, Jena, Germany; ³PhD Program in Computational Biology, Instituto Gulbenkian de Ciência, Oeiras, Portugal.

One of main challenges in the field of Metabolic Pathway Analysis is the computation of elementary flux modes (EFMs). Recently, we presented a theoretical method to enumerate the full set of EFMs in increasing number of reactions via linear integer programming (de Figueiredo et al., 2009). In contrast to traditional methods (Schuster et al., 2000; Schilling et al., 2000), our optimization-based method allows the computation of the shortest EFMs even at the genome-scale. In this talk, we show that, based on our previous method, a theoretical formulation for the enumeration of extreme pathways and a convex basis is also possible via linear integer programming. Details as to the application of our extended method to genome-scale metabolic networks are also presented.


Feedback cycles and Instability Causing Structures

Thomas Wilhelm, IFR Norwich, UK

Important information about metabolic networks can be obtained by analysing their stoichiometric network: the left eigenspace defines linear conservation relations and the right eigenspace (kernel) defines linear constraints in flux space and leads to definition of metabolic pathways, such as elementary flux modes (EFMs). We show how to use EFMs for robustness analyses (Wilhelm et al. 2004) and that our corresponding robustness measure for (sub)networks is more appropriate for evolutionary considerations than e.g. the related enzyme subsets.

We show that nonlinear conservation relations also exist. They cannot be found by stoichiometric network analysis. EFMs are based on the steady state assumption. We also
show that the additional demand for local stability of steady states leads to additional nonlinear constraints in flux space.

Recently we proposed a new topological network analysis method – the Instability Causing Structure Analysis (ICSA, Wilhelm 2007). ICSA is based on a General Jacobian concept that also allows unambiguous definition of positive and negative feedback loops. Positive feedbacks are associated with bistability and negative feedbacks with oscillations. We present minimal oscillating (Wilhelm & Heinrich 1995) and bistable (Wilhelm 2009) systems and we discuss the role of the contained feedback cycles in the context of necessary conditions for these phenomena.

We show that instability causing structures (ICSs) are always feedback cycles or combinations of feedback cycles, but not all feedback cycles are ICSs. Finally, we present the general procedure of systematically identifying all ICSs of a given system and we discuss limits of applicability of ICSA.