Chapter 2

Metabolic Network Reconstruction and Their Topological Analysis

Marie Beurton-Aimar, Tung Vu-Ngoc Nguyen, and Sophie Colombié

Abstract

This chapter focuses on the way to build a metabolic network and how to analyze its structure. The first part of this chapter describes the methods of the network model reconstruction from biochemical data found in specialized databases and/or literature. The second part deals with metabolic pathway analysis as a useful tool for better understanding the complex architecture of intracellular metabolism. The graph analysis and the stoichiometric network analysis are important approaches for understanding the network topology and consequently the function of metabolic networks. Among the methods presented, the Elementary Flux Modes analysis will be more detailed. Finally, we illustrate in this chapter an example of network reconstruction from heterotrophic plant cells metabolism and its topological analysis leading to a huge number of Elementary Flux Modes.

Key words Metabolic networks, Metabolic pathway analysis, Stoichiometry matrix, Topological analysis, Elementary Flux Modes, Plant cell metabolism

Abbreviations

AccoA	Acetyl coenzymeA
aKG	2-Oxoglutarate
Ala	Alanine
Asp	Aspartate
Cit	Citrate
coA	CoenzymeA
DHAP	Glycerone phosphate
E4P_p	Erythrose 4-phosphate
FA	Fatty acids
Fru	Fructose
F16bP	Fructose-1,6-biphosphate
Fum	Fumarate
F6P	Fructose-6-phosphate

Martine Dieuaide-Noubhani and Ana Paula Alonso (eds.), *Plant Metabolic Flux Analysis: Methods and Protocols*, Methods in Molecular Biology, vol. 1090, DOI 10.1007/978-1-62703-688-7_2,

GAP	Glyceraldehyde-3-phosphate
Glc	Glucose
Gln	Glutamine
Glu	Glutamate
G1P	Glucose-1-phosphate
G6P	Glucose-6-phosphate
HP	Hexose phosphate
Mal	Malate
OAA	Oxaloacetate
PEP	Phosphoenolpyruvate
PG	Phosphoglycerate
Pyr	Pyruvate
P5P	Pentose-5-phosphate
Ri5P	Ribose-5-phosphate
Ru5P	Ribulose-5-phosphate
Suc	Sucrose
Succ	Succinate
sucP	Sucrose phosphate
S7P	Sedoheptulose-7-phosphate
TP	Triose phosphate
X5P	Xylulose-5-phosphate
"_p"	For metabolites located in plastid
"_vac"	For metabolites located in vacuole

1 Introduction

A challenge in systems biology is to identify the relationship between structure, function and regulation in complex networks that can be reconstructed from genomic or biochemical data. Metabolic network reconstruction and simulation allows for an in depth insight into comprehending the molecular mechanisms. A reconstruction consists in merging metabolic pathways with their respective reactions and enzymes in the perspective of analyzing the entire network assimilated to a metabolic model. Dynamic mathematical modelling of large-scale networks meets difficulties as the mechanistic details and kinetic parameters are rarely available. Conversely, structural (topological) analyses require only reaction stoichiometries and reversibility information, which are often well-known.

This chapter describes two types of methods: network model reconstruction from biochemical data and stoichiometric network analysis as special class of structural (topological) analysis methods. Stoichiometric modelling has emerged in 1990 for microorganisms and has become an important approach in plant science only recently [1]. This approach is used to explore and to understand deeply the function of metabolic networks. The aim of this chapter

is to describe advantages and limitations of model reconstruction and stoichiometric analysis with respect to their potential applications for realistic biological network.

2 Reconstruction

A reconstruction involves collecting all of the relevant metabolic information of an organism and then compiling it in a way that makes sense for various types of analyses to be performed. A reconstruction often combines the relevant metabolic and genomic data of an organism. A reconstruction also allows for metabolic comparisons to be performed between various organisms of the same species as well as between different organisms.

The links between the genome and metabolism can be made by searching gene databases, such as KEGG or GeneDB, with a request by enzyme or protein names for particular genes. Nowadays, increasing work tend to generate the entire metabolic network from the available genomic data. These so-called genome-scale models are discussed at the end of the chapter and are developed in detail in Chapter 19 of this volume. Concerning classical reconstruction of one or several pathways, for instance to build the central metabolic network, international databases contain various metabolic pathways such as glycolysis, TCA cycle, pentose phosphate pathway, etc. But at this time, assembling different pathways to design a global network is mainly made by hand and it is a not trivial task. Plant models have been designed using metabolic databases, biochemical textbooks, and the primary literature, and are essentially confined to the well known pathways of central metabolism.

2.1 First Step of a Reconstruction: Finding Resources

Even a lot of genome annotations are available through free international databases, with cross-linked data, plant genomes are still rare. Today, the TAIR (http://www.arabidopsis.org/) Web site has listed about 20 projects in plant genetics (maize, rice, soybean, etc.) At this time, Arabidopsis and maize are fully annotated and recently, the Tomato Genome Consortium (http://solgenomics.net) has published in Nature the last version of their annotation [2].

To give a more detailed description of a few gene/enzyme/reaction/pathway databases that are crucial to a metabolic reconstruction, we can site:

1. *Kyoto Encyclopedia of Genes and Genomes* (KEGG, http://www.kegg.com/) is a bioinformatics database containing information on genes, proteins, reactions, and pathways. The 'KEGG Organisms' section, which is divided into eukaryotes and prokaryotes, encompasses many organisms (but less than 20 plant species) for which gene and DNA information can be searched

by typing in the enzyme of choice. This resource is useful to build associations between metabolism enzymes, reactions, and genes.

2. *BioCyc* (http://biocyc.org/) is a collection of 1,962 pathway/genome databases (version 16.1, 2012), with each database dedicated to one organism. Today about 20 databases of eukaryotes are available. For example, AraCyc is a highly detailed database on the genome and metabolic reconstruction of *Arabidopsis thaliana*, including thorough descriptions of signaling pathways and regulatory networks. The AraCyc database can serve as a paradigm and a model for any reconstruction. Additionally, MetaCyc, an encyclopedia of experimentally defined metabolic pathways and enzymes, contains 1,877 metabolic pathways and 10,247 metabolic reactions (2012).

The BioCyc database can be managed with the help of *Pathway Tools* framework, (http://bioinformatics.ai.sri.com/ptools/) a bioinformatics software package that assists in the construction of pathway/genome databases. Developed by Peter Karp and associates at the SRI International Bioinformatics Group, Pathway Tools comprises several separate units. First, PathoLogic takes an annotated genome for an organism and infers probable metabolic pathways to produce a new pathway/genome database. This can be followed by application of the Pathway Hole Filler, which predicts likely genes to fill "holes" (missing steps) in predicted pathways. Afterward, the Pathway Tools Navigator and Editor functions let users visualize, analyze, access, and update the database. Thus, using PathoLogic and encyclopedias like MetaCyc, an initial fast reconstruction can be obtained automatically, and then using the other units of Pathway Tools, a very detailed manual update, curation, and verification step can be carried out.

Other relevant databases where the request can be led by the protein name or the EC number (representing the catalytic function of the enzyme of interest) have to be mentioned:

3. *ENZYME* (http://enzyme.expasy.org/) is an enzyme nomenclature database (part of the ExPASY proteonomics server of the Swiss Institute of Bioinformatics). After searching for a particular enzyme on the database, this resource gives the reaction that is catalyzed. Additionally, ENZYME has direct links to various other gene/enzyme/medical literature databases such as KEGG, BRENDA, PUBMED, and PUMA2 to name a few.
4. *BRENDA* (http://www.brenda-enzymes.info/) is a comprehensive enzyme database to search an enzyme by name or EC number. Moreover, when an enzyme search is carried out, BRENDA provides a list of all organisms containing the particular enzyme of interest.

One can note that several inconsistencies exist between gene, enzyme, and reaction databases and published literature sources regarding the metabolic information of an organism. A reconstruction is a systematic verification and compilation of data from various sources that takes into accounts all of the discrepancies.

2.2 Next Step of the Reconstruction: Checking Data

After the initial stages of reconstruction, a systematic verification is made in order to track inconsistencies, i.e., that all the entries listed are correct and accurate. Furthermore, previous literature can be researched in order to support any information obtained from one of the many metabolic reaction and genome databases.

To achieve the reconstructed network, the collected pathways and reactions have to be properly connected. For instance pathways have to be connected by transport reactions between subcellular compartments. The presence or absence of certain reactions of the metabolism will affect the amount of reactants/products that are present for other reactions within the particular pathway. Francke et al. [3] provide an excellent example as to why the verification step of the project needs to be performed in significant detail. During a metabolic network reconstruction of *Lactobacillus plantarum*, the model showed that succinyl-CoA was one of the reactants for a reaction that was a part of the biosynthesis of methionine. However, an understanding of the physiology of the organism would have revealed that due to an incomplete tricarboxylic acid pathway, *Lactobacillus plantarum* does not actually produce succinyl-CoA, and the correct reactant for that part of the reaction was acetyl-CoA.

Therefore, systematic verification of the initial reconstruction will bring to light several inconsistencies that can adversely affect the final interpretation of the reconstruction, which is to accurately comprehend the molecular mechanisms of the organism. Then the network analysis suggests finding hypothetical routes between specific inputs and outputs that only emerge in the context of complex network.

2.3 Main Definitions

Before network analysis, we first need to define some terms related to the metabolic networks, composed with biochemical reactions.

The *stoichiometry* specifies the reactants, educts (substrates) and products participating in a reaction as well as the molar ratios in which they are consumed or produced. Stoichiometry is a quantitative relationship between substrates of a chemical reaction. The stoichiometric coefficient of a metabolite by convention is positive if it is produced when the reaction proceeds in the forward reaction, and negative otherwise.

The reaction *directionality* specifies the reversibility or not of the reaction. While a chemical reaction is thermodynamically reversible, in biochemical networks, some reactions are irreversible and proceed only in one direction due to biological constraints.

Knowledge on the reversibility of reactions allows constraining the number of feasible pathways in the network.

Reaction *kinetics* describes the dynamics of the reaction based on the reaction mechanism and the enzyme properties. More often, these characteristics are imprecise, context dependent, or unknown.

3 Analysis

The investigation of the structure of metabolic networks has recently attracted increasing interest. The formal way to deal with networks is to use graph modelling. The next paragraph gives basic knowledge about graph analysis.

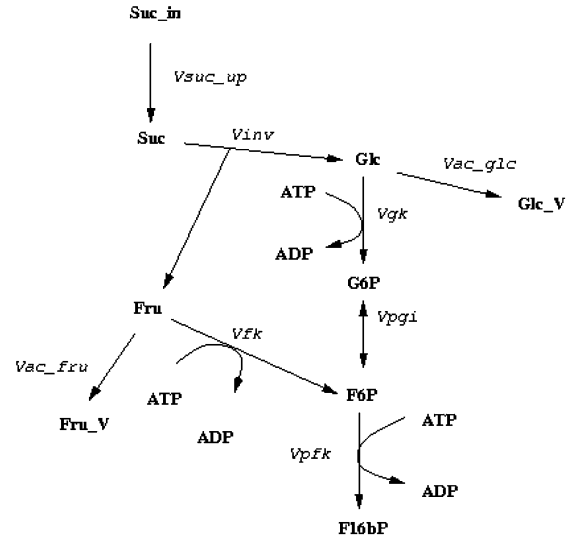
3.1 Graph Analysis

Several authors have proposed methods to analyze biological networks as classical graphs in computer science. Graph theory uses the scheme of network connectivities, which is a simplified representation of real reaction networks. The graph theory reaches back to the end of the eighteenth century, to solve the problem of finding the best way through Konigsberg crossing its seven bridges only once (Leonard Euler "Konigsberg bridge problem", 1736, see F. Schreiber for details in [4]). Nowadays, the widespread use of computers to analyze graphs has encouraged developments of algorithms to compute properties of large graphs. Mainly, topological analysis of a graph provides information about relationship between elements belonging to it. It is why topological properties are commonly computing to characterize graph organization. A graph $G = (N, E)$ consists of a set of nodes N and a set of edges E, where each edge is assigned to two nodes (not necessarily disjoint). To manage computation of graph properties, two main coding can be used: adjacency matrix or adjacency list. Figure 1a displays a simple example showing a simplified sucrose pathway, Fig. 1b gives the adjacency matrix built from this drawing. The adjacency list could be extract from this matrix following this rule, each element in the list is a set of two nodes linked by an edge.

3.1.1 Diameter and Degree of Nodes

Topology of graph can be studied with parameters such as the diameter or the degree of nodes. *The diameter* is the maximum distance of any pair of nodes. The average path length is the average distance between all pairs of nodes. The distance between two nodes i and j is the length of the shortest path between these two nodes, i.e., the minimal number of edges that needs to be traversed to travel from i to j. More details about equations to compute path length can be found in Steuer and Lopez [4]. It has been shown that, whatever the organism, large networks such as metabolic networks have a short average path length (around three reactions) between two metabolites in the network [5–7]. It is worth to note that networks like metabolic networks generally exhibit average

a



b

M \ R	*Vsuc_up*	*Vinv*	*Vgk*	*Vpgi*	*Vfk*	*Vpfk*	*Vac_glc*	*Vac_fru*
Suc_in	1	0	0	0	0	0	0	0
Suc	0	1	0	0	0	0	0	0
Glc	0	1	1	0	0	0	0	0
G6P	0	0	1	1	0	0	0	0
F6P	0	0	0	1	1	1	0	0
Fru	0	1	0	0	1	0	0	1
ATP	0	0	1	1	1	1	0	0
ADP	0	0	1	1	1	1	0	0
Fru_V	0	0	0	0	0	0	0	1
Glc_V	0	0	0	0	0	0	1	0
F16bP	0	0	0	0	0	1	0	0

Fig. 1 (**a**) Metabolic network of a simplified sucrose pathway. (**b**) The adjacency matrix built from this drawing

path length. Several authors, mainly coming from social sciences, have described properties of specific networks called small-world networks. To resume their description, in a small world network the average path length decreases slower or equal than the logarithm of the network size [4]. Metabolic networks can be considered as belonging to this category [7]. For example Fell et al. [5] have shown that a metabolic network of the TCA cycle posses such characteristics. More recently, Barabasi and Oltvai [8] have also studied the possibility to characterize biological networks as small world networks. They consider the degree distribution of nodes

which is more heterogeneous than those of a random network and the clustering coefficient to measure the cliquishness of the local neighborhoods. The average degree of nodes and the clustering coefficient are the two another parameters to analyze characterizing network topology. The degree of a node belonging to the graph G is equal to the number of adjacent edges to this node. The clustering coefficient measures the probability that two vertices with a common neighbor are connected. The clustering coefficient can be computed as following:

$$C_i = \frac{2E_i}{E_{\max}} \tag{1}$$

with

$$E_{\max} = k_i(k_i - 1) \tag{2}$$

where C_i is the clustering coefficient of the node n_i, and E_i is the number of edges; $E_{\max}$ is the maximum expected number of edges between the k_i neighbors of n_i. Then these measures are compared to those of a random network generated with the same number of nodes. Finding differences for both parameters, the average degree of nodes and the clustering coefficient between the random network and the studied network, signifies that a specific organization can be found in the studied network, led for example by functionality constraints.

3.1.2 Digraph Modelling

One complex characteristic of metabolic networks is that they are hypergraphs. Often enzymatic reactions imply more than one substrate and produce more than one metabolite (see the description of the simple sucrose pathway Fig. 1a in the previous section). A way to solve that is to convert the graph in a digraph that is a graph with two types of nodes. Petri net is a well known format of digraph. The graph is defined as a set of nodes called "places" for the molecules and a set of nodes called "transitions" for the reactions. Molecules are considered as discrete token and applying Gillepsie formula [9] the kinetics can be taken into account as probability to fire reactions. Srivastava et al. [10] gives an example of simulation with this tool. But even if the Petri net modelling provides an efficient simulation of the network behaviors, the proof that all behaviors have been explored has to be brought. Moreover, numerical values like reaction kinetics useful to validate the simulation are often missing.

3.2 Metabolic Pathway Analysis

Analysis of metabolic pathway has become increasingly important to assess inherent network properties in reconstructed biochemical reaction networks [11]. From the last 10 years, the amount of available data has increased drastically. With the development of recent systems biology approaches, the number of reactions taking into account for the design of one network has moved from tens to

one or several hundreds. In this case, it is not possible to identify by hands all the interactions through this kind of networks. Using matrix formalism, Seressiotis and Bailey [12] have suggested identifying pathways through the network using linear algebra algorithms. Before explaining more deeply different computational methods, the next paragraph describes the specific matrix which is the stoichiometry one, using reaction nodes and metabolite nodes.

3.2.1 The Stoichiometry Matrix

The stoichiometry of a complete system can be summarized in the matrix form, the so-called *stoichiometry matrix* denoted by the symbol N, only refers to the m internal metabolites of the network composed of n reactions. By convention, the stoichiometry matrix will have m rows and n columns. The stoichiometric matrix marks all interactions between metabolites and reactions:

- If a reaction i involves the metabolite j, the number of required molecules to process is given at the jth rows and the ith column.
- Else the value is zero.

Figure 2 gives the stoichiometry matrix for the given example in Fig. 1a. As previously mentioned a stoichiometric matrix takes into account only the internal metabolites. Following the metabolic assumptions, metabolic network can be considered at the steady state that is all the *internal metabolites* have to be balanced in a valid pathway. At the opposite, *external* metabolites are considered as always available and so unnecessary to be mentioned in the model. Designers have to pay attention to the decision to declare a metabolite as internal or external determines the frontiers of the analyzed network. Decision to balance or not a metabolite can drastically change the results of analysis.

To compute valid pathways, the stoichiometry matrix N is combined with the rate vector v to form a compact equation describing the rates of change of the molecular species S:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = N \cdot v \tag{3}$$

M\R	Vsuc_up	Vinv	Vgk	Vpgi	Vfk	Vpfk	Vac_fru	Vac_glc
Suc	1	−1	0	0	0	0	0	0
Glc	0	1	−1	0	0	0	0	−1
Fru	0	1	0	0	−1	0	−1	0
G6P	0	0	1	−1	0	0	0	0
F6P	0	0	0	1	1	−1	0	0

Fig. 2 The stoichiometry matrix of the network displayed in Fig. 1

3.2.2 Steady-State Assumption

A steady state is a situation in which all state variables are constant in spite of ongoing processes that strive to change them. For an entire system to be at steady state, i.e., for all state variables of a system to be constant, there must be a flow through the system. Several constraints have to be respected, there must be no accumulation of mass over the time period of interest and the same mass flow rate will remain constant in the flow path through each element of the system. Thermodynamic properties may vary from point to point, but will remain unchanged at any given point. Metabolism usually involves fast reactions and high turnover of metabolites compared to regulatory events. Therefore analysis of metabolic networks is often based on the assumption that metabolite concentrations and reactions rates are constant. This assumption leads to the fundamental mass balancing equation [13].

$$\frac{dS}{dt} = N \cdot v = 0 \tag{4}$$

By applying Metabolic Flux Analysis (MFA), one tries to shrink the possible solution space of Eq. 4 by measuring some of the reactions rates, such as uptakes or excretion rates, in steady state. Ideally, one unique solution remains for the actual flux distribution. In general MFA is useful to analyze specific flux distributions but it is not able to characterize the complete admissible steady state solution space.

With the metabolite balancing equation, resulting from steady state assumption, the space of all possible flux distributions in a reaction network is studied in a constraint-based modelling approach.

In order to deduce what the metabolic network suggests, recent research has centered on approaches such as Extreme Pathways (EPs), Elementary Flux Mode (EFM) analysis, Flux Balance Analysis (FBA) and a number of other constraint-based modelling methods. Pathway analysis deals with the discovery and analysis of feasible routes in metabolic networks.

3.3 Finding Feasible Pathways Through Network

EFMs and EPs closely match to the same approach; they are the smallest sub-networks that allow a metabolic reconstruction network to function at steady state [14, 15]. An EFM is the minimal biochemical pathway that, at steady state, catalyzes a set of net reactions between input(s) and output(s). EFMs characterize the complete space of admissible steady-state flux distributions by particular flux vectors. The EFMs method provides a rigorous formalism for describing and assessing metabolic pathways. According to Stelling et al. [15] EFMs can be used to understand cellular objectives for the overall metabolic network. Furthermore, EFMs analysis takes into account stoichiometries and thermodynamics when evaluating whether a particular metabolic route or network is feasible and likely for a set of proteins/enzymes [14].

Its ability to assess the functional and structural properties of metabolic networks means that EFM analysis is suitable for both biotechnology and physiology. EFMs can be used to recognize operational modes and cycles. EFM analysis has already been used for microbiological systems in metabolic engineering [16]. In plants, EFM analysis appears to be a useful theoretical tool for structural modelling. It has been successfully applied to biological systems characterized by a high redundancy level without any key (i.e., rate limiting) enzyme, e.g., the Calvin cycle [17], sucrose metabolism [18], and starch metabolism [19]. The main difference between EPs and EFMs is the way to code reversibility of reactions. The EPs method requires to double reversible reaction in two irreversible ones, one backward and one forward. Actually, each EP can be mapped onto a corresponding EFM while the inverse is not true. Moreover, if a reaction is deleted the subset of EFMs not involving this reaction is the complete set of EFMs in the reduced network [14]. In contrast the set of EPs needs to be recalculated whenever a partial reaction is removed or the direction modified. Considering the small network in Fig. 1a, four Elementary Flux Modes are computed and displayed in Fig. 3, with bold arrays.

Sucrose is the only one entry of the network, then the uptake of sucrose and its cleavage in glucose and fructose (corresponding to the reactions *Vsuc_up* and *Vinv*) are present in the four EFMs. Then these sugars can be stored in vacuole (reactions *Vac_fru* and *Vac_glc*, Fig. 3a) or both of them metabolized in fructose-1-6-bisphosphate (Fig. 3d). The EFMs in Fig. 3b, c are a mix of these two situations. Note that the irreversibility of the kinase reactions (*Vfk* and *Vgk*) constraints the network and prevent the reversible reaction *Vpgi* to be active in the forward the direction. But in real biological systems, two other reactions (catalyzed by the sucrose synthase and the sucrose phosphate synthase) are present and lead to a cycle of sucrose in this pathway.

3.4 Other Metabolic Analysis

3.4.1 Minimal Cut Sets

EFMs computation provides the set of feasible pathways of functional fluxes. Conversely, Minimal Cut sets (MCs) method allows computing a dual set, i.e., the set of reactions which disconnects the network and so disable the fluxes. Klamt and Gilles [20] and Ballerstein et al. [21] proposed an algorithm to compute the cut sets respecting the constraints of metabolic networks design: reaction reversibility/irreversibility and steady state assumption. Computing the MCs for the simple example described in Fig. 1a leads to this list of six MCs: 1: VSuc_up, 2: Vinv, 3: Vgk, Vac_glc, 4: Vpfk, Vac_glc, 5: Vfk, Vac_fru and 6: Vpfk, Vac_fru. In this simple network, the number of MCs is higher than that of EFMs. But this method takes advantage in the case of large networks. In this case, the number of MCs is expected to be smaller than the number of EFMs. However, at this time just a few networks have been explored with this approach.

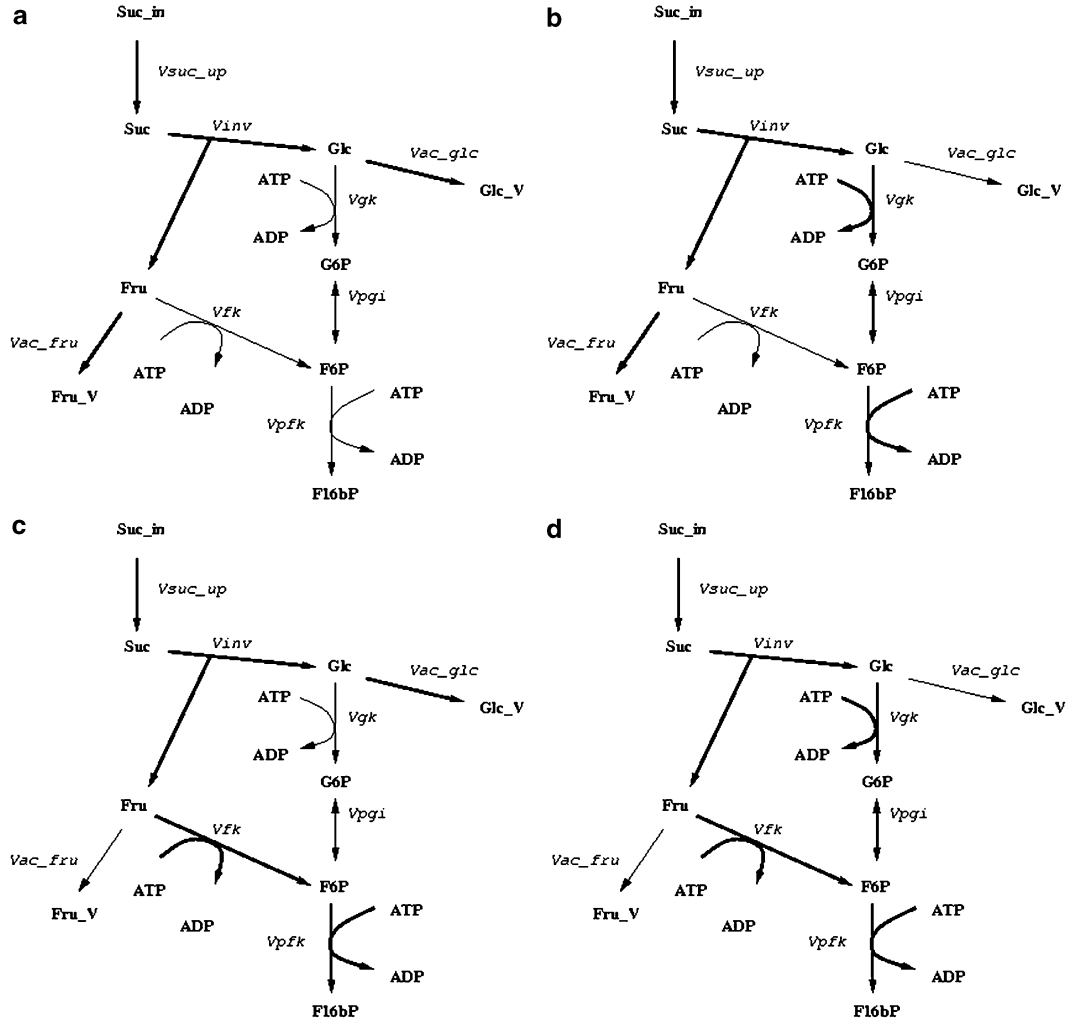


Fig. 3 The four Elementary Flux Modes of the network displayed in Fig. 1

3.4.2 Flux Balance Analysis

Another way to analyze metabolic networks is to simulate fluxes by Flux Balance Analysis (FBA). Linear programming is usually used to obtain the maximum potential of the objective function that you are looking at, and therefore, when using flux balance analysis, a single solution is found to the optimization problem [15]. In a flux balance analysis approach, exchange fluxes are assigned to those metabolites that enter or leave the particular network only. Those metabolites that are consumed within the network are not assigned any exchange flux value. Also, the exchange fluxes along with the enzymes can have constraints ranging from a negative to positive value (e.g., −10 to 10).

FBA can highlight the most effective and efficient pathway through the network in order to achieve a particular objective function. In addition, gene knockout studies can be performed

using flux balance analysis. The enzyme that correlates to the gene that needs to be removed is given a constraint value of 0. Then, the reaction that the particular enzyme catalyzes is completely removed from the analysis. Flux balance modelling which the main advantage is that it is relatively easy to scale up to cover very large networks can also be applied to study genome-scale metabolic models. FBA is explored deeply in Chapter 17 of this book.

3.4.3 Kinetic Modelling

In order to perform a dynamic simulation with such a network it is necessary to construct an ordinary differential equation system that describes the rates of change in each metabolite's concentration or amount. To this end, a rate law, i.e., a kinetic equation is required for each reaction. Often these rate laws contain kinetic parameters with uncertain values. In many cases it is desired to estimate these parameter values with respect to given time-series data of metabolite concentrations. The system is then supposed to reproduce the given data. For this purpose the distance between the given data set and the result of the simulation, i.e., the numerically or in few cases analytically obtained solution of the differential equation system is computed. The values of the parameters are then estimated to minimize this distance. One step further, it may be desired to estimate the mathematical structure of the differential equation system because the real rate laws are not known for the reactions within the system under study. This part is detailed in Chapter 16 of this book.

4 Example: Heterotrophic Plant Cells

Finally, we illustrate in this chapter an example of network reconstruction from heterotrophic plant cells metabolism and its topological analysis leading to a huge number of EFMs.

4.1 Description of the Heterotrophic Plant Cell Network

Let us first look at the metabolic network of heterotrophic plant cells described in Fig. 4 and appendix. The reconstruction of this metabolic network is detailed in our previous paper [22]. The network includes all the main pathways of the central carbon metabolism in plants: glycolysis (black), the TCA cycle (blue), the pentose phosphate pathway (pink), the starch and sucrose pathways (green) and the storage reactions towards the vacuole (brown) described in AraCyc and in relevant papers [23–25]. Irreversible reactions (see Appendix) are indicated by unidirectional arrows in Fig. 4. The glycolytic pathways are illustrated with reversible glycolysis in the cytosol and with irreversible glycolysis in plastids because amyloplasts lack fructose-1,6-bisphosphatase [26, 27]. Due to its autotrophic nature, the plant synthesizes its own respiratory substrates (mainly carbohydrates) which then serve as

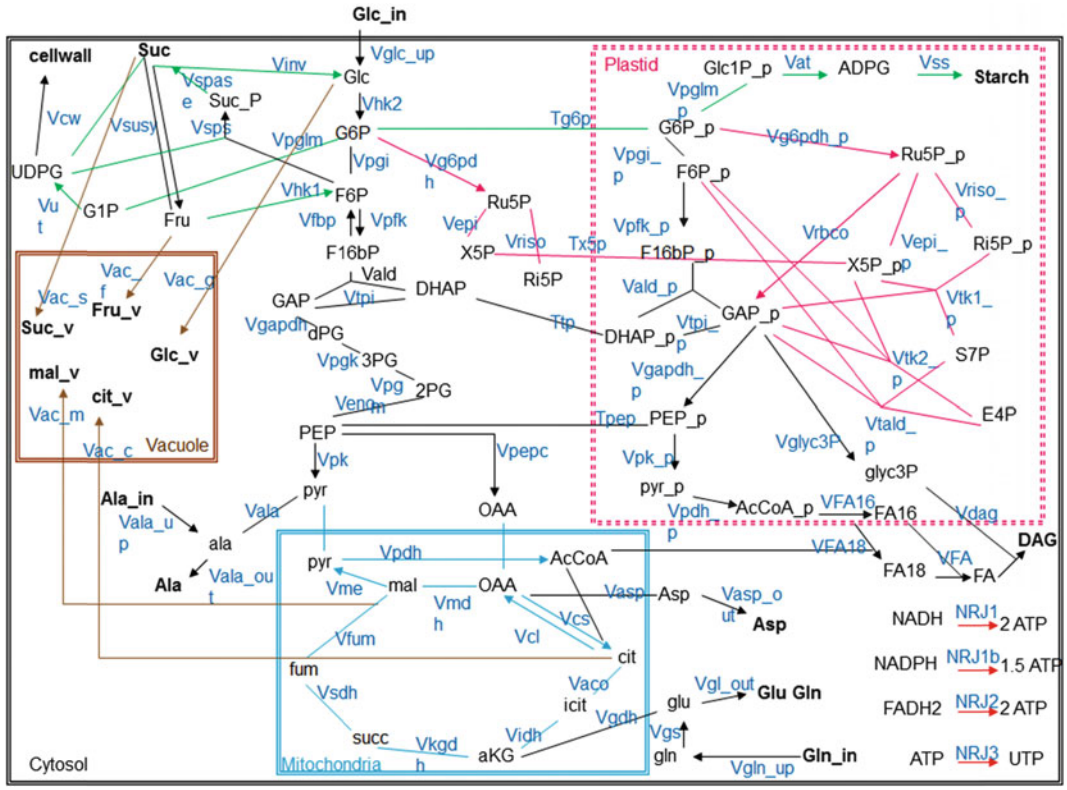


Fig. 4 Metabolic network of heterotrophic plant cells. From Beurton-Aimar et al. (2012) [22]. Freely available for redistribution

substrates for the TCA cycle. The TCA cycle provides precursors for several biosynthetic processes, such as nitrogen fixation and biosynthesis of amino acids [28]. The pentose phosphate pathway includes the irreversible oxidative branch, whereas the non-oxidative branch is reversible (recycling of pentose-phosphates from fructose phosphate and triose-phosphate). The enzymes of the oxidative branch, which also exist in cytosol, could lead to the synthesis of pentose phosphates in both cytosol and plastids [29, 30]. In the starch and sucrose pathways, sucrose is metabolized in cytosol, whereas starch is metabolized in plastids from imported hexose phosphates (G1P or G6P). Several effluxes are illustrated in Fig. 4: protein synthesis from several amino acids (glutamate and glutamine, aspartate and alanine), lipid synthesis (diacyl glycerol) from plastidial pyruvate and trioses, synthesis of cell wall polysaccharides from UDP-glucose, sugars (glucose, fructose and sucrose) and storage of organic acids (malate and citrate) in vacuoles. The energy reactions (NRJ, red arrows in Fig. 4) are essential to balance the cofactors of the system. Subcellular compartments, such as mitochondria and plastids, can lead to potentially reversible transport of metabolites such as G6P, X5P, PEP, and DHAP [30–33].

There are 70 different metabolites in this network at steady state. Among them, 15 metabolites are external, meaning that they are carbon sources or carbon sinks (nutrients, waste products, stored and excreted products, and precursors for further transformation). These are exogenous glucose and amino acids (glutamine and alanine), CO_2, sugars (sucrose, glucose and fructose) and organic acids (citrate and malate) stored in vacuoles, amino acids for protein synthesis (aspartate, alanine, glutamate and glutamine), cell wall polysaccharides, starch and lipids (cf. Appendix). Assuming that their concentration cannot affect the system, the other small molecules (e.g., oxygen, ammonium, phosphate, pyrophosphate and water) are not included in the metabolic network. The other 55 metabolites, including cofactors (ATP, NADH, NADPH, and $FADH_2$) are internal, which means that they are expected to be balanceable at steady state. For the sake of simplicity, oxidized equivalents (NAD, NADP, and FAD) and ADP are omitted in the reaction equations. As mentioned above, the network contains 78 reactions. The direction (reversibility or irreversibility) of the 78 reactions is derived from thermodynamic properties, 33 reactions are reversible.

4.2 Consistency and Properties of the Network

This step of analysis ensures that all the reactions present in the reconstruction are properly balanced. In order to validate the reconstruction, it is important to check the network consistency. That means to detect the blocked reactions or the missing elements which can compromise the use of the network model. The network consistency can be detected by analytical methods. Tools like CellNetAnalyzer (CNA, Matlab, http://www.mpi-magdeburg.mpg.de/projects/cna/cna.html) allow checking the network consistency. Similarly, the important reactions (the model is modified if these reactions are removed) or the correlated reactions (reactions operating together) can be detected first in the network, with analytical methods.

From the tools available for EFM computation, the original METATOOL software (module 5.0.3 beta in CellNetAnalyzer), a program with a graphical user interface for the analysis of metabolic networks based on Matlab (Mathworks, Inc.), was used for computation [34]. The result of the computation on the metabolic network of heterotrophic plant cells described in Fig. 4 is a set of 114,614 EFMs. Because each EFM represents a unique metabolic route, the 114,614 EFMs indicate a large number of metabolic routes that are potentially active at metabolic steady state. Thus, a robust network of high redundancy is obtained. Note that even though the number of EFMs is already quite big, Klamt and Stelling [35] showed that this number is actually smaller than the strict combinatorial computation of the metabolic reactions. Moreover, after MCS computing, 93,009 Minimal Cut Sets have been obtained providing a smaller set of data than the set of EFMs.

EFM analysis is very helpful to distinguish whether the EFM is producing or consuming ATP. Thus, it allows us to visualize futile cycles, as well as catabolic and anabolic EFMs. The analysis revealed that futile cycles exist between sucrose and hexose, as well as between hexose phosphates and triose phosphates. METATOOL software provides a list of external metabolites for each EFM. Although each EFM is unique, several EFMs lead to the same overall metabolite. This means that a specific external metabolite can be obtained from several different EFMs. This multiplicity is an indicator of the robustness of the metabolic network.

We used the results of the EFM analysis of the metabolic network of heterotrophic plant cells (about 80 reactions and 70 metabolites from the main pathways of the central carbon metabolism) in comparison with the experimental fluxes determined in *Brassica napus* embryos using 13C labelling measured with a glucose and organic/inorganic nitrogen source [36], by introducing the calculated coefficients of flux efficiency [22]. The main experimental results were cross-validated our EFM analysis so this work has shown that 13C-MFA experimental results can be qualitatively validated by the EFM analysis.

5 Conclusion

To sum up, a reconstruction that is fully accurate can lead to greater insight about understanding the functioning of the organism of interest. A reconstruction model serves as a first step to deciphering the complicated mechanisms. Metabolic network reconstruction and analysis can be used to understand how a cell functioning. Pathway analysis suggests finding hypothetical routes between specific inputs and outputs that only emerge in the context of complex network, and Elementary Flux Modes allow identifying all feasible routes in the network.

In conjunction with "classical" biochemical knowledge, the huge amount of data produced by the sequencing of complete genomes can be used for the reconstruction of metabolic networks. Metabolic models can be reconstructed at a genome-scale using all the reactions catalyzed by the enzymes encoded in an annotated genome (see Chapter 19). Henceforth, by sequence comparison methods and genome context methods, for greater than half of the newly sequenced genes, the physiological functions can be predicted. With such information available, pathway analysis (mainly FBA) can be performed at the scale of whole-cell models. However this remains a nontrivial task: and only few papers deal with genome_scale models of plant, with genome scale models [1, 37–39]. Several problems arise in the construction of metabolic models from genome annotation databases, including network gaps caused by incomplete or imprecise genome annotation,

mass-balance errors caused by reaction stoichiometry errors in the annotation database, or the presence of excess, non-functional reactions. Another difficulty is that genome annotation databases contain no information about reaction directionality.

Appendix

List of reactions corresponding to the network in Fig. 4:

Glc_up : Glc_in => Glc .
ala_up : ala_in => ala .
gln_up : gln_in => gln .
Vpgi : G6P <=> F6P .
Vhk1 : ATP + Fru => F6P .
Vhk2 : ATP + Glc => G6P .
Vpfk : ATP + F6P => F16bP .
Vfbp : F16bP => F6P .
Vald : F16bP <=> DHAP + GAP .
Vtpi : DHAP <=> GAP .
Vgapdh : GAP <=> dPG + NADH .
Vpgk : dPG <=> 3PG + ATP .
Vpgm : 3PG <=> 2PG .
Veno : 2PG <=> PEP .
Vpk : PEP => ATP + pyr .
Vpepc : CO2 + PEP => OAA .
Vpdh : pyr => CO2 + AccoA + NADH .
Vcl : cit + ATP => OAA + AccoA .
Vcs : OAA + AccoA => cit .
Vaco : cit <=> icit .
Vidh : icit <=> aKG + CO2 + NADH .
Vkgdh : aKG <=> CO2 + NADH + succ .
Vsdh : succ <=> FADH2 + fum .
Vfum : fum <=> mal .
Vmdh : mal <=> NADH + OAA .
Vme : mal => CO2 + NADH + pyr .
Vg6pdh : G6P => CO2 + Ru5P + NADPH .
Vepi : Ru5P <=> X5P .
Vriso : Ru5P <=> Ri5P .
Vpgi_p : G6P_p <=> F6P_p .
Vg6pdh_p : G6P_p => CO2 + Ru5P_p + NADPH .
Vepi_p : Ru5P_p <=> X5P_p .
Vriso_p : Ru5P_p <=> Ri5P_p .
Vtk1_p : S7P_p + DHAP_p <=> Ri5P_p + X5P_p .
Vtk2_p : F6P_p + DHAP_p <=> E4P_p + X5P_p .
Vtald_p : S7P_p + DHAP_p <=> E4P_p + F6P_p .
Vpfk_p : ATP + F6P_p => F16bP_p .
Vald_p : F16bP_p <=> GAP_p + DHAP_p .

Vtpi_p : GAP_p <=> DHAP_p .
Vgapdh_p : DHAP_p => ATP + NADH + PEP_p .
Vpk_p : PEP_p => ATP + pyr_p .
Vrbco : Ru5P_p + CO2 => 2 DHAP_p .
Vpdh_p : pyr_p => CO2 + AccoA_p + NADH .
Tg6p : G6P <=> G6P_p .
Ttp : DHAP <=> DHAP_p .
Tpep : PEP <=> PEP_p .
Tx5p : X5P <=> X5P_p .
Vpglm_p : G1P_p <=> G6P_p .
Vat : G1P_p + ATP => ADPG .
Vss : ADPG => starch .
Vpglm : G1P <=> G6P .
Vut : G1P + UTP => UDPG .
Vsusy : Fru + UDPG <=> Suc .
Vinv : Suc => Fru + Glc .
Vsps : F6P + UDPG => SucP .
Vspase : SucP => Suc .
Vala : pyr + glu <=> ala + aKG .
Vasp : OAA + glu <=> asp + aKG .
Vgdh : aKG + NADH <=> glu .
Vgs : gln + aKG + NADPH => 2 glu .
Vgl_out : glu => gl_out .
Vasp_out : asp => asp_out .
Vala_out : ala => ala_out .
Vcw : UDPG => CellWall .
VFA16 : 8 AccoA_p + 7 ATP + 14 NADPH => FA16 .
VFA18 : AccoA + FA16 + ATP + NADPH => FA18 .
VFA : FA16 + 2 FA18 => 3 FA .
Vglyc3P : DHAP_p + NADH + ATP => glyc3P .
Vdag : glyc3P + ATP + 2 FA => DAG .
Vac_g : Glc => Glc_v .
Vac_f : Fru => Fru_v .
Vac_s : Suc => Suc_v .
Vac_m : mal => mal_v .
Vac_c : cit => cit_v .
NRJ1 : NADH => 2 ATP .
NRJ1b : 2 NADPH => 3 ATP .
NRJ2 : FADH2 => 2 ATP .
NRJ3 : ATP => UTP .

References

1. Poolman M, Miguet L, Sweetlove LJ, Fell D (2009) A genome-scale metabolic model of Arabidopsis and some of its properties. Plant Physiol 151:1570–1781
2. The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485: 635–641

3. Francke C, Siezen RJ, Teusink B (2005) Reconstructing the metabolic network of a bacterium from its genome. Trends Microbiol 13(11):550–558
4. Junker BH, Schreiber F (eds) (2008) Analysis of biological networks. Global network properties. Wiley, Hoboken, pp 31–59
5. Fell DA, Wagner A (2000) The small-world of metabolism. Nat Biotechnol 18:1121–1122
6. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL (2000) The large-scale organization of metabolic network. Nature 407: 651–654
7. Wagner A, Fell DA (2001) The small-world inside large metabolic networks. Proc Roy Soc Lond B 268:1803–1810
8. Barabasi AL, Oltvai ZN (2004) Network biology. Nat Rev Genet 5:101–113
9. Gillepsie DT (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81(25):2340–2361
10. Srivatasta R, Peterson MS, Bentley WE (2001) Stochastic kinetic analysis of the *Escherichia coli* stress circuit using sigma(32)-targeted antisense. Biotechnol Bioeng 75 (1):120–129
11. Trinh CT, Wlaschin A, Srienc F (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl Microbiol Biotechnol 81:813–826
12. Seressiotis A, Bailey JE (1988) MPS: an artificially intelligent software system for the analysis and synthesis of metabolic pathways. Biotechnol Bioeng 31:587–602
13. Heinrich R, Schuster S (1996) The regulation of cellular systems. Springer, New York
14. Schuster S, Fell DA, Dandekar T (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol 18(3):326–32
15. Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED (2002) Metabolic network structure determines key aspects of functionality and regulation. Nature 420:190–193
16. Schuster S, Dandekar T, Fell DA (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol 17:53–60
17. Poolman MG, Fell DA, Raines CA (2003) Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. Eur J Biochem 270:430–439
18. Rohwer JM, Botha FC (2001) Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data. Biochem J 358:437–445
19. Poolman MG, Assmus HE, Fell DA (2004) Applications of metabolic modelling to plant metabolism. J Exp Bot 55:1177–1186
20. Klamt S, Gilles ED (2004) Minimal cut sets in biochemical reaction networks. Bioinformatics 20(2):226–234
21. Ballerstein K, von Kamp A, Klamt S, Haus U (2012) Minimal cut sets in a metabolic network are elementary modes in a dual network. Bioinformatics 28:381–387
22. Beurton-Aimar M, Beauvoit B, Monier A et al (2011) Comparison between elementary flux modes analysis and 13C-metabolic fluxes measured in bacterial and plant cells. BMC Syst Biol 5:95
23. Dieuaide-Noubhani M, Raffard G, Canioni P et al (1995) Quantification of compartmented metabolic fluxes in maize root-tips using isotope distribution from C13 labeled or C14 labeled glucose. J Biol Chem 270: 13147–13159
24. Rontein D, Dieuaide-Noubhani M, Dufourc EJ et al (2002) The metabolic architecture of plant cells. Stability of central metabolism and flexibility of anabolic pathways during the growth cycle of tomato cells. J Biol Chem 277:43948–43960
25. Schwender J, Ohlrogge JB, Shachar-Hill Y (2003) A flux model of glycolysis and the oxidative pentosephosphate pathway in developing Brassica napus embryos. J Biol Chem 278:29442–29453
26. Entwistle G, ap Rees T (1988) Enzymic capacities of amyloplasts from wheat (Triticum aestivum) endosperm. Biochem J 255:391–396
27. Borchert S, Harborth J, Schunemann D, Hoferichter P, Heldt HW (1993) Studies of the enzymic capacities and transport properties of pea root plastids. Plant Physiol 101: 303–312
28. Hill SA, ap Rees T (1994) Fluxes of carbohydrate metabolism in ripening bananas. Planta 192:52–60
29. Debnam PM, Emes MJ (1999) Subcellular distribution of enzymes of the oxidative pentose phosphate pathway in root and leaf tissues. J Exp Bot 50:1653–1661
30. Eicks M, Maurino V, Knappe S, Flügge UI, Fischer K (2002) The plastidic pentose phosphate translocator represents a link between the cytosolic and the plastidic pentose phosphate pathways in plants. Plant Physiol 128: 512–522
31. Kammerer B, Fischer K, Hilpert B et al (1998) Molecular characterization of a carbon

transporter in plastids from heterotrophic tissues: the glucose 6-phosphate/phosphate antiporter. Plant Cell 10:105–118

32. Fischer B, Kammerer M, Gutensohn B et al (1997) A new class of plastidic phosphate translocators: a putative link between primary and secondary metabolism by the phosphoenolpyruvate/phosphate antiporter. Plant Cell 9:453–462
33. Kang F, Rawsthorne S (1994) Starch and fatty acid synthesis in plastids from developing embryos of oilseed rape (*Brassica napus* L.). Plant J 6:795–805
34. Pfeiffer T, Sanchez-Valdenebro I, Nuno JC, Montero F, Schuster S (1999) METATOOL: for studying metabolic networks. Bioinformatics 15:251–257
35. Klamt S, Stelling J (2002) Combinatorial complexity of pathway analysis in metabolic networks. Mol Biol Rep 29:233–236
36. Junker BH, Lonien J, Heady LE, Rogers A, Schwender J (2007) Parallel determination of enzyme activities and in vivo fluxes in Brassica napus embryos grown on organic or inorganic nitrogen source. Phytochemistry 68: 2232–2242
37. de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK (2010) AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. Plant Physiol 152:579–589
38. Radrich K, Tsuruoka Y, Dobson P, Gevorgyan A, Swainston N, Schwartz JM (2010) Reconstruction of an in silico metabolic model of Arabidopsis thaliana through database integration. BMC Syst Biol 4:114
39. Saha R, Suthers P, Maranas C (2011) Zea mays irs1563: a comprehensive genome-scale metabolic reconstruction of maize metabolism. PLoS ONE 6(7):e21784