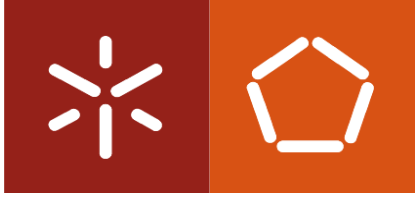


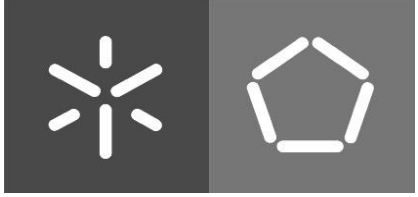
Universidade do Minho

Mestrado em Bioinformática

Brígida de Meireles

Exploring chemical conversions in metabolic networks by tracing atoms transition

Janeiro de 2015



Universidade do Minho

Mestrado em Bioinformática

Brígida de Meireles

Exploring chemical conversions in metabolic networks by tracing atoms transition

Supervisors:

Sónia Madalena Azevedo Carneiro, PhD

Rafael do Castro Carreira

Janeiro de 2015

ACKNOWLEDGMENTS

Agradeço aos meus orientadores toda a ajuda e apoio que me deram, pois sem eles eu não teria conseguido realizar este trabalho.
Aos meus amigos e familiares que sempre me aturaram e ajudaram quando precisei, principalmente aos meus pais, irmão, avós, Vinícius, Juliana, Ana, Sofia, Célia, Luciano, Liu, States, Zenha e Sara.

Muito obrigada a todos.

ABSTRACT

Understanding the metabolic processes that occur in a cell has been fundamental to rationally design organisms for the production of a desired metabolite. The improvement in biosynthesis of certain metabolites can be supported by tracing the atoms transition in different chemical conversions, and this way tackle which metabolic pathways are more efficient or most promising for manipulation.

In this work, a methodology to construct metabolic maps in which the carbon atom transitions are described was implemented. The carbon atom transition map for the central metabolism of *Escherichia coli* and *Actinobacillus succinogenes* were constructed. MetaCyc and KEGG databases were explored to obtain the metabolic information necessary to construct these maps and provided support for validation at some level.

RESUMO

Compreender os processos metabólicos que ocorrem numa célula tem sido essencial para o desenho de estirpes capazes de sintetizar produtos de elevado interesse. A síntese de um determinado metabolito pode ser avaliada pela transição de átomos ao longo das diversas conversões químicas numa ou mais vias metabólicas, de modo a identificar quais as mais eficientes ou mais promissoras para manipulação.

Neste trabalho foi implementada uma metodologia para a construção de mapas de transição dos átomos de carbono, que pode ser usado conjuntamente com modelos metabólicos. Neste caso, foram construídos os mapas de transição de átomos de carbono para o metabolismo central da *Escherichia coli* e da *Actinobacillus succinogenes.* As bases de dados MetaCyc e KEGG foram exploradas para obter as informações metabólicas necessárias à construção destes mapas.

LIST OF ABBREVIATIONS

^{13}C-MFA	^{13}C-metabolic flux analysis
2PG	2-phosphoglycerate
3PG	3-phosphoglycerate
6PG	6-phospho gluconate
AcCoA	Acetyl-CoA
Ace	Acetate
AcP	Acetyl phosphate
Acp	Acceptor
Ald	Acetaldehyde
ATM	Atom transition model
cis-Aco	cis-Aconitate
Cit	Citrate
CO2	Carbon dioxide
DHAP	Glycerone phosphate
E4P	Erythrose-4-phosphate
EtOH	Ethanol
F1,6P	Fructose-1,6-bisphosphate
F6P	Fructose-6-phosphate
FBA	Flux balance analysis
For	Formate
Fum	Fumarate
G1,3P	Glycerate-1,3-diphosphate
G1,5L6P	Glucono-1,5-lactone-6-phosphate
G3P	Glyceraldehyde-3-phosphate
G6P	Glucose-6-phosphate
GLC	Glucose
ID	Unique-identifier
Iso	Isocitrate
IUPAC	International Union of Pure and Applied Chemistry
KEGG	Kyoto Encyclopedia of Genes and Genomes

MAL	Malate
MFA	Metabolic flux analysis
NC-IUBMB	Nomenclature Committee of the International Union of Biochemistry and Molecular Biology
NCBI	National Center for Biotechnology Information
OAA	Oxaloacetate
PEP	Phosphoenolpyruvate
Pi	Orthophosphate
Pyr	Pyruvate
R5P	Ribose-5-phosphate
Rbu5P	Ribulose-5-phosphate
S7P	Sedoheptulose-7-phosphate
Suc	Succinate
SucCoA	Succinyl-CoA
TCA	Tricarboxylic acid cycle
X5P	Xylulose-5-phosphate
αKG	2-oxoglutarate

TABLE OF CONTENTS

1. SCOPE AND MOTIVATION

The cellular metabolism is a highly coordinated system consisting of two basic types of biomolecules: enzymes and metabolites, which interact to provide energy and building blocks for cellular growth, maintenance and other activities (Jeong et al. 2000; Nelson & Cox 2008). This system is often represented as a metabolic network that is a collection of biochemical reactions that represent metabolic conversions between chemical compounds. These networks can be used to explore the production of a biochemical compound from a given set of starting compounds through several biochemical reactions (Alberts et al. 2014). The analysis of those metabolic conversions is very important to understand the cellular metabolism and to explore novel metabolic pathways for the synthesis of chemical compounds. The possibility of promoting the production of a compound in a certain biological system is very appealing to the industry.

With advances in biochemical techniques, chemical reactions occurring in biological systems have been characterized and catalogued within metabolic databases, such as MetaCyc (Caspi et al. 2010) and KEGG (Kyoto Encyclopedia of Genes and Genomes (Goto et al. 2000)); and associated with metabolic pathways, like glycolysis, pentose phosphate and the tricarboxylic acid cycle, usually know as TCA (Papin et al. 2003). In addition, "-omics" data are expanding information regarding these molecular interactions, providing snapshots of the metabolism under specific conditions, increasing the biochemical and physiological information available (Durot et al. 2009).

In this work, the identification and investigation of chemical reactions and their metabolic information was central, thus databases and information resources mentioned above were key resources. Furthermore, the reconstruction of a metabolic network is based in the genomic, chemical and network information provided by these bioinformatics web resources, therefore by exploring those a compilation of a wide range of metabolic information may be fulfilled. The analysis of the reconstruction of a metabolic network is very important to define strategies to improve or inhibit the production of a metabolite.

For a better understanding of metabolic systems, the network structure and flux distribution analyses of metabolic networks have been explored (Stefan et al., 2012; Trang et al., 2012; Kurata et al., 2007; Chen et al. 2009). In particular, isotopically labeled substrates have been used to trace the transition of atoms and estimate the flux distribution in a metabolic network. The use of stable isotopes, such as ^{13}C, allows tracing the carbon atoms among various metabolic intermediates within a metabolic pathway or between pathways.

In such studies, the representation of atoms transitions in biochemical reactions is essential to investigate how the carbon labeling is distributed throughout the network. Therefore, a methodology for constructing atom transition maps corresponding to the metabolic reactions comprised in these networks was developed. These maps will, ultimately, integrate procedures for the flux analysis in model microorganisms (e.g. *Escherichia coli* (*E. coli*)) through ^{13}C-isotope labeling.

2. OBJECTIVES

This work aims to implement a methodology to construct atom transition maps that can be used to trace atom transition patterns in metabolic networks. To achieve this goal it is crucial to explore the chemical conversions associated with these metabolic activities and assess the changes in atom's position among the metabolic intermediaries involved in those conversions.

Databases such as KEGG and MetaCyc allow to extract specific data regarding the transition of atoms from substrates to products in biochemical reactions. The main objectives in this work were to construct a metabolic atom transition map for the central carbon metabolism of *E. coli,* which was compared to a previously published map, and then, to construct an atom transition network for the central carbon metabolism of *Actinobacillus succinogenes (A. succinogenes)*. The first case-study is intended to validate the implemented methodology to construct atom transition maps, while the second provides evidence for the application of this methodology to other case-studies.

3. INTRODUCTION

3.1. GENOME-SCALE MODELS

The cellular metabolism is a global network of all biochemical reactions, which is based on molecular interactions in metabolic processes (Nelson & Cox 2008). Systems biology studies the interactions between metabolites, enzymes and regulatory factors, and how they influence the behavior of a cell. Through computational methods, systems biology is able to model the metabolic system and predict the effect of these specific interactions in cells phenotype. The metabolism is very coordinated, complex and a dynamic process, where the enzymatic reactions convert metabolites into others that are necessary to support cell survival. Therefore the identification and representation of all metabolic activities of a system are fundamental to understand the metabolism of a cell and how it can be controlled. Using this information, the metabolic engineering attempts to select key metabolic targets and manipulate them to produce value-added compounds (Blazeck & Alper 2010).

Recent findings in the systems biology field support that robustness of many cellular processes is engrained in the dynamic interactions among its constituents, like proteins, DNA and small molecules. Improved capability to identify these constituents that participate in such a complex system has been fundamental. Genomic sequencing and metabolic information were necessary to integrate pathway-genome that provide connectivity maps of metabolic and cellular networks (Jeong et al. 2000). Through the progressive advances in genetics and high-throughput techniques, metabolic information has been growing, precisely characterized and catalogued (Papin et al. 2003; Bairoch et al. 2000). These catalogues has been helping to identify the constituents and how their interactions can be manipulated (Patil et al. 2004). Whether it is intended to increase or decrease the production of a particular component or metabolite in a cell, it is necessary to identify the metabolic processes involved and understand their limitations (Schilling et al. 1999) (Heath et al. 2010).

Many organisms have now their genomes sequenced, particularly microorganisms that are considered industrially relevant. For instance, *E. coli* is

the most characterized and studied bacterium that is used to produce recombinant proteins, such as biosynthetic human insulin (Goeddel et al. 1979). A first *E. coli* model was developed to predict the cellular phenotype under different conditions (Edwards & Palsson 2000). Similarly, the genome of *Saccharomyces cerevisiae* (*S. cerevisiae*), a yeast commonly used in industry for ethanol, alcoholic beverages, leavening production and also considered a food additive for human (Lin & Tanaka 2006), has been completely sequenced since 1997 (Goffeau et al. 1996). Since yeasts are capable to produce several compounds of industrial interest and are easy to manipulate and maintain on a large scale in laboratory, there was an interest to have a complete genome sequence and to assess their gene functions, namely metabolic capabilities. This information has been also used to reconstruct genome-scale metabolic models that are used to represent and model the cellular metabolism (Patil et al. 2004; Nielsen & Olsson 2002; Nelson & Cox 2008).

Due to their industrial relevance, these and other genome-scale metabolic models have been developed and widely applied within Metabolic Engineering projects. For instance, Cintolesi and co-workers (Cintolesi et al. 2012) have identified key metabolic targets in *E. coli* that can be manipulated in order to enhance the production of glycerol.

Metabolic Engineering approaches are used to find the best alterations in the complex metabolic network that may increase the production of a specific target, either by introducing genetic controls that allow to manipulate the level of expression of several enzyme-coding genes or deleting those genes (Blazeck & Alper 2010). These approaches have received increasing interest due to some successful applications in industrial biotechnology (Rocha et al. 2010; Patil et al. 2004). For example, microalgae have the ability to produce hydrocarbons and molecular hydrogen in large quantities (Zaslavskaia et al. 2001). For this reason it is considered a potential source for renewable fuels. With different strains and growth conditions, up to 75% of the biomass can be hydrocarbons (Gavrilescu & Chisti 2005). Another example is pharmaceutical synthesis. In the work developed by Savile *et al*. (Savile et al. 2010) a chemical process for an efficient

biocatalytic to product sitagliptin, an antidiabetic compound, in a large scale was replaced.

3.2. METABOLIC PATHWAYS AND DATABASES

The cellular metabolism is the collection of many interconnected reactions sequences that are regulated by some enzymes in order to provide the needs of a cell. The enzymes work together in sequential pathways to catalyze the reactions to obtain specifics metabolites. The enzymes are present as multi-enzyme complexes in the cell, responsible for important transformations in organic chemistry and biocatalysts, greatly enhancing the rate of specific reactions in the cell (Drauz et al. 2012; Nelson & Cox 2008).

The enzymes are classified according to the reactions they catalyze. The name of each enzyme has the suffix "-ase" preceded by the name of the main substrate or a term that describes its general activity. An international classification, also known as Enzyme Commission number (E.C. number), has be adopted to categorize enzymes into six classes:

1. Oxidoreductases: oxidation-reduction
   Catalyze the transfer of electrons (hydride ions or H atoms) acting on the reduction of C=O and C=C groups; the reductive amination of C=O; oxidation of C-H, C=C, C-N, and C-O; cofactor reduction/oxidation;
2. Transferases: chemical group transfers
   Catalyze the transfer of functional groups such as amino, acyl, phosphoryl, methyl, glycosyl, nitro, and sulfur-containing groups;
3. Hydrolases: hydrolytic bond cleavages
   Catalyze the hydrolysis of chemical bonds (i.e., transfer of functional groups to water): hydrolysis of esters, amides, lactones, lactams, epoxides, nitriles, etc;
4. Lyases: nonhydrolytic bond cleavages
   Catalyze the bonds by addition or removal of the functional groups (for example C=C, C=N and C=O);
5. Isomerases: changes in arrangements of atoms in molecules

Catalyze the transfer of groups within molecules to yield isomeric forms – isomerizations, such as racemizations, epimerizations, and rearrangement reactions. The isomerization has an important role in the overall chemistry of some pathways, for example in the glycolysis, because the rearrangement of the carbonyl and hydroxyl group. The isomers can differ in the arrangement of their carbonyl group (α and β) or are mirror images of each other (L and D);

6. Ligases: joining together two or more molecules

   Catalyze the condensation of two molecules. Similar to lyase but enzymatically active only when combined with ATP cleavage (Copeland 2004).

Despite various enzymatic studies, it is still difficult to define the properties of most enzymes, but it is already evident that enzyme catalyzes chemical conversion of functional groups involving distinct mechanisms that are mostly conserved among biological systems (Ramanathan & Agarwal 2011).

Most of the required knowledge provided by experimental studies on metabolic networks to construct a metabolic model is scattered in literature and databases (see Table 3.1 and in Appendix A for further information). Particularly, databases are repositories with well-characterized pathways, reactions, enzymes, etc., that span most of the metabolic capabilities that we can find in nature (Blazeck & Alper 2010). Metabolic databases (Table 3.1) normally dispose biochemical reactions in pathways or modules that contextualize the metabolic capabilities of different organisms.

Many databases allow to access different levels of information, like KEGG that contains genomic, chemical and functional information associated with each molecular level within an organism system. This allows having a more comprehensive overview of the metabolic structure and coordination of most processes occurring in a particular system (Kanehisa et al. 2014).

Through the cataloging of the results of biochemical experiments, the reactions are associated to one or more pathways, like glycolysis, pentose phosphate and the TCA pathways, the metabolites are organized by reactions

and enzymes according to their catalytic role involving those metabolites (Papin et al. 2003; Bairoch 2000).

Table 3.1 – Databases most explored for metabolic data. Each database is represented with their name, their web address and with a small description.

Metabolic databases		
CheBI	Ebi.ac.uk/chebi	Dictionary of molecular entities (constitutionally or isotopically distinct atom, molecule, ion, ion pair, radial, radial ion, complex…) focused on "small" chemical compounds. This database incorporates an ontological classification (the relationships between molecular entities or classes of entities and their parents/children are specific) and follow the rules of international scientific bodies, IUPAC and NC-IUBMB;
PubChem	Pubchem.ncbi.nlm.nih.gov	Repository for chemical structures and their biological test results. It is organized as three linked databases: Substance contains contributed sample descriptions; Compound, contains unique chemical structures derived from subtracts; and BioAssay, contains biological tests. This database can be accessed through the NCBI Entrez system.
KEGG	Genome.jp/KEGG	Bioinformatic resource for understanding the functional meaning and the utilities of the cell or the organism from its genome information. This database contains information regarding molecular interaction networks like pathways and complexes – protein network (Pathway database), genes and proteins generated by genome projects – gene universe (Genes database) and biochemical compounds and reactions – chemical universe (Compound/Reaction database).
MetaCyc	Metacyc.org	Universal reference of metabolic pathways, verified experimentally, and enzymes information curated from the scientific literature. Is a non-redudant database of small-molecule metabolism and contains pathways involved in both primary and secondary metabolism, and their association to metabolites, reactions, enzymes and genes.

3.3. FLUX ANALYSIS

With the advent of "omics" technologies, like genomics, transcriptomics, proteomics, metabolomics and fluxomics, a new systems-wide knowledge has been generated. Systems biology attempts to understand the cellular metabolic networks and predict how metabolism behaves in particular conditions through mathematical modeling and computational methods (*in silico* simulation) (Park et al. 2009; Paul Lee et al. 2010). The increasingly large amounts of data

concerning metabolic networks provide the information required to computational identification of pathways with biological relevance, which are necessary to applications like metabolic engineering, metabolic network analysis and metabolic network reconstruction (Heath et al. 2010).

In order to adapt the huge quantity of data and integrate the genetic and biochemical processes, computer models are generated. The stoichiometric model is an example and a pioneer approach (Sanford et al., 2002).

Two main mathematical representations of metabolic models have been used to describe the cellular metabolism: kinetic models and stoichiometric models. Briefly, kinetic models consider the dynamic proprieties of metabolic reactions, combining stoichiometric information and kinetic parameters. On the other hand, the stoichiometric models define the metabolic network as a set of stoichiometric equations that represents all intracellular biochemical reactions (Christensen et al. 2002; Patil et al. 2004).

To quantitatively analyze the cellular metabolism it is necessary to understand how enzymes catalyze the reactions that compose the cellular metabolism. The catalysts can be controlled at the enzymatic level through inhibitors, often metabolites being produced by the catalyst itself. This and other regulatory layers need to be represented when the global control of cellular metabolism is described (Copeland 2004). Unfortunately, for many enzymes the availability of information regarding kinetic parameters is scarce, which is essential to estimate catalytic rates – fluxes (Winter & Krömer 2013).

A crucial feature to find meaningful metabolic pathways is the atom mapping. The atom mapping track the atoms through the metabolic networks and the tracking results are essential to obtain insights about meaningful and functional pathways. One of the applications of atom mapping is to assess the transition of a selected atom, usually carbon atoms, from one compound into another. It is useful to select pathways biologically relevant because eliminates false connections. The atoms mapping can also track atoms in the isotope labeling experiments that are used to elucidate metabolic pathways (Latendresse et al. 2012; Heath et al. 2010).

Without the measurement of the fluxes, the ability to quantitatively analyze the metabolism is limited. As a result, different approaches have been developed to estimate fluxes, such as constraint based metabolic flux analysis (MFA) and flux balance analysis (FBA) (Winter & Krömer 2013).

Flux analysis combines metabolic fluxes measurements with stoichiometric network models to determine the distribution of metabolic fluxes (Sauer 2006; McKinlay et al. 2007). MFA and FBA approaches assume a steady state to examine phenotypic alterations in a microorganism under different environmental and genetic conditions (Rocha et al. 2010). The MFA has been used to characterize the flux distribution, through internal fluxes estimation combining isotope labeling techniques and mathematical analysis. Normally, FBA is applied to estimate the cellular growth rate under certain conditions through a linear objective function (Covert et al. 2001). But, FBA can also use alternative objective function, such as the maximization of ATP or reducing power, and the maximization of a single biochemical reaction. Therefore, the objective function depending on the purpose can be applied to maximize the growth rate or minimize one product formation (Park et al. 2009).

FBA is able to determine an optimum metabolic flux distribution, without requiring details of enzyme kinetics or metabolite concentrations, in a genome-scale network, but has some limitations to predict fluxes for reversible reactions and parallel reaction (Wiechert 2001).

The metabolic fluxes are related to the relative isotopic abundances of different metabolites. Typically, carbon isotopes, either radioactive ^{14}C or stable ^{13}C (most used) are used to estimate these fluxes using mapping matrices, which trace those isotopes, also known as isotopomers, that are differently labeled variants of the same molecule. Originally it was exemplified as atom-mapping matrices that trace the carbon atom transitions from reactants to products (Latendresse et al. 2012). Most recently, it was simplified as isotopomer mapping matrices that enumerate all possible product isotopomers that can be created from each reactant isotopomers (Ravikirthi et al. 2011; Zupke & Stephanopoulos 1994).

^{13}C-metabolic flux analysis (^{13}C-MFA) has been developed to overcome some of the limitations found in FBA and stoichiometric MFA, namely the measurement of intracellular metabolic fluxes, the determination of fluxes in parallel pathways and reversible reactions (Crown & Antoniewicz 2013). With this approach, the input from carbon atom transitions along with measurements of metabolic fluxes provides information to impose restrictions to the optimization problem, allowing to obtain more consistent and reliable prediction of flux distributions (Wiechert 2001; Stephanopoulos et al. 1998). In the past few years, experiments with isotopic labelling have shown to be very valuable to analyze and discover metabolic pathways (Niittylae et al. 2009). Some examples are the work by Lanfermann and co-workers that uses ^{13}C-labelled precursors and ^{18}O-water to elucidate the metabolism of the amino acids and to prove the activity of an oxygenase in the pathway to sotolon (Lanfermann et al. 2014); also the work developed by Rattray and co-workers, where they explore the possible biosynthetic pathways of ladderane lipids in anammox bacteria through the use of 2-^{13}C-labelled acetate (Rattray et al. 2009).

Carbon isotope-labelling experiments have been valuable to estimate metabolic fluxes (Jouhten et al. 2009). For instance, Jouhten *et al*. used a computational carbon path analysis to reconstruct the biosynthetic pathways of amino acids for *Trichoderma reesei* (*T. reesei)*. The intracellular metabolic flux ratios in the central carbon metabolism were determined using a partial ^{13}C labelling and metabolic flux ratio analysis (Jouhten et al. 2009).

Another example, McKinlay *et al*. (2007) grew *Actinobacillus succinogenes* (*A. succinogenes*) with [1-^{13}C] glucose and used these data to infer metabolic fluxes based on the mass isotopomer distributions of amino acids, organic acids and glycogen monomers. With this experiment, important aspects of the *A. succinogenes* metabolism were found, namely the carbon flux distribution toward succinate production and the formation of alternative products and their biosynthetic pathways (McKinlay et al. 2007).

Software tools and resources emerged to integrate the high throughput data, the system complexity and the knowledge, for metabolic analysis (Weitzel et al. 2013; Quek et al. 2009; Zamboni et al. 2005). OptFlux is an example of a powerful

computational tool for *in silico* metabolic engineering applications. On top of this software, the CBFA tool has been developed (Carreira et al. 2014). This tool can be used to perform stoichiometric metabolic model simulations under different environmental and genetic conditions, in order to trace desired compounds, improve their production and analyze the resulting flux distribution, given a set of flux measurements and constraints through the application of MFA methods (Rocha et al. 2010).

3.4. CASE-STUDIES

3.4.1. *ESCHERICHIA COLI*

E. coli is a prokaryotic model-organism, perhaps the best characterized and well-studied bacterium with relevance for industry. This organism is present in the lower intestine of humans, facility grown in a laboratory setting and is readily ease to be manipulated genetically. Due to the interest in this gram-negative bacterium, *in silico* modeling of its cellular behavior has been helpful (Milne et al. 2009). Most of the metabolic data for this organism was experimentally verified and so, the current reconstructions of *E. coli* metabolic network are the best representation of the metabolic capabilities of this organism (McCloskey et al. 2013; Edwards & Palsson 2000; Edwards et al. 2001). These models have been used to design *E. coli* strains through *in silico* analyses to increase the production of target metabolites, like lycopene, lactic acid, ethanol, succinic acid, L-valine, L-threonine and more (Feist & Palsson 2008). *E. coli* K-12 is the most widely studied strain of *E. coli* and so, a reference for this species which is used in numerous works (Richmond et al. 1999).

3.4.2. *ACTINOBACILLUS SUCCINOGENES*

In this work the metabolism of the *Actinobacillus succinogenes* (*A. succinogenes* or *Asus*), a gram-negative capnophilic (i.e. grows with high concentrations of carbon dioxide), bacterium isolated from the bovine's rumen was analyzed. This bacterium is one of the most promising succinate producers,

can accumulate large quantities of succinate and so such is considered an industrially relevant biocatalyst (Guettler et al. 1999; Mckinlay et al. 2010). Succinate is an attractive chemical platform for the production of various value-added derivatives, such as 1,4-butanediol (precursor of "stronger-than-steel" plastics), ethylenediamine disuccinate (a biodegradable chelator), diethyl succinate (a green solvent for replacement of methylene chloride), and adipic acid (nylon precursor) (Mckinlay et al. 2005). On rumen, it is a relevant metabolic intermediate and also a source of energy for many bacteria (Mckinlay et al. 2010).

Among several microorganisms, *A. succinogenes* is one of the best producers described so far. This organism to produce highest concentrations of succinate also produces a formate and acetate. Since succinate is the product of interest, several studies have been done to improve the homofermentative production of succinate by decreasing or inhibiting the production of other by-products (Lin et al. 2008).

Succinic acid is an end-product of microbial fermentation. Metabolic engineering combined with genetic engineering explore the different pathways and due to its peculiar ability to produce this acid naturally from a vast source of carbons (Lin et al. 2008), this organism can be used for analyzing the metabolic control effects of CO_2 and the reducing power on flux distribution between PEP, pyruvate, OAA and malate (McKinlay et al. 2007).

The glucose-6-phosphate is catabolized in PEP and then converted in fermentation products that are split in two branches: C_3 pathway (leading to formate, acetate, ethanol) and C_4 pathway (leading to succinic acid); with malic enzyme and OAA decarboxylase catalyze reversible reactions between these pathways (Mckinlay et al. 2010).

The metabolic map of this organism will be very important to MFA and to apply engineering designs for understand and optimize chemical production. For a better understanding of how the metabolism of *A. succinogenes* works, ^{13}C labeling experiments to measure intracellular metabolic fluxes can be applied (Mckinlay et al. 2005).

4. METHODOLOGY

A Java-based generic tool was implemented to map the transition of carbon atoms in biochemical reactions, such as the ones in the central carbon metabolism of *E. coli* and *A. succinogenes*. The previously published carbon atom transitions model for *E. coli* (Latendresse, Malerich, Travers, & Karp, 2012) served as a reference in this work and was compared to the reconstructed model for *E. coli* using the implemented methodology as a proof-of-concept. The carbon atom transitions model for the central carbon metabolism of *A. succinogenes* was then constructed comprising the following metabolic pathways: glycolysis, pentose phosphate pathway, TCA cycle and pyruvate metabolism.

To facilitate and automate the construction of such atom transition models, a Java algorithm was designed and implemented in order to extract, process and compile information from different biological databases (i.e. KEGG and MetaCyc) and construct the atom transitions for the defined set of metabolic reactions. Procedures are detailed next and Figure 4.1 shows a schematic illustration of the overall procedure:

1. **Definition of the reactions set**. The set of reactions associated with the previously mentioned pathways were assembled. For each reaction a unique identifier, either a KEGG ID or a MetaCyc ID was associated (Figure 4.2). If a KEGG ID is associated, the corresponding MetaCyc ID was retrieved using a previously compiled file with a list of reactions and the corresponding cross-references to MetaCyc ID. Otherwise, a MetaCyc ID could be directly associated by manual curation.

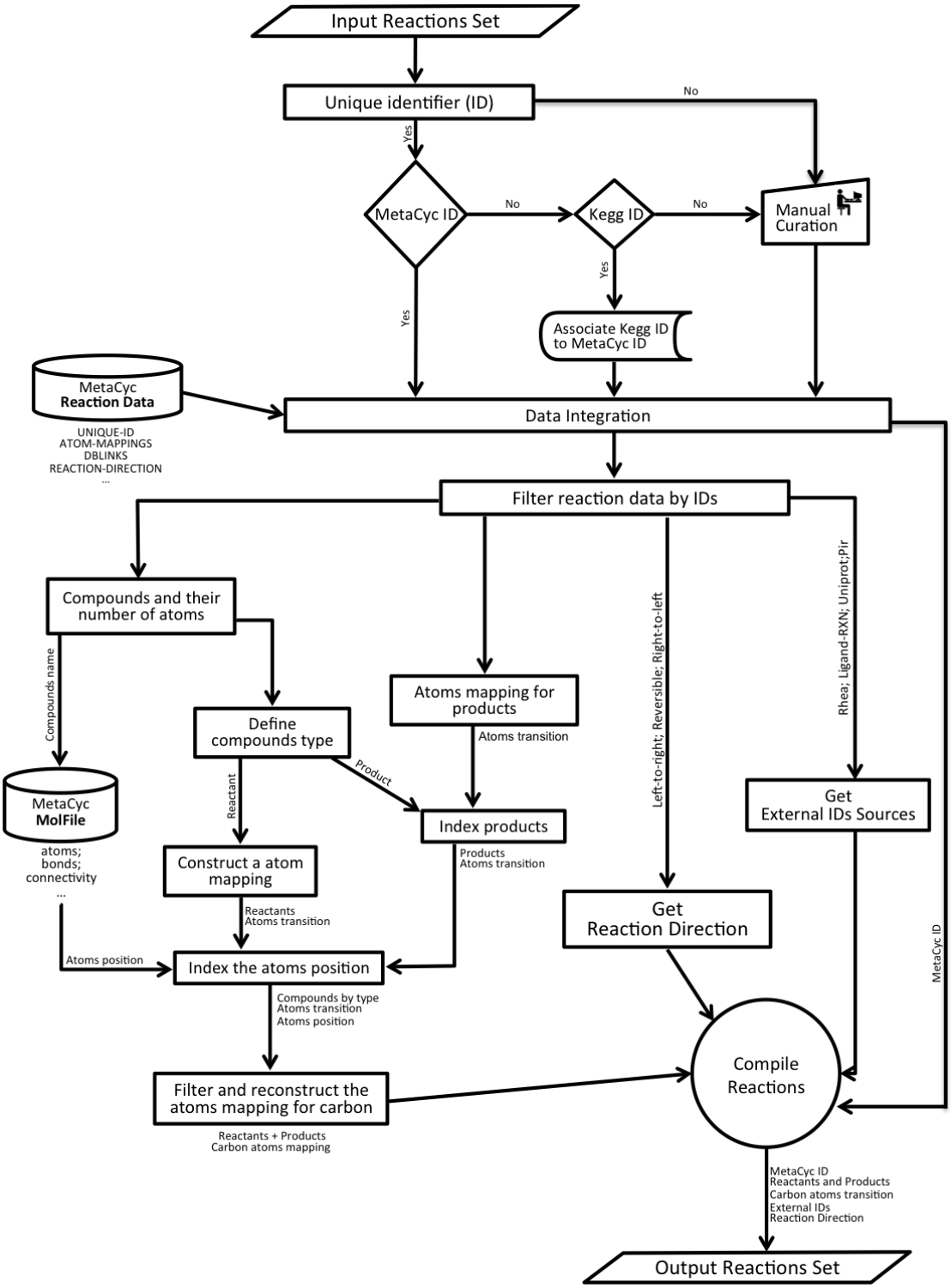


Figure 4.1 – Methodology pipeline. The reactions set was used as an input file, where a list of reactions with a corresponding ID, either KEGG ID or MetaCyc ID was used to extract and obtain the required information from MetaCyc files (e.g., "Reaction file" and "MolFile"). If an ID is not provided, manual curation to attribute a MetaCyc ID is required. For each reaction it is retrieved the reactants, products, atoms mapping for products, directionality and IDs from others databases (external IDs sources). The reactants and products with their number of atoms and their atoms position (acquired with the respective "MolFile") are extracted and, using the atoms mapping for products the carbon atoms mapping is reconstructed. The MetaCyc ID, the reactants the products, the carbon atom transitions, the direction of the reaction and the external IDs sources obtained, were organized for each reaction. All reactions mapped are complied and added to the input file.

1	2
R02506	CINNAMOYL-COA-REDUCTASE-RXN
R02505	PINOSYLVIN-SYNTHASE-RXN
R02504	TRIMETAPHOSPHATASE-RXN
R01862	ARYL-ACYLAMIDASE-RXN
R08570	DHDOGALDOL-RXN
R02509	NN-DIMETHYLFORMAMIDASE-RXN
R08572	GKI-RXN
R08573	ERYTHRULOSE-REDUCTASE-RXN
R01869	OROTATE-REDUCTASE-NADH-RXN
R01860	AMINOBENZOATE-DECARBOXYLASE-RXN
R01850	HYDROGEN-SULFIDE-S-ACETYLTRANSFERASE-RXN

Figure 4.2 – File format of the corresponding KEGG IDs and the MetaCyc IDs for each reaction. This *csv* format file is divided in two columns, the first with KEGG ID (1) and the second the corresponding MetaCyc ID (2).

2. **Collection of data from KEGG** (Version updated in September 1, 2014) **and MetaCyc** (Version 18.1 updated in June 23, 2014). Information regarding the chemical conversions and enzymes associated with the defined reactions set was retrieved. A "Reactions file" and "Molfiles" from MetaCyc were used to retrieve information associated with reactions previously defined. The "Reactions file" contains the reaction atom–atom mapping and is structured as follows (see Figure 4.3): each reaction is identified by a unique-identifier (1) with several attributes like atoms mapping (2), reactants (by order), products (by order), external links (3) and reactions directionality (4). Each compound has a corresponding "Molfile" (Figure 4.4), which is provided only by MetaCyc, that contains information regarding atoms, bonds, connectivity and coordinates of a molecule (depicted as 2 in Figure 2).

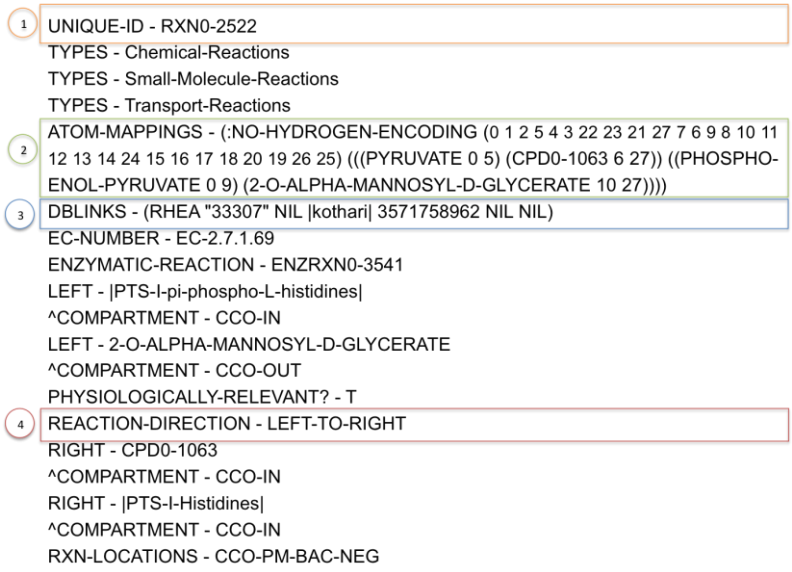


Figure 4.3 – Example of the information associated to each reaction in the Reactions file. The information contained in the file follows the same structure comprising the following attributes: the unique-identifier (1), the atoms mapping (2), the reactants (2), the products (2), the external links (3) and the reaction directionally (4), which are the attributes of interest.

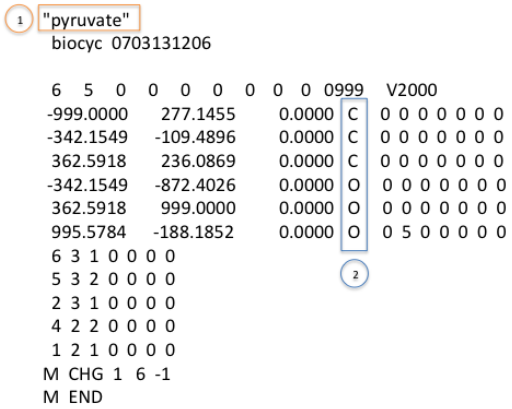


Figure 4.4 – Example of a Molfile structure. All the compounds have a corresponding Molfile comprising the following information: the name of the compound (1) in the first line; and two blocks of information corresponding to the atoms position, bonds, connectivity and coordinates of a molecule. The information used in this work is depicted by (2) corresponding to the atoms position.

3. **Data integration (Figure 4.5a)**. For each reaction in the input file, the reaction ID was retrieved. It is performed an automatically cross-reference to attribute a MetaCyc ID (if the reaction ID retrieved didn't belong to MetaCyc). If no cross-references were provided was made a manual research to attribute a MetaCyc ID. The MetaCyc ID is used to

extract and process information from MetaCyc. Reactions that have no corresponding MetaCyc ID were not integrated.

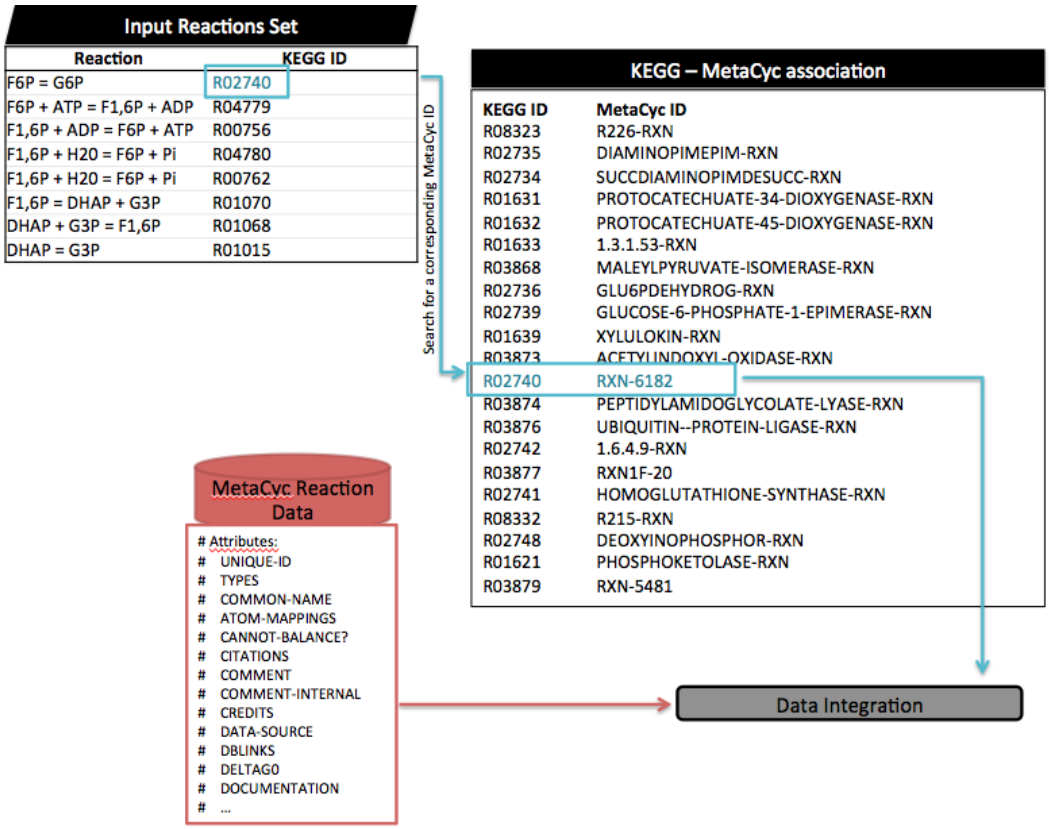


Figure 4.5a – Illustration of a carbon atom's mapping for the reaction with KEGG ID: R02740. The input file, containing a set of reactions, is imported and verified if a KEGG ID is provided. For the first reaction (R02740) an automatically association is performed, since a MetaCyc ID is associated (MetaCyc ID: RXN-6182). The corresponding data in the "Reaction File" is searched for this ID.

4. **Data filtration (Figure 4.5b).** The atoms mapping for products, reactants, products, reaction's directionality and the external sources for each reaction with MetaCyc ID in the "Reactions file" were retrieved:

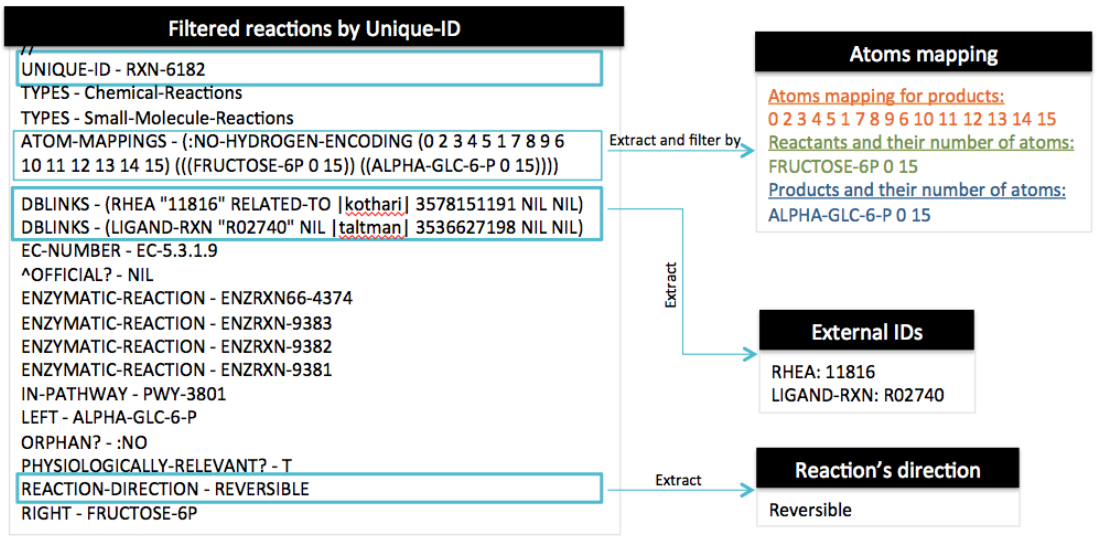


Figure 4.5b – Illustration of a carbon atom's mapping for the reaction with KEGG ID: R02740. The information available for this reaction is then retrieved. As illustrated, the "ATOM-MAPPINGS" has the information regarding: the atoms mapping for the product (0 2 3 4 5 1 7 8 9 6 10 11 12 13 14 15; the reactant ID (FRUCTOSE-6P) with its number of atoms (0 to 15); the product ID (ALPHA-GLC-6-P) with its number of atoms (0 to 15); the "DBLINKS" containing the external IDs for different databases, such as RHEA and LIGAND-RXN; the "REACTION-DIRECTION" with the directionally of the reaction, in this case the reaction is reversible.

a. **Get reaction direction**. The directionality of a reaction can be assigned either as right-to-left, left-to-right, reversible or with no information.

b. **Extract external IDs sources**. The external IDs of a reaction can be from Rhea, Ligand-RXN, Uniprot and/or Pir. A reaction can have more than one external ID and several IDs from the same source, so the external IDs were distributed by source and grouped.

c. **Atoms transition maps (Figure 4.5c).**
   i. **Identification of carbon atoms positions**. Compounds in a reaction were either identified as reactants or products, and, for each, the position of the first and last atoms was retrieved. Using the "MolFiles" for each compound, all atoms position was identified.

ii. **Mapping carbon positions of reactants**. To build the carbon atoms' mapping for reactants, it was verified the position of the first and last atom of each reactant and, according to its atoms position, the atoms were sequentially added. Yet, only the carbon atoms had their position identified to construct the map.

iii. **Mapping carbon positions of products.** To build the carbon atoms' mapping for products, it was verified the position of the first and last atom of each product and, according to the atoms mapping extracted, the positions identified with carbon atoms in the map for reactants were added and the map was reconstructed. If exist two or more atoms mapping in the "Reaction file", by default only the first is extracted.

iv. **Building the carbon atom transitions map.** For each carbon position in the reactant map was designated a letter (by order) and, the same letter was designated to the same position in the product map. The reactants, the products and the respective maps were compiled to build a reaction carbon atom transitions map.

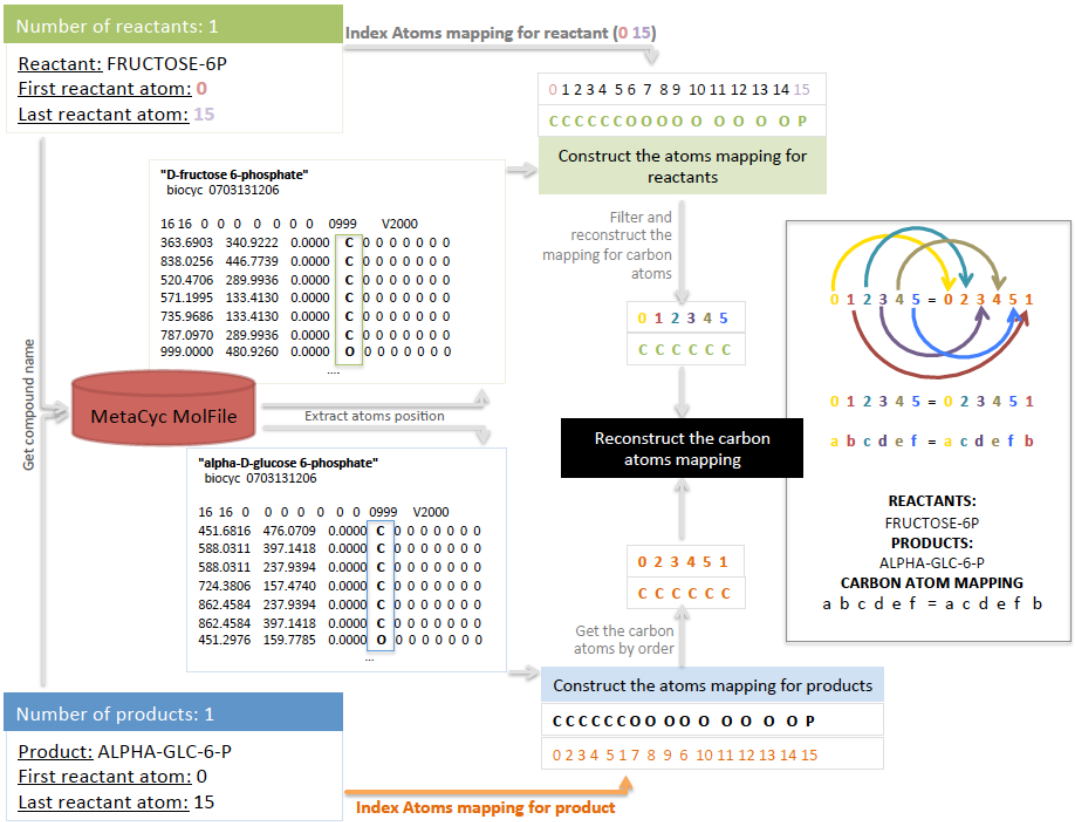
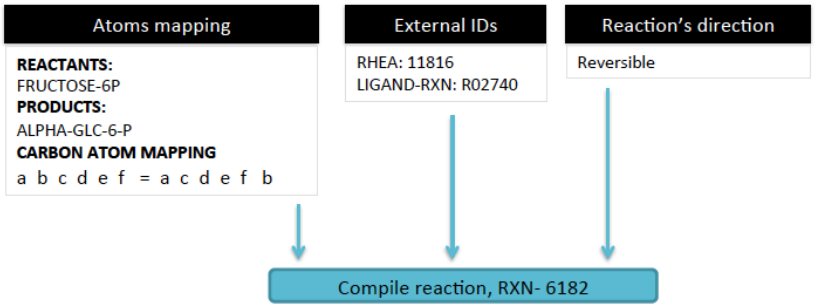


Figure 4.5c – Illustration of a carbon atom's mapping for the reaction with KEGG ID: R02740. The name of the reactant and the product is used to search their corresponding "MolFile", and then the atoms positions are identified. To construct the carbon atoms map for reactants, the position of the first and last atom of FRUCTOSE-6P that is sequentially (0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15) is indexed to the atoms position (C C C C C C O O O O O O O O P). The positions of the carbons are filtered and the map is reconstructed taking in account only the carbon atoms and their numeral position. To construct the carbon atoms map for products, the atoms mapping are extracted (0 2 3 4 5 1 7 8 9 6 10 11 12 13 14 15) and indexed to the atoms position (C C C C C C O O O O O O O O P), and the position identified with carbon atoms in the map of reactants are filtered and extracted orderly. For each carbon position in the reactant map a letter has been designated (by order) and, the same letter was designated to the same position in the product map (0 = a; 1 = b; 2 = c; 3 = d; 4 = e; 5 = f). The FRUCTOSE-6P and the ALPHA-GLC-6-P have the respective maps (a b c d e f = a c d e f b) compiled to build a reaction carbon atom transitions map.

5. **Output file (Figure 4.5d)**. All the reactions with an associated MetaCyc ID were complied and added to the metabolic information input file. The output file is the input file with additional information – reactants and products, carbon atom transition map, the external IDs and reactions directionality.



Reaction	KEGG ID	MetaCyc ID	Reactants	Products	Carbon Atom Transition	External IDs	Direction
F6P = G6P	R02740	RXN-6182	FRUCTOSE-6P	ALPHA-GLC-6-P	abcdef = acdefb	RHEA: 11816 LIGAND-RXN: R02740	Reversible
F6P + ATP = F1,6P + ADP	R04779						
F1,6P + ADP = F6P + ATP	R00756						
F1,6P + H20 = F6P + Pi	R04780						
F1,6P + H20 = F6P + Pi	R00762						
F1,6P = DHAP + G3P	R01070						
DHAP + G3P = F1,6P	R01068						
DHAP = G3P	R01015						

Figure 4.5d – Illustration of a carbon atom's mapping for the reaction with KEGG ID: R02740. The reaction and their reactants and products, carbon atoms mapping, external IDs and directionality are complied and when all reactions are mapped, they are added to the input file to compile the output file.

All these results were validated by manually inspecting the constructed transition maps and information presented in literature and MetaCyc's webpage.

5. RESULTS AND DISCUSSION

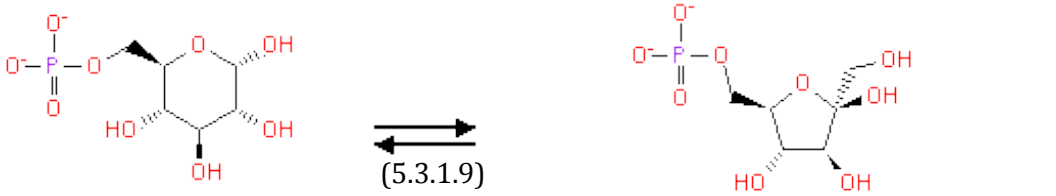
5.1. CARBON ATOMS MAPPING MODEL OF *E. COLI*

A list of 57 reactions and their carbon atom transitions associated with central carbon metabolic pathways for *E. coli* was used to validate the developed algorithm (Quek et al. 2009; Ravikirthi et al. 2011). For each reaction a KEGG or MetaCyc ID was associated. However, for most reactions associated to amino acid biosynthetic pathways it was not possible to associate an MetaCyc ID, as these corresponded to lumped reactions, for example the reaction: oxaloacetate = aspartate (see Table 5.1).

Table 5.1 – Reactions presents in the input file and their association with an identifier.

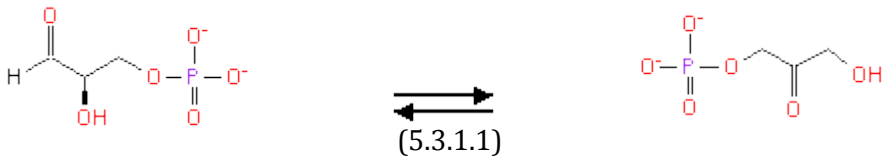
Database identifier	Number of reactions	Percentage
MetaCyc and correspondent KEGG ID	27	49.10 %
None	30	50. 90 %
Total	**57**	**100 %**

From the 57 only 27 reactions were associated to a MetaCyc ID and for those it was possible to construct a carbon atom transition and compare it with the published carbon atom transitions. This comparison was performed by manual inspection to guarantee that no errors were included. From the comparison of the 25 reactions only 2 reactions presented a different carbon atom transitions. The two reactions were then inspected (Figures 5.1 and 5.2).



D-glucose 6-phosphate **β-D-fructofuranose 6-phosphate**

Figure 5.1 – Reaction schema of RXN-6182: D-glucose 6-phosphate ↔ β-D-fructofuranose 6-phosphate,, adapted from MetaCyc database. The enzyme glucose-6-phosphate isomerase (E.C. number 5.3.1.9) catalyzes the reversible isomerization of D-glucose 6-phosphate to β-D-fructofuranose 6-phosphate.



D-glyceraldehyde 3-phosphate **dihydroxyacetone phosphate**

Figure 5.2 – Reaction schema of TRIOSEPISOMERIZATION-RXN: D-glyceraldehyde 3-phosphate ↔ dihydroxyacetone phosphate, adapted from MetaCyc database. The enzyme triose-phosphate isomerase (E.C. number 5.3.1.1), catalyzes the reversible isomerization of D-glyceraldehyde 3-phosphate to dihydroxyacetone phosphate.

Both are isomerization reactions involving an intramolecular rearrangement of the reactants by the redistribution of electrons within the specific functional groups, yielding isomeric forms. As such, it is plausible that differences in the carbon transitions are due to dissimilar molecular conformations. These stereoisomers can assume different positions in space and according to these rearrangements the position of the carbons can change. In the reaction RXN-6182, the glucose-6-phosphate isomerase transfers the carbonyl group of D-glucose 6-phosphate forming an isomer, β-D-fructofuranose 6-phosphate. In the validation model it is not specified which of the enantiomer it is considered (i.e., D or L) or which epimer it is (i.e., α or β), however this reaction was initially associated to the MetaCyc ID RXN-6182. A second reaction ID (PHOSPHOGLUCMUT-RXN) that it is also associated with this reaction (glucose 6-phosphate ↔ fructofuranose 6-phosphate) was thereafter used and the carbon atom transition resulted is the same as the validation file. Since this reaction has a crucial role in the chemistry of glycolytic pathway, because of the rearrangement of the carbonyl and hydroxyl groups are a prelude to the next steps of the glycolytic pathway, it is important to consider the correct reactants conformation (Nelson & Cox 2008).

In the reaction TRIOSEPISOMERIZATION-RXN (Figure 5.2), the enzyme triose-phosphate isomerase catalyzes the interconversion of D-glyceraldehyde 3-phosphate in dihydroxyacetone phosphate. Using the same method as in the previous example, a manual inspection was performed to attribute an identifier to a reaction with a reactant with a different conformation, but no reaction in the database MetaCyc was found.

If the set of reactions contains isomerization reactions they need to be identified for when was associated an atom mapping a manual curation be made, in order to confirm if the atom mapping associated is correct.

After this first validation test, a core metabolic model of *E. coli* was used to map the carbon atom transitions of 74 reactions, where each reaction is identified by a "CoreModel ID" and has information regarding the reactants, products and directionality. Reactions of the input model were manually assigned to a metabolic pathway and to a MetaCyc or a KEGG ID.

Figure 5.3 shows how the 74 reactions in the core model were assigned to individual pathways. 19 reactions were associated with the transport of extracellular metabolites

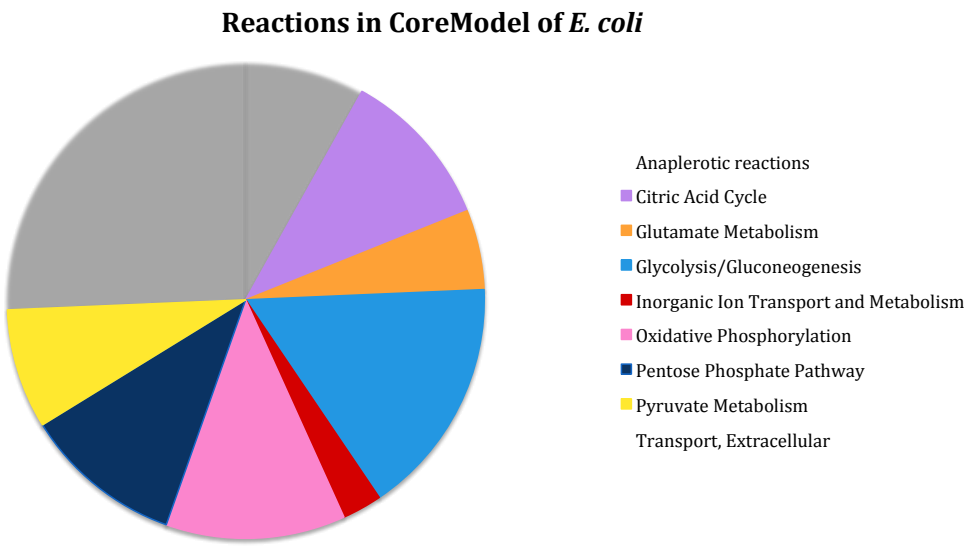


Figure 5.3 – Representation of the relative number of reactions assigned to individual pathways.

The others pathways have all their reactions associated to a MetaCyc, unless the inorganic ion transport, 2 reactions are related to this pathway but only 1 has a MetaCyc ID associated, and the metabolism and the oxidative phosphorylation, in this pathway 8 of 9 reactions has a MetaCyc identifier.

From the 74 reactions, only 49 reactions were associated to a KEGG ID and 48 were automatically linked by cross-references to a MetaCyc ID. To the 26 reactions without a MetaCyc ID, a manual search in order to attribute more

identifiers was performed. In total, 63 reactions were associated to a MetaCyc ID and 49 reactions had an associated KEGG ID, indicating that data integration between these two data sources is still poor (Figure 5.4).

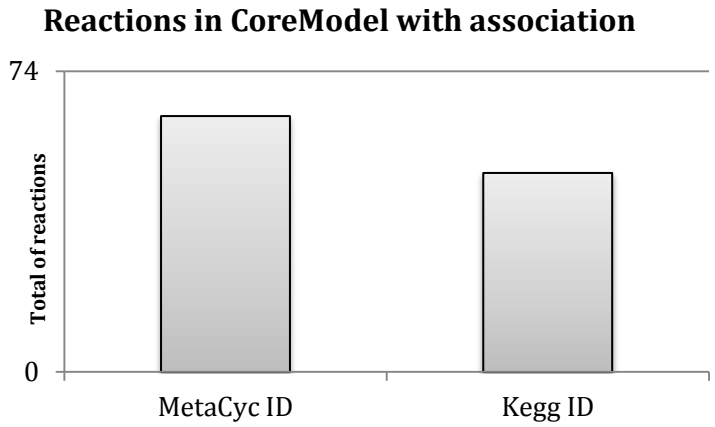


Figure 5.4 – Quantity of reactions in the CoreModel with an identifier from MetaCyc and KEGG associated. To a total of 74 reactions presents in the input CoreModel, 63 have an identifier from MetaCyc and 49 an identifier from KEGG.

Although it might seem disadvantageous to use KEGG IDs to link reactions to MetaCyc IDs, this algorithm was designed to take advantage of cross-links often contained in genome-scale models. Because MetaCyc IDs have not been extensively used in genome-scale models as KEGG IDs, this may speed up the initial step of the proposed procedure by simplifying the manual curation step.

Table 5.2 – Reactions of the core model of *E. coli* and the associated carbon atom transitions constructed using the developed algorithm. Reactions in the CoreModel were associated to a metabolic pathway, MetaCyc ID, KEGG ID and the carbon atom transitions.

CoreModel ID	MetaCyc ID	KEGG ID	Reaction	Carbon Atoms Mapping
Anaplerotic reactions				
R_PPCK	PEPCARBOXYKIN-RXN	R00341	ATP + OAA -> ADP + PEP + CO2	abcd = abd + c
R_ICL	ISOCIT-CLEAV-RXN	R00479	Icit -> GLX + SUC	abcdef = df + abce
R_PPC	PEPCARBOX-RXN	R00345	H +Pi + OAA <-> H20 + CO2 + PEP	abc + d = abdc
R_ME1	1.1.1.39-RXN	R00214	NADH + PYR + CO2 -> NAD + MAL	abc+ d = abdc
R_ME2	MALIC-NADP-RXN	R00216	NADP + MAL -> NADPH + PYR + CO2	abc + d = abdc
R_MALS	MALSYN-RXN	R00472	H20 + AcCoA + GLX -> H + CoA + MAL	abcdefghijklmnopqrstuvw + xy = cbdefghijkmnopqrstuvw + axly
Citric Acid Cycle				
R_FUM	FUMHYDR-RXN	R01082	H2O + FUM <-> MAL	abcd = badc
R_ICDHyr	ISOCITDEH-RXN	R00267	NADPH + alphaKG + CO2 <-> NADP + Icit	abcdefghijklmnopqrstu + vwxyz + A = abcdefghijklmnopqrstu + wvyxAz
R_SUCOAS	SUCCCOASYN-RXN	R00405	ADP + SucCoA + Pi <-> ATP + CoA + SUC	abcdefghijklmnopqrstuvwxz = baefghijklmnqrstuvwxz + cdop
				abcdefghijklmnopqrstuvwxz = baefghijklmnqrstuvwxz + dcpo
R_ACONTa	ACONITATEDEHYDR-RXN	R01325	CIT <-> H20 + cis-Aco	abcdef = bafdce
R_ACONTb	ACONITATEHYDR-RXN	R01900	H2O + cis-Aco <-> Icit	abcdef = bceafd
R_AKGDH	2-OXOGLUTARATE-SYNTHASE-RXN	R01197	CoA + alphaKG + NAD -> SucCoA + CO2 + NADH	abcdefghijklmnopqrstu + vwxyz = bawvcdefghijklyxmnopqrstu + z
R_CS	CITSYN-RXN	R00351	H20 + AcCoA + OAA -> H + CoA + CIT	abcdef =bd + afce
R_MDH	MALATE-DEH-RXN	R00342	MAL + NAD <-> H + NADH + OAA	abcd = abcd
Glutamate Metabolism				
R_GLUSy	GLUTAMATESYN-RXN	R00114	H + GLN + NADPH + alphaKG -> NADP + 2.0*GLU	abcde + fghij = fghij + abcde
R_GLUDy	GLUTDEHYD-RXN	R00248	NADP + GLU + H2O <-> H + NADPH + NH4	abcdefghijklmnopqrstu + vwxyz = abcdefghijklmnopqrstu + vwxyz
R_GLUN	GLUTAMIN-RXN	R00256	GLN + H2O -> GLU + NH4	abcde = abcde
R_GLNS	GLUTAMINESYN-RXN	R00253	ATP + GLU + NH4 -> ADP + H +GLN + Pi	abcdefghij + klmno = abcdefghij + klmno
Glycolysis/Gluconeogenesis				
R_PDH	PYRUVDEH-RXN	R00209	NAD + CoA + PYR -> CO2 + NADH + AcCoA	abcdefghijklmnopqrstu + vwx = vbacdefghijwklmnopqrstu + x
R_TPI	TRIOSEPISOMERIZATION-RXN	R01015	DHAP <-> G3P	abc = abc

R_PGI	RXN-6182	R02740	G6P <-> F6P	abcdef = acdefb
R_PGK	PHOSGLYPHOS-RXN	R01512	ATP + 3PG <-> ADP + G1,3P	abc = abc
R_PGM	3PGAREARR-RXN	R01518	2PG <-> 3PG	abc = abc
	RXN-15513			abc = abc
R_GAPD	GAPOXNPHOSPHN-RXN	R01061	NAD + G3P + Pi <-> H + NADH + G1,3P	abc = bca
R_FBA	F16ALDOLASE-RXN	R01068	F1,6P <-> G3P + DHAP	abcdef = abc + def
R_PFK	6PFRUCTPHOS-RXN	R04779	ATP + F6P -> ADP + H + F1,6P	abcdef = abcdef
R_ENO	2PGADEHYDRAT-RXN	R00658	2PG <-> H2O + PEP	abc = abc
R_PYK	PEPDEPHOS-RXN	R00200	ADP + H + PEP -> ATP + PYR	abc = abc
R_PPS	PEPSYNTH-RXN	R00199	ATP + H2O + PYR -> AMP + 2.0*H + Pi + PEP	abcdefghij + klm = abcdefghij + klm
R_FBP	F16BDEPHOS-RXN	R00762	H2O + F1,6P -> Pi + F6P	abcdef = abcdef
Inorganic Ion Transport and Metabolism				
R_PIt2r			H_e + Pi_e <-> H + Pi	
R_NH4t	RXN-9615		NH4_e <-> NH4	
Oxidative Phosphorylation				
R_ADK1	ADENYL-KIN-RXN	R00127	ATP + AMP <-> 2.0*ADP	abcdefghij + klmnopqrst = klmnopqrst + abcdefghij
R_SUCDi	RXN-14971		q8 + SUC -> FUM + q8h2	abcd = badc
R_FRD7	RXN-14970		FUM + q8h2 -> q8 + SUC	abcd = badc
R_NADH16	NADH-DEHYDROG-A-RXN	R02163	4.0* H + NADH + q8 .0 -> 3.0*H_e + NAD + q8h2	abcdefghijklmn + opqrstuvwxyzABCDEFGHI = abcdefghijklmn + opqrstuvwxyzABCDEFGHI
R_ATPM	3.6.3.52-RXN	R00086	ATP + H2O -> ADP + H + Pi	abcdefghij = abcdefghij
R_ATPS4r	TRANS-RXN-249		ADP + 4.0*H_e + Pi <-> ATP + 3.0*H + H2O	
R_NADTRHD	PYRNUTRANSHYDROGEN-RXN	R00112	NADPH + NAD -> NADP + NADH	
R_THD2	TRANS-RXN0-277		NADP + 2.0*H_e + NADH -> 2.0*H + NADPH + NAD	
R_CYTBD		R09504	2.0*H + q8h2 + 0.5*O2 -> H2O + 2.0*H_e + q8	
Pentose Phosphate Pathway				
R_PGL	6PGLUCONOLACT-RXN	R02035	H2O + G1,5L6P -> H + 6PG	abcdef = abcdef
R_G6PDH2r	GLU6PDEHYDROG-RXN	R02736	NADP +G6P <-> H + NADPH + G1,5L6P	abcdef = abcdef
R_TKT1	1TRANSKETO-RXN	R01641	R5P + X5P <-> G3P + S7P	abcde + fghij = jgi + fbhdaec

R_TKT2	2TRANSKETO-RXN	R01067	E4P + X5P <-> G3P + F6P	abc + defghi = adcb + efigh
R_GND	RXN-9952	R01528	NADP + 6PG -> NADPH + CO2 + Rub5P	abcde + f = bdecaf
R_RPI	RIB5PISOM-RXN	R01056	R5P <-> Rub5P	abcde = acbde
R_TALA	TRANSALDOL-RXN	R01827	G3P + S7P <-> F6P - E4P	abcd + efghij = heg + fbjdica
R_RPE	RIBULP3EPIM-RXN	R01529	Rub5p <-> X5P	abcde = acbde
			Pyruvate Metabolism	
R_ALCD2x	ALCOHOL-DEHYDROG-RXN	R00754	NAD + EtOH <-> H + NADH + AcAld	ab = ab
R_PTAr	PHOSACETYLTRANS-RXN	R00230	AcCoA + Pi <-> CoA + AcP	ab = ab
R_PFL	PYRUVFORMLY-RXN	R00212	CoA + PYR -> FOR + AcCoA	abc = ab + c
R_LDH_D	DLACTDEHYDROGNAD-RXN	R00704	NAD + LAC <-> H + NADH + PYR	abcdefghijklmnopqrstu + vwx = abcdefghijklmnopqrstu + vwx
R_ACKr	ACETATEKIN-RXN	R00315	ATP + Ace <-> ADP + AcP	abcdefghij + kl = abcdefghij + kl
R_ACALD	ACETALD-DEHYDROG-RXN	R00228	NAD + CoA + AcAld <-> H + NADH + AcCoA	ab = ab
			Transport, Extracellular	
R_FORt2			H_e + FOR_e -> H + FOR	
R_GLUt2r			H_e + GLU_e <-> GLU + H	
R_PYRt2r			H_e + PYR_e <-> H + PYR	
R_GLNabc	ABC-12-RXN		ATP + H2O + GLN_e -> ADP + H +GLN + Pi	abcdef = abcdef
R_MALt2_2	TRANS-RXN0-451		2.0*H_e + MAL_e -> 2.0*H + MAL	abcd = abcd
R_ACALDt			AcAld_e <-> AcAld	
R_H2Ot			H2O_e <-> H2O	
R_ETOHt2r	TRANS-RXN0-546		H_e + EtOH_e <-> H + EtOH	ab = ab
R_FUMt2_2			2.0*H_e + FUM_e -> 2.0*H + FUM	
R_CO2t	TRANS-RXN0-545		CO2_e <-> CO2	
R_O2t	TRANS-RXN0-474		O2_e <-> O2	
R_FORti			FOR -> FOR_e	
R_FRUpts2			FRU_e + PEP -> PYR + F6P	
R_ACt2r	TRANS-RXN0-571		H_e + Ace_e <-> H + Ace	ab = ab
R_SUCCt2_2	TRANS-RXN-121		SUC_e + 2.0*H_e -> 2.0*H + SUC	abcd = abcd

R_GLCpts		GLC_e + PEP -> PYR + G6P	
R_SUCCt3		H_e + SUC -> SUC_e + H	
R_D_LACt2	TRANS-RXN0-515	H_e + LAC _e <-> H + LAC	abc = abc
R_AKGt2r	TRANS-RXN-23	H_e + alphaKG_e <-> H + alphaKG	abcde =abcde

From the carbon atom transitions map of the CoreModel, two reactions posed some concerns: R_SUCOAS and R_PGM. R_SUCOAS has associated two different atoms mapping. By default, only the first atom mapping would be extracted, but since these were inspected manually, the two atom mapping were then extracted. The other reaction R_PGM may be associated to two reactions with different MetaCyc ID, 3PGAREARR-RXN and RXN-15513. This reaction: 2-phospho-D-glycerate ↔ 3-phospho-D-glycerate (2PG <-> 3PG) can be catalyzed by a phosphoglycerate mutase (2,3-diphosphoglycerate-dependent), which has the identifier RXN-15513, or by a phosphoglycerate mutase (2,3-diphosphoglycerate-independent), which the identifier is 3PGAREARR-RXN. This model belongs to *E. coli* and according to information in MetaCyc, some bacteria, in particular Gram-negative, require as cofactor 2,3-diphospho-D-glycerate. So, is most probable that this reaction is catalyzed by the phosphoglycerate mutase (2,3-diphosphoglycerate-dependent) than the phosphoglycerate mutase (2,3-diphosphoglycerate-independent). Since the two enzymes can catalyze this reaction, the two reactions from MetaCyc were associated to the reaction with the CoreModel ID: R_PGM.

The algorithm processed the 63 reactions with an associated MetaCyc ID, as illustrated in Figure 5.5. This figure shows the distribution of reactions by pathway that have an associated carbon atom transition and those without an associated carbon atom transitions. To most reactions from central pathways it was possible to associate a MetaCyc ID and a carbon atom transition, only a few reactions in the others pathways (most related to the transport extracellular) were not associated to a MetaCyc ID and thus, a carbon atoms mapping.

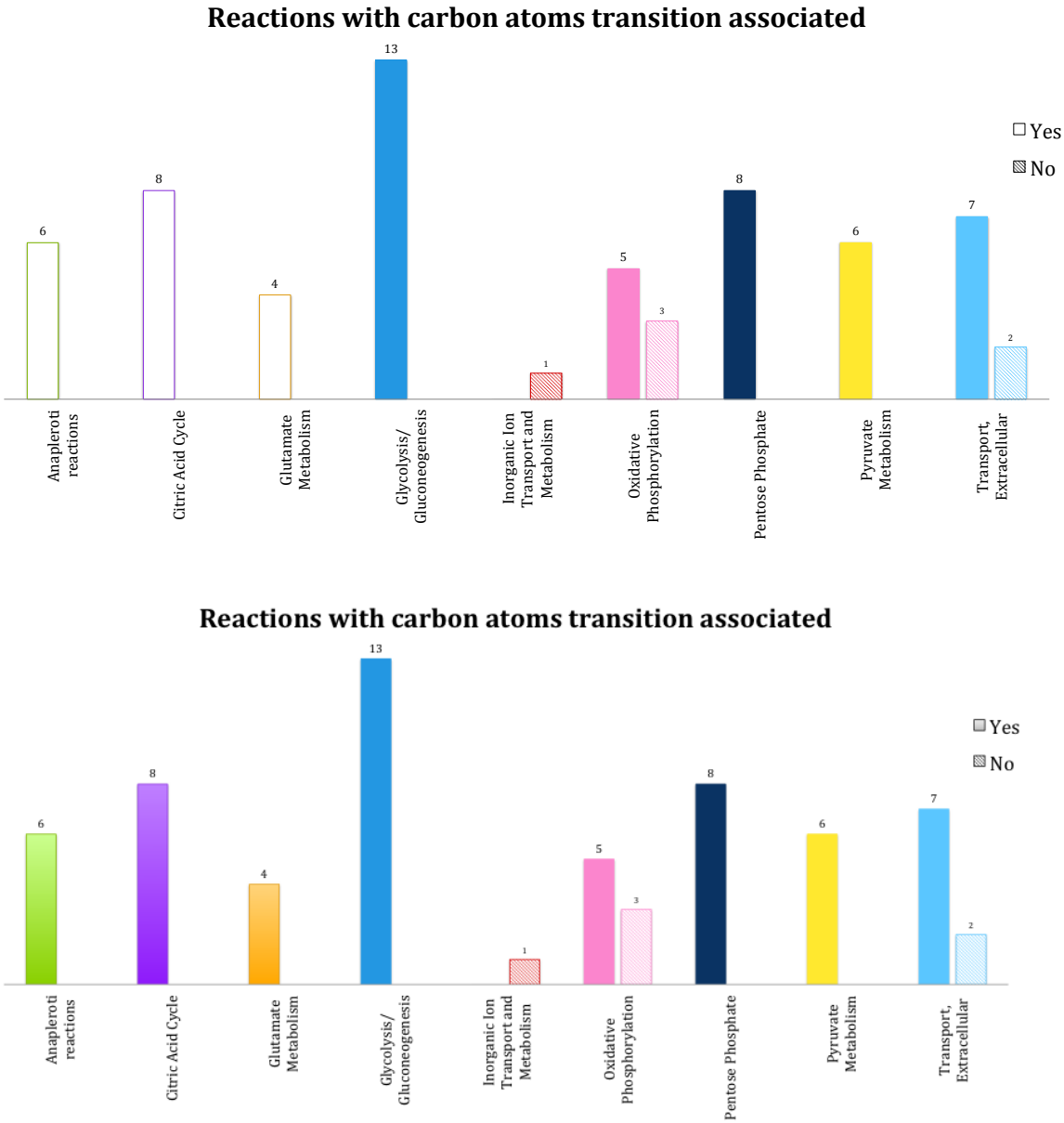


Figure 5.5 – Distribution of each reaction with a MetaCyc ID by pathway that have or not a carbon atom transitions associated. Each bar has a color according to the pathway and the number in top of each bar correspond to the number of reactions that have a carbon atom transitions (bar is full filled) or that don't have a carbon atom transitions (bar isn't full filled).

5.2. CARBON ATOMS' MAPPING MODEL OF *A. SUCCINOGENES*

Based on the published central metabolic model for *A. succinogenes* (McKinlay et al. 2007; Mckinlay et al. 2010; Förster et al. 2003; Guettler et al. 1999), a new metabolic model with mapped carbon atom transitions was constructed. Reactions considered to represent the central metabolic model were associated

to the following metabolic pathways: glycolysis pathway, pentose phosphate pathway, citrate cycle and pyruvate metabolism.

55 reactions were manually curated to attribute a KEGG ID and 43 reactions were then automatically associated to the respective MetaCyc ID. The carbon atom mapping, reactions directionality and IDs from others databases were then extracted and associated to reactions. For the 9 reactions with no cross-reference and the others 3 reactions that hadn't an identifier, it was performed a manual exploration in MetaCyc database for each one and, it was possible associate a MetaCyc ID for 3 reactions (Table 5.3).

Table 5.3 – Reactions in the model and their association with an identifier.

Database identifier	Number of reactions	Percentage
MetaCyc ID	3	5.45 %
KEGG ID	9	16.35 %
MetaCyc and correspondent KEGG ID	43	78.20 %
Total	**55**	**100 %**

The list of reactions with a corresponding MetaCyc ID was processed by the algorithm and the carbon atom mapping, the directionality and the ID from others databases, were added to the model (see Table B.1 and Table B.2). As shown in Table 5.4, most reactions have an atom mapping associated, except the reaction with MetaCyc ID SUCCCOASYN-RXN that has two atom mappings (this reaction is the same reaction that have two atom mapping associated in the core model of *E. coli,* but for this case only the first atom mapping is recorded) and 7 reactions have none.

Table 5.4 – Reactions with a MetaCyc ID associated and their associated atoms mapping.

Atom mapping	Number of reactions
None	7
One	38
Two or more	1
Total	**46**

The pathways glycolysis, pentose phosphate pathway, citrate cycle and pyruvate metabolism provided by KEGG were highlighted where the reactions mapped of the central metabolic model occur (see Appendix C).

To support the constructed model the central metabolic model provided by McKinlay and co-workers was used as basis (Mckinlay et al. 2010). The reactions and their metabolic information were considered reliable for the design of a carbon atom transition network for *A. succinogenes*. In the model, to each reaction a KEGG ID, a MetaCyc ID and a carbon atom transition is displayed. The ATP, ADP, NAD and NADH are irrelevant for the carbon atom mapping but were represented to indicate when a reaction spends or not spends energy. The Figure 5.7 shows in detail the final designed model for the carbon atom transition network for *A. succinogenes.*

The developed model can be used to trace the transition of carbon atoms of desired compounds. As exemplified in Figure 5.7, if the isocitrate is the compound of interest, a possible pattern of reactions to manipulate the production of this product is highlighted with carbon atoms colored with red. However the pattern chosen has a gap, between the compounds fructose-1,6-bisphosphate and glyceraldehyde-3-phosphate. There are two possible patterns, one reaction (fructose-1,6-bisphosphate → glyceraldehyde-3-phosphate) or two reactions involving another compound (fructose-1,6-bisphosphate → glycerone phosphate → glyceraldehyde-3-phosphate). None of them have a carbon atom transition that can connect the pattern so, it was assumed that the carbon atoms don't change their previously order. This means that the position of the carbon atom (b) in the reaction glucose-6-phosphate → fructose-1,6-bisphosphate is the same when the pattern restarted in the reaction glycerone phosphate → glyceraldehyde-3-phosphate.

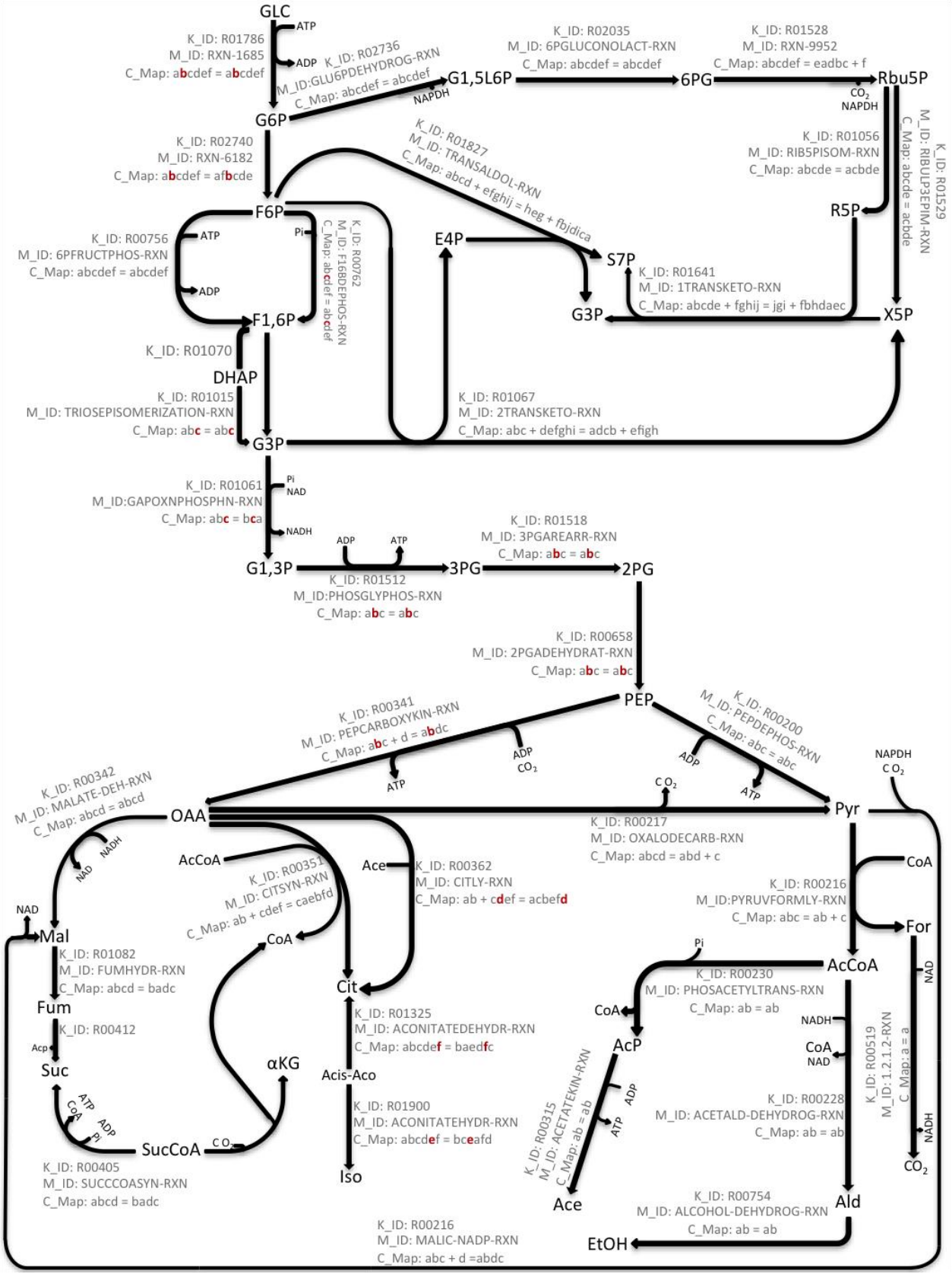

GLC
ATP
ADP
K_ID: R01786
M_ID: RXN-1685
C_Map: abcdef = abcdef
K_ID: R02736
M_ID:GLU6PDEHYDROG-RXN
C_Map: abcdef = abcdef
NAPDH
G1,5L6P
K_ID: R02035
M_ID: 6PGLUCONOLACT-RXN
C_Map: abcdef = abcdef
6PG
K_ID: R01528
M_ID: RXN-9952
C_Map: abcdef = eadbc + f
Rbu5P
CO2
NAPDH
G6P
K_ID: R02740
M_ID: RXN-6182
C_Map: abcdef = afbcde
K_ID: R01827
M_ID: TRANSALDOL-RXN
C_Map: abcd + efghij = heg + fbjdica
K_ID: R01056
M_ID: RIB5PISOM-RXN
C_Map: abcde = acbde
K_ID: R01529
M_ID: RIBULP3EPIM-RXN
C_Map: abcde = acbde
R5P
F6P
K_ID: R00756
M_ID: 6PFRUCTPHOS-RXN
C_Map: abcdef = abcdef
ATP
ADP
Pi
K_ID: R00762
M_ID: F16BDEPHOS-RXN
C_Map: abcdef = abcdef
E4P
S7P
K_ID: R01641
M_ID: 1TRANSKETO-RXN
C_Map: abcde + fghij = jgi + fbhdaec
G3P
X5P
F1,6P
K_ID: R01070
DHAP
K_ID: R01015
M_ID: TRIOSEPISOMERIZATION-RXN
C_Map: abc = abc
K_ID: R01067
M_ID: 2TRANSKETO-RXN
C_Map: abc + defghi = adcb + efigh
G3P
K_ID: R01061
M_ID:GAPOXNPHOSPHN-RXN
C_Map: abc = bca
Pi
NAD
NADH
G1,3P
ADP
ATP
K_ID: R01512
M_ID:PHOSGLYPHOS-RXN
C_Map: abc = abc
3PG
K_ID: R01518
M_ID: 3PGAREARR-RXN
C_Map: abc = abc
2PG
K_ID: R00658
M_ID: 2PGADEHYDRAT-RXN
C_Map: abc = abc
PEP
K_ID: R00341
M_ID: PEPCARBOXYKIN-RXN
C_Map: abc + d = abdc
ATP
ADP
CO2
K_ID: R00200
M_ID: PEPDEPHOS-RXN
C_Map: abc = abc
ADP
ATP
NAPDH
CO2
K_ID: R00342
M_ID: MALATE-DEH-RXN
C_Map: abcd = abcd
OAA
CO2
Pyr
K_ID: R00217
M_ID: OXALODECARB-RXN
C_Map: abcd = abd + c
NAD
NADH
AcCoA
K_ID: R00351
M_ID: CITSYN-RXN
C_Map: ab + cdef = caebfd
Ace
K_ID: R00362
M_ID: CITLY-RXN
C_Map: ab + cdef = acbefd
CoA
K_ID: R00216
M_ID:PYRUVFORMLY-RXN
C_Map: abc = ab + c
For
NAD
Mal
K_ID: R01082
M_ID: FUMHYDR-RXN
C_Map: abcd = badc
Cit
Pi
AcCoA
K_ID: R00230
M_ID: PHOSACETYLTRANS-RXN
C_Map: ab = ab
Fum
K_ID: R00412
Acp
Suc
K_ID: R01325
M_ID: ACONITATEDEHYDR-RXN
C_Map: abcdef = baedfc
CoA
AcP
NADH
CoA
NAD
K_ID: R00519
M_ID: 1.2.1.2-RXN
C_Map: a = a
αKG
Acis-Aco
ATP
ADP
K_ID: R00315
M_ID: ACETATEKIN-RXN
C_Map: ab = ab
K_ID: R00228
M_ID: ACETALD-DEHYDROG-RXN
C_Map: ab = ab
NADH
CO2
ATP
CoA
ADP
Pi
SucCoA
CO2
K_ID: R01900
M_ID: ACONITATEHYDR-RXN
C_Map: abcdef = bceafd
Iso
Ace
K_ID: R00754
M_ID: ALCOHOL-DEHYDROG-RXN
C_Map: ab = ab
Ald
K_ID: R00405
M_ID: SUCCCOASYN-RXN
C_Map: abcd = badc
K_ID: R00216
M_ID: MALIC-NADP-RXN
C_Map: abc + d =abdc
EtOH

Figure 5.7 – *A. succinogenes* carbon atom transition network based on KEGG Pathways and *A. succinogenes* central metabolic network on McKinlay *et al*.. Reactions name represents the two databases: K_ID, reaction identifier on KEGG; M_ID, reaction identifier on MetaCyc; C_Map, carbon atom transition for each reaction. Metabolites: 2PG, 2-phosphoglycerate; 3PG, 3-phosphoglycerate; 6PG, CPD-2961; AcCoA, acetyl-CoA; Ace, acetate; AcP, acetyl phosphate; Acp, acceptor; Ald, acetaldehyde; cis-Aco, cis-Aconitate; Cit, citrate; CO2, carbon dioxide; DHAP, glycerone phosphate; E4P, erythrose-4-phosphate; EtOH, ethanol; F1,6P, fructose-1,6-bisphosphate; F6P, fructose-6-phosphate; For, formate; Fum, fumarate; G1,3P, glycerate-1,3-diphosphate; G1,5L6P, glucono-1,5-lactone-6-phosphate; G3P, glyceraldehyde-3-phosphate; G6P, glucose-6-phosphate; GLC, glucose; Iso, isocitrate; Mal, malate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; Pi, orthophosphate; Pyr, pyruvate; R5P, ribose-5-phosphate; Rbu5P, ribulose-5-phosphate; S7P, sedoheptulose-7-phosphate; Suc, succinate; SucCoA, succinyl-CoA; X5P, xylulose-5-phosphate; αKG, 2-oxoglutarate. The atoms that are colored with dark red are an example of a tracing with carbon atom.

6. CONCLUSIONS

The atom mapping of biochemical reactions may have many applications, in particular, the identification of new biochemical pathways. The algorithm developed during this work allows to trace most atom transitions in a metabolic network and to construct atom transition maps that might be successfully implemented in ^{13}C metabolic flux analysis. The algorithm was applied to trace atom transition patterns in the central metabolism of *E. coli* and in the central metabolism of *A. succinogenes*.

The exploration of KEGG and MetaCyc was very instructive for specific data regarding the transition of atoms from substrates to products in biochemical reactions. With the metabolic information available in those databases, it was possible to build a reliable model. However, some reactions were found not having a carbon atom mapping associated, which limits the applicability of those carbon atom transitions.

This algorithm does not provide a complete curation-free approach because if the reaction does not have an identifier (ID) provided by KEGG or MetaCyc, it is necessary a manual inspection to associate an identifier. In a large-scale application this can be a problem because the manual research will consume much time compared to an automatically research.

All the manual labor in this work was very important to increase the efficiency and the robustness of the method developed. Once again, the manual curation was fundamental to attribute an identifier from MetaCyc or KEGG when they weren't provided in the input file. Some cross-referenced errors can occur when the association is made and the manual curation was useful to correct these errors.

The metabolic information, in particular the reactions, obtained in the databases are generic and as the method can be implemented regardless of the organism or the pathway, the transition of the carbon atoms resulting from the algorithm can be applied to any organism or metabolic network.

For a future work, the methodology implemented that has resulted in an accurate carbon atom transition model, can be expanded if some improvements are made, such as automatic integration to reduce manual curation.

The carbon atom model resulted from the algorithm developed can be incremented in a metabolic engineering tool to track desired compounds through the metabolic pathways. The resulting carbon atom maps can be used to perform tasks of ^{13}C metabolic flux analysis in order to better investigate key aspects of the studied organisms, namely in the case of *A. succinogenes*, which has not been extensively explored as *E. coli.* For example, the atoms transition map resulted for *A. succinogenes* can be used to inspect the partitioning of carbon flux between C_3 and C_4 pathways.

7. BIBLIOGRAPHY

Alberts, B. et al., 2014. *Molecular Biology of the Cell* 6th ed., Garland Science.

Bairoch, a, 2000. The ENZYME database in 2000. *Nucleic acids research*, 28(1), pp.304–305.

Bairoch, A., Universitaire, C.M. & Servet, M., 2000. The ENZYME database in 2000. , 28(1), pp.304–305.

Blazeck, J. & Alper, H., 2010. Systems metabolic engineering: genome-scale models and beyond. *Biotechnol J*, 5(7), pp.647–659.

Carreira, R. et al., 2014. CBFA : phenotype prediction integrating metabolic models with constraints derived from experimental data. , 8(123), pp.1–10.

Caspi, R. et al., 2010. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic acids research*, 38(Database issue), pp.D473–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2808959&tool=pmcentrez&rendertype=abstract [Accessed December 5, 2014].

Christensen, B., Gombert, A.K. & Nielsen, J., 2002. Analysis of flux estimates based on 13C-labelling experiments. *European Journal of Biochemistry*, 269, pp.2795–2800.

Cintolesi, A. et al., 2012. Quantitative analysis of the fermentative metabolism of glycerol in Escherichia coli. *Biotechnology and Bioengineering*, 109(1), pp.187–198.

Copeland, R., 2004. *Enzymes: a practical introduction to structure, mechanism, and data analysis* 2th ed., Wiley-VCH, Inc.

Covert, M.W. et al., 2001. Metabolic modeling of microbial strains in silico. *Trends in biochemical sciences*, 26(3), pp.179–86. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11246024.

Crown, S.B. & Antoniewicz, M.R., 2013. Publishing 13C metabolic flux analysis studies: a review and future perspectives. *Metabolic engineering*, 20, pp.42–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24025367 [Accessed September 26, 2014].

Drauz, K., Groger, H. & May, O., 2012. *Enzyme Catalysis in Organic Synthesis* 3rd ed., Wiley-VCH Verlag GmbH & Co. KGaA. Available at: http://link.springer.com/chapter/10.1007/978-1-4612-2630-7_4 [Accessed December 17, 2014].

Durot, M., Bourguignon, P.-Y. & Schachter, V., 2009. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS microbiology reviews*, 33(1), pp.164–90. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2704943&tool=pmcentrez&rendertype=abstract [Accessed January 27, 2014].

Edwards, J.S., Ibarra, R.U. & Palsson, B.O., 2001. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nature biotechnology*, 19, pp.125–130.

Edwards, J.S. & Palsson, B.O., 2000. The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America*, 97, pp.5528–5533.

Feist, A.M. & Palsson, B.Ø., 2008. The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. *Nature biotechnology*, 26(6), pp.659–667.

Förster, J. et al., 2003. Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome research*, 13(2), pp.244–53. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=420374&tool=pmcentrez&rendertype=abstract [Accessed July 11, 2014].

Gavrilescu, M. & Chisti, Y., 2005. Biotechnology-a sustainable alternative for chemical industry. *Biotechnology advances*, 23(7-8), pp.471–99. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15919172 [Accessed July 11, 2014].

Goeddel, D. V et al., 1979. Expression in Escherichia coli of chemically synthesized genes for human insulin. , 76(1), pp.106–110.

Goffeau, A. et al., 1996. Life with 6000 Genes. *Science*, 274, pp.546–551.

Goto, S., Nishioka, T. & Kanehisa, M., 2000. LIGAND: chemical database of enzyme reactions. *Nucleic acids research*, 28(1), pp.380–2. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=102410&tool=pmcentrez&rendertype=abstract [Accessed February 20, 2014].

Guettler, M. V, Rumler, D. & Jainf, M.K., 1999. succinic-acid-producing strain f rorn the bovine rurnen. , (1 999).

Heath, A.P., Bennett, G.N. & Kavraki, L.E., 2010. Finding metabolic pathways using atom tracking. *Bioinformatics (Oxford, England)*, 26(12), pp.1548–55. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2881407&tool=pmcentrez&rendertype=abstract [Accessed January 13, 2015].

Jeong, H. et al., 2000. The large-scale organization of metabolic networks. *Nature*, 760(1990), pp.651–654.

Jouhten, P. et al., 2009. 13C-metabolic flux ratio and novel carbon path analyses confirmed that Trichoderma reesei uses primarily the respirative pathway also on the preferred carbon source glucose. *BMC systems biology*, 3, p.104. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2776023&tool=pmcentrez&rendertype=abstract [Accessed December 3, 2014].

Kanehisa, M. et al., 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research*, 42(Database issue), pp.D199–205. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3965122&tool=pmcentrez&rendertype=abstract [Accessed July 10, 2014].

Lanfermann, I., Krings, U. & Berger, R.G., 2014. Isotope labelling experiments on the formation pathway of 3-hydroxy-4 , 5-dimethyl-2 (5 H) - furanone from L -isoleucine in cultures of Laetiporus sulphureus. , (November 2013), pp.233–239.

Latendresse, M. et al., 2012. Accurate atom-mapping computation for biochemical reactions. *Journal of chemical information and modeling*, 52(11), pp.2970–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22963657.

Lin, S.K.C. et al., 2008. Substrate and product inhibition kinetics in succinic acid production by Actinobacillus succinogenes. *Biochemical Engineering Journal*, 41(2), pp.128–135. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1369703X08001162 [Accessed December 27, 2014].

Lin, Y. & Tanaka, S., 2006. Ethanol fermentation from biomass resources: current state and prospects. *Applied microbiology and biotechnology*, 69(6), pp.627–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16331454 [Accessed July 9, 2014].

McCloskey, D., Palsson, B.Ø. & Feist, A.M., 2013. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. *Molecular systems biology*, 9(661), p.661. Available at: http://msb.embopress.org/content/9/1/661.abstract.

Mckinlay, J.B. et al., 2010. A genomic perspective on the potential of Actinobacillus succinogenes for industrial succinate production. *BMC Genomics*, 11(1), p.680. Available at: http://www.biomedcentral.com/1471-2164/11/680.

McKinlay, J.B. et al., 2007. Determining Actinobacillus succinogenes metabolic pathways and fluxes by NMR and GC-MS analyses of 13C-labeled metabolic product isotopomers. *Metabolic engineering*, 9(2), pp.177–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17197218 [Accessed December 4, 2014].

Mckinlay, J.B., Zeikus, J.G. & Vieille, C., 2005. Insights into Actinobacillus succinogenes Fermentative Metabolism in a Chemically Defined Growth Medium Insights into Actinobacillus succinogenes Fermentative Metabolism in a Chemically Defined Growth Medium. , 71(11).

Milne, C.B. et al., 2009. Accomplishments in genome-scale in silico modeling for industrial and medical biotechnology. *Biotechnology Journal*, 4, pp.1653–1670.

Nelson, D. & Cox, M., 2008. *Lehninger Principles of Biochemistry* W. H. Free., Available at: http://www.amazon.com/Lehninger-Principles-Biochemistry-David-Nelson/dp/071677108X [Accessed December 17, 2014].

Nielsen, J. & Olsson, L., 2002. An expanded role for microbial physiology in metabolic engineering and functional genomics : moving towards systems biology 1. , 2, pp.175–181.

Niittylae, T. et al., 2009. Comparison of Quantitative Metabolite Imaging Tools and Carbon-13 Techniques for Fluxomics D. A. Belostotsky, ed. *Methods Mol Biol*, 553(1), pp.355–372.

Papin, J. a et al., 2003. Metabolic pathways in the post-genome era. *Trends in biochemical sciences*, 28(5), pp.250–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12765837 [Accessed January 27, 2014].

Park, J.M., Kim, T.Y. & Lee, S.Y., 2009. Constraints-based genome-scale metabolic simulation for systems metabolic engineering. *Biotechnology advances*, 27(6), pp.979–88. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19464354 [Accessed December 10, 2014].

Patil, K.R., Akesson, M. & Nielsen, J., 2004. Use of genome-scale microbial models for metabolic engineering. *Current opinion in biotechnology*, 15(1), pp.64–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15102469 [Accessed November 13, 2014].

Paul Lee, W.-N. et al., 2010. Tracer-based metabolomics: concepts and practices. *Clinical biochemistry*, 43(16-17), pp.1269–77. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2952699&tool=pmcentrez&rendertype=abstract [Accessed December 17, 2014].

Quek, L.-E. et al., 2009. OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis. *Microbial cell factories*, 8, p.25. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2689189&tool=pmcentrez&rendertype=abstract [Accessed October 11, 2014].

Ramanathan, A. & Agarwal, P.K., 2011. Evolutionarily conserved linkage between enzyme fold, flexibility, and catalysis. *PLoS Biology*, 9(11).

Rattray, J.E. et al., 2009. Carbon isotope-labelling experiments indicate that ladderane lipids of anammox bacteria are synthesized by a previously undescribed, novel pathway. *FEMS Microbiology Letters*, 292, pp.115–122.

Ravikirthi, P., Suthers, P.F. & Maranas, C.D., 2011. Construction of an E. Coli genome-scale atom mapping model for MFA calculations. *Biotechnology and bioengineering*, 108(6), pp.1372–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21328316 [Accessed October 16, 2014].

Richmond, C.S. et al., 1999. Genome-wide expression profiling in Escherichia coli K-12. *Nucleic acids research*, 27(19), pp.3821–3835.

Rocha, I. et al., 2010. OptFlux: an open-source software platform for in silico metabolic engineering. *BMC systems biology*, 4, p.45. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2864236&tool=pmcentrez&rendertype=abstract.

Sauer, U., 2006. Metabolic networks in motion: 13C-based flux analysis. *Molecular systems biology*, 2, p.62. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1682028&tool=pmcentrez&rendertype=abstract [Accessed July 22, 2014].

Savile, C.K. et al., 2010. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science (New York, N.Y.)*, 329(5989), pp.305–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20558668 [Accessed July 10, 2014].

Schilling, C.H. et al., 1999. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnology progress*, 15(3), pp.296–303. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10356246.

Stephanopoulos, G.N., Aristidou, A.A. & Nielsen, J., 1998. *Metabolic Engineering*, Elsevier. Available at: http://www.sciencedirect.com/science/article/pii/B9780126662603500170 [Accessed February 2, 2015].

Weitzel, M. et al., 2013. 13CFLUX2 - High-performance software suite for 13C-metabolic flux analysis. *Bioinformatics*, 29(1), pp.143–145.

Wiechert, W., 2001. 13C metabolic flux analysis. *Metabolic engineering*, 3, pp.195–206. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11461141.

Winter, G. & Krömer, J.O., 2013. Fluxomics - connecting 'omics analysis and phenotypes. *Environmental microbiology*, 15(7), pp.1901–16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23279205 [Accessed July 29, 2014].

Zamboni, N., Fischer, E. & Sauer, U., 2005. FiatFlux--a software for metabolic flux analysis from 13C-glucose experiments. *BMC bioinformatics*, 6, p.209.

Zaslavskaia, L. a et al., 2001. Trophic conversion of an obligate photoautotrophic organism through metabolic engineering. *Science (New York, N.Y.)*, 292(June), pp.2073–2075.

Zupke, C. & Stephanopoulos, G., 1994. Modeling of Isotope Distributions and Intracellular Fluxes in Metabolic Networks Using Atom Mapping Matrices. , pp.489–498.

APPENDIX A – DATABASES

Table A.1 - Most explored databases for DNA sequence and genome annotation.

DNA sequence and genome annotation databases		
DDBJ	ddbj.nig.ac.jp	General nucleotide sequence
EMBL	ebi.ac.uk/embl	General nucleotide sequence
GenBank	ncbi.nlm.nih.gov/Genbank	General nucleotide sequence
SEED	seed-viewer.theseed.org	Integrated system for analysis and annotation of genomes using functional subsystems

Table A.2 – Most explored databases for protein and enzymes information

Protein and enzyme databases		
BRENDA	brenda-enzymes.info	Comprehensive enzyme information system
ENZYME	expasy.ch/enzyme	Enzyme nomenclature database providing information for all enzymes and a corresponding EC number
UniProt	ebi.ac.uk/uniprot	Universal Protein Resource with extra protein sequences and annotations from SwissProt , tremble andPIR

Table A.3 – Most explored repositories databases for experimental data

Experimental data repositories		
Array Express	ebi.ac.uk/aerep	Repository of microarray data
GEO	ncbi.nlm.nih.gov/geo	Repository of microarray data
ASAP	asap.ahabs.wisc.edu	Repository with results of functional genomics experiments for selected bacterial species
E. coli multi-omics DB	ecoli.iab.keio.ac.jp	Dataset of transcriptomic, proteomic, metabolomic, and fluxomic experiments for E. coli K12
PubMed	pubmed.org	Database on biomedical literature

Table A.3 – Most explored repositories databases for metabolic models

Metabolic model repositories		
BiGG	bigg.ucsd.edu	Reconstructed genome-scale metabolic models
BioModels	ebi.ac.uk/biomodels	Mathematical models of biological systems

APPENDIX B – DATA RESULTED FROM ALGORITHM FOR *ASUS*

Table B.1 – Carbon atoms' mapping model

KEGG ID	MetaCyc ID	Reaction	Carbon Atom Transition	Enzyme
		Glycolysis and Entner Doudoroff		
		GLC_EX = GLC6P		
R01786	GLUCOKIN-RXN	GLC + ATP = G6P + ADP		2.7.1.2
R02740	RXN-6182	F6P = G6P	abcdef = acdefb	5.3.1.9
	PGLUCISOM-RXN	F6P = G6P	abcdef = acdefb	5.3.1.9
	RXN-13720	F6P = G6P		5.3.1.9
R04779		F6P + ATP = F1,6P + ADP		2.7.1.1
R00756	6PFRUCTPHOS-RXN	F1,6P + ADP = F6P + ATP	abcdef = abcdef	2.7.1.1
R04780		F1,6P + H20 = F6P + Pi		3.1.3.1
R00762	F16BDEPHOS-RXN	F1,6P + H20 = F6P + Pi	abcdef = abcdef	3.1.3.1
R01070		F1,6P = DHAP + G3P		4.1.2.1
R01068	F16ALDOLASE-RXN	DHAP + G3P = F1,6P	abc + def = becadf	4.1.2.1
R01015	TRIOSEPISOMERIZATION-RXN	DHAP = G3P	abc = abc	5.3.1.1
R01061	GAPOXNPHOSPHN-RXN	G3P + Pi +NAD+ = G1,3P + NADH	abc = bca	1.2.1.12
R01512	PHOSGLYPHOS-RXN	ATP + G1,3P = ADP + 3PG	abc = abc	2.7.2.3
R01518	3PGAREARR-RXN	2PG = 3PG	abc = abc	5.4.2.1
R00658	2PGADEHYDRAT-RXN	PEP + H2O = 2PG	abc = abc	4.2.1.1
R00200	PEPDEPHOS-RXN	PEP + ADP = PYR + ATP	abc = abc	2.7.1.40
R00209	Pyruvdeh-RXN; RXN0-1134; RXN0-1133; RXN0-1132	NAD+ + COA + PYR = NADH + ACCOA + CO2	abcdefghijklmnopqrstu + vwx = vbacdefghijwklmnopqrstu + x	1.2.4.1; 1.8.1.4; 2.3.1.12
R01196	PYRUFLAVREDUCT-RXN	ACCOA + CO2 = PYR + CoA		1.2.7.1
R02036	PGLUCONDEHYDRAT-RXN	GLC6P = 6PGC	abcdef = becadf	4.2.1.12
R05605	KDPGALDOL-RXN	6PGC = G3P + PYR	abcdef = cbe + adf	4.1.2.14
		Pyruvate and Glyoxylate and dicarboxylate		
R00217	OXALODECARB-RXN	PYR + CO2 = OAA	abc + d = abdc	4.1.1.3; 1.1.1.38; 4.1.1.3
R00212	PYRUVFORMLY-RXN	CoA + PYR = AcCoA + For	abc = ab + c	2.3.1.54
R00228	ACETALD-DEHYDROG-RXN	NAD + CoA + Ald = NADH + AcCoA	ab = ab	1.2.1.10
R00754	ALCOHOL-DEHYDROG-RXN	NADH + Ald = NAD + EtOH	ab = ab	1.1.1.1

R00230	PHOSACETYLTRANS-RXN	CoA + AcP = AcCoA + Pi	ab = ab	2.3.1.8
R00315	ACETATEKIN-RXN	ADP + AcP = ATP + Ace		2.7.2.1; 2.7.2.15
R00519	1.2.1.2-RXN	NAD + For = NADH + CO2	a = a	1.2.1.2
		Pentose-phosphate pathway		
R02736	GLU6PDEHYDROG-RXN	G6P + NADP = G1,5L6P + NADPH	abcdef = abcdef	1.1.1.49
R02035	6PGLUCONOLACT-RXN	G1,5L6P = 6PG	abcdef = abcdef	3.1.1.31
R01528	RXN-9952	6PG + NADP = Rbu5P + CO2 + NADPH	abcdef = eadbc + f	1.1.1.44
R01056	RIB5PISOM-RXN	Rbu5P = R5P	abcde = bdeca	5.3.1.6
R01641	1TRANSKETO-RXN	X5P + R5P = G3P + S7P	abcde + fghij = ebd + afcgjhi	2.2.1.1
R01827	TRANSALDOL-RXN	G3P + S7P = E4P + F6P	abc + defghij = jeig + bdcahf	2.2.1.2
R01067	2TRANSKETO-RXN	E4P + X5P = G3P + F6P	abcd + efghi = ifh + bedcag	2.2.1.1
R01830	Não encontra	F6P + G3P = E4P + X5P		2.2.1.1
R01529	RIBULP3EPIM-RXN	Rbu5P = X5P	abcde = abcde	5.1.3.1
		TCA cycle		
R00351	CITSYN-RXN	AcCoA + OAA = CIT + CoA	ab + cdef = caebfd	2.3.3.1
R00362	CITLY-RXN	Ace + OAA = CIT		4.1.3.6
R00342	MALATE-DEH-RXN	NADH + OAA = NAD + Mal	abcd = abcd	1.1.1.37; 1.1.1.299
R01082	FUMHYDR-RXN	Fum + H2O = Mal	abcd = badc	4.2.1.2
R00412	SUCC-FUM-OXRED-RXN	Suc + Acp = Fum + RedAcp		1.3.99.1
R00405	SUCCCOASYN-RXN	ADP + SucCoA + Pi= ATP + CoA + Suc	abcdefghijklmnopqrstuvwxz = baefghijklmnqrstuvwxz + cdop	6.2.1.5
			abcdefghijklmnopqrstuvwxz = baefghijklmnqrstuvwxz + dcpo	6.2.1.5
R01197				1.2.7.3
R02570	RXN0-1147	CoA + DSsucci = SucCoA + EnzN6		2.3.1.61
R07618				1.8.1.4
R03316				1.2.4.2
R00621				1.2.4.2
R00268	RXN-8642	alphaKG + CO2 = Osuc	abcde + f = badcfe	1.1.1.42
R01899	RXN-9951	NADPH + Osuc = NADP + Icit	abcdef = abcdef	1.1.1.42
	ISOCITDEH-RXN	NADP + Icit = NADPH + alphaKG + CO2	abcdef = badcf + e	1.1.1.42
R00709				1.1.1.41; 1.1.1.286
R01900	ACONITATEHYDR-RXN	cis-Aco + H2O = Icit	abcdef = bceafd	4.2.1.3
R01325	ACONITATEDEHYDR-RXN	Cit = cis-Aco + H2O	abcdef = bafdce	4.2.1.3

R00216	MALIC-NADP-RXN	MAL = PYR + CO2	abcd = abd + c	1.1.1.40
R00341	PEPCARBOXYKIN-RXN	PEP + CO2 = OAA	abc + d = abdc	4.1.1.49

Table B.2 – Reactions' directionality and external Identifiers (IDs)

Reaction KEGG	Reaction MetaCyc	Reaction	Direction	Uniprot IDs	Rhea IDs	Ligand IDs
Glycolysis and Entner Doudoroff						
		GLC_EX = GLC6P				
R01786	GLUCOKIN-RXN	GLC + ATP = G6P + ADP	LEFT-TO-RIGHT		17825	R01786
R02740	RXN-6182	F6P = G6P	REVERSIBLE		11816	R02740
	PGLUCISOM-RXN	F6P = G6P	REVERSIBLE	P06744, P28718, Q7LZP0, O83488, Q9JTW1, Q59000, Q9PMD4, O25781, O84382, Q9JSS6, P81181, P08059, P50309, P13376, P13375, P12709, P0A6T1, P06745, P13377, P12341, P18240, P29333, P34796, P34797, P54240, P54242, Q59088, P78033, P52983, P49105, P42862, P42863, Q9SB57, O82058, O82059, O61113, P78917, Q9RMC1, Q9X670	11816	R00771
	RXN-13720	F6P = G6P				
R04779		F6P + ATP = F1,6P + ADP				
R00756	6PFRUCTPHOS-RXN	F1,6P + ADP = F6P + ATP	LEFT-TO-RIGHT	Q7M3F7, Q7M4J2, P12382, Q27665, P52034, Q7M4K9, P43863, P20275, P47457, Q01813, Q07636, P16861, P16862, P00512, P0A796, P06999, P08237, P00511, P70927, Q7M3F5, Q9TWY0, P30835, Q03215, Q03216, P80019, Q59214, Q27705, P75476, P72830, Q55988, Q49084, O08333	16109	R00756
R04780		F1,6P + H20 = F6P + Pi				
R00762	F16BDEPHOS-RXN	F1,6P + H20 = F6P + Pi	NULL	P09199, P19112, P22780, P09467, Q9JUL6, Q9CIU9, P09202, Q45597, P45292, Q9PP83, P19912, P19911, Q9HGZ4, P09201, P0A993, P00636, P23014, P27994, P14766, P09195, Q7M0S7, P25851, Q05079, P46275, P46276, Q43139, P00637, O65827, Q42796, P46267, Q07204, P22418	11064	R00762
R01070		F1,6P = DHAP + G3P				

R01068	F16ALDOLASE-RXN	DHAP + G3P = F1,6P	REVERSIBLE	Q7LZE8, Q7M2K6, P07764, Q9URB4, P14223, Q07159, P07752, P14540, P07341, P0AB71, P04075, P05062, P09972, P05064, P05063, P22197, P00883, P27995, P05065, P00884, P09117, P17784, P29356, P16096, P08440, P44429, P13243, O51401, P47269, Q9CED4, Q9JW15, Q59100, Q59101, Q7M4Z5, Q7M4Z4, P19537, Q01516, Q01517, Q91384, P52210, Q42690, P53447, Q7LZE9, P53818, P46257, P46256, Q42476, P75089, O22486, Q40677, O65581, Q9SVJ6, P93565, P50923, O04975, P36580, P53444	14729	R01068
R01015	TRIOSEPISOMERIZATION-RXN	DHAP = G3P	REVERSIBLE	P19118, P21820, P48501, P62002, O27120, P27876, Q9PMQ6, P19583, Q9UXX2, P47721, P47670, P50918, P43727, Q58923, O28965, Q59182, O59536, P36204, Q9JW31, P04828, P00943, P00942, P00940, P60175, P0A858, P60174, P00941, P15426, P17751, P00939, P12863, P07669, P35144, P48494, Q7M4X7, P29613, P30741, Q01893, P48499, P48496, P46226, P46225, P48492, Q56738, Q7LZE5, P46711, O32757, O74067, P48491	18585	R01015
R01061	GAPOXNPHOSPHN-RXN	G3P + Pi +NAD+ = G1,3P + NADH	REVERSIBLE	P07486, P08477, P17878, Q01651, Q27890, Q27820, Q7M188, Q7LZR1, Q58546, P55971, P46795, P07487, Q01558, Q7M187, O25902, Q9JWT8, P29272, P47543, O83816, P20445, P26517, P09124, P00362, P00358, P00359, P00360, P00356, P0A9B2, P0A9B6, P17721, P04406, P17244, P04796, P19089, P26518, P26988, P04970, P17329, P17330, P17331, P08439, P00357, P16858, P26521, P26520, P26519, P04797, P25861, P00361, P09317, P22512, P22513, P10097, P17819, P28844, P08735, P09316, Q9JX51, O67161, P44304, O34425, P64178, Q46450, P50321, P50322, P34917, P34918, Q07234, P24748, P24746, P24749, P24751, P24750, Q60143, Q64467, P46406, Q59800, P80534, Q9UW96, P32809, P32810, P20287, P35143, P25858, P10618,	10300	R01061

				Q7M517, Q7M516, Q09054, Q43247, P23722, P17729, Q01597, P32637, P29497, P32638, P32635, P32636, Q59309, P34783, Q00584, Q42671, P34920, P39460, Q37265, Q37264, Q01077, P54270, Q48335, Q43833, Q59906, P46713, P75358, Q7M2K2, Q43359, O68075, Q08060, O49222, P34922, O04106, P49644, O32755, P34924, O59841, P78958, P54118		
R01512	PHOSGLYPHOS-RXN	ATP + G1,3P = ADP + 3PG	REVERSIBLE	P11977, P09411, P09041, P07205, P16617, Q37743, Q58058, P56154, O29119, Q01655, Q9URB3, P47542, O27121, Q9PMQ5, P40924, P43726, O66519, Q9JWS8, Q9CIW1, P36204, Q59181, P50319, P50310, P51903, P18912, P41757, P27362, P24269, P00560, P25055, Q01604, P00558, P07377, P07378, P14828, P09404, P29408, P20972, P20971, P24590, P29409, P33161, P29405, P50317, P50315, P29407, P61884, Q42542, P50318, P46712, P78018, Q49073, Q42961, Q42962, O81394, P41758, O32756, P38667, P08966, P08967, P0A799, P09188, P14228, P08891, P08892, P08893, P12782, P12783	14801	R01512
R01518	3PGAREARR-RXN	2PG = 3PG	REVERSIBLE	P30792, P52832, P44865, P30798, P62707, P39773, P47669, Q9CEU3, P56196, Q9JTF2, Q9PI71, Q9CIM0, P00950, P18669, P15259, P16290, P35167, P33158, Q06464, P36623, P35494, P37689, P35493, Q9VAN7, Q42908, Q12326, P53531, P51379, P75167, P72649, P74507, Q49006, O24246, Q9X519	15901	R01518
R00658	2PGADEHYDRAT-RXN	PEP + H2O = 2PG	REVERSIBLE	P04764, P07322, P07323, P06733, P19140, P25704, P30575, P33675, P47647, Q9JU46, P37869, Q9URB2, Q9CIT0, Q9CHS7, P43806, P42448, P42848, Q05524, P42897, P40370, P51913, P26300, Q42887, P25696, P00924, P00925, P0A6P9, P29201, P09104, P21550, P08734, Q7M4Y6, P15429, P13929, P15007, P17182, P17183, P26301, P31683, Q54274, P42896, Q27727, Q9UWJ5, P48285, Q12007, P42222, P75189, P77972, Q49059, P42895,	10164	R00658

				Q42971, Q43130, O69174		
R00200	PEPDEPHOS-RXN	PEP + ADP = PYR + ATP	REVERSIBLE	Q7M034, P11979, P11980, Q07637, P34038, P43924, Q57572, P0AD61, P19680, Q9PIB0, P80885, Q9UYU6, Q9JWX8, P47458, Q46078, P30614, P22200, Q27788, P51182, P51181, P31865, P00549, P00548, P30613, O75758, P12928, O30853, P30615, P30616, Q02499, P22360, P21599, P14618, Q42954, Q40545, P52480, P52489, P78031, Q55863, P73534, O65595, Q42806, Q43117, Q10208	18157	R00200
R00209	Pyruvdeh-RXN; RXN0-1134; RXN0-1133; RXN0-1132	NAD+ + COA + PYR = NADH + ACCOA + CO2				
R01196	PYRUFLAVREDUCT-RXN	ACCOA + CO2 = PYR + CoA				
R02036	PGLUCONDEHYDRAT-RXN	GLC6P = 6PGC	LEFT-TO-RIGHT	P0ADF6, P21909, P56111, Q9JTV9, Q02139, Q44327	17277	R02036
R05605	KDPGALDOL-RXN	6PGC = G3P + PYR	LEFT-TO-RIGHT	P44480, P0A955, P00885, Q9JR44, O25729, Q9ZKB4, O83578, Q9WXS1, P50846, P38448, Q55872, P94802	17089	R05605
Pyruvate and Glyoxylate and dicarboxylate						
R00217	OXALODECARB-RXN	PYR + CO2 = OAA	LEFT-TO-RIGHT	P13187, P13155, P13156, Q58628	15641	R00217
R00212	PYRUVFORMLY-RXN	CoA + PYR = AcCoA + For	REVERSIBLE	P32674, O26446, P75793, P42632, O28823, O32797, Q46266, P09373, P37836	11844	R00212
R00228	ACETALD-DEHYDROG-RXN	NAD + CoA + Ald = NADH + AcCoA	REVERSIBLE	Q52434, P0A9Q7, P77580, P71866, Q52060, Q52016, O85978	23288	R00228
R00754	ALCOHOL-DEHYDROG-RXN	NADH + Ald = NAD + EtOH	REVERSIBLE		25290	R00754
R00230	PHOSACETYLTRANS-RXN	CoA + AcP = AcCoA + Pi	REVERSIBLE	P38503, P47541, P45107, Q9PPL9, Q9CF22, P0A9M8, Q9RY77, P75359, P73662	19521	R00230
R00315	ACETATEKIN-RXN	ADP + AcP = ATP + Ace	REVERSIBLE	Q9JTM0, P38502, P37877, Q9PPL8, Q9CE36, P47599, Q9CE35, Q9JT07, P44406, P0A6A3, P73162	11352	R00315, R00315
R00519	1.2.1.2-RXN	NAD + For = NADH + CO2	REVERSIBLE		15985	R00519
Pentose-phosphate pathway						
R02736	GLU6PDEHYDROG-RXN	G6P + NADP = G1,5L6P + NADPH	LEFT-TO-RIGHT	P11411, P12646, Q00612, P54996, P21907, P54547, O51581, P0AC53, P11413, P29686,	15841	R02736

				P44311, P56110, Q9JTW0, O25730, O51240, O66787, P05370, P11410, P11412, Q9R5T2, P48828, P37986, Q8IKU0, Q27741, P48826, Q42919, P37830, Q9FY99, Q49700, P73411, Q8L743, Q43793, O65856, Q43839, O81978, O24357, O24358, O24359, O22404, O22405, O22406, Q9FJI5, Q9LK23		
R02035	6PGLUCONOLACT-RXN	G1,5L6P = 6PG	LEFT-TO-RIGHT	P63337	12556	R02035
R01528	RXN-9952	6PG + NADP = Rbu5P + CO2 + NADPH	LEFT-TO-RIGHT		10116	R01528
R01056	RIB5PISOM-RXN	Rbu5P = R5P	REVERSIBLE	P0A7Z0, Q9CDI7, P44725, Q58998, Q9JTM5, Q9PP08, P37351, P74234, Q55766	14657	R01056
R01641	1TRANSKETO-RXN	X5P + R5P = G3P + S7P	REVERSIBLE	P29401, P33570, P21725, Q58094, Q9JTR1, Q9CF56, P21726, P47312, Q9PM31, P43757, Q58092, P45694, Q52723, Q9Z475, P34736, P33315, Q42675, Q42676, Q42677, P46708, P75611, P73282, Q49047, O20250, O78327, Q9URM2, Q9RFB7, P23254, P27302, P29277, P22976	10508	R01641
R01827	TRANSALDOL-RXN	G3P + S7P = E4P + F6P	REVERSIBLE	P17441, P17440, P37837, P45055, Q9JSU1, P0A867, Q9PIL5, Q58370, P34214, P0A870, P15019, P53228, P51778, P72797, O04894, P78865	17053	R01827
R01067	2TRANSKETO-RXN	E4P + X5P = G3P + F6P	REVERSIBLE	P22976, P29401, P33570, P21725, Q58094, Q9JTR1, Q9CF56, P21726, P47312, Q9PM31, P43757, Q58092, P45694, Q52723, Q9Z475, P34736, P33315, Q42675, Q42676, Q42677, P46708, P75611, P73282, Q49047, O20250, O78327, Q9URM2, Q9RFB7, P23254, P27302, P29277	27626	R01067
R01830		F6P + G3P = E4P + X5P				
R01529	RIBULP3EPIM-RXN	Rbu5P = X5P	REVERSIBLE	P40117, Q9CEB9, P0AG07, Q9PI57, Q9JUA9, P45455, Q43157, Q43843, P74061, O23782, P51012	13677	R01529
TCA cycle						
R00351	CITSYN-RXN	AcCoA + OAA = CIT + CoA	REVERSIBLE	Q9JRA5, Q9JQX0, Q9CHQ6, P39120, Q9PLZ5, P20902, P51033, P39119, P42457, P18789,	16845	R00351, R00351

				P20903, P51038, P34085, P34575, Q43175, P43635, O24259, O32705, P00890, P08679, Q47237, P20115, P26491, P00889, P20901, P09948, P21553, Q9JRA5, Q9JQX0, P31660, Q9CHQ6, P39120, Q9PLZ5, P20902, P51033, P39119, P42457, P18789, P20903, P51038, P34085, P34575, Q43175, P43635, O24259, O32705, P00890, P08679, P0ABH7, Q47237, P20115, P26491, P00889, P20901, P09948, P21553		
R00362	CITLY-RXN	Ace + OAA = CIT	LEFT-TO-RIGHT	P02903, P44459, P44460, P44461, P75726, Q9CGA9, Q9CGA8, P69330, Q9CGA7, Q9KTU0, P17725, P45413, Q9RLT6, Q9RLT5, Q9RLT4	10760	R00362
R00342	MALATE-DEH-RXN	NADH + OAA = NAD + Mal	REVERSIBLE	Q07841, P33163, P44427, Q9PHY2, P22133, P17505, P32419, P25077, P14152, Q8R1P0, P19446, P17783, P04636, P58408, P49814, Q93ZA7, Q7M4Y9, Q7M4Z0, P10887, P11386, P19983, P19981, P19979, P19977, P19982, P19978, P19980, P16142, P46487, P46488, Q43744, Q59202, Q55383, Q42972, O81278, O81279, O65363, O65364, O81609, Q43743, Q42686, P93106, Q04820, O48903, O48904, O48905, O48906, O24047, Q9XTB4, P50917, Q49981, Q9ZP05, Q9ZP06, Q9SN86, P11708, Q07841, P33163, P44427, Q9PHY2, P22133, P17505, P32419, P25077, P14152, Q8R1P0, P19446, P17783, P04636, P58408, Q58820, P49814, Q93ZA7, Q7M4Y9, Q7M4Z0, P10887, P11386, P19983, P19981, P19979, P19977, P19982, P19978, P19980, P16142, P46487, P46488, Q43744, Q59202, Q55383, Q42972, O81278, O81279, O65363, O65364, O81609, Q43743, Q42686, P93106, Q04820, O48903, O48904, O48905, O48906, O24047, Q9XTB4, P50917, Q49981, Q9ZP05, Q9ZP06, Q9SN86	21432	R00342
R01082	FUMHYDR-RXN	Fum + H2O = Mal	LEFT-TO-RIGHT	Q04718, Q58034, O69294, Q9JTE3, Q9JTR0, Q58690, O25883, O84863, Q51404, Q7M4Z3, P39461, Q55674, P93033, Q43180, O94552, O66271, Q60022, P07343, P08417, P10173,	12460	R01082

				P14408, Q04718, Q58034, O69294, Q9JTE3, P14407, Q9JTR0, Q58690, O25883, O53446, O84863, Q51404, Q7M4Z3, P39461, Q55674, P93033, Q43180, O94552, O66271, Q60022, P07343, P08417, P05042, P0AC33, P10173, P14408		
R00412	SUCC-FUM-OXRED-RXN	Suc + Acp = Fum + RedAcp				
R00405	SUCCCOASYN-RXN	ADP + SucCoA + Pi= ATP + CoA + Suc		O28098, P53594, O28733, P45101, P45102, Q9JUT0, Q58643, O26663, P80886, Q9PHY1, Q9JUS9, P80865, Q9PHY0, O67729, O67546, P53593, P0AGE9, P0A836, O82662	17661	R00405
R01197						
R02570	RXN0-1147	CoA + DSsucci = SucCoA + EnzN6	REVERSIBLE			R02570
R07618						
R03316						
R00621						
R00268	RXN-8642	alphaKG + CO2 = Osuc	LEFT-TO-RIGHT			R00268
R01899	RXN-9951	NADPH + Osuc = NADP + Icit	LEFT-TO-RIGHT		25588	R01899
	ISOCITDEH-RXN	NADP + Icit = NADPH + alphaKG + CO2	REVERSIBLE			
R00709						
R01900	ACONITATEHYDR-RXN	cis-Aco + H2O = Icit	REVERSIBLE		22144	R01900
R01325	ACONITATEDEHYDR-RXN	Cit = cis-Aco + H2O	LEFT-TO-RIGHT	P16276, Q01059, Q9TSA1, Q9JTI5, P37032, Q9PP88, P09339, Q9CHQ5, Q9JT05, Q99798, P28271, P21399, Q42669, P19414, P20004, P49609, P74582, Q42560, Q9SZ36, O04916, P16276, Q01059, Q9TSA1, Q9JTI5, P37032, P81291, Q9PP88, P36683, O53166, P25516, P09339, Q9CHQ5, Q9JT05, Q99798, P28271, P21399, Q42669, P19414, P20004, P49609, P74582, Q42560, Q9SZ36, O04916	10228	R01325
R00216	MALIC-NADP-RXN	MAL = PYR + CO2	LEFT-TO-RIGHT	P12628, P06801, P13697, P16243, O48656, P22178, P93139, P34105, P28227, P36444, P40927, P37223, P48163, Q16798, O50015,	18253	R00216

				O04935, O04936, P37222, Q42888, Q42889		
R00341	PEPCARBOXYKIN-RXN	PEP + CO2 = OAA	LEFT-TO-RIGHT	P10963, Q9PP01, P43923, P22259, P42066, P49292, O49547, Q9T074, Q39294, Q9RC49	18617	R00341

APPENDIX C – KEGG'S PATHWAYS

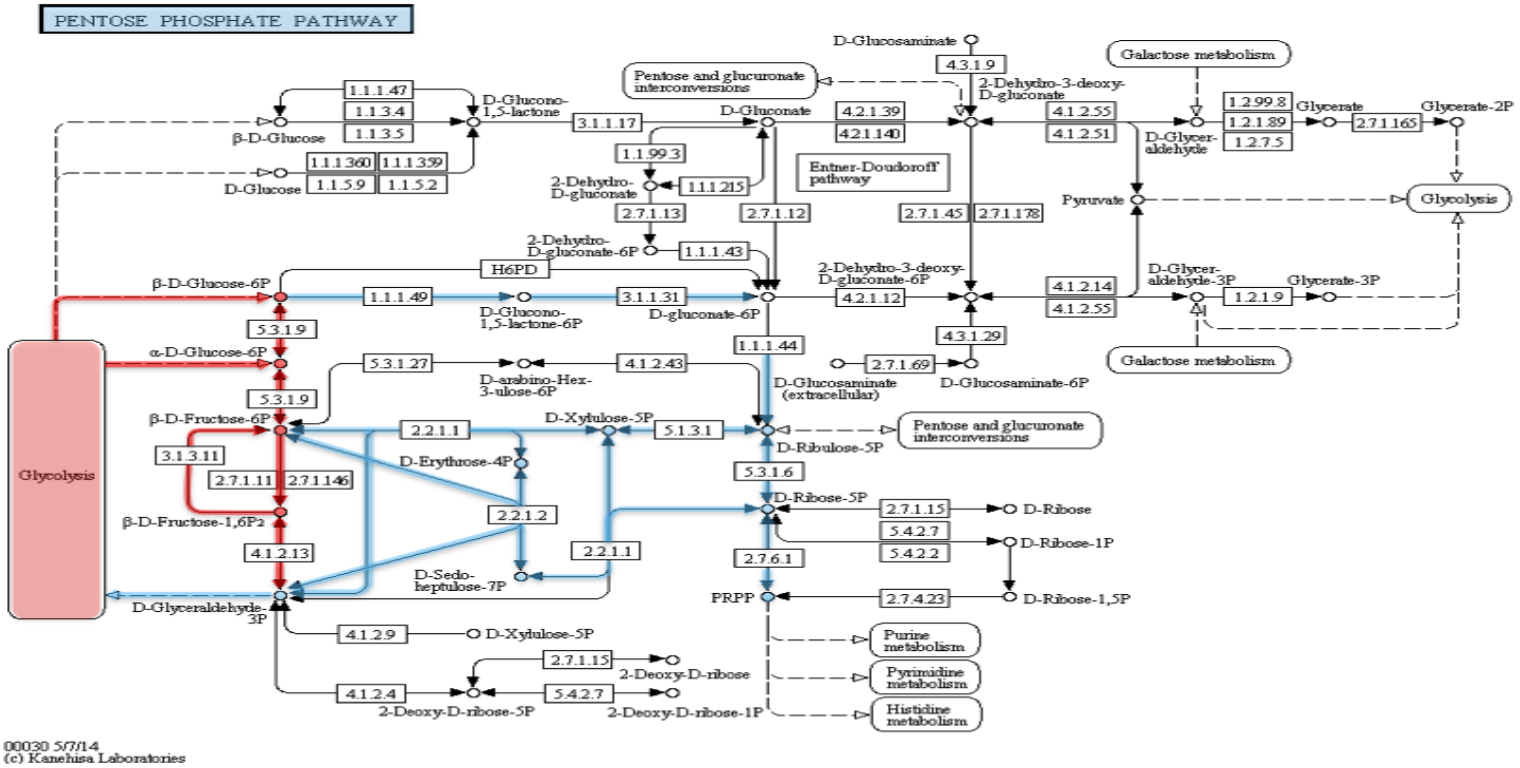


Figure C.1 – Pentose phosphate pathway from KEGG Pathway, adapted for *A. succinogenes*. Colored dots and arrows represent the compounds and the reactions pathway (in this order) to build the atom transition map. Blue, represent the compounds and the reactions of interest in Pentose phosphate pathway; red, identifies the compounds and the reactions from Glycolysis/Gluconeogenesis pathway that interact in Pentose phosphate pathway.

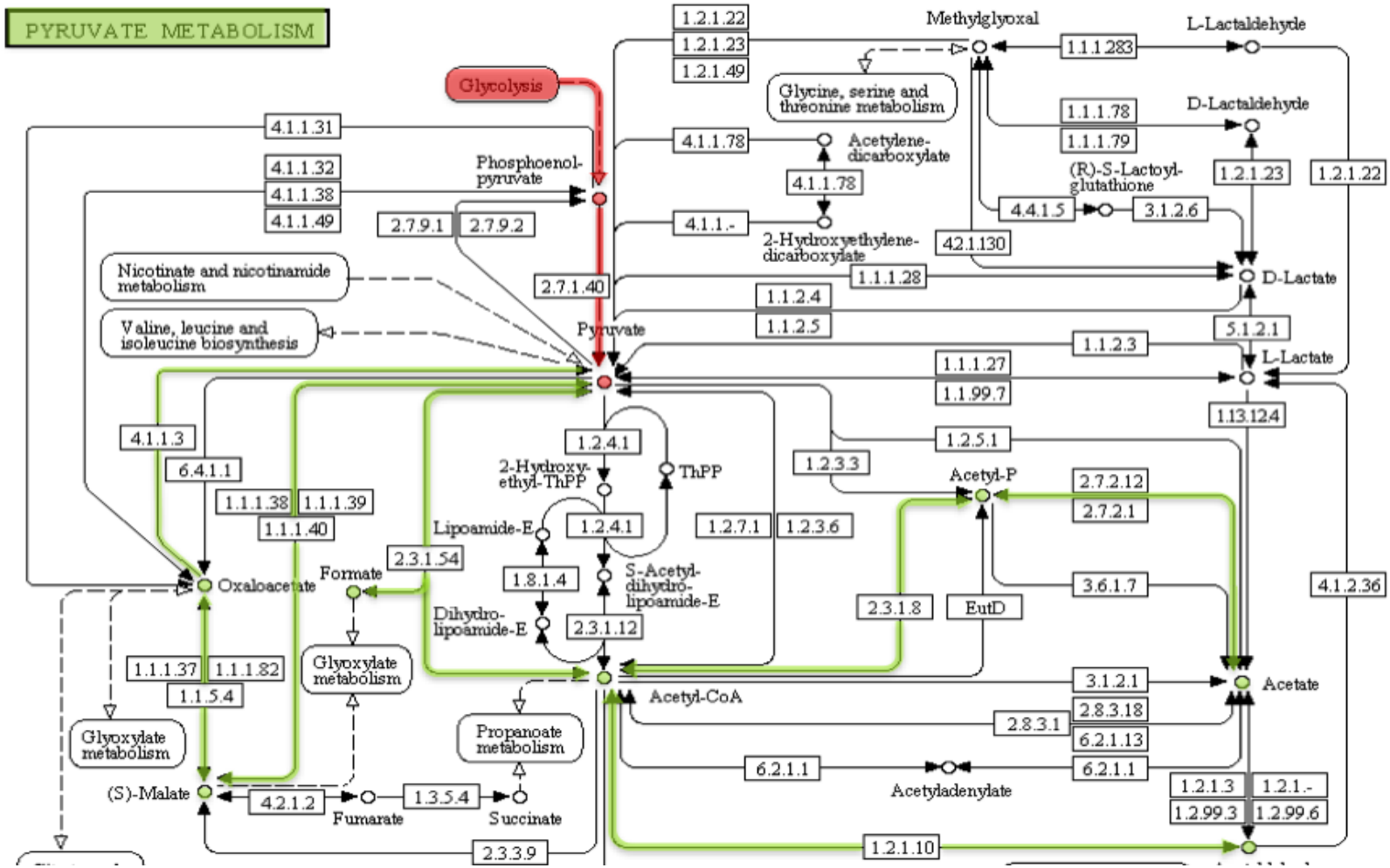


Figure C.2 – Pyruvate metabolism from KEGG Pathway, adapted for *A. succinogenes*. Colored dots and arrows represent the compounds and the reactions pathway (in this order (QUERO DIZER POR ESTA ORDEM)) to build the atom transition map. Green, represent the compounds and the reactions of interest in Pyruvate metabolism; red, identifies the compounds and the reactions from Glycolysis/Gluconeogenesis pathway that interact in Pyruvate metabolism.

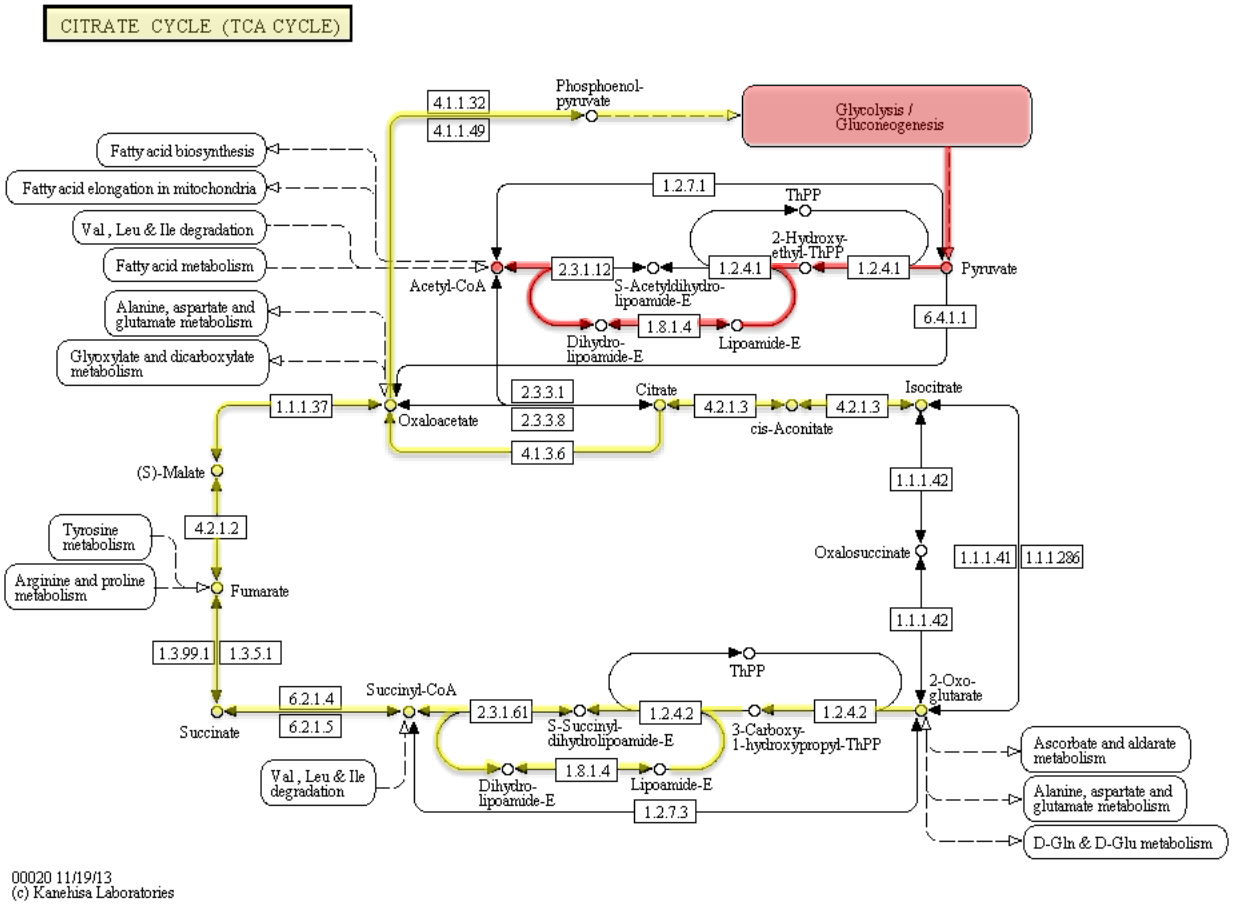

CITRATE CYCLE (TCA CYCLE)
Phosphoenol-pyruvate
4.1.1.32
4.1.1.49
Glycolysis / Gluconeogenesis
Fatty acid biosynthesis
Fatty acid elongation in mitochondria
Val , Leu & Ile degradation
Fatty acid metabolism
Alanine, aspartate and glutamate metabolism
Glyoxylate and dicarboxylate metabolism
1.2.7.1
ThPP
2-Hydroxy-ethyl-ThPP
Acetyl-CoA
2.3.1.12
S-Acetyldihydro-lipoamide-E
1.2.4.1
1.2.4.1
Pyruvate
1.8.1.4
Dihydro-lipoamide-E
Lipoamide-E
6.4.1.1
1.1.1.37
Oxaloacetate
2.3.3.1
2.3.3.8
4.1.3.6
Citrate
4.2.1.3
cis-Aconitate
4.2.1.3
Isocitrate
(S)-Malate
1.1.1.42
Tyrosine metabolism
4.2.1.2
Arginine and proline metabolism
Fumarate
Oxalosuccinate
1.1.1.41
1.1.1.286
1.1.1.42
1.3.99.1
1.3.5.1
ThPP
2-Oxo-glutarate
6.2.1.4
6.2.1.5
Succinyl-CoA
Succinate
2.3.1.61
S-Succinyl-dihydrolipoamide-E
1.2.4.2
3-Carboxy-1-hydroxypropyl-ThPP
1.2.4.2
Ascorbate and aldarate metabolism
Val , Leu & Ile degradation
1.8.1.4
Dihydro-lipoamide-E
Lipoamide-E
Alanine, aspartate and glutamate metabolism
1.2.7.3
D-Gln & D-Glu metabolism
00020 11/19/13
(c) Kanehisa Laboratories

Figure C.3 – Citrate cycle (TCA cycle) from KEGG Pathway, adapted for *A. succinogenes*. Colored dots and arrows represent the compounds and the reactions pathway (in order) to build the atom transition map. Yellow, represent the compounds and the reactions of interest in TCA cycle; red, identifies the compounds and the reactions from Glycolysis/Gluconeogenesis pathway that interact in TCA cycle.

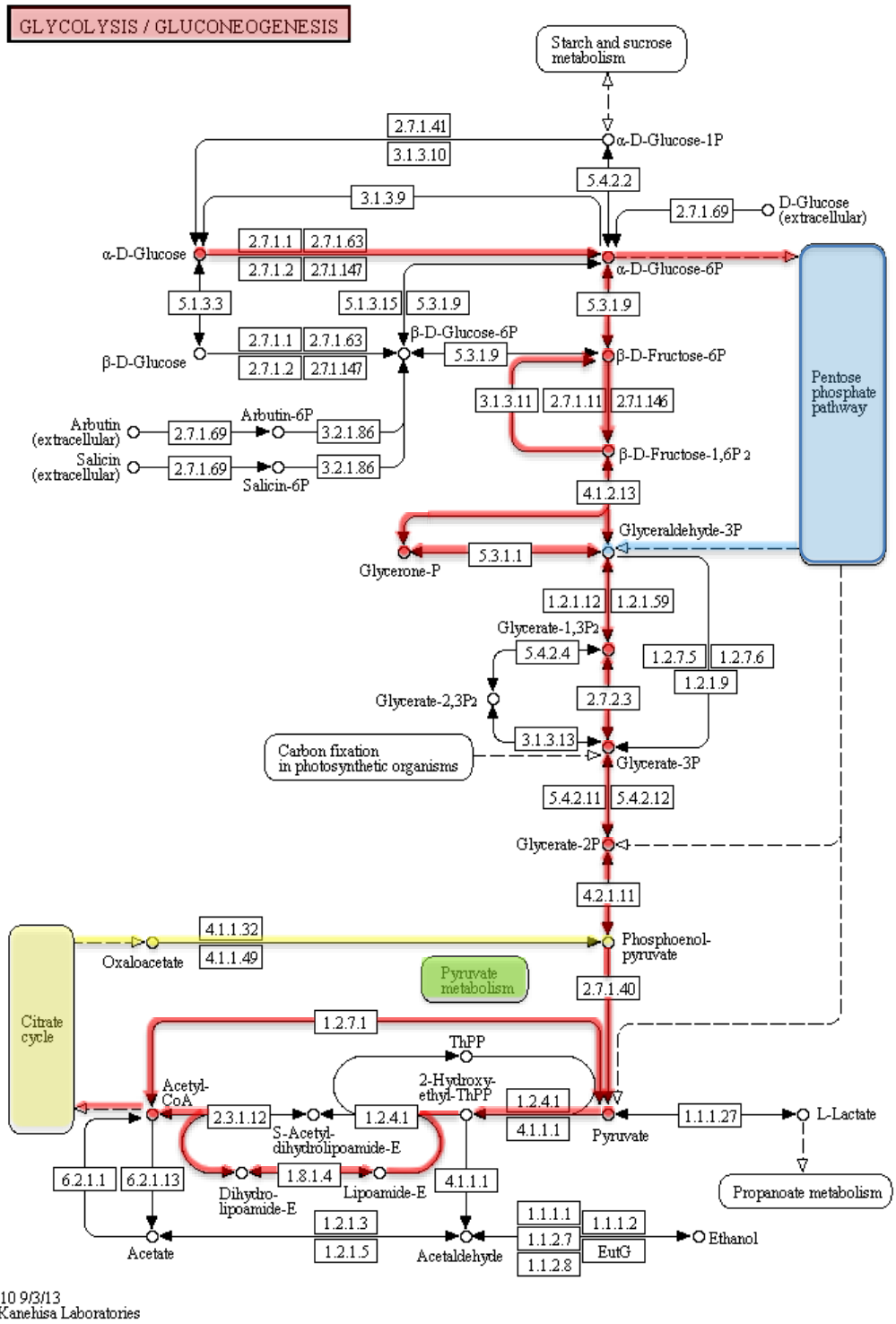


Figure C.4 – Glycolysis/Gluconeogenesis pathway from KEGG Pathway, adapted for *A. succinogenes*. Colored dots and arrows represent the compounds and the reactions pathway (in order) to build the atom transition map. Red represents the compounds and the reactions of interest in Glycolysis/Gluconeogenesis pathway; blue, identifies the compound from Pentose phosphate pathway that interacts in Glycolysis/Gluconeogenesis pathway; yellow, identifies the compound and the reaction from Pentose phosphate pathway that interacts in Glycolysis/Gluconeogenesis pathway; green represents the Pyruvate metabolism.