Reconstruction and Use of Microbial Metabolic Networks: the Core *Escherichia coli* Metabolic Model as an Educational Guide

Jeffrey D. Orth, R.M.T. Fleming, Bernhard Ø. Palsson*

June 2, 2009

Contents

---

*Mailing Address: Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, Mail Code 0412, La Jolla, CA 92093-0412. Email Address: palsson@ucsd.edu Phone: 858-534-5668

1   Introduction

Biochemical network reconstructions have become popular tools in systems biology [131]. There are different types of reconstructions representing various types of biological networks (metabolic, regulatory, transcription/translation), although metabolic network reconstructions have been the most extensively used. Early reconstructions of *Escherichia coli* metabolism were small, containing only central metabolic reactions [190]. Today, there are many genome-scale reconstructions available, based on all known metabolic genes in the annotated genome of an organism along with other data sources [47, 55, 126, 157]. Metabolic network reconstructions are biochemically, genetically and genomically (BiGG) structured databases of biochemical reactions and metabolites. They contain information such as exact reaction stoichiometry, reaction reversibility, and the relationships between genes, proteins, and reactions. Reconstructed networks serve as flexible BiGG knowledge bases [57], storing curated information in a useful format while allowing for content to be updated based on new research. Although many organisms have similar central metabolic networks, there can be differences even between two closely related organisms. Network reconstructions are therefore organism specific [145]. For most applications of network reconstructions, it is necessary to convert the network into a mathematical model. Metabolic models are usually formulated as a stoichiometric matrix, while regulatory network models are often formulated as Boolean networks.

There have been many practical uses of network reconstructions. Some reconstructions have been used as tools to study bacterial evolution. The effects on metabolism of adding or removing genes from the network can be simulated, enabling studies of horizontal gene transfer [142], adaptation to new environments [141], and evolution to minimal genomes [143]. Reconstructions can also be used for analysis of network properties. In these studies, methods have been developed to determine the interactions between different sets of reactions and compounds, improving our understanding of the organisms. Some examples include identification of alternate optimal network states [112], identification of sets of coupled reactions [22], and studies of the states of regulatory networks [7, 162]. It has also been determined by simulating thousands of growth conditions that *E. coli* contains a set of common, high-flux backbone reactions [3]. Network reconstructions have been extensively used to study the phenotypic behavior of wild-type and mutant stains under a variety of conditions, linking genotypes with phenotypes. These predictions have been verified to match the results of experimental studies [51]. Such phenotypic simulations have allowed for the prediction of growth after genetic manipulations [171, 173], growth after adaptive evolution [148], and prediction of essential genes [90]. Another promising use of reconstructions is in the discovery of unknown biological features. By comparing experimental data such as growth phenotypes [150], metabolic flux measurements [77], or gene essentiality [107] to model based predictions, missing content in reconstructions can be identified. The reconstructions can be updated and new biological knowledge can be elucidated. Finally, network reconstructions have proven to be very useful for metabolic engineering and synthetic biology [132]. Because of the capacity of models to be used to predict growth and metabolite secretion phenotypes, its is possible to predict the genetic interventions most likely to produce a strain with desired properties [62]. Model based algorithms can even predict non-intuitive designs that couple production of desired metabolites to cell growth [23, 135].

This chapter serves as an introduction to metabolic and regulatory network reconstructions and models, and gives a complete description of the core *E. coli* metabolic model. This model can be encoded and analyzed in any computational format (such as MATLAB or Mathematica) based on the information given in this chapter. The core *E. coli* model is a small scale model that can be used for educational purposes. It is meant to be used by senior undergraduate and first year graduate students learning about constraint-based modeling and systems biology. This model has enough reactions and pathways to enable interesting and insightful calculations, but it is also simple enough that the results of such calculations can be easily understood. This model is also useful for testing and evaluating new constraint-based analysis methods, since its small scope makes troubleshooting and interpretation of results easier.

2   The 4-Step Process of Building Genome-Scale Metabolic Reconstructions

The construction of a genome-scale metabolic network is a long term process (from many months to several years, depending on the size of the network) consisting of four major steps, each requiring the use of different types of bio-

logical data (Figure 1). In the first step, an organism's annotated genome is used to generate a draft reconstruction. Second, the draft reconstruction is curated in a long process that involves the study of many highly specific data sources. In the third step, the reconstruction is converted to a mathematical model, and model based simulations can be compared to phenotypic data. In the fourth and final step, high-throughput data can be integrated with the model, allowing for biological discovery and iterative model refinement. After the model is complete, it can be used for a variety of purposes (Section 6).

2.1   Initial Reconstruction Based on an Annotated Genome

The first step in building any new genome scale reconstruction, whether of metabolic or regulatory networks, is to use the genome annotation for the desired organism to generate an initial list of functions. Genomes have been annotated to some degree for hundreds of microbial organisms, and they provide several types of valuable information. First, they contain the genome sequence of an organism, from which open reading frames (ORFs) can be identified. The function of each ORF can then be determined through a variety of methods [151]. The strongest evidence for the function of a particular ORF usually comes from direct biochemical analysis, such as isolation and characterization of the function of an enzyme. *E. coli* is a very well studied model organism and many of its ORFs have been experimentally characterized [93]. Unfortunately, for many other organisms, little biochemical data is available. To identify the ORFs in these organisms (and also to identify uncharacterized ORFs in well studied organisms), their genomic sequences are compared to the genomes of other organisms to identify homologous genes. *In silico* methods can also be used for annotation, including methods that identify genes based on protein-protein interactions, transcriptomics, phylogenetic profiles, protein fusion, and operon clustering [177]. These methods typically allow for 40-70% of the genes in a new genome to be annotated. When a high quality genome annotation is not available for a particular organism, it becomes more challenging to build a reconstruction of that organism on a genome scale.

There are organism specific databases for some genome annotations, including EcoCyc [93] for *E. coli* and Saccharomyces Genome Database (SGD) [32] and Comprehensive Yeast Genome Database (GYGD) [68] for yeast. For many other microbial organisms EntrezGene [111], Comprehensive Microbial Resource (CMR) [138], Genome Reviews [178], and Integrated Microbial Genomes (IMG) [114] contain useful genomic information. All of the metabolic genes identified in the genome annotation for the desired organism can be assembled into an initial parts list. From this initial list of genes, an initial list of enzymes and the reactions they catalyze can be constructed by mapping each gene to one or more reactions according to information from a database. The data used for this mapping can be included in the genome annotations or it can be obtained from metabolic databases such as KEGG [92], BRENDA [168], ENZYME [5], MetaCyc [106], the SEED [45], or TransportDB [155]. Most databases include EC (Enzyme Commission) numbers that can be used to easily identify the enzymes and reactions associated with a particular gene. With the appropriate databases, the reactions known to be associated with each gene can be identified. Some of these databases will be more useful for certain organisms than others. The process of building an initial reconstruction from a genome annotation and reaction information from databases can be performed manually, or it can be partially or fully automated. Tools available for building draft reconstructions include SimPheny (Genomatica, Inc., San Diego, CA), PathoLogic [130], and PRIAM [33]. SimPheny is a commercial software platform for building and analyzing metabolic constraint based models. It can download the annotated genome of an organism and provide a framework for manually associating metabolic genes with reactions. SimPheny is also useful in the other stages of building reconstructions, as it contains tools for manual curation and model quality control and quality assurance, as well as tools for performing simulations and analyzing experimental data. PathoLogic, part of the Pathway Tools software system, is tool for mapping genes to reactions in an automated manner. It requires a fully annotated genome, and uses EC numbers, Gene Ontology terms [4], or the annotated gene names to predict which reactions are associated with a particular gene. It can then predict which pathways are present in an organism by comparing the predicted set of reactions to a reference database such as MetaCyc. PRIAM is another automated method that identifies enzymes in any genome sequence. This program uses all of the known sequences for any individual enzyme in the ENZYME database to identify the characteristic sequence modules of that enzyme. Specific rules that can identify an enzyme based on which modules are present in a sequence are then formulated. PRIAM forms these modules and rules for every enzyme in the database, and then uses scoring matrices to identify modules in the genome of interest. It uses the rules to predict which enzyme is associated with every gene in the genome. This algorithm can be very useful because it does not require a fully annotated genome. The result of the initial mapping process is a draft reconstruction that lists most of the
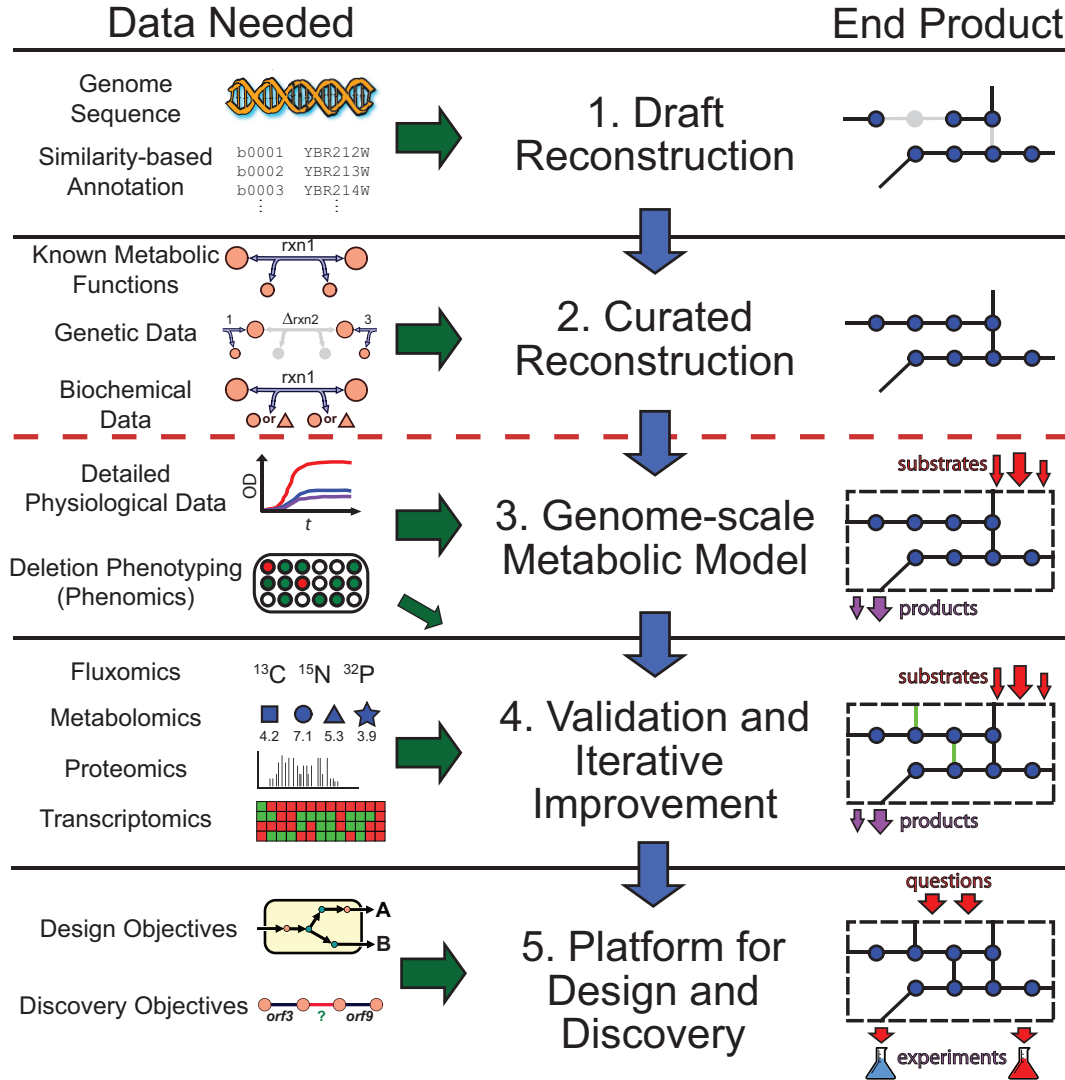
Figure 1: The phases and data utilized for generating a metabolic reconstruction. Genome scale metabolic reconstruction can be summarized in five major phases, each of the latter phases building off the previous. Also characteristic of the reconstruction process is the iterative refinement of reconstruction content that is driven by experimental data and occurs in the three latter phases. For each phase, specific data types are necessary and these range from high-throughput data types, e.g. metabolomics, to detailed studies characterizing individual components, e.g. biochemical data for a particular reaction. For example, the genome annotation can provide a parts list of a cell, whereas genetic data can provide information about the contribution of each gene product towards a phenotype, e.g. when removed or mutated. The product generated from each reconstruction phase can be utilized and applied to examine a growing number of questions with the final product having the broadest applications. Adapted from [58].

metabolic genes and reactions in an organism with reasonable accuracy.

2.2 Curation of the Initial Reconstruction

The next step in the reconstruction process is to manually curate the initial reconstruction. Any reconstruction resulting from a fully automated procedure will be incomplete and in some cases incorrect. Some reactions will be missing because it is not known which genes encode their enzymes, leaving gaps in pathways. Other reactions may be mistakenly included due to incorrect genome annotation or nonspecific information in databases. Reactions in the reconstruction may have incorrect or unbalanced stoichiometries or cofactor usage, as these attributes are often unique to enzymes in specific organisms [145]. Gene-protein-reaction associations must also be included to formally connect reactions with one or more functional protein, and every protein with one or more genes [149]. To correct any mistakes and improve the reconstruction, a researcher must manually curate the list of reactions using data from many different sources. Organism specific textbooks and databases are very useful for this purpose. The genome scale *E. coli* reconstruction *i*AF1260 [55] relied heavily on the textbook *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology [121]. Unfortunately, such texts are usually not available for less well-studied organisms. Literature data, from both primary and review articles, is also extremely useful. These articles can contain useful and specific information about reaction stoichiometry and directionality and can indicate the presence of many reactions with unknown genes. There are many different types of studies that can be useful, including enzyme assays, gene knockout studies, metabolomic studies including flux measurements, and protein localization studies. Often the information from these sources cannot be found in databases. The manual curation process is extremely labor intensive, usually requiring the study of hundreds of literature sources over a period of several months or years.

2.3 Conversion of Reconstruction to a Computational Model

After a high quality reconstruction has been assembled, it must be converted into a genome scale constraint-based model to be further analyzed [131]. A reconstruction is a BiGG knowledge base, a list of stoichiometrically balanced reactions and their associated genes and proteins. A model is a network in a mathematical format with defined system boundaries and constraints on the reactions [190]. While a reconstruction is unique to an organism, many different models (e.g. condition specific) can be derived from a reconstruction. Metabolic network models are usually encoded in a stoichiometric matrix, in which each unique metabolite is represented by a row in the matrix and each reaction is represented by a column. The entries in each column are the stoichiometric coefficients of the metabolites in a reaction, with negative coefficients for consumed metabolites and positive coefficients for produced metabolites. The properties of this matrix can be investigated through various constraint-based analysis methods, including flux balance analysis [52, 190, 191, 192]. This method uses linear optimization to identify optimal reaction flux distributions of the network given a set of minimum and maximum reaction rates and an objective, such as maximum cellular growth. In order to simulate growth with a genome scale metabolic model, a biomass reaction is needed. This is an organism-specific reaction that drains specified metabolites from the network, representing the metabolic precursors that contribute to biomass [190, 191]. To construct a biomass reaction, the relative amounts of nucleic acids, lipids, proteins, and other macromolecules of an organism must be known. These macromolecules can then be broken down into building blocks such as nucleotides or amino acids. The relative amounts of each of these building blocks form the stoichiometric coefficients of the biomass reaction (Section 4.9). Experimentally determined growth data must also be used to determine the amount of energy needed for growth and for non-growth associated maintenance functions, representing the energy demands of the cell [55]. Once the biomass reaction has been constructed, flux balance analysis [190, 191] can be used to predict optimal growth rates under many different conditions [192]. A completed genome scale metabolic model can be used to assess the quality of the reconstruction. There must be continuous pathways to every biomass precursor or the *in silico* cell cannot grow. Other gaps in the network can be identified by unreachable reactions or metabolites [48]. Another common test of a new model is to compare growth simulations under different conditions to actual growth data [51].

2.4 Model Validation and Iterative Improvement

Genome scale metabolic models can be used to map many types of experimental data to a biological network, allowing for the integration of different data types. This data can be compared to model predictions, and the discrepancies can lead to discovery of new reactions and pathways. High throughput screens for growth on different

Figure 2: An illustration of an *in silico* knockout of glucose-6-phosphate isomerase, *PGI* [65], which catalyzes the conversion of D-glucose-6-phosphate, *g6p* to D-fructose-6-phosphate, *f6p*. The Δpgi *in silico* model predicts that loss of glucose-6-phosphate isomerase, also known as phosphoglucose isomerase, results in use of the pentose phosphate pathway as the primary route of glucose catabolism, as has been observed experimentally [82].

media conditions can be used to reveal previously unknown substrate uptake pathways. *In vivo* knockout screens and synthetic lethal screens can be compared to *in silico* predicted knockout effects (Figure 2), with discrepancies indicating the presence of alternative metabolic pathways [151]. Metabolomic data can predict the presence of metabolites not accounted for in a reconstruction, necessitating the addition of new production and utilization pathways. Proteomic and transcriptomic data can be used to suggest the genes and proteins that fill gaps in a reconstruction [18]. As new biological features and capabilities are discovered, the model can be improved by incorporating this new data. The updated model can then be used to probe different pathways and features, leading to an iterative cycle of discovery and model improvement.

3 Building Genome Scale Regulatory Network Reconstructions

The process of reconstructing transcriptional regulatory networks is not as well developed as the process for building metabolic reconstructions. To date, there are only a few examples of genome scale regulatory reconstructions [15, 78, 199]. The oldest of these is *i*MC1010, which contains the regulators of *E. coli* metabolism [38]. This model was built by a process similar to that used for metabolic reconstructions, relying on a variety of data types including the genome annotation and literature sources. A gene expression study was then conducted in which gene expression under aerobic and anaerobic conditions was compared in several different transcription factor knockout strains. By analyzing the discrepancies between the experimental results and the model predictions, the model was iteratively updated and improved.

Several automated methods for inferring transcriptional regulatory networks have been developed recently [15, 54, 81, 170]. Progress in high throughput experimental methods is allowing for transcriptional regulatory reconstructions to be assembled in a top-down, automated manner [28, 29, 31]. The connectivity of an organism's transcriptional regulatory network can be determined by performing ChIP-chip experiments (chromatin immuno-precipitation followed by microarray hybridization). In these studies, all of the DNA binding sites of a particular

transcription factor on an entire genome under a particular set of conditions can be identified *in vivo*. First, proteins are fixed to genomic DNA in living cells and then the DNA is extracted and sheared into small fragments. Next, the fragments with a particular transcription factor bound are filtered out using antibodies, and the fragments are identified by hybridization to a DNA microarray. When ChIP-chip experiments are repeated under a variety of conditions, all of the binding sites of a transcription factor can be found, identifying all of the genes regulated by that transcription factor.

A set of ChIP-chip experiments must be run for every transcription factor to elucidate an entire regulatory network. As useful as these experiments are, they do not reveal what the effect of a transcription factor on its targets is, or how different transcription factors interact. The direct and indirect effects of transcription factors can be determined by expression profiling of strains with those transcription factor genes knocked out [83]. Performing so many high throughput experiments is a very expensive and time consuming process, but improvements in parallel sequencing technologies should improve the effectiveness of these approaches. ChIP-seq experiments (chromatin immunoprecipitation followed by DNA sequencing) are an example of this trend [28, 29, 30, 70, 172].

As with metabolic reconstructions, regulatory reconstructions must be converted to computational models to utilize their predictive potential. There are several different model types that can be used. The most common modeling framework is the Boolean model, which represents the connections between genes or other variables as logical rules. Boolean models can qualitatively describe the functions of a regulatory network and make accurate predictions of behavior [39]. An equivalent structure to the Boolean model is the 'regulation matrix' [67]. By reformulating a Boolean model as a matrix, more advanced mathematical analysis is possible, and all of the possible expression states of the model can be sampled. Newer regulatory network models are being formulated in structures similar to the stoichiometric matrix, allowing them to be interrogated with constraint-based analysis methods.

4 Description of the Core *E. coli* Metabolic Reconstruction

Here, we summarize the biochemistry of each reaction in the *E. coli* core metabolic model. An overview of the reactions is given in Figure 4. Where possible, reactions are described in the context of their major functional pathway. Occasionally, reactions participate in more than one pathway and in such cases this is highlighted. The *E. coli* core model is based on the first stoichiometric reconstruction of *E. coli* fueling pathways [190]. A metabolic reconstruction consists of a set of data for each metabolite and reaction and metabolite at a uniform level of detail. Further biochemical knowledge, treated in various levels of detail may be found in EcoSal Section 3, 'Metabolism and Metabolic Fluxes.' [121]. The aim here is to summarize the information encoded by the *E. coli* core model.

Metabolites and reactions are given both full names and abbreviations. Metabolite abbreviations are lowercase, and extracellular metabolites are denoted with the suffix '[e]', e.g. extracellular acetate is abbreviated as ac[e]. This reconstruction does not distinguish between the periplasmic space and the extracellular medium. All metabolites that are not denoted as extracellular are cytosolic. In many reconstructions, cytosolic metabolites use the suffix '[c]', but here this is omitted for clarity. In the figures describing the metabolic network, such as Figure 4, cytosolic metabolites are represented by orange circles and extracellular metabolites are yellow circles. Reaction abbreviations are uppercase and italicized. For example, acetaldehyde dehydrogenase (acetylating) is abbreviated as *ACALD*. There are several common suffixes used in the reaction abbreviations, including '*abc*' (ATP-Binding Cassette transporter), '*i*' (irreversible), '*r*' (reversible), and '*t*' (transport). Most reactions are named after the enzymes that catalyze them.

In the figures, metabolic reactions are represented as blue arrows, and the reaction abbreviations are inside blue boxes with yellow outlines. Certain reactions are assumed to be effectively irreversible when thermodynamic considerations are taken into account [1, 56]. Briefly, if the *in vivo* change in Gibbs energy for a biochemical reaction is highly negative, then it is assumed that net flux is always in the forward direction. In all figures, effectively irreversible reactions are denoted with arrowheads in one direction only. The tables associated with each pathway also indicate reactions that are effectively irreversible with the symbol $\rightarrow$ in the reaction equation.

Published gene-protein-reaction associations are shown adjacent to each reaction. The reconstruction associates each reaction with an enzyme or an enzyme complex. Some reactions may be catalyzed by more than one enzyme and this distinction is also represented, e.g. Figure 3. Some reactions are known to occur in *E. coli* but the corresponding gene has yet to be identified. These orphan reactions are discussed further in Section 6.3. Each protein is associated with a gene name and genomic locus. The genomic locus ties a reaction to one particular unique nucleotide sequence in the 4,639,675 nucleotide base *Escherichia coli* K-12 MG1655 circular chromosome [159]. The functional genomic structure, such as operon structure or transcriptional units, are not specifically
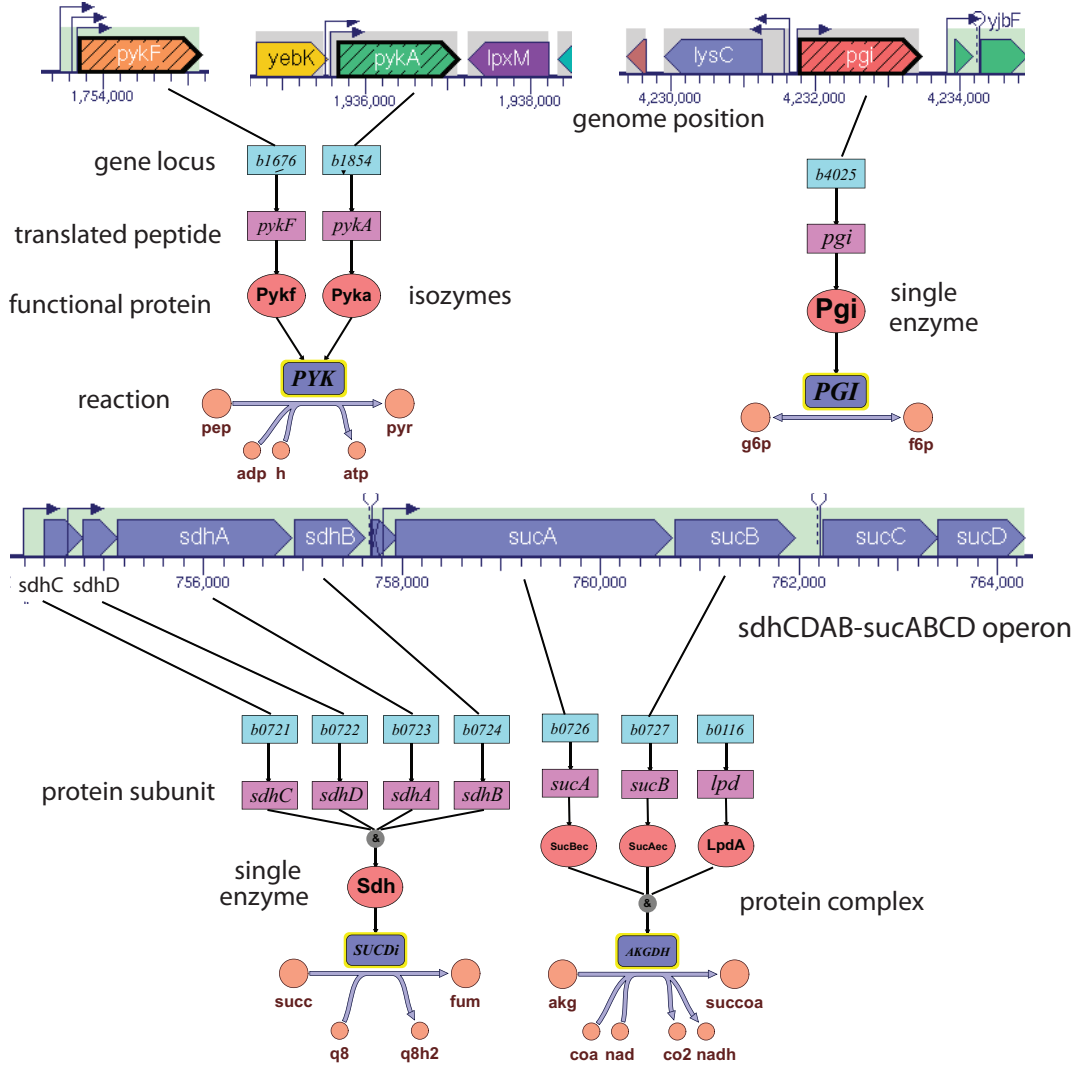
Figure 3: An example of a gene-protein-reaction associations from the *E. coli* core model. Genes are represented by boxes and designated by their locus name, translated peptides are represented by purple boxes, functional proteins are represented by red ovals, and reactions are labeled with blue boxes. For isozymes, two different proteins are connected to the same reaction. For proteins with multiple peptide subunits, the peptides are connected with an '&' above the protein. For complexes of multiple functional proteins, the proteins are connected with an '&' above the reaction. The genomic context of some of these genes is highlighted. Certain genes for the same reaction, e.g. *pykF* and *pykA*, are encoded by genes in operons widely separated on the genome. Operons are represented by shaded rectangles around one or more genes. Genes are represented by rectangles with one side pointed to denote the direction of the sense strand. Other operons contain multiple genes which encode protein subunits in a large protein. In this case the same *sdhCDAB-sucABCD* operon that codes for the *SUCDi* proteins [41] also codes for two proteins of the 2-oxoglutarate dehydrogenase enzyme complex, *AKGDH*. Genome context figures created using Pathway Tools Genome Browser from EcoCyc [93, 97].

represented in the *E. coli* core reconstruction but are represented in more comprehensive models of *E. coli* [184]. The charge of each reaction is included in the core model in order to determine the proper chemical formula in *E. coli*. These charges were determined by using the pK_a of each metabolite at a pH of 7.2 [55].

The reactions and pathways in the core model were chosen to represent the most well known and widely studied metabolic pathways of *E. coli*. These pathways are often the subjects of textbook chapters and should be familiar to most readers with a basic biochemistry background. As much as possible, the reactions were taken directly from the *i*AF160 genome scale *E. coli* reconstruction [55]. Some pathways, such as the electron transport system, were greatly simplified in order to limit the scope of the model and ensure that every reaction is understandable. This model contains a total of 72 metabolites and 95 reactions. There are 20 extracellular metabolites and 52 intracellular metabolites, with a total of 54 unique metabolites (most extracellular metabolites are just extracellular versions of intracellular metabolites). There are 20 exchange reactions, one for each extracellular metabolite. The model also has 25 transport reactions, 49 metabolic reactions, and one biomass reaction (Section 4.9).
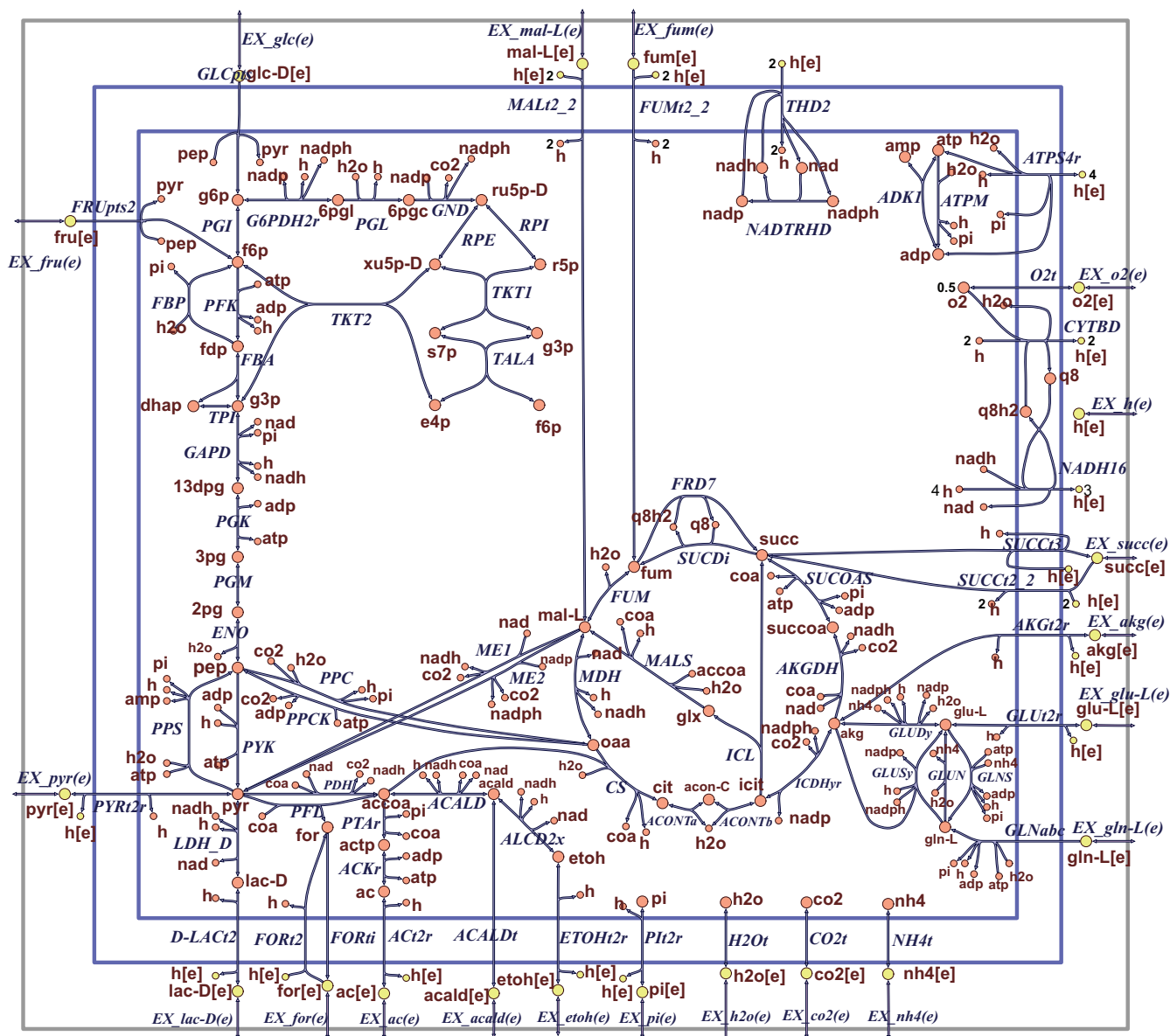
Figure 4: An overview of core *E. coli* metabolism. The outer gray box represents the boundary between the model and the environmental source of substrates and sink for waste metabolites. The outer and inner blue boxes represent the outer and inner surface of the cytoplasmic membrane. The periplasmic space is outside the scope of the core model. Cytosolic metabolites are represented by orange circles and extracellular metabolites are represented by yellow circles.

4.1 Glycolysis

The phosphoenolpyruvate:carbohydrate phosphotransferase protein complex senses carbohydrates in the environment and actively translocates them across the inner cytoplasmic membrane [179]. A phosphotransferase, alternatively known as a kinase, is a type of enzyme that transfers phosphate groups from high-energy donor molecules, such as phosphoenolpyruvate, to specific target substrates in a process termed phosphorylation. Certain proteins of the phosphoenolpyruvate:carbohydrate phosphotransferase complex are carbohydrate specific, but in each case, transport is driven by hydrolysis of the phospho group of phosphoenolpyruvate, which is also used to phosphorylate the carbohydrate.

In the reaction D-glucose transport via phosphoenolpyruvate:pyruvate phosphotransferase system, *GLCpts*, phosphoenolpyruvate donates the phospho group to glucose to form D-glucose-6-phosphate and the dephosphorylated remainder of phosphoenolpyruvate is pyruvate [154]. The same general procedure generates D-fructose-6-phosphate from fructose in the reaction fructose transport via PEP:Pyr PTS (f6p generating), *FRUpts2* [50]. The interconversion of D-glucose-6-phosphate and D-fructose-6-phosphate is catalyzed by glucose-6-phosphate isomerase, *PGI* [65]. Isomers are metabolites with the same molecular formula but different structural formulas, and an isomerase denotes an enzyme that catalyzes such a structural rearrangement. D-glucose-6-phosphate is a precursor for synthesis of the carbohydrate energy reserve glycogen. Glycogen stores accumulate when growth is limited by the availability a nutrient other than a carbon source, e.g. nitrogen. D-fructose-6-phosphate is a precursor for biosynthesis of UDP-N-acetyl muramic acid, which is essential for cell wall synthesis in bacteria. The cell wall preserves cellular integrity and helps to counter the high internal osmotic pressure that results from the relatively high intracellular solute concentration of solutes compared to the environment. flux balance

Phosphofructokinase, *PFK*, as its suffix suggests, catalyzes the phosphorylation of D-fructose-6-phosphate to form D-fructose-1-6-bisphosphate driven by adenosine triphosphate (ATP) hydrolysis [42, 160]. This reaction is effectively thermodynamically irreversible but a stoichiometrically distinct reverse transformation of D-fructose-1-6-bisphosphate to D-fructose-6-phosphate, liberating inorganic phosphate, is catalyzed by fructose-bisphosphatase, *FBP* [79]. The role of this anapleurotic reaction is discussed further in Section 4.4. Fructose-bisphosphate aldolase, *FBA*, splits the 6 carbon D-fructose-1-6-bisphosphate into two 3-carbon molecules, dihydroxyacetone-phosphate and glyceraldehyde-3-phosphate [2, 6, 79, 186]. Triose-phosphate isomerase, *TPI*, rapidly and reversibly structurally rearranges dihydroxyacetone-phosphate to glyceraldehyde-3-phosphate [140]. The reaction is catalyzed so efficiently that its flux is limited only by the rate the substrate can diffuse into the enzyme's active site [180]. A linear sequence of four reversible reactions catalyzed by glyceraldehyde-3-phosphate dehydrogenase, *GAPD* [17], phosphoglycerate kinase, *PKG* [122], phosphoglycerate mutase, *PGM* [64], and enolase, *ENO* [100], converts glyceraldehyde-3-phosphate to phosphoenolpyruvate. An intermediate metabolite in this sequence, 3-phospho-D-glycerate, is a precursor for biosynthesis of the amino acids, serine, glycine and cysteine, as well as vitamin B9, also known as folate. This sequence of reactions also reduces one nicotinamide adenine dinucleotide (NAD^+) to form NADH and yields one high energy currency metabolite, ATP. Strictly speaking it is more accurate to refer to ATP as a high chemical potential metabolite since chemical potential contains an internal energy and a entropic contribution [96].

In the final step of glycolysis, pyruvate kinase, *PYK*, catalyzes the transfer of the phospho group of phosphoenolpyruvate to ADP resulting in the production of pyruvate and ATP [66, 119]. A cursory glance at Figure 4 reveals that pyruvate is an important precursor involved in many pathways. It is a direct precursor for the amino acids alanine, valine and leucine. It can also be converted into acetyl-CoA, which provides carbon for the tricarboxylic acid cycle, as discussed in Section 4.3. Pyruvate can also be converted into lactate as part of fermentation, discussed in Section 4.6. *E. coli* can also grow on 3- or 4-carbon sources, such as lactate or malate, but in this situation, proximal 6-carbon metabolites in the glycolytic pathway are still required as precursors for biosynthesis. With only 3- or 4-carbon sources in the environment, the glycolytic pathway can actually be reversed to produce net flux from pyruvate to glucose-6-phosphate. This reversal of glycolytic flux is referred to as gluconeogenesis (Latin: gluco = glucose, neo = new, genesis = creation). The two reactions of glycolysis that are effectively irreversible, catalyzed by phosphofructokinase, *PFK*, and pyruvate kinase, *PYK*, are replaced with two reversing anapleurotic reactions, catalyzed by fructose-bisphosphatase, *FBP*, and phosphoenolpyruvate synthase, *PPS*, respectively. Phosphoenolpyruvate synthase, *PPS* [120], catalyzes the conversion of pyruvate to phosphoenolpyruvate and in the process hydrolyzes one ATP[36].
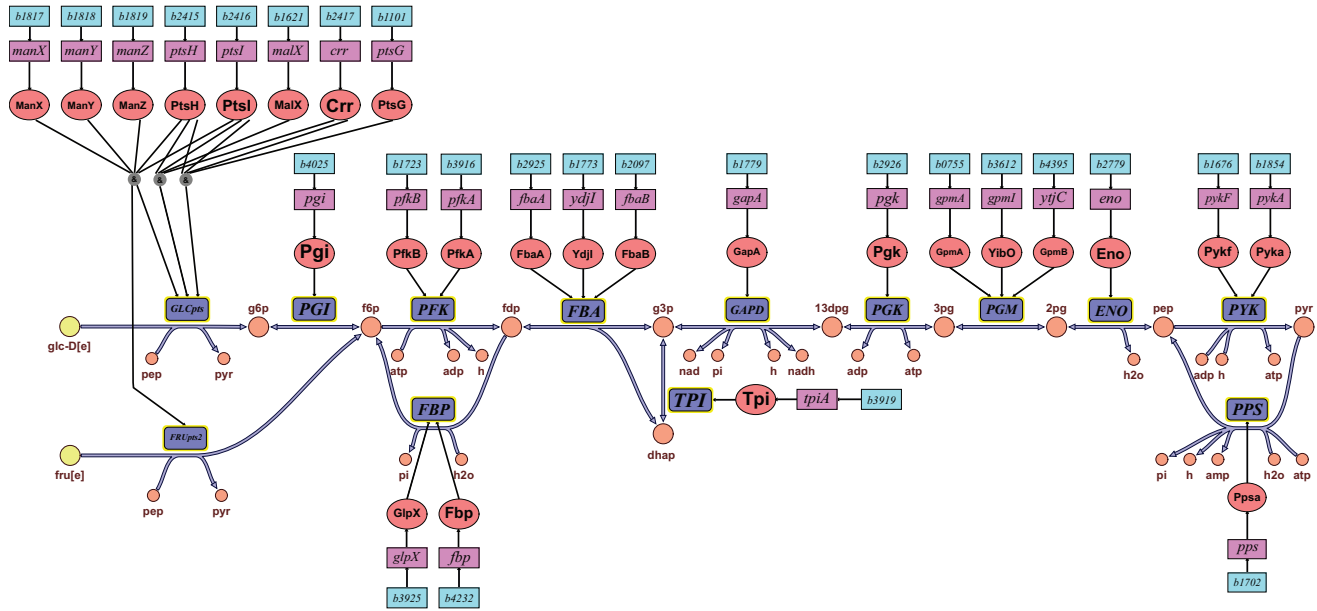
Figure 5: Glycolysis consists of a sequence of reactions that converts sugars, such as glucose or fructose, into a series of precursors for biosynthesis, terminating with pyruvate (linear sequence of reactions along the top). There is also the concomitant production of a relatively small amounts of ATP and NADH. Gluconeogenesis refers to the production of sugars by reversing the flux through thermodynamically reversible reactions in the glycolytic pathway, bypassing reactions that are effectively thermodynamically irreversible.

Abbr.	Reaction	Equation
GLCpts	D-glucose transport via PEP:Pyr PTS	glc-D[e] + pep→g6p + pyr
PGI	glucose-6-phosphate isomerase	g6p $\rightleftharpoons$ f6p
FRUpts2	Fructose transport via PEP:Pyr PTS (f6p generating)	fru[e] + pep → f6p + pyr
PFK	phosphofructokinase	atp + f6p → adp + fdp + h
FBP	fructose-bisphosphatase	fdp + h2o → f6p + pi
FBA	fructose-bisphosphate aldolase	fdp $\rightleftharpoons$ dhap + g3p
TPI	triose-phosphate isomerase	dhap $\rightleftharpoons$ g3p
GAPD	glyceraldehyde-3-phosphate dehydrogenase	g3p + nad + pi $\rightleftharpoons$ 13dpg + h + nadh
PGK	phosphoglycerate kinase	3pg + atp $\rightleftharpoons$ 13dpg + adp
PGM	phosphoglycerate mutase	2pg $\rightleftharpoons$ 3pg
ENO	enolase	2pg $\rightleftharpoons$ h2o + pep
PYK	pyruvate kinase	adp + h + pep → atp + pyr
PPS	phosphoenolpyruvate synthase	atp + h2o + pyr → amp + 2 h + pep + pi

Abbr.	Metabolite	Formula	Charge
glc-D	D-Glucose	$C_6H_{12}O_6$	0
g6p	D-Glucose-6-phosphate	$C_6H_{11}O_9P$	-2
fru	D-Fructose	$C_6H_{12}O_6$	0
f6p	D-Fructose-6-phosphate	$C_6H_{11}O_9P$	-2
fdp	D-Fructose-1,6-bisphosphate	$C_6H_{10}O_{12}P_2$	-4
dhap	Dihydroxyacetone-phosphate	$C_3H_5O_6P$	-2
g3p	Glyceraldehyde-3-phosphate	$C_3H_5O_6P$	-2
13dpg	3-Phospho-D-glyceroyl-phosphate	$C_3H_4O_{10}P_2$	-4
3pg	3-Phospho-D-glycerate	$C_3H_4O_7P$	-3
2pg	D-Glycerate-2-phosphate	$C_3H_4O_7P$	-3
pep	Phosphoenolpyruvate	$C_3H_2O_6P$	-3
pyr	Pyruvate	$C_3H_3O_3$	-1

4.2 Pentose Phosphate Shunt

The primary function of the pentose phosphate shunt is to provide the 5-carbon and 4-carbon biosynthetic precursors, respectively alpha-D-ribose-5-phosphate and D-erythrose-4-phosphate. Alpha-D-ribose-5-phosphate is a precursor for biosynthesis of the the amino acid histidine, the purines ATP, GTP, dATP and dGTP, and the pyrimidines, UTP, CTP, dCTP and dTTP. ATP and GTP are the primary energy currency metabolites for metabolism and macromolecular biosynthesis, respectively. dATP and dGTP are their derivatives which are essential for DNA synthesis. UTP and CTP are essential for macromolecular biosynthesis. dTTP and dCTP are also essential for DNA synthesis. D-erythrose-4-phosphate along with phosphoenolpyruvate, are required for biosynthesis of the amino acids tryptophan, tyrosine and phenylalanine. Alpha-D-ribose-5-phosphate and D-erythrose-4-phosphate can be produced by either of two parallel pathways. Either via the decarboxylating, oxidative pathway or through the non-oxidative pathway. In anaerobic conditions there is a greater flux through the non-oxidative pathway than the oxidative pathway [88, 89].

Decarboxylation is any chemical reaction in which a carboxyl group, -COOH is removed from a compound as carbon dioxide, CO_2. The decarboxylating, oxidative pathway is effectively irreversible, consumes D-glucose-6-phosphate, and after three reactions catalyzed by glucose-6-phosphate dehydrogenase, *G6PDH2r* [139], 6-phosphogluconolactonase, *PGL* [185], and phosphogluconate dehydrogenase, *GND* [193], produces D-ribulose-5-phosphate. These reactions also reduce two pyrimidines, nicotinamide adenine dinucleotide phosphate ($NADP^+$) to NADPH. This oxidative branch of the pentose phosphate shunt is important for production of reducing power in the form of NADPH. However, the pentose phosphate shunt is not the only source of NADPH [40] and the reactions catalyzed by NAD(P) transhydrogenase, *THD2*, isocitrate dehydrogenase (NADP), *ICDHyr*, and malic enzyme (NADP), *ME2*, can also supply *E. coli* with NADPH. D-ribulose-5-phosphate is then reversibly structurally rearranged into the precursor alpha-D-ribose-5-phosphate in the reaction ribose-5-phosphate isomerase, *RPI* [53].

The non-oxidative reversible rearrangement of the glycolytic sugar monophosphates to the pentose phosphate shunt sugar monophosphates represents the simplest possible mechanism for creating these precursors [115], but does not contribute to reducing power. This rearrangement requires three steps, as illustrated in Figure 6. First transketolase, *TKT2* [84, 176], catalyzes the conversion of a 6-carbon, D-Glucose-6-phosphate, plus 3-carbon, Glyceraldehyde-3-phosphate, into a 5-carbon, D-Xylulose-5-phosphate, plus a 4-carbon precursor, D-Erythrose-4-phosphate. Then transaldolase, *TALA*, catalyzes the conversion of this 4-carbon plus another original 6-carbon, into a 7-carbon, sedoheptulose-7-phosphate, plus the original 3-carbon. Then, the multifunctional enzyme transketolase, *TKT1*, catalyzes the conversion of this 7-carbon plus the 3-carbon into two different 5-carbons, D-Xylulose-5-phosphate and the precursor alpha-D-Ribose-5-phosphate. In addition to ribose-5-phosphate isomerase, ribulose 5-phosphate 3-epimerase, *RPE* [110], provides another reversible catalytic link between the oxidative and non-oxidative branches since it interconverts D-Xylulose-5-phosphate and D-Ribulose-5-phosphate. An epimerase is an enzyme that catalyzes the stereochemical inversion of molecular configuration about an asymmetric carbon atom in a substrate having more than one center of asymmetry. A cyclical operation of the pentose phosphate pathway could lead, stoichiometrically, to the complete degradation of sugar. Some cyclic use of the shunt has been postulated for a wild type strain [95]. However, phosphofructokinase mutants are almost completely unable to grow on glucose [43], leading one to conclude that the pentose phosphate shunt does not function cyclically.
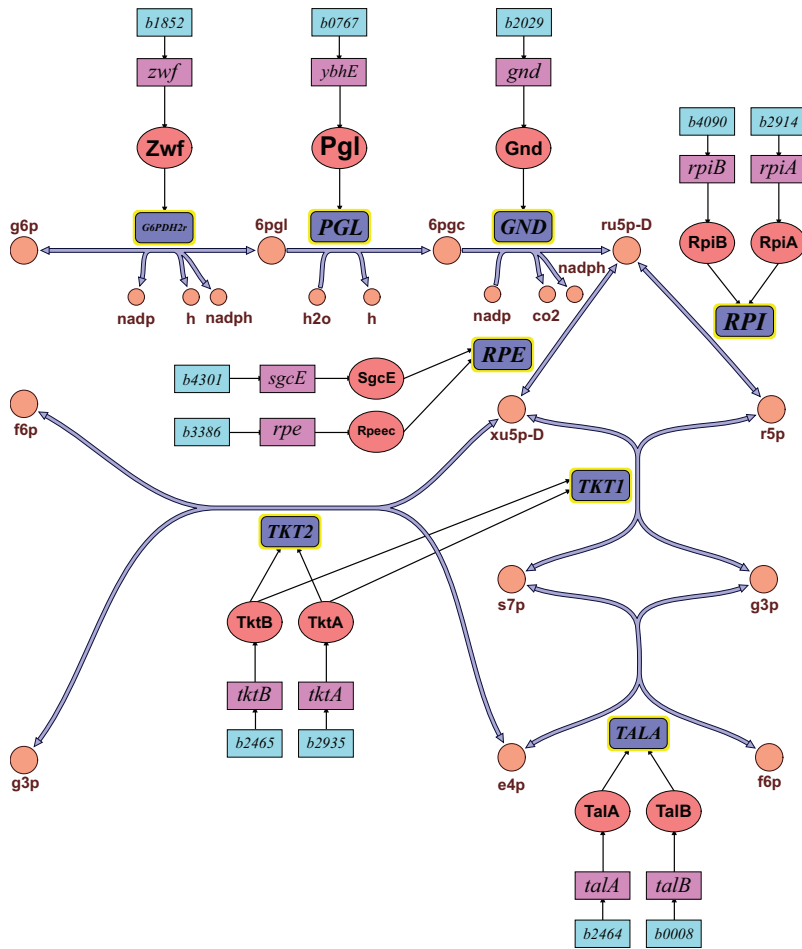
Figure 6: The pentose phosphate shunt. The irreversible, decarboxylating, oxidative pathway runs along the top and below is the reversible set of rearrangements which constitute the non-oxidative pathway.

Abbr.	Reaction	Equation
G6PDH2r	glucose 6-phosphate dehydrogenase	g6p + nadp $\rightleftharpoons$ 6pgl + h + nadph
PGL	6-phosphogluconolactonase	6pgl + h2o $\rightarrow$ 6pgc + h
GND	phosphogluconate dehydrogenase	6pgc + nadp $\rightarrow$ co2 + nadph + ru5p-D
RPI	ribose-5-phosphate isomerase	r5p $\rightleftharpoons$ ru5p-D
TKT2	transketolase	e4p + xu5p-D $\rightleftharpoons$ f6p + g3p
TALA	transaldolase	g3p + s7p $\rightleftharpoons$ e4p + f6p
TKT1	transketolase	r5p + xu5p-D $\rightleftharpoons$ g3p + s7p
RPE	ribulose 5-phosphate 3-epimerase	ru5p-D $\rightleftharpoons$ xu5p-D

Abbr.	Metabolite	Formula	Charge
g6p	D-Glucose-6-phosphate	$C_6H_{11}O_9P$	-2
6pgl	6-phospho-D-glucono-1-5-lactone	$C_6H_9O_9P$	-2
6pgc	6-Phospho-D-gluconate	$C_6H_{10}O_{10}P$	-3
ru5p-D	D-Ribulose-5-phosphate	$C_5H_9O_8P$	-2
r5p	alpha-D-Ribose-5-phosphate	$C_5H_9O_8P$	-2
f6p	D-Fructose-6-phosphate	$C_6H_{11}O_9P$	-2
g3p	Glyceraldehyde-3-phosphate	$C_3H_5O_6P$	-2
xu5p-D	D-Xylulose-5-phosphate	$C_5H_9O_8P$	-2
s7p	Sedoheptulose-7-phosphate	$C_7H_{13}O_{10}P$	-2
e4p	D-Erythrose-4-phosphate	$C_4H_7O_7P$	-2

4.3   Tricarboxylic Acid (TCA) Cycle

The tricarboxylic acid cycle is a well-studied pathway with a variety of functions depending on the environment. After an overview to orientate the reader, we discuss each reaction in more detail. During aerobic growth on 6-carbon sugars such as glucose, the tricarboxylic acid cycle functions almost exclusively to create the precursors, oxaloacetate, 2-oxoglutarate (also commonly called alpha-ketoglutarate), and succinyl-CoA for amino acid and heme biosynthesis. Heme is an organic group containing iron with a crucial role in the electron transport chain (Section 4.5). The aerobic production of precursors is carried out primarily by the oxidative arm of the tricarboxylic acid cycle, from oxaloacetate through to 2-oxoglutarate. This is the counter-clockwise, lower part of the cycle in Figure 7. The full tricarboxylic acid cycle, continuing counter-clockwise in Figure 7, can totally oxidize acetyl-CoA, but is only functional during aerobic growth on acetate or fatty acids. Then the cycle can produce all the energy and reducing potential needed to support growth. Cycle intermediates are still required as biosynthetic precursors, so inward flux from an anapleurotic pathway, the glyoxylate shunt, is required to maintain the pool of dicarboxylic intermediates, see Figure 8. Under anaerobic conditions, the tricarboxylic acid cycle functions not as a cycle, but as two separate pathways. The oxidative pathway, again, the counterclockwise lower part of the cycle in Figure 7, forms the precursor 2-oxoglutarate. The reductive pathway, the clockwise upper part of the cycle in Figure 7, forms the precursor succinyl-CoA. *E. coli* can grow in an environment where the only carbon substrate is one of the tricarboxylic acid cycle intermediates. This is represented by proton symport transport reactions that translocate either 2-oxoglutarate, succinate, fumarate or L-malate into the cell.

In aerobic conditions, pyruvate dehydrogenase, *PDH* in Figure 4.6, catalyzes the synthesis of acetyl-CoA from pyruvate and coenzyme-A with concomitant reduction of NAD^+ to NADH. Coenzymes are small organic non-protein molecules that carry chemical groups between enzymes. Acetyl-CoA conveys carbon atoms within the acetyl group to the tricarboxylic acid cycle to be oxidized for energy production. This reaction is carried out by a large multienzyme complex containing 12 AceE dimers, 6 LpdA dimers, and a 24 AceF core [153]. Citrate synthase, *CS* [124], catalyzes the condensation reaction of the 2-carbon acetate residue from acetyl-CoA and a molecule of 4-carbon oxaloacetate to form the 6-carbon citrate, with the release of coenzyme-A. Next, the reactions aconitase A, *ACONTa*, and aconitase B, *ACONTb* , isomerize citrate to isocitrate via cis-aconitate [20, 146]. Isocitrate sits at a branch point in the tricarboxylic cycle, where carbon flux either continues in the oxidative branch, catalyzed by isocitrate dehydrogenase, *ICDHyr* [24], or is diverted into the glycoxylate cycle, catalyzed in the first reaction by isocitrate lyase, *ICL*. The glycoxylate cycle is discussed further in Section 4.4.

Isocitrate dehydrogenase catalyzes the oxidative decarboxylation of isocitrate, producing 2-oxoglutarate and CO_2 while reducing $NADP^+$ to NADPH. 2-oxoglutarate provides a carbon backbone for synthesis of glutamate and glutamine, the central metabolites in nitrogen metabolism, discussed in Section 4.7. In aerobic conditions, 2-oxoglutarate dehydrogenase, *AKGDH*, an enzyme of the same family as pyruvate dehydrogenase, *PDH*, catalyzes a three step mechanism consisting of decarboxylation of 2-oxoglutarate, producing CO_2, reduction of NAD^+ to NADH, and transfer of coenzyme-A to the decarboxylated carbon to synthesize succinyl-CoA. 2-oxoglutarate dehydrogenase is a large enzyme complex of the same family as pyruvate dehydrogenase, containing 12 SucA enzymes, an Lpd dimer, and 24 SucB enzymes [137]. Succinyl-CoA, together with 3-phospho-D-glycerate, is a precursor for heme.

During aerobic growth on acetate or fatty acids, the tricarboxylic acid cycle continues, counterclockwise in Figure 7, with the sequential reactions catalyzed by succinyl-CoA synthetase, *SUCOAS* [19], succinate dehydrogenase, *SUCDi* [35], fumarase, *FUM* [10, 61, 198], and malate dehydrogenase, *MDH* [181], until oxaloacetate is formed, completing the full tricarboxylic acid cycle. Succinate dehydrogenase, *SUCDi*, catalyzes the only irreversible reaction in this set. When this set is used in reverse, clockwise in Figure 7, as the reductive pathway of the tricarboxylic acid cycle, a reversing anapleurotic reaction is catalyzed by fumarate reductase, *FRD7* [34]. During aerobic growth, continuing counterclockwise from succinyl-CoA in Figure 7, succinyl-CoA synthetase, *SUCOAS*, generates ATP by utilizing the energetic bond between succinate and coenzyme A, produced by the earlier step in the cycle, catalyzed by 2-oxoglutarate dehydrogenase, *AKGDH*.

Succinate dehydrogenase, *SUCDi*, is a multiprotein enzyme complex which straddles the cytoplasmic membrane allowing it to couple the tricarboxylic acid cycle to the electron transport chain, discussed in Section 4.5 [201]. The net effect of the succinate dehydrogenase complex is to catalyze the oxidation of succinate to fumarate, by donating electrons to the electron transport chain, which reduces ubiquinone-8, q8, to ubiquinol-8, q8h2. Ubiquinol-8 is then released from the enzyme complex and free to diffuse through the cytoplasmic membrane to interact with subsequent enzymes of the electron transport chain. Succinate dehydrogenase is only induced in aerobically grown cells, whereas the enzyme which catalyzes the reverse reaction, fumarate reductase, *FRD7*, is only induced in anaerobically grown cells when flux is necessary through the reductive arm of the tricarboxylic acid cycle, the clockwise upper portion

of Figure 7. During anaerobic growth, succinate can be excreted from the cell as a waste metabolite using succinate transport out via proton antiport, *SUCCt3*. We discuss anaerobic metabolism further in Section 4.6. In aerobic conditions, after succinate is converted fumarate, fumarase, *FUM*, reversibly catalyzes the conversion of fumarate plus water into L-malate [198]. Finally, to complete the cycle, malate dehydrogenase, *MDH*, reversibly catalyzes the conversion of malate into oxaloacetate while reducing NAD^+ to NADH.

Table 1: Tricarboxylic acid cycle reactions and metabolites.

Abbr.	Reaction	Equation
CS	citrate synthase	accoa + h2o + oaa → cit + coa + h
ACONTa	aconitase (half-reaction A, Citrate hydro-lyase)	cit ⇌ acon-C + h2o
ACONTb	aconitase (half-reaction B, Isocitrate hydro-lyase)	acon-C + h2o ⇌ icit
ICDHyr	isocitrate dehydrogenase (NADP)	icit + nadp ⇌ akg + co2 + nadph
AKGDH	2-Oxoglutarate dehydrogenase	akg + coa + nad → co2 + nadh + succoa
SUCOAS	succinyl-CoA synthetase (ADP-forming)	atp + coa + succ → adp + pi + succoa
FRD7	fumarate reductase	fum + q8h2 → q8 + succ
SUCDi	succinate dehydrogenase (irreversible)	q8 + succ → fum + q8h2
FUM	fumarase	fum + h2o ⇌ mal-L
MDH	malate dehydrogenase	mal-L + nad ⇌ h + nadh + oaa
AKGt2r	2-oxoglutarate reversible transport via symport	akg[e] + h[e] ⇌ akg + h
SUCCt3	succinate transport out via proton antiport	h[e] + succ → h + succ[e]
SUCCt2_2	succinate transport via proton symport (2 H)	2 h[e] + succ[e] → 2 h + succ
FUMt2_2	Fumarate transport via proton symport (2 H)	fum[e] + 2 h[e] → fum + 2 h
MALt2_2	Malate transport via proton symport (2 H)	2 h[e] + mal-L[e] → 2 h + mal-L

Abbr.	Metabolite	Formula	Charge
accoa	Acetyl-CoA	$C_{23}H_{34}N_7O_{17}P_3S$	-4
cit	Citrate	$C_6H_5O_7$	-3
acon-C	cis-Aconitate	$C_6H_3O_6$	-3
icit	Isocitrate	$C_6H_5O_7$	-3
akg	2-Oxoglutarate	$C_5H_4O_5$	-2
succoa	Succinyl-CoA	$C_{25}H_{35}N_7O_{19}P_3S$	-5
fum	Fumarate	$C_4H_2O_4$	-2
mal-L	L-Malate	$C_4H_4O_5$	-2
oxa	Oxaloacetate	$C_4H_2O_5$	-2

Figure 7: The tricarboxylic acid cycle reactions with gene-protein-reaction associations. The oxidative pathway of the tricarboxylic acid cycle runs counterclockwise in the lower part of the cycle, from oxaloacetate, oaa, through 2-oxoglutarate, akg. Continuing counterclockwise from 2-oxoglutarate, the full tricarboxylic acid cycle can totally oxidize acetyl-CoA, but is only functional during aerobic growth on acetate or fatty acids. Under anaerobic conditions, the tricarboxylic acid cycle functions not as a cycle, but as two separate pathways. The oxidative pathway, again, the counterclockwise lower part of the cycle, still forms the precursor 2-Oxoglutarate. However, the reductive pathway, the clockwise upper part of the cycle, forms the precursor succinyl-CoA. The reactions in the center of the cycle, *ICL* and *MALS*, are part of the glycoxylate cycle, discussed in Section 4.4.

4.4 Glycoxylate Cycle, Gluconeogenesis and Anapleurotic Reactions

The glycoxylate cycle consists of some of the aforementioned reactions in the tricarboxylic acid cycle as well as other reactions used only by the glycoxylate cycle, see Figure 8. During growth on tricarboxylic acid intermediates, the glycoxylate cycle is upregulated for the conversion of tricarboxylic acid intermediates into glycolytic intermediates. The glycoxylate cycle is essential for growth on carbon sources such as acetate or fatty acids, since it can convert the precursor acetyl-CoA into glycolytic intermediates without loss of carbon to carbon dioxide. Acetyl-CoA is made up of acetate, a proton, and coenzyme-A. The glycoxylate cycle overlaps with the tricarboxylic acid cycle from the incorporation of acetyl-CoA to the production of isocitrate at the aforementioned branch point in the oxidative arm of the tricarboxylic acid cycle. Isocitrate lyase, *ICL*, competes with the tricarboxylic acid cycle enzyme, isocitrate dehydrogenase, *ICDHyr*, for isocitrate. A lyase denotes an enzyme which cleaves carbon-carbon bonds, so as it's name suggests, isocitrate lyase catalyzes the cleavage of 6-carbon isocitrate into 4-carbon succinate and 2-carbon glycoxylate [80], the latter giving the cycle its name. Malate synthase, *MALS*, then catalyzes the condensation of glyoxylate with another acetyl-CoA, yielding malate [118]. Succinate generated in the first step can continue along the tricarboxylic acid cycle to eventually form oxaloacetate.

Growth on 4-carbon dicarboxylic acid intermediates of the tricarboxylic acid cycle, such as malate, requires that the cell be able to produce phosphoenolpyruvate for gluconeogenesis. Gluconeogenesis refers to the reversal of flux through the glycolytic pathway. There are two pathways existing to fulfill these phosphoenolpyruvate demands [74, 75]. One pathway involves the conversion of malate to pyruvate by malic enzyme, *ME1* or *ME2* [86, 113], followed by the synthesis of phosphoenolpyruvate from pyruvate by phosphoenolpyruvate synthase, *PPS* [125]. Malic enzyme, *ME1*, reduces one molecule of NAD^+ to NADH while converting malate to pyruvate. A second parallel reaction, malic enzyme (NADP), *ME2*, reduces one molecule of $NADP^+$ to NADPH. Although malic enzyme, *ME1*, can use NAD^+ and malic enzyme, *ME2*, can use NADP as a cofactor, it is observed that in the growth on 4-carbon dicarboxylic acids, only the malic enzyme with NAD^+ as a cofactor, *ME1*, can be used. The reason is probably because any surplus NADPH produced cannot easily give up electrons via the electron transfer system [74], so it appears that NADPH is produced only to the extent required for biosynthesis.

The other pathway from the tricarboxylic acid cycle to glycolytic intermediates is the conversion of oxaloacetate to phosphoenolpyruvate by the action of phosphoenolpyruvate carboxykinase, *PPCK* [46]. Phosphoenolpyruvate carboxykinase catalyzes the reverse reaction to the anapleurotic enzyme, phosphoenolpyruvate carboxylase, *PPC* [91], discussed further in Section 4.4. The former reaction consumes a high-energy phosphate bond and produces CO_2, whereas the latter releases inorganic phosphate and consumes CO_2. Although the reactions catalyzed by the enzymes phosphoenolpyruvate carboxykinase, *PPCK*, and malic enzyme, *ME1* and *ME2*, are thermodynamically reversible, physiologically they are found to operate unidirectionally. Evidence for the unidirectional operation of phosphoenolpyruvate carboxykinase, *PPCK*, producing the 3-carbon phosphoenolpyruvate, and malic enzyme producing the 3-carbon pyruvate, comes from the observation that mutants deficient in phosphoenolpyruvate carboxylase are able to grow on 4-carbon compounds but not on 3-carbon compounds [104, 105]. Furthermore, mutants deficient in phosphoenolpyruvate carboxykinase are able to grow on 3-carbon compounds but not on 4-carbon compounds. In summary, the glycoxylate cycle and gluconeogenic reactions allow *E. coli* to grow on 3-carbon and 4-carbon compounds by avoiding the loss of carbon as carbon dioxide from the tricarboxylic acid cycle, providing a pathway for generation of glycolytic intermediates from tricarboxylic acid cycle intermediates, and reversing the carbon flux through glycolysis to produce essential precursors for biosynthesis.

Anapleurotic reactions replenish tricarboxylic acid cycle intermediates drained off for biosynthesis. The tricarboxylic acid cycle operating cyclically can completely oxidize acetate to carbon dioxide without net consumption or production of intermediates. However, intermediates of the tricarboxylic cycle such as oxaloacetate and 2-oxoglutarate are consumed in the production of macromolecules. Tricarboxylic acid cycle intermediate generation from the glycolytic metabolites is accomplished by the irreversible carbon dioxide-fixing conversion of the 3-carbon phosphoenolpyruvate to the 4-carbon oxaloacetate, catalyzed by the enzyme phosphoenolpyruvate carboxylase, *PPC* (Figure 8). Carboxylation denotes a chemical reaction in which a carboxylic acid group is introduced in a substrate. In fact, carbon dioxide has been found to be essential to the growth of the cell [189, 156]. Normally, cells pick up the necessary carbon dioxide from the carbonate buffer in their media.
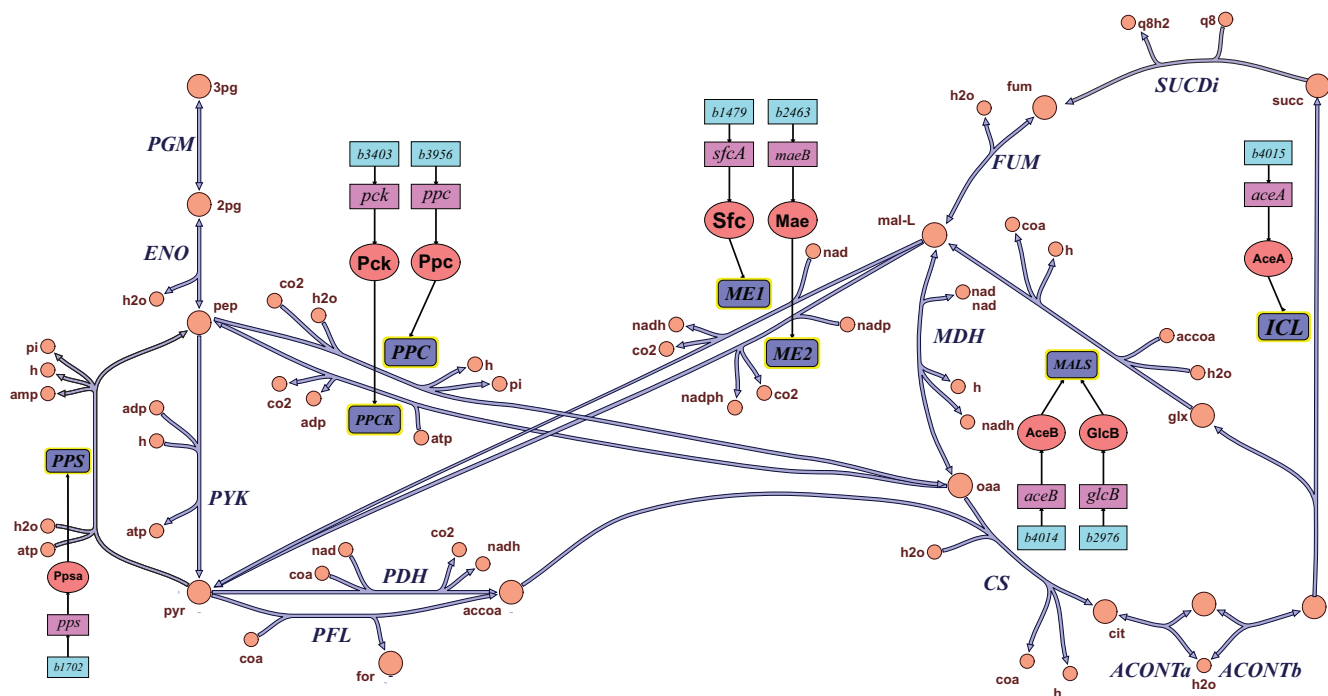
Abbr.	Reaction	Equation
ICL	isocitrate lyase	icit → glx + succ
MALS	malate synthase	accoa + glx + h2o → coa + h + mal-L
ME1	malic enzyme (NAD)	mal-L + nad → co2 + nadh + pyr
ME2	malic enzyme (NADP)	mal-L + nadp → co2 + nadph + pyr
PPS	phosphoenolpyruvate synthase	atp + h2o + pyr → amp + 2 h + pep + pi
PPCK	phosphoenolpyruvate carboxykinase	atp + oaa → adp + co2 + pep
PPC	phosphoenolpyruvate carboxylase	co2 + h2o + pep → h + oaa + pi

Abbr.	Metabolite	Formula	Charge
glx	Glyoxylate	C_2HO_3	-1

Figure 8: Glycoxylate cycle, gluconeogenic and anapleurotic reactions. The glycoxylate cycle is essential for growth on carbon sources such as acetate or fatty acids, since it can convert the precursor acetyl-CoA into glycolytic intermediates without loss of carbon to carbon dioxide. Acetate enters the glycoxylate cycle via acetyl-CoA, accoa, at two points, lower center left entering the citrate synthase reaction, *CS*, and in the center of the tricarboxylic cycle, entering the malate synthase reaction, *MALS*. Tricarboxylic acid intermediates are then converted into gluconeogenic precursors via either malic enzyme, ME1 and ME2, or phosphoenolpyruvate carboxykinase, *PPCK*. Gluconeogenesis then synthesizes 6-carbon sugars with the aid of reactions, such as phosphoenolpyruvate synthase, *PPS*, which effectively reverses an otherwise thermodynamically unfavorable glycolytic reaction. Phosphoenolpyruvate synthase, *PPS*, is an anapleurotic reaction which uses glycolytic intermediates to replenish the precursors drained from the tricarboxylic acid cycle for biosynthesis.

4.5　Oxidative Phosphorylation and Transfer of Reducing Equivalents

Energy is required to drive endergonic processes such as biosynthesis, polymerization, active transport of substrate into the cell against concentration gradients, maintaining internal pH, and motility. There are two main mechanisms for the production of energy, substrate level phosphorylation, and the electron transport chain. Certain fueling pathways are net producers of energy, where ATP is formed by a reaction between ADP and a phosphorylated intermediate of a fueling pathway. This is known as substrate level phosphorylation and examples of such reactions are phosphoglycerate kinase, PGK, and pyruvate kinase, PYK, in glycolysis, and succinyl-CoA synthetase, $SUCOAS$, in the tricarboxylic acid cycle. These substrate level phosphorylations occur in reactions distal to the cleavage of glucose into two trioses by fructose-bisphosphate aldolase, FBA, so each molecule of glucose can potentially lead to the net generation of four molecules of ATP.

The second mechanism for generation of energy is the electron transport chain which produces the bulk of the cell's ATP under aerobic conditions. The detailed study of the thermodynamics of the electron transport chain spawned the field of bioenergetics [196]. Mitchell's chemiosmotic theory describes the mechanism by which electron transport is coupled to the generation of ATP, see EcoSal Section 3.2.1 Energy Transduction by Ion Currents. The electron transport chain translocates protons, H^+, from the cytoplasm, across the cytoplasmic membrane into the the periplasmic space. Since the cytoplasmic membrane is effectively impermeable to protons and hydroxyl ions, OH^-, this establishes a difference in concentration of protons, and a difference in electrical charge, across the cytoplasmic membrane. This thermodynamic potential difference gives rise to a proton motive force which can be utilized to drive a myriad of endergonic reactions, such as synthesis of high energy currency metabolites, such as ATP.

This core *E. coli* model condenses the sequence of steps in the electron transport chain into two reactions. First, an NADH dehydrogenase, *NADH16*, catalyzes the oxidation of NADH to form NAD$^+$ while removing 4 protons from the cytoplasm, translocating 3 protons to the periplasmic space and combining the the fourth with a proton plus two electrons from NADH with ubiquinone-8 to form the reduced ubiquinol-8 [167, 175]. In the model, the protons are translocated into the extracellular medium, but this is a reasonable approximation given the pH of the periplasm and extracellular medium is the same [197].

Ubiquinone-8 and ubiquinol-8 are oil soluble coenzymes which diffuse freely within the lipid environment of the cytoplasmic membrane, see Figure 9. The next condensed step is when cytochrome oxidase, $CYTBD$, catalyzes the oxidation of ubiquinol-8 back to ubiquinone-8, which drives the translocation of two more protons into the extracellular space [103]. The two spare electrons are then combined with two cytoplasmic protons and an oxygen atom to form water. Oxygen is known as a terminal electron acceptor, and this requirement for oxygen is reflected in the complementary physiological term aerobic respiration, for the process we have described. Oxygen spontaneously diffuses from the environment into the cell down a concentration gradient, $O2t$, and represents an exogenous source of terminal electron acceptor. In *E. coli*, fermentation refers to growth without an exogenous terminal electron acceptor, and is discussed in Section 4.6. Exogenous refers to something generated outside a system, (Greek: exo = outside, gen = generation). The opposite is endogenous, something generated within a system.

The thermodynamic potential difference established across the cytoplasmic membrane by the electron transport chain can be used for a variety of purposes, for example (i) synthesis of ATP, (ii) transport of substrate metabolites against a concentration gradient (symport and antiport transport reactions), (iii) maintenance of a steady intracellular pH in the face of fluctuating environmental pH [63], and (iv) driving biological propellers known as flagella for cell motility. The enzyme ATP synthase, $ATPS4r$, catalyzes the synthesis of ATP from ADP, forming a high energy phosphate bond by coupling it to the proton motive force established by the electron transport chain [25, 94]. The exact number of high energy phosphate bonds, P, that are generated per oxygen molecule, O$_2$, used as a terminal acceptor is known as the P/O ratio. This varies depending on periplasmic pH and other environmental conditions, but we fix the P/O ratio stoichiometrically to 1.25. ATP maintenance requirement, $ATPM$, is not a real biochemical reaction. It is included for modeling purposes since the scope of the *E. coli* core model does not extend to all of the reactions that consume ATP in the cell. Adenylate kinase, $ADK1$, is a phosphotransferase enzyme that catalyzes the interconversion of adenine nucleotides, and plays an important role in cellular energy homeostasis [11, 21].

Catabolism is the set of metabolic pathways which break down molecules into smaller units and release energy. Anabolism is the set of metabolic pathways that construct molecules from smaller units. These reactions are endergonic and therefor require an input of energy. NADH is used for the catabolic activities of the cell, for instance driving the export of protons into the periplasmic space in the electron transport chain in the reaction catalyzed by NADH dehydrogenase, *NADH16*. In contrast NADPH is essential for anabolic metabolism such as the biosynthesis of building blocks for polymerization reactions from precursor metabolites produced by the fueling pathways of core

metabolism. Maintaining the proper balance between anabolic reduction charge, NADPH/NADP$^+$, and catabolic reduction charge, NADH/NAD$^+$, is achieved by reactions catalyzed by transhydrogenase enzymes. NAD(P) transhydrogenase, *THD2*, catalyzes the transfer of a hydride ion from NADH to create NADPH, in a reaction coupled to the proton motive force [12]. A hydride ion is a negative ion of hydrogen, H^-. The opposite transfer, of a hydride ion from NADPH, to create NADH, is catalyzed by another enzyme, NAD$^+$ transhydrogenase, *NADTRHD*, but it is not coupled to the translocation of protons [163]. These pair of reactions effectively allow transfer of reducing equivalents between anabolic and catabolic reduction charge.

Figure 9: NAD(P) transhydrogenase, *THD2,* and NAD$^+$ transhydrogenase, *NADTRHD*, catalyze the transfer reducing equivalents between anabolic and catabolic reduction charge (far left). The electron transport chain (right), generates an electrochemical potential between the cytoplasm and periplasm. Oil soluble ubiquinone-8 ubiquinol-8 freely diffuse within the lipid membrane separating the periplasm and cytoplasm. Enzyme complexes which involve proton translocation straddle the cytoplasmic membrane and are in contact with periplasm and cytoplasm. ATP synthase, *ATPS4r*, catalyzes the synthesis of ATP from ADP using the electrochemical potential between the cytoplasm and the periplasm (middle).

Abbr.	Name	Equation
NADH16	NADH dehydrogenase (ubiquinone-8 & 3 protons)	4 h + nadh + q8 → 3 h[e] + nad + q8h2
CYTBD	cytochrome oxidase bd (ubiquinol-8: 2 protons)	2 h + $\frac{1}{2}$ o2 + q8h2 → h2o + 2 h[e] + q8
O2t	o2 transport via diffusion	o2[e] ⇌ o2
ATPS4r	ATP synthase (four protons for one ATP)	adp + 4 h[e] + pi ⇌ atp + h2o + 3 h
ATPM	ATP maintenance requirement	atp + h2o → adp + h + pi
ADK1	adenylate kinase	amp + atp ⇌ 2 adp
THD2	NAD(P) transhydrogenase	2 h[e] + nadh + nadp → 2 h + nad + nadph
NADTRHD	NAD transhydrogenase	nad + nadph → nadh + nadp

Abbr.	Name	Formula	Charge
q8	Ubiquinone-8	$C_{49}H_{74}O_4$	0
q8h2	Ubiquinol-8	$C_{49}H_{76}O_4$	0
nad	Nicotinamide-adenine-dinucleotide (NAD$^+$)	$C_{21}H_{26}N_7O_{14}P_2$	-1
nadh	Nicotinamide-adenine-dinucleotide-reduced	$C_{21}H_{27}N_7O_{14}P_2$	-2
nadp	Nicotinamide-adenine-dinucleotide-phosphate	$C_{21}H_{25}N_7O_{17}P_3$	-3
nadph	Nicotinamide-adenine-dinucleotide-phosphate-reduced	$C_{21}H_{26}N_7O_{17}P_3$	-4
atp	Adenosine-5'-triphosphate	$C_{10}H_{12}N_5O_{13}P_3$	-4
adp	Adenosine diphosphate	$C_{10}H_{12}N_5O_{10}P_2$	-3
amp	Adenosine monophosphate	$C_{10}H_{12}N_5O_7P$	-2
h	Proton	H	+1

4.6 Fermentation

During aerobic respiration, oxygen is the terminal electron acceptor for the electron transport chain, which yields the bulk of ATP required for biosynthesis. Anaerobic respiration refers to respiration without molecular oxygen, where the terminal electron acceptor can be, for example, fumarate. In addition to anaerobic respiration, *E. coli* can generate ATP by substrate level phosphorylation alone which is referred to as fermentation. Glycolysis results in the net production of 2 ATP per glucose by substrate level phosphorylation, but this is low compared to 17.5 ATP per glucose for aerobic respiration. The substrates of fermentation are typically sugars, so during fermentative growth, each cell must maintain a large magnitude flux through glycolysis to generate sufficient ATP to drive the constitutive biosynthesis, polymerization, and assembly reactions required for growth. This necessitates a large magnitude efflux of fermentative end products since there is insufficient ATP to assimilate all carbon as biomass. Approximately 10% of carbon substrate is assimilated due to the poor energy yield of fermentation. Glycolysis also produces 2 molecules of NADH for each glucose. NADH must be reoxidized by fermentation in order to regenerate NAD^+ to maintain the oxidation-reduction balance of the cell.

There are various fermentative pathways in *E. coli*, denoted by the end products excreted into the environment. Homolactic fermentation refers to the oxidation of NADH to form NAD^+ by reducing pyruvate to lactate. This reaction is catalyzed by D-lactate dehydrogenase, *LDH_D* [49, 87], which is active only when the pH of the environment drops due to increased concentrations of other acidic fermentative end products, such as acetate, formate and succinate. Homolactic fermentation is a primitive method for regeneration of NAD^+ and can only be used for growth on sugars. Mixed acid fermentation is a more complex fermentation strategy which produces the organic acids formate, acetate, acetaldehyde, ethanol and fumarate. The later is a tricarboxylic acid cycle intermediate which may be reduced to succinate when it is used as a terminal electron acceptor by the electron transport chain during anaerobic respiration. This is part of the clockwise operation of the reductive arm of the tricarboxylic acid cycle in Figure 7.

Like homolactic fermentation, pyruvate is the entry metabolite to mixed acid fermentation. Under anoxic conditions, pyruvate formate lyase, *PFL*, catalyzes the non oxidative cleavage of pyruvate into acetyl-CoA and formate, with the incorporation of coenzyme-A into acetyl-CoA [102, 166]. Acetyl-CoA can be then either be converted to the fermentative end products acetate or ethanol. Both pathways involve two step reversible mechanisms. In the conversion to acetate, phosphotransacetylase, *PTAr* [182], catalyzes transfer of a phosphate group onto the acetyl moiety of acetyl-CoA, to form acetyl-phosphate and release coenzyme-A to be recycled for the previous reaction catalyzed by pyruvate formate lyase, *PFL*. Then acetate kinase, *ACKr*, catalyzes the conversion of acetyl-phosphate to acetate, in the process forming a much needed high energy phosphate bond by converting ADP to ATP [174]. The reactions catalyzed by phosphotransacetylase and acetate kinase are thermodynamically reversible. This allows *E. coli* to grow aerobically on acetate by reversing the flux through this fermentative pathway. During the conversion of acetyl-CoA to ethanol, two molecules of NADH are reoxidized, one by the first reaction catalyzed by acetaldehyde dehydrogenase (acetylating), *ACALD* [59], and the second by the subsequent reaction catalyzed by alcohol dehydrogenase (ethanol), *ALCD2x* [98].

Mixed acid fermentation is more flexible than homolactic fermentation since ethanol is more reduced than carbohydrates, whereas acetate is more oxidized than ethanol. This allows *E. coli* to grow anaerobically on substrates that are more reduced than sugars, e.g. sorbitol, and substrates that are more oxidized than sugars, e.g. glucuronic acid, by adjusting the relative flux through the acetate producing or ethanol producing pathways. In contrast, homolactic fermentation allows anaerobic growth only on sugars. The end products of each fermentation pathway exit the cell along a concentration gradient and in the process transport a proton from the cytoplasm to the periplasmic space.
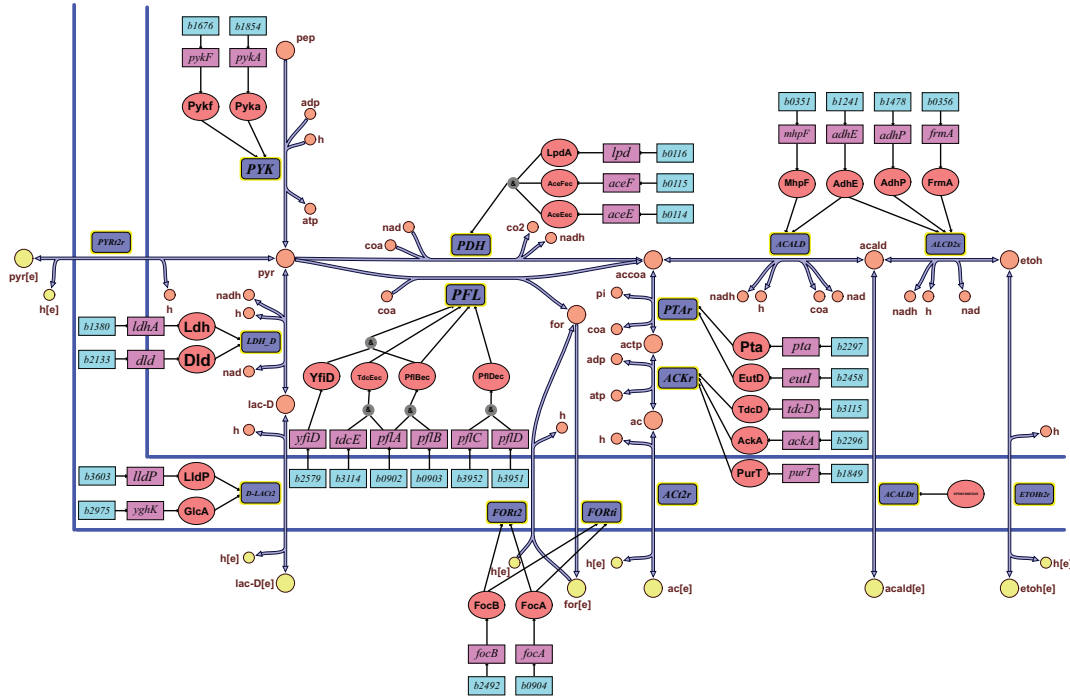
Figure 10: Fermentation begins with pyruvate for both the homolactic fermentative pathway (lower left), and the mixed acid fermentative pathways (lower right). Homolactic fermentation refers to the oxidation of NADH to form NAD^+ by reducing pyruvate, pyr, to lactate, lac-D. Mixed acid fermentation is a more complex fermentation strategy which produces the organic acids formate (for), acetate (ac), acetaldehyde (acald), and ethanol (etoh).

Abbr.	Reaction	Equation
LDH_D	D lactate dehydrogenase	lac-D + nad $\rightleftharpoons$ h + nadh + pyr
D_LACt2	D-lactate transport via proton symport	h[e] + lac-D[e] $\rightleftharpoons$ h + lac-D
PDH	pyruvate dehydrogenase	coa + nad + pyr $\rightarrow$ accoa + co2 + nadh
PFL	pyruvate formate lyase	coa + pyr $\rightarrow$ accoa + for
FORti	formate transport via diffusion	for $\rightarrow$ for[e]
FORt2	formate transport via proton symport	for[e] + h[e] $\rightarrow$ for + h
PTAr	phosphotransacetylase	accoa + pi $\rightleftharpoons$ actp + coa
ACKr	acetate kinase	ac + atp $\rightleftharpoons$ actp + adp
ACALD	acetaldehyde dehydrogenase (acetylating)	acald + coa + nad $\rightleftharpoons$ accoa + h + nadh
ALCD2x	alcohol dehydrogenase (ethanol)	etoh + nad $\rightleftharpoons$ acald + h + nadh
ACt2r	acetate reversible transport via proton symport	ac[e] + h[e] $\rightleftharpoons$ ac + h
ACALDt	acetaldehyde reversible transport	acald[e] $\rightleftharpoons$ acald
ETOHt2r	ethanol reversible transport via proton symport	etoh[e] + h[e] $\rightleftharpoons$ etoh + h

Abbr.	Metabolite	Formula	Charge
lac-D	D-Lactate	$C_3H_5O_3$	-1
for	Formate	CHO_2	-1
actp	Acetyl-phosphate	$C_2H_3O_5P$	-2
ac	Acetate	$C_2H_3O_2$	-1
acald	Acetaldehyde	C_2H_4O	0
etoh	Ethanol	C_2H_6O	0

4.7   Nitrogen Metabolism

Nitrogen is the fourth most abundant atom in *E. coli* and enters the cell either by ammonium ion uptake, *NH4t*, or as a moiety within organic molecules, such as glutamine or glutamate. Glutamate is an extremely abundant metabolite, with a measured concentration of 100.55 μmol gDW^{-1} and production rate of 4.77 mmol gDW^{-1} hr^{-1}. Glutamine is less abundant, with a measured concentration of 3.92 μmol gDW^{-1} and production rate of 3.36 mmol gDW^{-1} hr^{-1} [202]. Glutamine is actively transported across the inner cytoplasmic membrane by an ATP binding cassette transporter, *GLNabc*, which couples the hydrolysis of one phosphate bond to the import of one molecule of glutamine [200]. Glutamate is also actively transported across the inner cytoplasmic membrane by the energetically more efficient method of proton symport, *GLUt2r* [195].

In *E. coli*, assimilation refers to the incorporation of an environmental substrate into a cell by fueling reactions. Amination is the process by which an amine group is introduced into an organic molecule. An amine group refers to one type of moiety that contains a basic nitrogen atom. Direct ammonia assimilation is catalyzed by NADPH specific reductive amination of 2-oxoglutarate to glutamate by glutamate dehydrogenase (NADP), *GLUDy* [202]. Indirect ammonia assimilation is by a cyclic pair of sequential reactions catalyzed by glutamine synthetase, *GLNS* [158], and glutamate synthase (NADPH), *GLUSy* [202]. Glutamine synthetase first catalyzes the assimilation of ammonia by converting glutamate, with one amino moiety, into glutamine, with two amino moieties. Then glutamate synthase, *GLUSy*, catalysis the transfer of the amide group of glutamine to 2-oxoglutarate, generating two molecules of glutamate. The net result of the indirect and direct pathways is the synthesis of one glutamate molecule by combining one each of 2-oxoglutarate and ammonia. In turn, glutamate is a precursor for the biosynthetic pathways that lead to the amino acids proline and arginine.
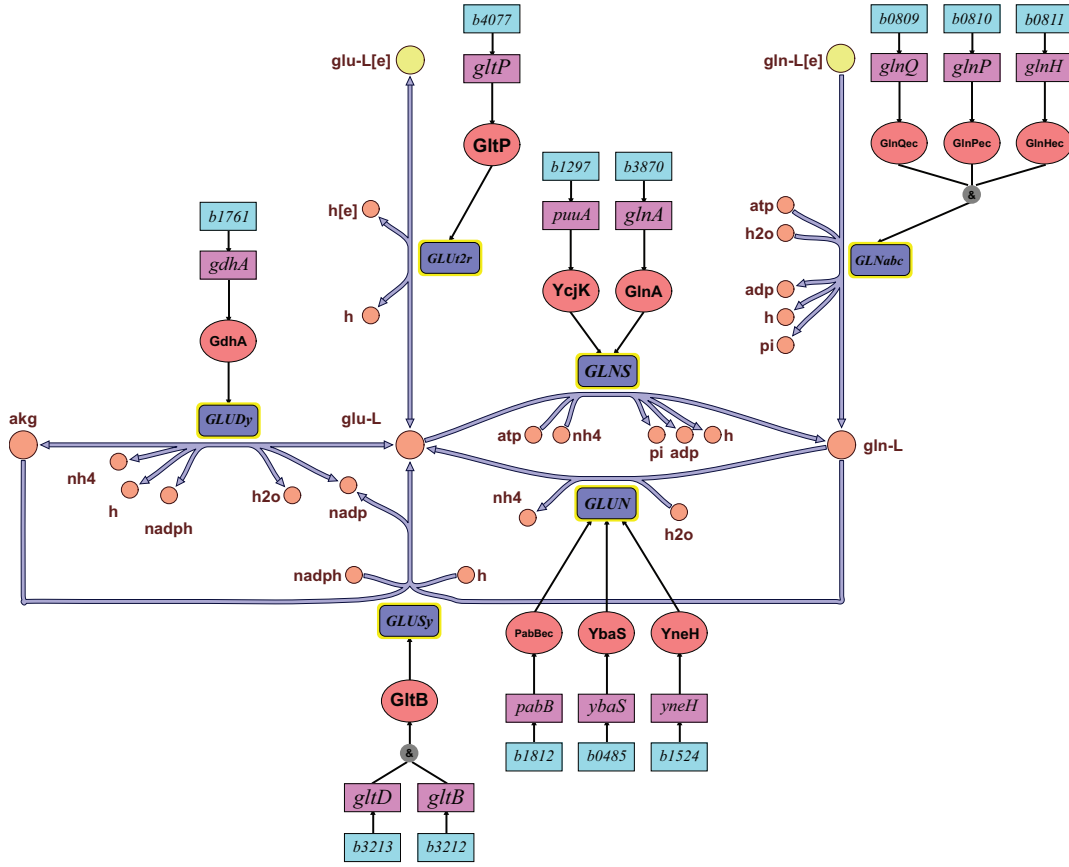
Figure 11: Nitrogen metabolism in *E. coli*. Glutamate dehydrogenase (NADP), *GLUDy*, catalyzes direct ammonia assimilation by NADPH specific reductive amination of 2-oxoglutarate to glutamate. Indirect ammonia assimilation is by a cyclic pair of sequential reactions, beginning with glutamine synthetase, *GLNS*, which catalyzes the addition of an amino moiety to glutamate to form glutamine. Then glutamate synthase (NADPH), *GLUSy,* catalyzes the transfer of this amino moiety to 2-oxoglutarate, generating two molecules of glutamate. The net result is amination of 2-oxoglutarate to glutamate, as in the direct pathway.

Abbreviation	Reaction	Equation
GLNabc	L-glutamine transport via ABC system	atp + gln-L[e] + h2o → adp + gln-L + h + pi
GLUt2r	L-glutamate transport via proton symport	glu-L[e] + h[e] ⇌ glu-L + h
GLUDy	glutamate dehydrogenase (NADP)	glu-L + h2o + nadp ⇌ akg + h + nadph + nh4
GLNS	glutamine synthetase	atp + glu-L + nh4 → adp + gln-L + h + pi
GLUSy	glutamate synthase (NADPH)	akg + gln-L + h + nadph → 2 glu-L + nadp
GLUN	glutaminase	gln-L + h2o → glu-L + nh4

Abbr.	Metabolite	Formula	Charge
glu-L	L-Glutamate	$C_5H_8NO_4$	-1
gln-L	L-Glutamine	$C_5H_{10}N_2O_3$	0
nh4	Ammonium	H_4N	+1
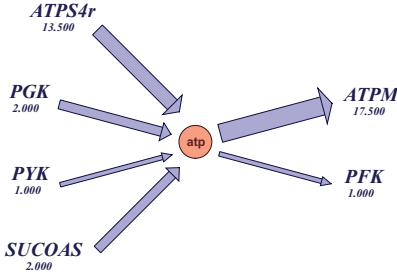
Figure 12: All of the reactions that produce or consume ATP in a particular flux distribution. The total amount of ATP produced is exactly equal to the amount of ATP consumed, so $\frac{d[ATP]}{dt} = 0$.

4.8   Mathematical Formulation of Core *E. coli* Metabolism

The core *E. coli* metabolic reconstruction can be converted into a mathematical model by encoding all of the metabolites and reactions in a stoichiometric matrix, $\mathbf{S} \in \mathbb{Z}^{m,n}$. The stoichiometric matrix for the core *E. coli* model has $m = 72$ rows, each corresponding to a metabolite, and $n = 95$ columns, each corresponding to a reaction. The coefficients within a single column of the stoichiometric matrix represent the stoichiometry of a single reaction. A negative stoichiometric coefficient indicates the number of molecules of a particular metabolite consumed in that reaction. A positive stoichiometric coefficient represents the number of molecules of a metabolite produced in that reaction. Since each reaction typically involves only a few metabolites, the stoichiometric matrix is sparse, consisting mostly of zero coefficients. Each reaction must be balanced with respect to consumption and production of both chemical elements and electrical charge. The exception to this rule are the columns of the *E. coli* core stoichiometric matrix which correspond to exchange reactions at the boundary of the model and its environment. Each exchange reaction represents the influx or efflux of an extracellular metabolite between the model and the environment. Therefore a metabolite leaving the model is represented with a positive exchange flux, and the opposite for a metabolite entering the model. The biomass reaction, representing the synthesis of biomass, is also represented by a column of the stoichiometric matrix. The biomass reaction may be considered a special type of exchange reaction, and is discussed further in Section 4.9.

The net flux through all of the 95 reactions in the model can be mathematically represented by a 95 dimensional vector, $\mathbf{v} \in \mathbb{R}^n$. By convention, net flux is represented with the unit: milli-mole, per gram dry weight, per hour (mmol gDW^{-1} hr^{-1}). The dot product of the stoichiometric matrix and the net flux vector gives an m dimensional vector of changes in metabolite concentration over time, $\frac{d\mathbf{x}}{dt} \in \mathbb{R}^m$. During balanced growth, when considering a large population of cells, it is reasonable to assume that all metabolite concentrations are constant, and therefore we have the fundamental system of m equations for mass conservation at steady state

$$\mathbf{S} \cdot \mathbf{v} = \frac{d\mathbf{x}}{dt} \equiv \mathbf{0} \tag{1}$$

Let the first row of the stoichiometric matrix correspond to ATP, then consider the meaning of the first equation in the system of equations 1,

$$\mathbf{S}_{1,1}\mathbf{v}_1 + \mathbf{S}_{1,2}\mathbf{v}_2 + \ldots + \mathbf{S}_{1,n}\mathbf{v}_n = \frac{d[ATP]}{dt} = 0 \tag{2}$$

Since the stoichiometric coefficients represent the number of ATP molecules consumed or produced in each reactions in the model, and the net fluxes represent the rates for each respective reaction, then equation 2 balances the total net production and consumption of ATP (Figure 12).

The net flux for all biochemical reactions is bounded by diffusion limitations [60]. In addition, experimental data on maximum reaction rates, substrate uptake rates, or waste secretion rates may be used to further tighten these bounds, mathematically represented by

$$\mathbf{v}_{lb} \leq \mathbf{v} \leq \mathbf{v}_{ub}$$

For reactions that are considered to be effectively thermodynamically irreversible in the forward direction, then the convention is to set the lower bound on a reaction rate to zero [56]. In any realistic stoichiometric model there are more columns than rows in the stoichiometric matrix. Therefore there are an insufficient number of equations to specify a unique net flux in equation 1. In order to predict a biologically meaningful net flux, a technique termed flux

balance analysis can be used [190, 191]. Flux balance analysis uses linear programming to optimize a biologically motivated objective function, subject to steady state mass balance and bounds on net fluxes to predict an *in vivo* flux. This can be expressed as the linear programming problem

$$
\begin{aligned}
maximize: &\quad \mathbf{c}^T \cdot \mathbf{v} \\
subject\ to: &\quad \mathbf{S} \cdot \mathbf{v} = \mathbf{0} \\
&\quad \mathbf{v}_{lb} \leq \mathbf{v} \leq \mathbf{v}_{ub}
\end{aligned}
$$

The objective function, $\mathbf{c}^T \cdot \mathbf{v}$, can maximize or minimize any linear combination of reaction fluxes. In flux balance analysis of *E. coli*, the biomass reaction is often the sole reaction to be maximized by the objective function since maximum growth is evolutionarily favored in genetically heterogeneous culture [148]. The publicly available COBRA Toolbox software for MATLAB [169] may be used to make numerical predictions of *in vivo* net flux by implementing a mathematical model as a computational model [58].

4.9   Biomass Reaction

In order to represent growth, the core *E. coli* model includes a biomass reaction which drains precursor metabolites from the network at stoichiometrically fixed relative rates while also producing several by-product metabolites [55], see Table 2 and Figure 13. These precursors are used to produce the lipids, proteins, nucleic acids, and other macromolecules required to replicate a cell. To determine these metabolites and their quantity, we used the dry weight composition data for an average *E. coli* B/r cell growing exponentially at 37^oC under aerobic conditions in glucose minimal medium, with an approximate doubling time of 40 min having a dry cell weight of 2.81013 grams, see EcoSal Chapter: Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. Since most of the subunits of the cellular macromolecules, such as nucleic acids and amino acids, are not present in the core model, they could not be directly accounted for in the biomass reaction. The metabolites in the core model that those macromolecular subunits are synthesized from are included instead. These are precursor metabolites. For example, the amino acid L-alanine is synthesized from pyruvate and L-glutamate, so both of these metabolites are consumed in the biomass reaction. Several metabolites are actually produced by the biomass reaction. ADP, protons, and inorganic phosphate are produced by the hydrolysis of ATP in the balanced reaction atp + h2o → adp + h + pi. 2-oxoglutarate is produced during the synthesis of amino acids, when L-glutamate transfers its amino group to another compound in a transamination reaction. Coenzyme A is produced when acetyl-CoA is consumed, and NAD^+ is reduced to NADH and NADPH is oxidized to $NADP^+$ during biomass synthesis.

Additional energetic requirements exist for growth beyond what is needed to generate the macromolecular content of the cell. These energetic maintenance requirements are for growth associated maintenance (e.g., protein polymerization costs) and non-growth-associated maintenance (e.g., membrane leakage). To represent growth associated maintenance, ATP is converted to ADP at 59.81 mmol gDW^{-1}, accounting for energy used in cell division and other growth processes. Non growth associated maintenance is **not** part of the biomass reaction. Instead, it is represented with a lower bound of 8.39 mmol gDW^{-1} hr^{-1} on the ATP maintenance reaction (*ATPM*), simulating energy used for protein turnover and other processes that do not change with growth.

Table 2: In the biomass reaction, 23 different metabolites are consumed or produced in order to simulate growth. The metabolites that are consumed have negative coefficients and the metabolites that are produced have positive coefficients.

Abbr.	Metabolite	Stoichiometry
3pg	3-Phospho-D-glycerate	-1.496
accoa	Acetyl-CoA	-3.7478
adp	ADP	+59.81
akg	2-Oxoglutarate	+4.1182
atp	ATP	-59.81
coa	Coenzyme A	+3.7478
e4p	D-Erythrose 4-phosphate	-0.361
f6p	D-Fructose 6-phosphate	-0.0709
g3p	Glyceraldehyde 3-phosphate	-0.129
g6p	D-Glucose 6-phosphate	-0.205
gln-L	L-Glutamine	-0.2557
glu-L	L-Glutamate	-4.9414
h	H+	+59.81
h2o	H2O	-59.81
nad	Nicotinamide adenine dinucleotide	-3.547
nadh	Nicotinamide adenine dinucleotide - reduced	+3.547
nadp	Nicotinamide adenine dinucleotide phosphate	+13.0279
nadph	Nicotinamide adenine dinucleotide phosphate - reduced	-13.0279
oaa	Oxaloacetate	-1.7867
pep	Phosphoenolpyruvate	-0.5191
pi	Phosphate	+59.81
pyr	Pyruvate	-2.8328
r5p	alpha-D-Ribose 5-phosphate	-0.8977

5   Boolean Core *E. coli* Transcriptional Regulation

In addition to the metabolic reconstruction, the core *E. coli* model also contains a Boolean representation of part of the associated transcriptional regulatory network. This network is a modified subset of the genome scale transcriptional regulatory reconstruction *i*MC1010 [38]. In response to external and internal stimuli, *in silico* transcription factors either activate or repress genes associated with metabolic reactions. This regulation improves the predictive fidelity of the metabolic model by imposing additional context dependent constraints on certain reactions. The transcriptional regulatory reconstruction consists of a set of Boolean rules that dictate whether a gene is either fully induced or fully repressed. If the genes associated with an enzyme or transport protein/complex are repressed, then *in silico* flux is set to zero for the corresponding reaction. The solution space of the network shrinks when these additional constraints are imposed. Reactions that are not used due to regulatory effects are thus restricted, so when using flux balance analysis, the optimal flux distribution will be consistent with known regulation. This optimal flux distribution may be different from the flux distribution of an unregulated model. In this case, the flux distribution of the unregulated model violated at least one regulatory constraint, making it biologically unrealistic. The use of computationally implemented Boolean rules in a genome scale model has been shown to lead to more accurate flux balance analysis predictions [38].

A gene is considered to be induced when evaluation of the corresponding Boolean rule gives 'true'. In contrast, a gene is considered to be repressed, when evaluation of the corresponding Boolean rule gives 'false'. Boolean logic is used to evaluate each Boolean rule. For example, consider the enzyme phosphoenolpyruvate synthase, *PPS*, which catalyzes the first step of gluconeogenesis, the conversion of pyruvate to phosphoenolpyruvate. The gene for phosphoenolpyruvate synthase is *pps* and its Boolean rule is simply 'FruR'. That is, if FruR is 'true' then the *pps* gene is induced allowing *in silico* flux through the reaction catalyzed by phosphoenolpyruvate synthase, *PPS*. FruR is a transcriptional regulator which is active when the cytoplasmic concentration of D-fructose-1,6-bisphosphate,
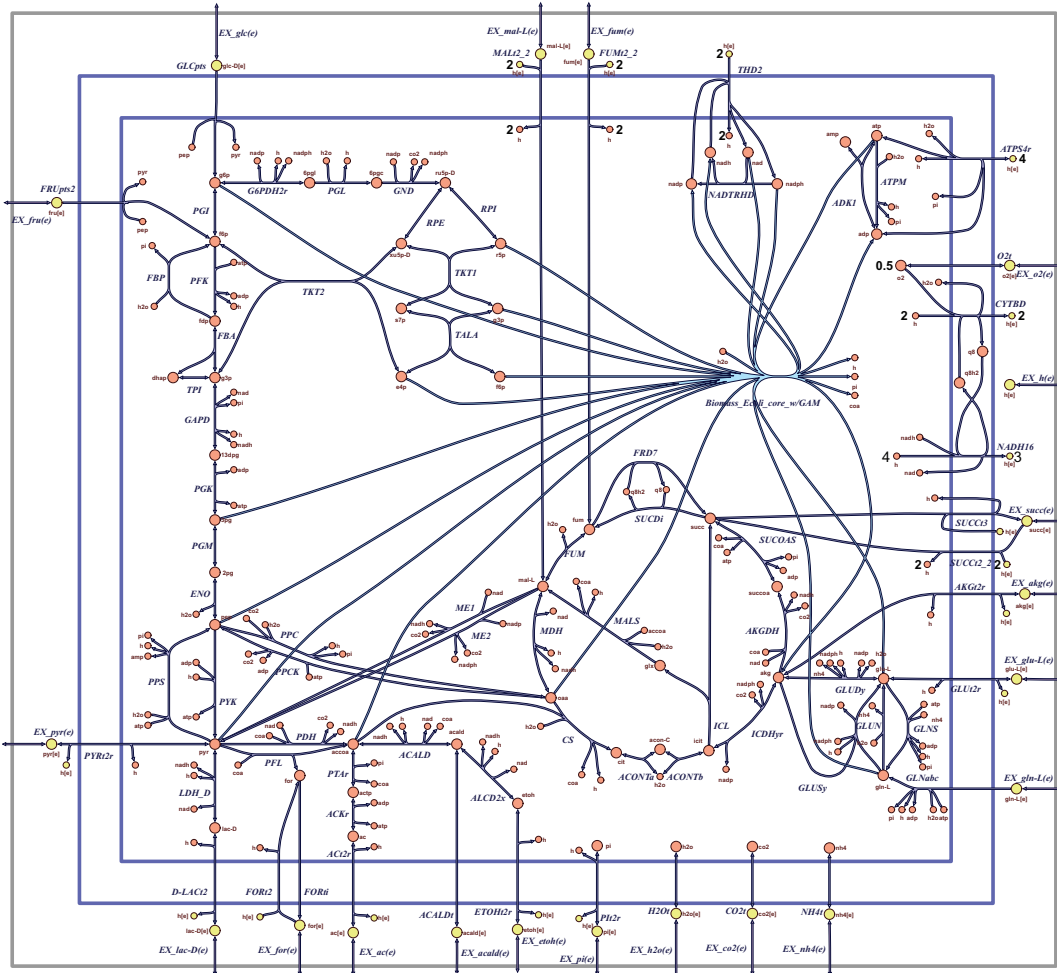
Figure 13: The biomass reaction (light blue) of the *E. coli* core model shown in the context of the entire model. This is a single reaction which consumes and produces metabolites from many different pathways in the model. In order to grow, *E. coli* must be able to produce all of the compounds that are consumed. The metabolites that are consumed enter the reaction from the left, and the by-products that are produced leave the reaction to the right. The biomass reaction drains the precursors and energy/redox carriers 3-phospho-D-glycerate, acetyl-CoA, ATP, D-erythrose-4-phosphate, D-fructose-6-phosphate, Glyceraldehyde-3-phosphate, D-glucose-6-phosphate, L-glutamine, L-glutamate, H_2O, NAD^+, NADPH, oxaloacetate, phosphoenolpyruvate, pyruvate, and alpha-D-ribose-5-phosphate from the network while producing ADP, 2-oxoglutarate, coenzyme-A, H^+, NADH, $NADP^+$, and phosphate.

FDP, is low [13]. The Boolean rule for FruR is 'NOT surplusFDP'. That is, if there is no surplus of D-fructose-1,6-bisphosphate, then FruR is 'true' and therefore the *pps* gene is induced, allowing *in silico* gluconeogenic flux through the reaction catalyzed by phosphoenolpyruvate synthase, *PPS*. In contrast, if *surplusFDP* is 'true', then FruR is 'false' and therefore *pps* is repressed.

Regulatory conditions, such as *surplusFDP*, are variables that represent a complex regulatory rule for a transcription factor that cannot be accurately represented with only one variable. The regulatory rule for *surplusFDP* is '((NOT FBP) AND (NOT (TKT2 OR TALA OR PGI))) OR fru[e]'. If *fru[e]* is 'true', then *surplusFDP* is 'true', independent of the state of the other variables. If fructose-bisphosphatase, *FBP*, is 'false', and any one of transketolase, *TKT2*, transaldolase, *TALA*, or glucose-6-phosphate isomerase, *PGI*, is 'false', then *surplusFDP* is 'true', therefore *FruR* is 'false' and *pps* is repressed. By using Boolean logic, all rules in a regulatory network can be reduced to either 'true' or 'false', and ultimately this dictates whether each metabolic gene is induced or repressed. Not every gene in the metabolic network is controlled by the regulatory network, so the unregulated genes are assumed to always be active, and their fluxes are never constrained to zero. Table 3 lists the Boolean regulatory rules for regulated metabolic genes, and Table 4 lists the Boolean regulatory rules for transcription factors and regulatory conditions in the transcriptional regulatory network. An abstract overview of part of the regulatory network is depicted in Figure 14.

5.1 Regulation of Anoxic Growth

Under anoxic conditions, the transcription factors ArcA and Fnr act as global regulators that induce many different genes needed for fermentation and growth without oxygen. However, the principal function of ArcA and Fnr is to repress genes not required when oxygen is abundant [85]. When oxygen availability is reduced, the drop in redox potential signals the phosphorylation and thereby activation of the global regulator ArcA. Only in anaerobic conditions is the global transcriptional regulator Fnr is also activated. When there is oxygen in the cytoplasm, an oxidized 4Fe-4S cluster in Fnr inactivates it. The regulatory rule 'NOT o2[e]' is the same for ArcA and Fnr. Therefore when o2[e] is false, ArcA and Fnr are true, representing activation.

Both ArcA and Fnr induce fermentative genes such as *pflA*, *pflB*, *pflD*, and *pflC*, coding for pyruvate formate lyase, *PFL* [164], and *focA* and *focB* coding for formate transport, *FORt2* and *FORti*. The regulatory rule for *pflC* is 'ArcA or Fnr'. Therefore if either ArcA or Fnr are true, then *pflC* is true, representing expression. Fnr also induces *tdcD* coding for the fermentative enzyme acetate kinase, *ACKr*. Anoxic conditions also induce *adhE* encoding the enzymes for both acetaldehyde dehydrogenase (acetylating), *ACALD*, and alcohol dehydrogenase (ethanol), *ALCD2x*. The regulatory rule for *adhE* is '(NOT o2[e]) OR (NOT (o2[e] AND FruR)) OR Fis', representing its indirect induction when oxygen is absent or derepression by the inactivity of the transcriptional regulator FruR. Fis is discussed in Section 5.5.

In the absence of a common electron acceptor such as oxygen or nitrate, the reactions of the tricarboxylic acid cycle no longer operate as an energy producing cycle. Instead they function as two separate biosynthetic pathways. Beginning at oxaloacetate, a reductive pathway via fumarate reductase, *FRD7*, producing succinyl-CoA, and an oxidative pathway producing 2-oxoglutarate. In the absence of oxygen, ArcA represses a number of genes in the tricarboxylic acid cycle that are unnecessary for its cyclic operation, including *sdhA*, *sdhB*, *sdhC*, and *sdhD*, coding for succinate dehydrogenase (irreversible), *SUCDi* [41, 133], in addition to *fumA* and *fumC* coding for fumarase, *FUM* [134], and *mdh* coding for malate dehydrogenase, *MDH*. The regulatory rule for *sdhA-D*, '(NOT (ArcA OR Fnr)) OR Crp OR Fis' indicates that these genes are expressed when ArcA and Fnr are false, or when either of the transcriptional regulators Crp or Fis are true.

The cell adapts to changing environmental oxygen conditions by utilizing different isozymes of fumarase, *FUM*. Both Fnr and ArcA repress *fumA* coding for fumarase A [134] but Fnr induces *fumB* coding for the isozyme fumarase B which has greater affinity for malate as a substrate [187]. The regulatory rule for *fumA* is 'NOT (ArcA OR Fnr)' so the *fumA* isozyme is expressed when ArcA and Fnr are false. The regulatory rule for *fumB* is 'Fnr OR Crp OR DcuR' so the *fumB* isozyme is expressed when any one of Fnr, Crp or DcuR is true. The transcriptional regulator DcuR is discussed in Section 5.3. Fnr compliments induction by DcuR of *frdA*, *frdB*, *frdC*, and *frdD*, coding for fumarate reductase, *FRD7* [188]. In anoxic conditions, fumarate reductase, *FRD7*, catalyzes the reverse reaction to that catalyzed by succinate dehydrogenase, *SUCDi*, in oxic conditions. The regulatory rule for *frdA-D* is 'Fnr OR DcuR' meaning that these genes are expressed when either Fnr or DcuR are activated.

Both ArcA and Fnr downregulate oxidative phosphorylation by repressing the *nuoA* - *nuoN* operon coding for NADH dehydrogenase (ubiquinone-8 & 3 protons), *NADH16* [14, 71]. This reaction is used for aerobic respiration
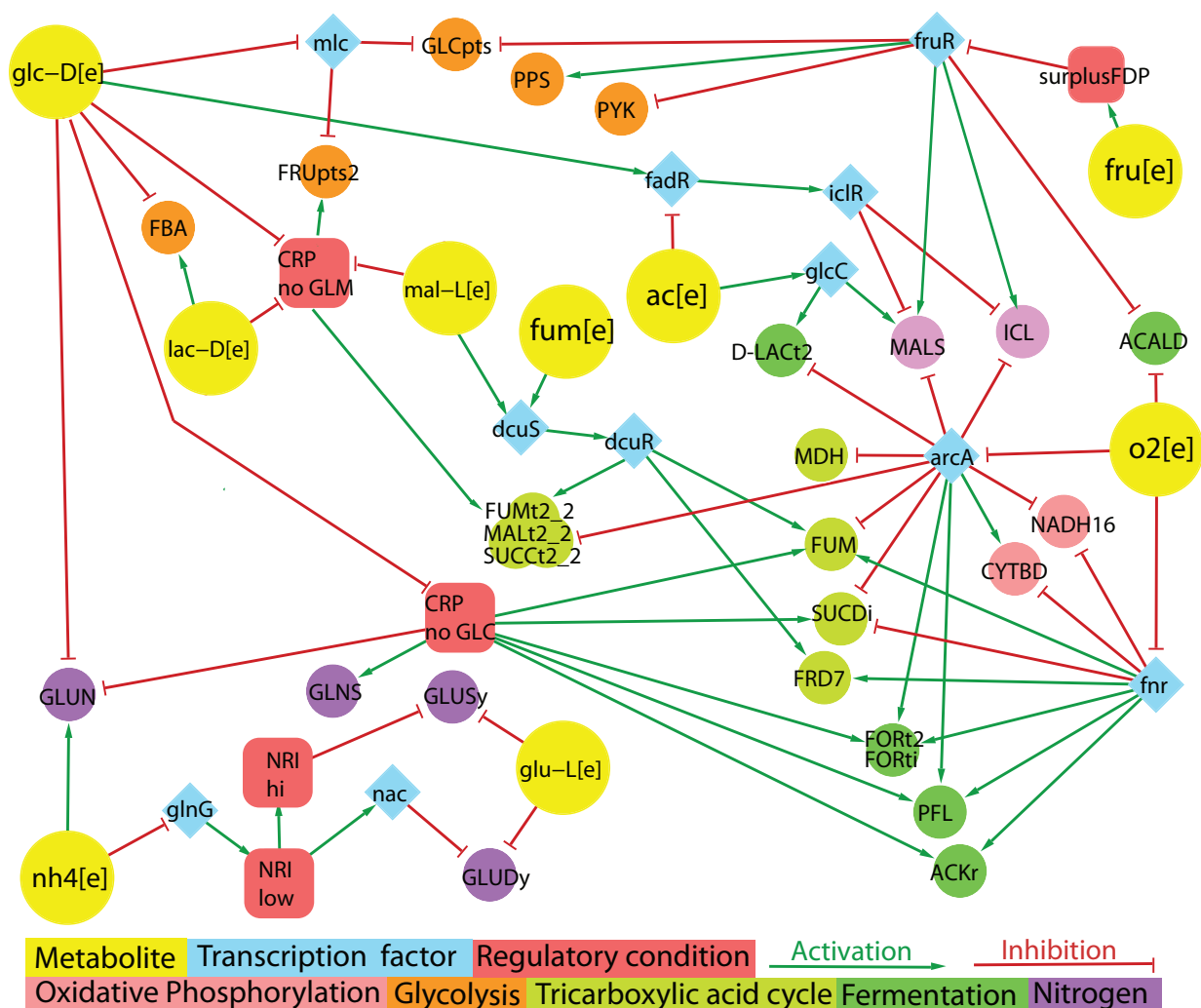
Figure 14: An schematic overview of part of the *E. coli* core regulatory network from an environmental stimulus-response perspective. For the sake of clarity, reaction abbreviations are displayed rather than the gene(s) which code for the associated protein or protein complex. The Boolean state of *in silico* reactions, true = flux, or false = zero flux, may be determined by traversing the activation or inhibition links, beginning at an environmental metabolite stimulus (yellow circles). An anoxic environment is represented by o2[e] = false. Oxygen inactivates Fnr, represented by an inhibition link from o2[e] to Fnr. When o2[e] is false, then Fnr is true indicating that Fnr is active in the absence of oxygen. Active Fnr then induces or represses the expression of genes for various reactions. Induction of the genes for fumarase, *FUM*, by Fnr, is represented by an activation link from Fnr to *FUM*. Repression of the genes for oxidative phosphorylation by Fnr is represented by the inactivation links from Fnr to NADH dehydrogenase, *NADH16*, and cytochrome oxidase, *CYTBD*, i.e. activation assigns the target node the same Boolean status as the source node whereas inhibition sets the Boolean status of the target to the opposite of the source. The central role of the global transcriptional regulators ArcA, Fnr, Crp, and FruR is evident from the number of reactions under their control. Boolean regulatory rules also encode combinatorial regulation of the same gene by multiple transcription factors. For details see Table 3.

and is therefore redundant in anoxic conditions. The corresponding regulatory rule for *nuoA-N* is '(NOT (ArcA OR Fnr))'. When oxygen is present at low concentration ArcA induces *cydA* and *cydB,* coding for the reaction cytochrome oxidase bd (ubiquinol-8: 2 protons), *CYTBD* [188]. In fully anoxic conditions these same genes *cydA* and *cydB* coding for cytochrome oxidase bd (ubiquinol-8: 2 protons), *CYTBD*, are repressed by Fnr [188]. The regulatory rule for *cydA* and *cydB* is '(NOT Fnr) OR ArcA'. Both Fnr and ArcA are true in the absence of oxygen and in the *E. coli* core model this conflicting regulatory rule for *cydA* and *cydB* must be adjusted to represent anoxic/low oxygen/oxic conditions. Modeling analogue variables in regulatory models, is a topic of ongoing research [72]. ArcA downregulates the glycoxylate cycle by repressing *aceB* and *glcB*, coding for malate synthase, *MALS*, and *aceA* coding for isocitrate lyase, *ICL* [136]. ArcA also represses a number of transporters including *glcA* and *lldP* coding for D-lactate transport via proton symport, *D_LACt2*, and *dctA* coding for the aerobic transporter for fumarate, malate, and succinate, *FUMt2_2*, *MALt2_2*, and *SUCCt2_2* [44]. Many of the latter genes are also regulated by other transcription factors giving rise to a combinatorial expansion in the total number of network states.

5.2 Catabolite Repression

In a medium containing glucose and another sugar substrate, such as lactate or malate, *E. coli* preferentially catabolizes glucose until it is depleted, thereafter switching to catabolism of the less desirable substrate [127]. The repression of enzymes for catabolism of a less desirable substrate by the presence of a desirable substrate is generally termed catabolite repression [161]. Cytoplasmic concentrations of the cofactor cyclic adenosine monophosphate, cAMP, allosterically control the activity of the cAMP receptor protein, Crp [16]. The presence of glucose reduces the cytoplasmic concentration of Crp-cAMP which is necessary for induction of various genes necessary for catabolism of less desirable substrates. To represent regulation by Crp-cAMP in a Boolean manner, extra regulatory condition variables are necessary in order to represent Crp-cAMP under different conditions. The *CRPnoGLM* regulatory condition is false when either glucose (*glc-D[e]*), malate (*mal-L[e]*) or lactate (*lac-D[e]*) are present in the media. In this condition the gene *dctA* is repressed, which codes for the fumarate, malate, and succinate transporters, *FUMt2_2*, *MALt2_2*, and *SUCCt2_2* [44]. The regulatory rule for *dctA* is 'CRPnoGLM AND (NOT ArcA) AND DcuR'. Therefore ArcA must also be false and DcuR must also be true for *dctA* to be expressed. When the *CRPnoGLM* regulatory condition is false, the genes *manX*, *manY*, and *manZ* are repressed down regulating fructose transport via PEP:Pyr PTS, *FRUpts2*. Note that *manX*, *manY*, and *manZ* can also code for subunits of D-glucose transport via PEP:Pyr PTS, *GLCpts*. The latter can also be encoded by *ptsH*, *ptsI*, *malX*, *crr* and *ptsG*, providing an independent means of glucose transport.

When glucose is absent, the regulatory condition *CRPnoGLC* is true representing activation of the global regulator Crp-cAMP. *CRPnoGLC* induces the genes *sdhA*, *sdhB*, *sdhC*, *sdhD*, coding for succinate dehydrogenase, *SUCDi* [41]. *CRPnoGLC* also upregulates fermentation by inducing *pflA*, *pflB*, and *tdcE* coding for pyruvate formate lyase, *PFL*, inducing *focA* and *focB* coding for formate transport, *FORt2* and *FORti*, inducing *tdcD* coding for acetate kinase, *ACKr* [165]. The regulatory rule for *focA* and *focB* is 'ArcA OR (Fnr AND CRPnoGLC)' therefore these genes are induced when ArcA is true, or when both Fnr and *CRPnoGLC* are true. When *CRPnoGLC* is true, this induces *fumB* coding for fumarase, *FUM*, in the tricarboxylic acid cycle. *CRPnoGLC* represses *glnA* coding for glutaminase, *GLUN*, but induces *yneH* coding for glutamine synthase, *GLNS*. Another transcription factor that is activated when glucose is absent is Mlc. Mlc represses *ptsG*, *manX*, *manY*, and *manZ*, which all code for subunits of the glucose and fructose phosphotransferase systems, *FRUpts2* and *GLCpts* [144]. The regulatory rule for *manX-Z* is 'CRPnoGLM OR (NOT Mlc)', indicating that these transport genes are induced when *CRPnoGLM* is true or Mlc is false.

5.3 Growth on Acetate or C$_4$-Dicarboxylate Compounds

The gene *glcC* codes for a transcription factor that is induced when acetate (*ac[e]*) is present in the media [136]. The regulatory rule for GlcC is therefore simply 'ac[e]'. GlcC induces *glcA* and *glcB* which code for D-lactate transport via proton symport, *D_LACt2*, and malate synthase, *MALS*, respectively [136]. The regulatory rule for *glcA* and *glcB* is '(NOT ArcA) AND GlcC' so these genes are induced when acetate and oxygen are present. The dual transcriptional regulator FadR is activated by the presence of glucose or the absence of acetate in the media, and it is a regulator of fatty acid synthesis. FadR activates transcription of the gene *iclR*, which is also a transcription factor [73]. IclR represses transcription of the genes *aceB* and *aceA* [37], which code for the malate synthase, *MALS*, and isocitrate lyase, *ICL*, enzymes, respectively. The overall effect of *fadR* is thus to inhibit the
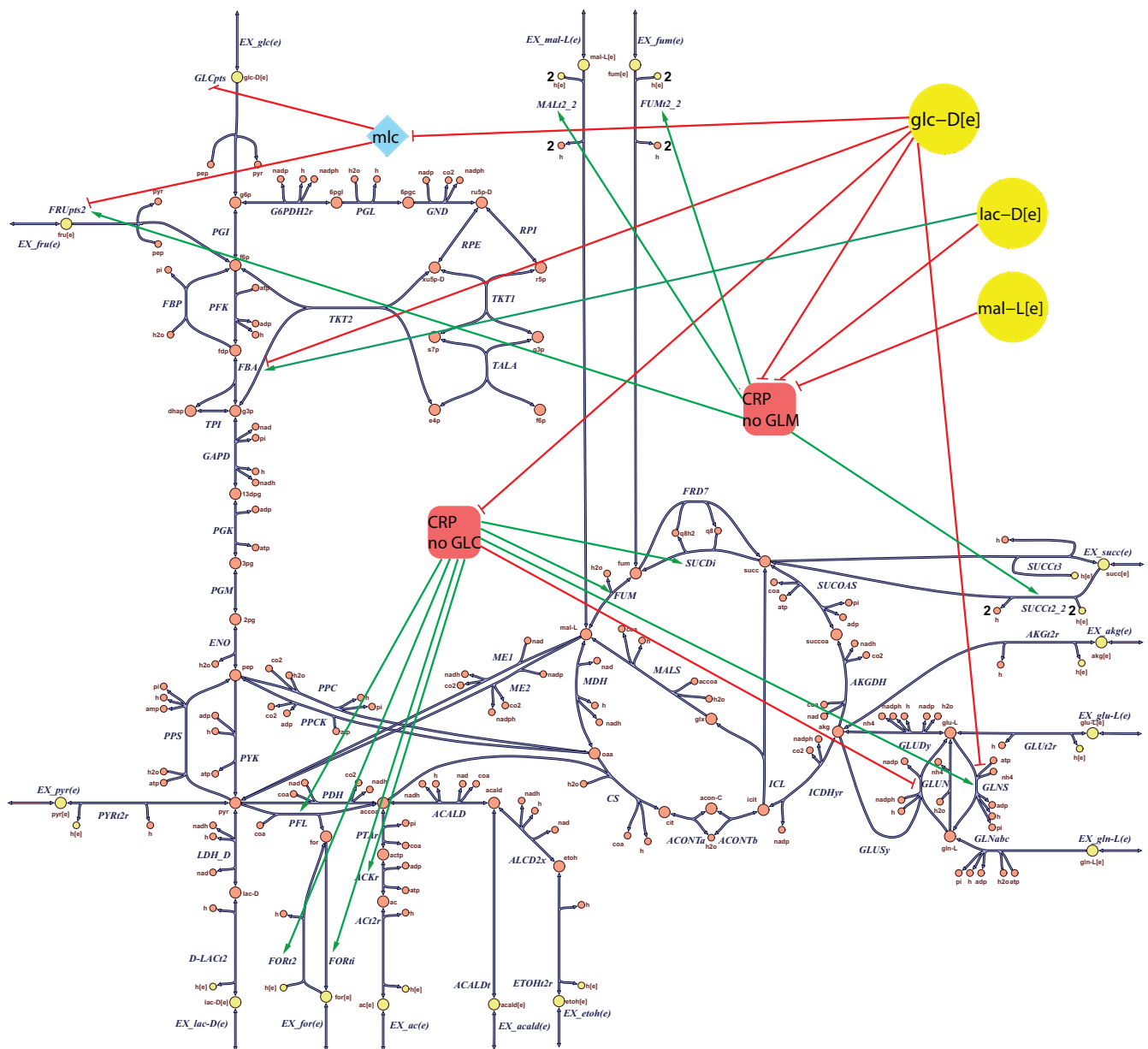
Figure 15: Overview of the effects of catabolite repression on the core *E. coli* model. When the metabolites glucose, (glc-D[e]), lactate (lac-D[e]), or malate (mal-L[e]) are absent from the media, the conditions *CRPnoGLM* and *CRPnoGLC* are activated, regulating a variety of metabolic reactions.

glyoxylate cycle when the cell is not consuming acetate. The genes *aceB* and *aceA* have the regulatory rule '(NOT IclR) AND ((NOT ArcA) OR FruR)' indicating that the glycoxylate cycle is active when glucose and oxygen are present or when glucose is present and acetate is absent.

The response to C_4-dicarboxylate compounds is regulated by the two component system *dcuR* and *dcuS*. *dcuS* codes for a sensor histidine kinase that is activated when fumarate (*fum[e]*), L-malate (*mal-L[e]*), or succinate (*succ[e]*) are present in the media [69]. Once activated, the DcuS protein phosphorylates DcuR, activating it. DcuR then induces the gene *dctA* [44], which codes for the transporter for fumarate, malate, and succinate, *FUMt2_2*, *MALt2_2*, and *SUCCt2_2*, allowing these metabolites to be consumed. DcuS also upregulates *fumB* [187], which codes for the fumarase B isozyme. Finally, DcuS upregulates *frdA*, *frdB*, *frdC*, and *frdD* [69, 188], which are subunits of the fumarate reductase enzyme, *FRD7*. These two reactions are responsible for interconversion of succinate, fumarate, and malate. The regulatory rule for *frdA-D* is 'Fnr OR DcuR' indicating that either when oxygen is present or fumarate, L-malate, or succinate is present in the media, then fumarate reductase is induced.

5.4   Glycolysis vs Gluconeogenesis

When cytoplasmic concentrations of D-Fructose-1-6-bisphosphate, fdp, are low, the dual transcriptional regulator FruR represses glycolytic and fermentative enzymes yet simultaneously induces gluconeogenic genes [13]. In contrast, surplus D-Fructose-1-6-bisphosphate binds to FruR dislodging it from its binding sites thereby derepressing glycolytic and fermentative enzymes and deactivating gluconeogenic enzymes. Since current constraint-based models do not explicitly model metabolite concentrations, the regulatory condition *surplusFDP* is used to indicate conditions of excess D-Fructose-1-6-bisphosphate. The *surplusFDP* condition is met when fructose, fru[e], is present in the media or the reactions *FBP* and any of *TKT2*, *TALA*, or *PGI* have zero flux. When *surplusFDP* is false then FruR is true, thereby repressing glycolytic and fermentative enzymes and inducing gluconeogenic enzymes.

FruR induces *ptsG* coding for D-glucose transport via PEP:Pyr PTS, *GLCpts*, but represses *pykF* coding for the glycolytic enzyme pyruvate kinase, *PYK* [183]. FruR also downregulates fermentation by repressing *adhE* coding for acetaldehyde dehydrogenase (acetylating), *ACALD*, and alcohol dehydrogenase (ethanol), *ALCD2x* [116, 117]. FruR upregulates the glycoxylate cycle by inducing *aceA* coding for isocitrate lyase, *ICL*, and inducing *aceB* coding for malate synthase, *MALS* [37]. FruR also upregulates gluconeogenesis by inducing *pps* which encodes posphoenolpyruvate synthase *PPS* [123]. In summary, FruR is capable of reversing the flow of carbon to replenish glycolytic intermediates as sensed by the level of D-Fructose-1-6-bisphosphate.

The dual transcriptional regulator *pdhR* (pyruvate dehydrogenase complex regulator) downregulates pyruvate dehydrogenase when the pyruvate concentration in the cell is low [147]. The Boolean regulatory rule for PdhR is 'NOT surplusPYR' (Figure 17). High pyruvate concentration is represented by the variable *surplusPYR*, which is true when there is no flux through *ME1* or *ME2*, and no flux through either one of *GLCpts*, *PYK*, *PFK*, *LDH_D*, or *SUCCt2_2*. The Boolean rule for *surplusPYR* is '(NOT (ME2 OR ME1)) AND (NOT (GLCpts OR PYK OR PFK OR LDH_D OR SUCCt2_2))'. PdhR inhibits the genes *aceE* and *aceF* which both code for subunits of the pyruvate dehydrogenase complex, *PDH*. The regulatory rule for *aceE* and *aceF* is '(NOT PdhR) OR Fis' meaning that these genes are repressed when PdhR is true, meaning that pyruvate dehydrogenase is repressed when cytoplasmic pyruvate concentration is low.

5.5   Regulation by Fis

The DNA binding protein Fis regulates the expression of many genes by bending the genomic DNA, changing its topological structure. It has been shown to directly or indirectly regulate 21% of the genes in *E. coli*, and 894 Fis-associated regions of the genome have been identified [29]. Transcription of *fis* is activated when the cell is in exponential growth phase. Fis then activates the genes *aceE*, *aceF*, coding for two of the subunits of the pyruvate dehydrogenase enzyme, *PDH*. It also activates transcription of the succinate dehydrogenase, *SUCDi*, subunits *sdhC*, *sdhD*, *sdhA*, *sdhB* [41]. Fis can also activate *adhE*, which codes for an enzyme that carries out both the acetaldehyde dehydrogenase (acetylating), *ACALD*, and alcohol dehydrogenase (ethanol), *ALCD2x*, reactions [116]. When modeling the balanced steady state growth typical of exponential phase, the state of Fis is always set to true. Other growth phases can be modeled with the full *E. coli* Boolean regulatory model [38].
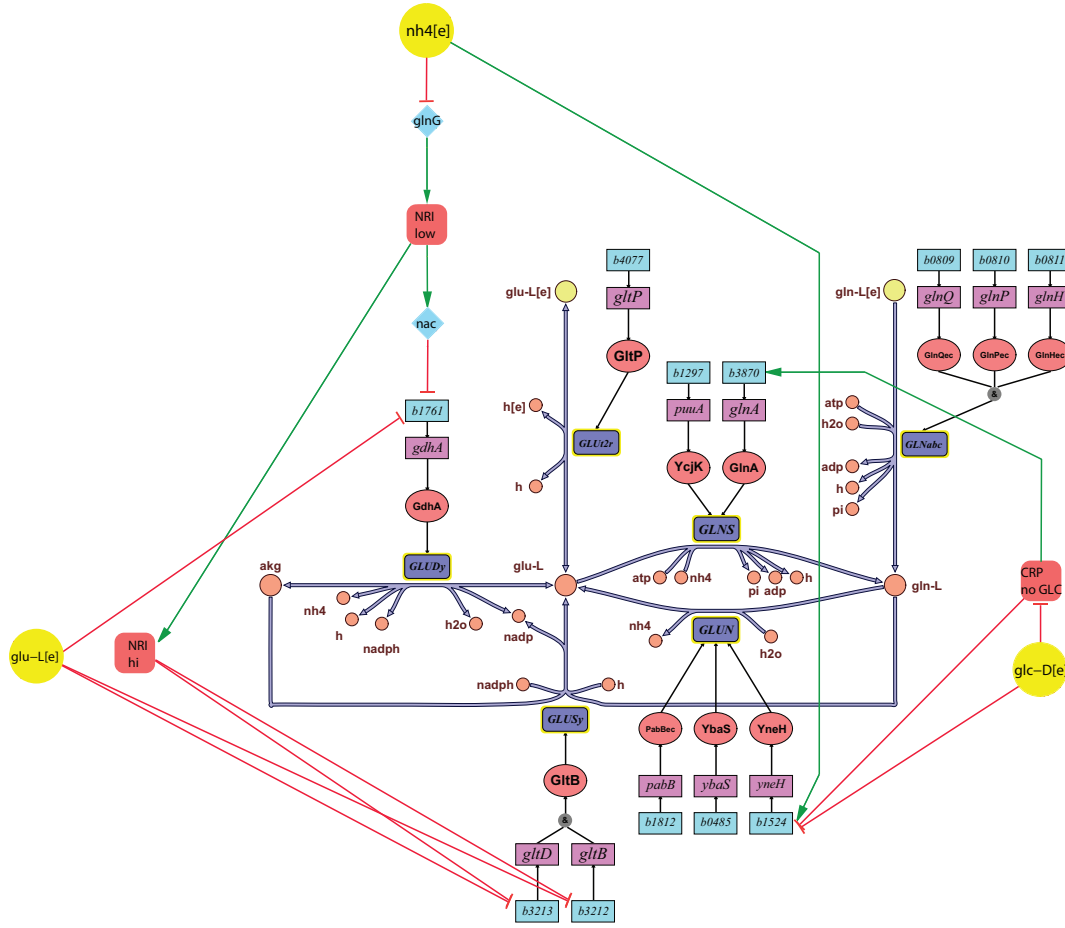
Figure 16: Regulation of nitrogen metabolism in the *E. coli* core model. Extracellular ammonium, nh4[e], activates the low and high level nitrogen responses, *NRI_low* and *NRI_hi*, which, along with extracellular glutamate, glu-L[e], inhibit the reactions glutamate dehydrogenase, *GLUDy*, and glutamate synthase, *GLUSy*. Glutaminase, *GLUN*, is also activated by extracellular ammonium. Extracellular glucose, glc-D[e], through *CRPnoGLC*, inhibits glutamine synthetase, *GLNS*, and glutaminase, *GLUN*.

5.6   Nitrogen Regulation

The response to low nitrogen concentration in *E. coli* is a complex process. There is a fast (low-level) response and a slower (high-level) response, represented in the Boolean regulatory model with the regulatory condition variables *NRI_low* and *NRI_hi*. The low-level response is activated by the transcription factor GlnG. GlnG is activated by low extracellular ammonium, nh4[e], concentration and regulates nitrogen levels by induction or repression of many different genes [108]. The high-level nitrogen response is activated by the low-level response, so *NRI_hi* is always activated after *NRI_low* is activated.

As a whole, the low and high level nitrogen responses conserve nitrogen by decreasing glutamate production. *NRI_low* induces transcription of the transcription factor Nac (nitrogen assimilation control). The product of this gene then represses *gdhA* [26], which codes for the glutamate dehydrogenase enzyme, *GLUDy*, which produces glutamate from 2-oxoglutarate. *NRI_hi* represses the genes *gltB* and *gltD* [121], which code for the subunits of the glutamate synthase enzyme, *GLUSy*. This enzyme produces glutamate through an alternate mechanism, so *NRI_hi* leads to further nitrogen conservation.
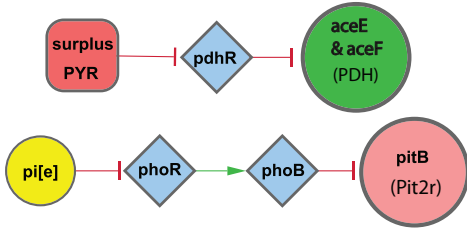
Figure 17: A schematic representation of cytoplasmic pyruvate concentration and phosphorus transport in the *E. coli* core model. Transcription factors are indicated by blue diamonds. The genes *aceE* and *aceF* both code for subunits of the pyruvate dehydrogenase complex, *PDH*. The *pitB* gene codes for the phosphate transporter, *PIt2r*.

5.7 Phosphorus Regulation

Phosphorus uptake is regulated by the two-component system *phoR/phoB* [76]. *phoR* codes for a sensor kinase that is phosphorylated when extracellular inorganic phosphate is not present. The phosphorylated enzyme is activated, and it phosphorylates the transcriptional regulator PhoB. Phosphorylated PhoB then represses the *pitB* gene, which codes for the phosphate transporter, *PIt2r*. As indicated in Figure 17, the regulatory rule for *pitB* is 'NOT PhoB' therefore *pitB* is true when phoB is false, and inorganic phosphate is present. The overall effect of phosphorus regulation is to downregulate the phosphate transport reaction, *PIt2r*, when no extracellular inorganic phosphate is present.

Table 3: Regulatory rules for metabolic genes in the core model.

Gene	Locus	Reaction Abbreviation	Regulatory Rule
aceA	b4015	ICL	(NOT IclR) AND ((NOT ArcA) OR FruR)
aceB	b4014	MALS	(NOT IclR) AND ((NOT ArcA) OR FruR)
aceE	b0114	PDH	(NOT PdhR) OR Fis
aceF	b0115	PDH	(NOT PdhR) OR Fis
adhE	b1241	ACALD and ALCD2x	(NOT o2[e]) OR (NOT (o2[e] AND FruR)) OR Fis
cydA	b0733	CYTBD	(NOT Fnr) OR ArcA
cydB	b0734	CYTBD	(NOT Fnr) OR ArcA
dctA	b3528	FUMt2_2, MALt2_2 and SUCCt2_2	CRPnoGLM AND (NOT ArcA) AND DcuR
focA	b0904	FORt2 and FORti	ArcA OR (Fnr AND CRPnoGLC)
focB	b2492	FORt2 and FORti	ArcA OR (Fnr AND CRPnoGLC)
frdA	b4154	FRD7	Fnr OR DcuR
frdB	b4153	FRD7	Fnr OR DcuR
frdC	b4152	FRD7	Fnr OR DcuR
frdD	b4151	FRD7	Fnr OR DcuR
fumA	b1612	FUM	NOT (ArcA OR Fnr)
fumB	b4122	FUM	Fnr OR CRPnoGLC OR DcuR
fumC	b1611	FUM	NOT ArcA
gdhA	b1761	GLUDy	NOT (Nac OR glu-L[e])
glcA	b2975	D_LACt2	(NOT ArcA) AND GlcC
glcB	b2976	MALS	(NOT ArcA) AND GlcC
glnA	b3870	GLNS	CRPnoGLC
gltB	b3212	GLUSy	(NOT (NRI_hi AND glu-L[e]))
gltD	b3213	GLUSy	(NOT (NRI_hi AND glu-L[e]))
lldP	b3603	D_LACt2	NOT ArcA
manX	b1817	FRUpts2 and GLCpts	CRPnoGLM OR (NOT Mlc)
manY	b1818	FRUpts2 and GLCpts	CRPnoGLM OR (NOT Mlc)
manZ	b1819	FRUpts2 and GLCpts	CRPnoGLM OR (NOT Mlc)
mdh	b3236	MDH	NOT ArcA
nuoA	b2288	NADH16	(NOT (ArcA OR Fnr))
nuoB	b2287	NADH16	(NOT (ArcA OR Fnr))
nuoC	b2286	NADH16	(NOT (ArcA OR Fnr))
nuoE - N	b2276-b2285	NADH16	(NOT (ArcA OR Fnr))
pflA	b0902	PFL	ArcA OR (Fnr AND CRPnoGLC)
pflB	b0903	PFL	ArcA OR (Fnr AND CRPnoGLC)
pflC	b3952	PFL	ArcA OR Fnr
pflD	b3951	PFL	ArcA OR Fnr
pitB	b2987	PIt2r	NOT PhoB
pps	b1702	PPS	FruR
ptsG	b1101	GLCpts	(NOT Mlc) OR (NOT FruR)
pykF	b1676	PYK	NOT FruR
sdhA	b0723	SUCDi	(NOT (ArcA OR Fnr)) OR CRPnoGLC OR Fis
sdhB	b0724	SUCDi	(NOT (ArcA OR Fnr)) OR CRPnoGLC OR Fis
sdhC	b0721	SUCDi	(NOT (ArcA OR Fnr)) OR CRPnoGLC OR Fis
sdhD	b0722	SUCDi	(NOT (ArcA OR Fnr)) OR CRPnoGLC OR Fis
tdcD	b3115	ACKr	CRPnoGLC OR Fnr
tdcE	b3114	PFL	CRPnoGLC OR Fnr
yneH	b1524	GLUN	(NOT glc-D[e]) OR (nh4[e] AND (NOT CRPnoGLC))

Table 4: Regulatory rules for transcriptional regulators and regulatory conditions.

Regulator	Locus	Regulatory Rule
ArcA	b4401	NOT o2[e]
DcuR	b4124	dcuS
DcuS	b4125	succ[e] OR fum[e] OR mal-L[e]
FadR	b1187	glc-D[e] OR (NOT ac[e])
Fis	b3261	Biomass_Ecoli_core_w_GAM
Fnr	b1334	NOT o2[e]
FruR	b0080	NOT surplusFDP
GlcC	b2980	ac[e]
GlnG	b3868	NOT nh4[e]
IclR	b4018	fadR
Mlc	b1594	NOT glc-D[e]
Nac	b1988	NRI_low
PdhR	b0113	NOT surplusPYR
PhoB	b0399	phoR
PhoR	b0400	NOT pi[e]
CRPnoGLC	b3357	NOT glc-D[e]
CRPnoGLM	b3357	NOT (glc-D[e] OR mal-L[e] OR lac-D[e])
NRI_hi		NRI_low
NRI_low		glnG
surplusFDP		((NOT FBP) AND (NOT (TKT2 OR TALA OR PGI))) OR fru[e]
surplusPYR		(NOT (ME2 OR ME1)) AND (NOT (GLCpts OR PYK OR PFK OR LDH_D OR SUCCt2_2))

6 Uses of Metabolic Models

Once an accurate metabolic reconstruction is converted into a computational model, it may be used for a growing number of applications. Such models have been utilized to address a broad spectrum of basic and practical applications in five main categories: studies of evolutionary processes, analysis of network properties, interpretation of phenotypic screens, model-directed discovery and metabolic engineering (Figure 18). Metabolic reconstructions are a common denominator in the systems analysis of metabolic functions. As evident from Sections 4 and 5, a wealth of biological knowledge is encoded in a manually curated network reconstruction. Section 6.1 describes how this knowledge may then be used to predict capabilities of a network which emerge from the interaction of multiple components. Similarly, the knowledge encoded in a model may be used to study the process of bacterial evolution [58]. Applications include the interpretation of experimental adaptive evolution [141], horizontal gene transfer [141, 142] and evolution to minimal metabolic networks [143]. Network reconstructions also provide a context for integration of high throughput data from multiple complementary experiments, as described in Section 6.2. Discovery of the biochemical function of previously uncharacterized genes using metabolic models is discussed in Section 6.3. A growing application of genome scale metabolic reconstructions is the prediction of optimal mutant strains in synthetic biological and engineering settings, Section 6.4.

6.1 Understanding Metabolic Capabilities

Flux balance analysis and other constraint-based methods can be used to analyze the capabilities of a metabolic network [131]. In flux balance analysis, the biomass reaction can be optimized (maximized) using linear programming software to simulate growth [192]. The result is an *in silico* prediction of steady state flux through each reaction in the model including a prediction of the maximum balanced growth rate of the cell. Growth can be simulated under many different conditions, such as aerobic or anaerobic conditions, or growth on glucose or other substrates. Different conditions are simulated by changing the constraints on the exchange reactions. Exchange reactions, at the boundary of the model and the environment, act as sources or sinks of substrate and waste metabolites (Figure 4). For example, anaerobic conditions are simulated when the lower bound of the O_2 exchange reaction, *EX_o2(e)*, is constrained to zero flux, allowing no O_2 to enter the system. To simulate different media, the exchange reactions for metabolites present in the media are constrained to have a lower bound equal to their desired uptake rate, and all metabolites not present will have exchange reactions with lower bounds constrained to zero.

Growth was simulated using the core *E. coli* metabolic model (without regulation) with glucose as the only organic substrate, under both aerobic and anaerobic conditions, and the resulting flux distributions are shown in Figure 19. Under aerobic conditions with a glucose uptake rate (*EX_glc(e)*) of 10 mmol gDW^{-1} hr^{-1}, the growth rate is 0.87 hr^{-1}. The flux through the electron transport chain is high, and no organic by-products are secreted. Under anaerobic conditions with the same glucose uptake rate, the growth rate is 0.21 hr^{-1}. The electron transport chain and most of the TCA cycle are not used at all, and formate, acetate, and ethanol are all secreted. When the regulatory model is combined with the metabolic model, the reactions *FORt2*, *FORti*, *FUMt2_2*, *ICL*, *MALS*, *MALt2_2*, *PFL*, and *SUCCt2_2* are inactivated under aerobic conditions when growing on glucose. However, since none of these reactions are used in the optimal flux distribution for growth on glucose, the growth rate is not affected. Under anaerobic conditions, the reactions *D_LACt2*, *FUMt2_2*, *ICL*, *MALS*, *MALt2_2*, *MDH*, *NADH16*, and *SUCCt2_2* are inactivated. These reactions are also not part of the optimal flux distribution, so anaerobic growth rate is still 0.21 hr^{-1}. These results highlight the predictive capacity of assuming an optimal growth rate, given a particular environmental condition.

The maximum yields of important cofactors such as ATP, NADH, and NADPH can also be determined using flux balance analysis [190]. By constraining the glucose uptake rate to exactly -1 mmol gDW^{-1} hr^{-1} and setting the ATP maintenance reaction (*ATPM*) as the objective to be maximized, the yield of ATP from glucose can be calculated. *ATPM* is a stoichiometrically balanced reaction that drains ATP from the network. To determine the maximum yields of NADH and NADPH, similar balanced drain reactions must be added to the network and set as objectives. The maximum yields of these cofactors are given in Table 5. These yields are limited by the balancing of protons and by the stoichiometry of the network. See Table 6 for examples of COBRA Toolbox [169] commands for performing these simulations.
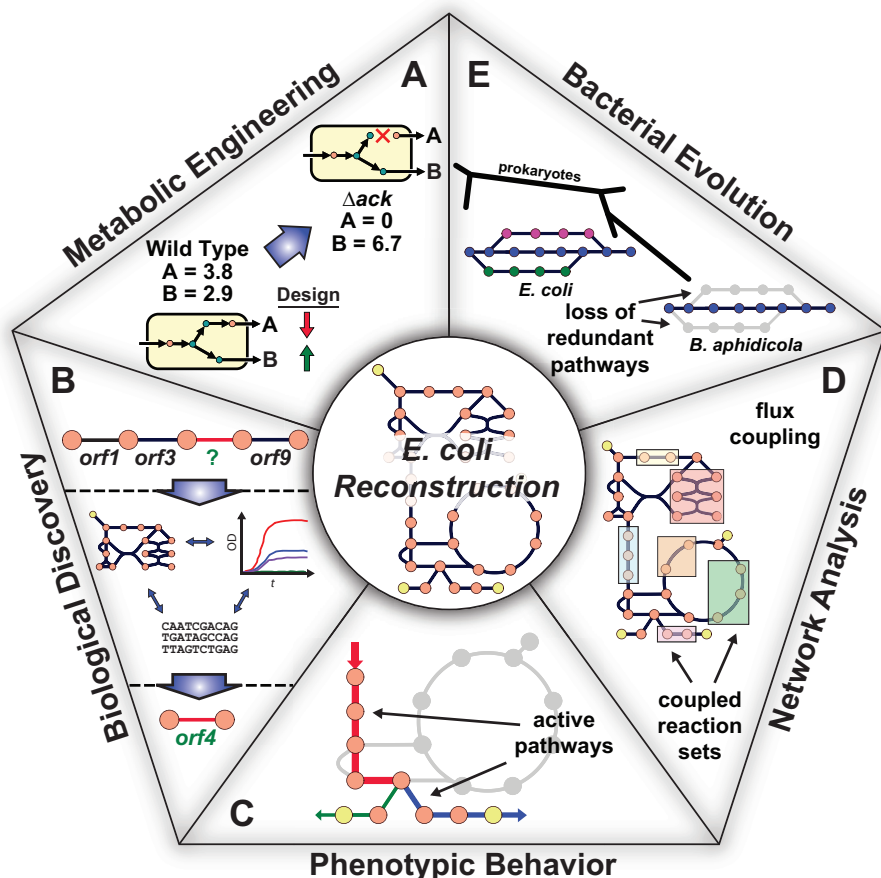
Figure 18: Metabolic models have practical applications in metabolic engineering, model-directed discovery, interpretations of phenotypic screens, analysis of network properties and studies of evolutionary processes. (A) Metabolic engineering, e.g. predicting the effect from a loss-of-function mutation as part of *in silico* strain design to overproduce desired products. (B) Biological discovery involves prospective use of biochemical and genetic information included in the metabolic network along with additional data types to drive discovery, e.g. predicting genes responsible for orphan reactions. (C) Phenotypic studies, e.g. computational analyses of gene, metabolite and reaction essentiality. (D) Network analysis, e.g. finding coupled reaction fluxes across different growth conditions. (E) Evolutionary studies have used metabolic models in order to interpret adaptive evolution events, horizontal gene transfer and minimal metabolic network evolution. From [58].

Table 5: Maximum cofactor production from glucose, aerobically. PPS is the percentage of the glucose uptake flux that enters the pentose phosphate shunt.

Cofactor	Yield	PPS	Constraint
ATP	17.5	0%	H^+
NADH	10.0	0%	Energy, Stoichiometry
NADPH	8.778	300%	Energy, Stoichiometry
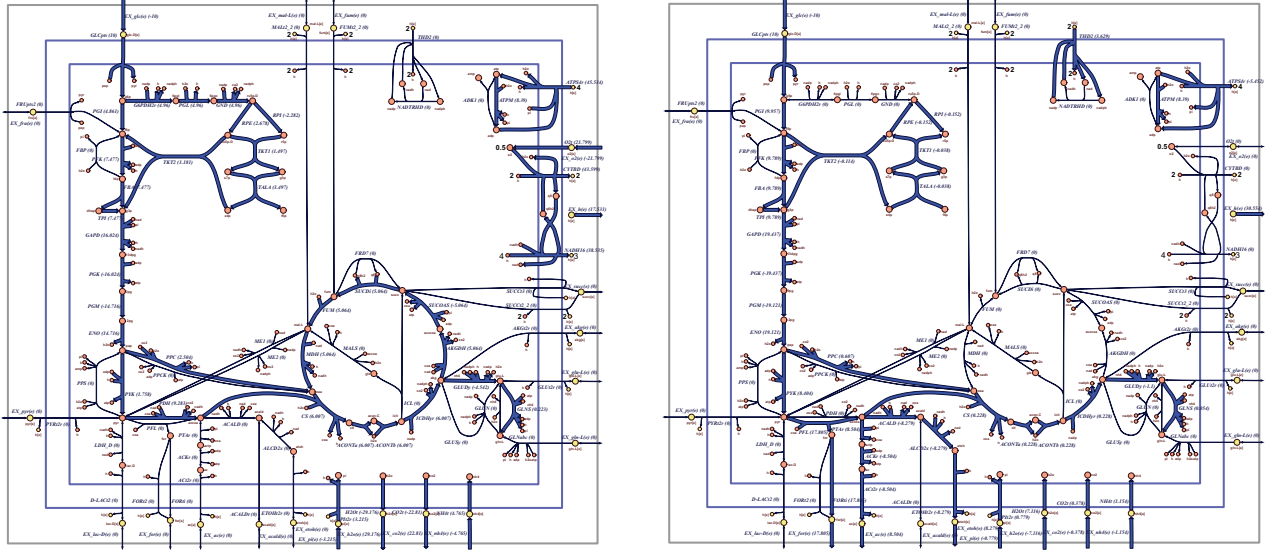
Figure 19: Growth on glucose under aerobic (left) and anaerobic (right) conditions. Reactions with thick blue arrows have nonzero fluxes, and reactions with thin black arrows carry zero flux.

Table 6: Example COBRA Toolbox commands for performing flux balance analysis [169].

Action	Command
change bounds for anaerobic growth	model = changeRxnBounds(model,'EX_o2(e)',0,'l');
change bounds for aerobic growth	model = changeRxnBounds(model,'EX_o2(e)',-1000,'l');
change glucose uptake rate to 10 mmol gDW^{-1} hr^{-1}	model = changeRxnBounds(model,'EX_glc(e)',-10,'l');
simulate maximum growth by FBA	solution = optimizeCbModel(model);
simulate maximum growth of regulated model	[FBAsols,DRgenes,constrainedRxns,cycleStart,states] = optimizeRegModel(model);
change objective to maximum ATP yield	model = changeObjective(model,'ATPM');
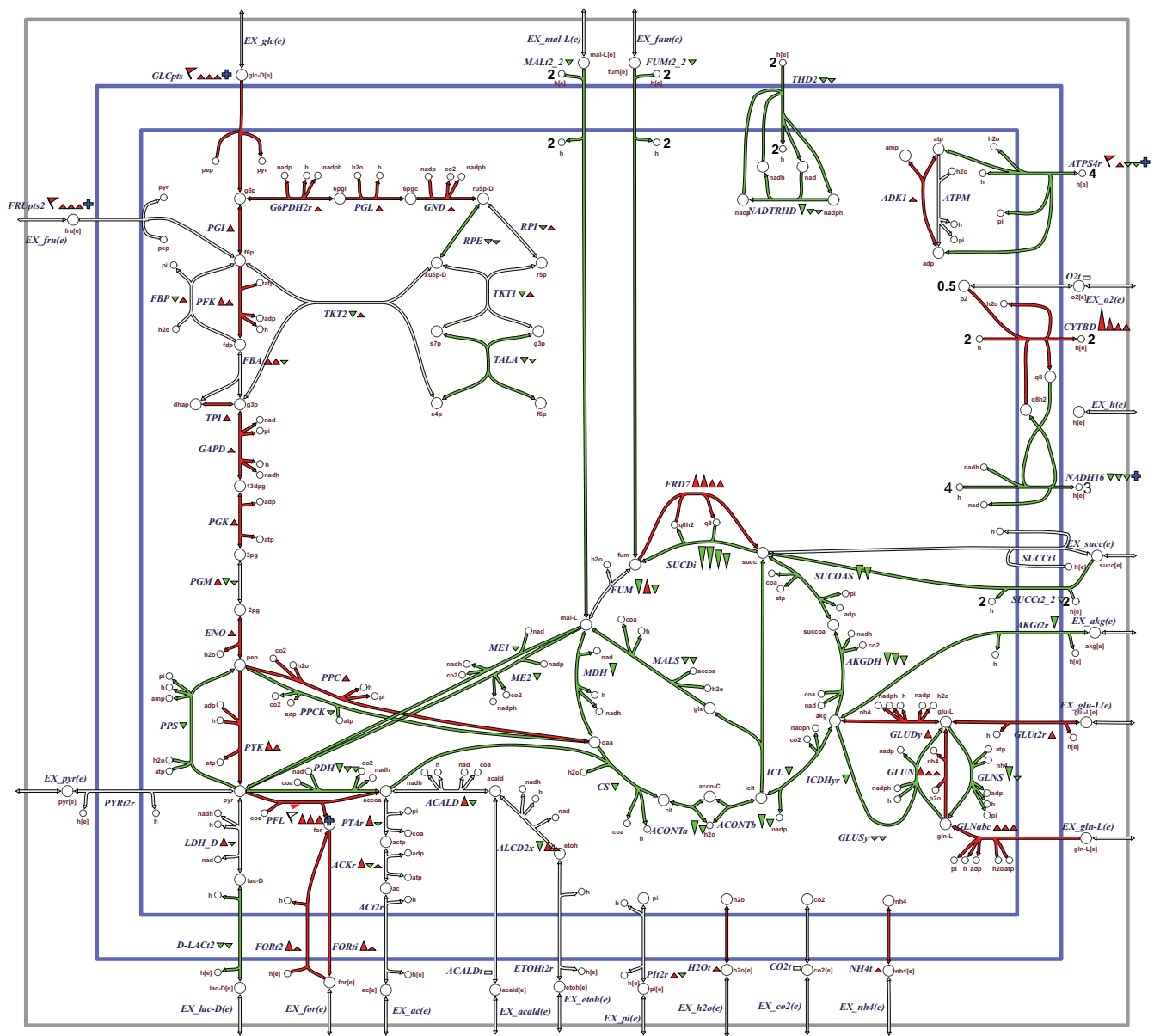
Figure 20: Gene expression data comparing anaerobic and aerobic growth mapped to the core *E. coli* model. Triangles next to reaction names represent the genes associated with each reaction. Red genes and reactions are upregulated under anaerobic conditions and green genes and reactions are downregulated.

6.2 High Throughput Data Analysis

A metabolic network can be used as a tool to analyze different types of high throughput data, including gene expression data. This data can be mapped to the network, providing a context in which to interpret the results. In this example, gene expression data for *E. coli* comparing aerobic growth to anaerobic growth on glucose [38] was mapped to the core *E. coli* model (Figure 20). In this figure, the triangles next to each reaction name represent the genes associated with each reaction. The gene-protein-reaction interactions in the core model were used to determine which reactions were upregulated or downregulated based on the gene regulation. When plotted against the network map, it is clear that glycolysis and the oxidative branch of the pentose phosphate pathway are upregulated, while the TCA cycle is downregulated. These results do not fully agree with the fluxes predicted by flux balance analysis in Section 6.1, which is expected as gene expression and metabolic fluxes are not trivially quantitatively related. Nevertheless, such experimental data can be used to identify repressed genes and hence regulatory rules for a particular condition. Setting the flux to zero for reactions catalyzed by repressed genes, allows for more biochemically realistic flux balance analysis.

A recently developed computational algorithm called GIMME [9] can be used to interpret gene expression data using constraint-based models. This algorithm uses data from microarrays to form context-specific models by adding reactions for which the associated gene expression levels are above a specified threshold. Additional reactions are then added to meet a known cellular function such as growth or secretion of a metabolite for the conditions in which the microarray data was gathered. These reactions are assigned an inconsistency score based on their agreement with the gene expression data. The GIMME algorithm and the *E. coli* core metabolic model were used along with 170 *E. coli* microarrays covering a large number of growth conditions and genetic perturbations to find which conditions are most consistent with the network topology needed for secretion of 12 different organic metabolites. This analysis reveals which conditions are most amenable to production of certain metabolites. Figure 21 shows the consistency scores for each context-specific model created. As we would expect, the common anaerobic secretion products such as ethanol, acetate, and formate [192] are more consistent with the arrays from anaerobically grown strains than with aerobic strains.

6.3 Discovery

Metabolic and regulatory models can be used to facilitate biological discovery. Computational predictions can be compared to experimental measurements under many different conditions, and when there are disagreements, it is because the model may be incomplete or incorrect in some way. These disagreements can be analyzed, aiding in the discovery of new biological features. Usually, the most up to date genome scale models are used for discovery purposes. The core *E. coli* model is not suitable for discovery because it intentionally lacks most known metabolic reactions and has a limited scope.

Most often, models have been used to characterize unknown ORFs in an organism's genome. An algorithm that combines computational analysis of the genome scale metabolic model *i*JR904 [152] with experimental growth phenotype screening was recently used to identify eight unknown ORFs in *E. coli* [151]. First, *E. coli* was grown under many different minimal media conditions in a high throughput screen with different carbon and nitrogen sources. The growth phenotypes were qualitatively compared to growth phenotypes predicted by flux balance analysis using the genome scale *E. coli* metabolic model. The comparison highlighted 50 conditions were growth is possible *in vivo* but not possible *in silico*. For every growth phenotype that could not be explained by the model, an optimization algorithm was used to determine the minimum number of reactions that needed to be added to the model from a universal database of reactions (from KEGG [92]) to make growth possible.

In this optimization algorithm, the stoichiometric matrix, representing *E. coli* reactions and metabolites, was used in addition to a second matrix, containing all the reactions from the universal database, and a third matrix, containing exchange reactions for metabolites not included in *i*JR904. For 26 of the 50 tested conditions, at least one set of universal database and/or exchange reactions were predicted to be necessary for *in silico* growth. In some growth conditions, up to 15 different sets of reactions were predicted to each allow *in silico* growth. These new reaction sets served as hypotheses for the identities of unknown ORFs, and several of these hypotheses were experimentally investigated in more detail.

The algorithm predicted that transport reactions needed to be added to the model to allow growth on propionate and 5-keto-D-gluconate. Eight genes that were predicted to be transporters, by homology search, were tested experimentally. It was found that *E. coli* cannot grow on propionate without the gene *putP*, and cannot grow on 5-keto-D-gluconate without *idnT*, indicating that these genes do code for transporters. The algorithm also predicted
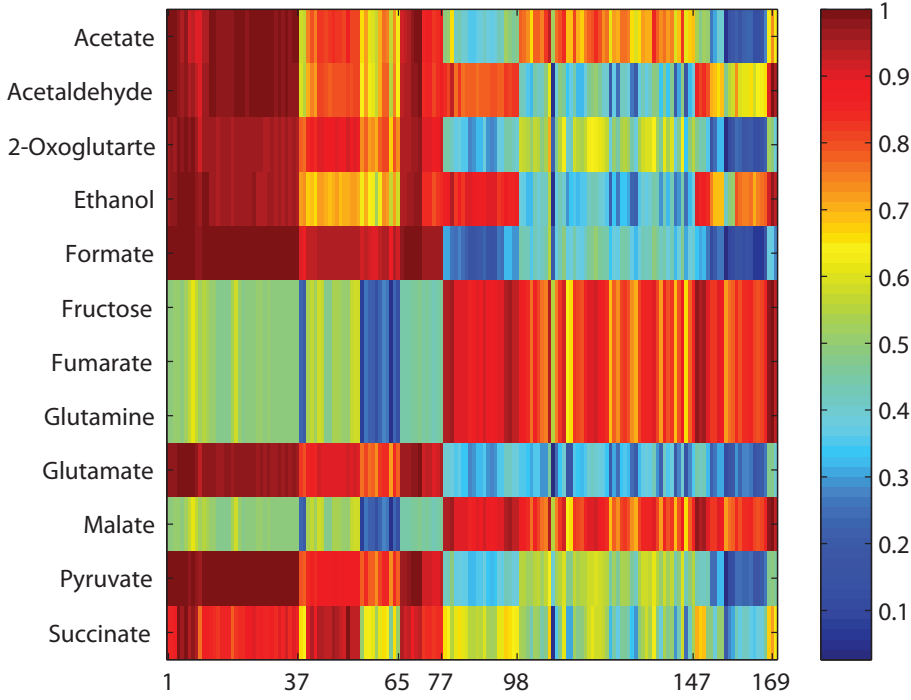
Figure 21: Normalized consistency scores for the secretion of 12 different products from the core *E. coli* model as computed from 170 microarrays [9]. A high consistency score is given when the microarray data is in good agreement with the network topology necessary for the secretion of the product of interest. Each row has been normalized so that consistency scores within rows can be compared, but comparisons cannot be made between rows. Arrays 1-37 are wild type regulatory gene-deletion strains of *E. coli* grown on glucose under anaerobic conditions. Arrays 38-65 represent strains in which metabolic genes were knocked out followed by anaerobic growth on glucose. Array 66-77 are from regulatory gene-deletion mutants and wild type strains grown anaerobically on glucose with nitrate used as the terminal electron acceptor. Arrays 78-98, 148-170 glucose aerobic wild type and gene deletions of both regulatory proteins and metabolic enzymes. Arrays 99-147 are from wild type strains which were under aerobically in lactate. Consistency scores are higher for production of most of the anaerobic fermentation products when using microarray data from anaerobic conditions, indicating that these conditions are more likely to lead to production of these products.

Table 7: Orphan reactions in the core *E. coli* reconstruction. Orphan reactions are reactions that are known to exist but are catalyzed by enzymes without an assigned ORF.

Abbreviation	Full Name	Equation
ACt2r	acetate reversible transport via proton symport	ac[e] + h[e] $\rightleftharpoons$ ac + h
ETOHt2r	ethanol reversible transport via proton symport	etoh[e] + h[e] $\rightleftharpoons$ etoh + h
PYRt2r	pyruvate reversible transport via proton symport	h[e] + pyr[e] $\rightleftharpoons$ h + pyr
SUCCt3	succinate transport out via proton antiport	h[e] + succ $\rightleftharpoons$ h + succ[e]

that growth on D-malate is possible with a transport reaction and enzymatic conversion to succinate. *E. coli* strains lacking the genes *dctA*, *yeaT*, or *yeaU* are unable to grow on D-malate. Through analysis of gene expression data and biochemical assays, it was found that *dctA* is the transporter, *yeaU* is the enzyme, and *yeaT* is a regulator of the enzyme.

In this same study, genes involved in growth on galactonate γ-lactone were identified through gene expression data, even though the algorithm was unable to identify the necessary reactions. The genome scale model was still useful, however, because it demonstrated that no known genes were responsible for growth on this substrate. Orphan reactions, reactions that are known to exist but are catalyzed by unknown enzymes, provide another way to identify unknown ORFs in microbial organisms. In the core *E. coli* model, there are four orphan transport reactions, listed in Table 7. Algorithms that consider the phylogenetic profiles [27] or coexpression and clustering on the genome [99] of the reactions adjacent the orphan reactions in the network have been shown to be successful in assigning tentative ORFs to such reactions.

Recently, an entirely new pathway for pyrimidine catabolism was discovered in *E. coli* using a combination of a subsystem approach to genome annotation and experimental validation [109]. The subsystem approach to genome annotation and pathway analysis allows integrated knowledge of existing biochemical network structure and genomic structure to be projected across the entire collection of diverse species with completely sequenced genomes [129]. In addition to establishing which organisms implement one or the other functional variants of a subsystem this approach helps to reveal gaps in knowledge (missing genes) and potential new players (predicted genes) [128].

6.4   Synthetic Biology and Metabolic Engineering

Because of the predictive capabilities of constraint-based metabolic models, they can be used in synthetic biology and metabolic engineering applications [58, 8]. In particular, constraint-based models are useful because they can predict when a particular metabolite will be overproduced and secreted. Gene knockouts and knockins can be simulated by removing or adding reactions to the network, and the behavior of the modified network can be predicted by flux balance analysis. These model based predictions are more accurate than predictions based in intuition and knowledge of gene functions because flux balance analysis considers the complicated interacting effects a knockout has on all pathways simultaneously.

Growth coupled designs are a particularly promising class of metabolically engineered strains. When the product of interest in one of these strains is produced at a higher rate, the growth rate of the strain also increases. This is unlike most strain designs, in which the growth rate decreases as flux is diverted to a metabolic by-product. Growth coupled designs are evolutionarily stable, as their production rates actually increase as mutations that increase growth rate accumulate. The growth coupling of designs can be visualized using a production envelope, a graph that shows the solution space of a model in the dimensions of growth rate and the exchange reaction of a particular metabolite. When the metabolite is growth coupled, its minimum exchange rate at the maximum possible growth rate will be greater than zero. OptKnock is a bilevel linear programming optimization algorithm that uses a constraint-based model to identify sets of reaction knockouts that couple the production of a metabolite to growth [23]. It seeks to simultaneously maximize growth rate and product secretion rate. This algorithm was used to identify growth coupled designs for many metabolites in the core *E. coli* model under anaerobic conditions, and the production envelopes of some growth coupled designs are shown in Figure 22.
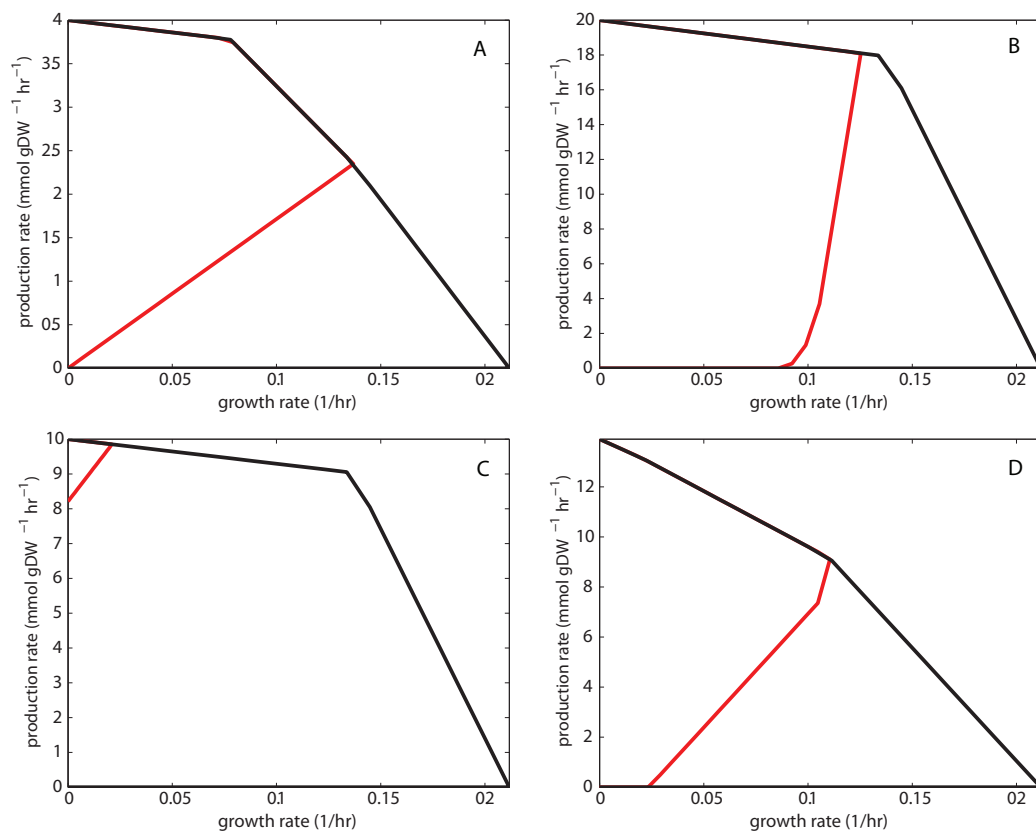
Figure 22: Production envelopes for OptKnock designs for the production of 2-oxoglutarate (A), D-lactate (B), pyruvate (C), and succinate (D). Black envelopes are wild-type, and red envelopes are for knockout strains. Growth is possible at any point inside the production envelope, but over time *E. coli* will adapt to the highest possible growth rate [62].

7 Summary

Currently, the wealth of biochemical information exceeds the scope and depth of even the largest network reconstructions. The scope of the *E. coli* core model represents perhaps the most well characterized fraction of the latest *E. coli* genome scale metabolic model [56]. Nevertheless, the results of future genetic studies which establish the genes corresponding to orphan reactions need to be incorporated as part of an iterative cycle of development [58]. The regulatory network for *E. coli* core metabolism is rapidly being discovered using high throughput protein-DNA binding assays [31, 101, 194, 203]. The incorporation of this new knowledge, with a more comprehensive study of the biochemical literature on core metabolic regulation, would result in a significant expansion in the number and complexity of Boolean regulatory rules in the *E. coli* core model.

Biochemical network reconstructions are biochemically, genetically and genomically structured databases. These reconstructions rely heavily on information retrieved from biochemical characterization of reactions and their substrates. Genetic studies are essential to identify the open reading frame(s) which encode the enzyme(s) responsible for each catalytic reaction, giving rise to gene-protein-reaction associations. More broadly, species specific biochemical and molecular biological studies are essential to provide a sound experimental basis for the components and reactions which are the key elements in any reconstruction. The utility of biochemical network reconstructions is driven by the fact that they can be transformed into a computational model. In turn the computational model can be applied to address an increasingly wide variety of biological questions [58], including bacterial evolution [143], analysis of network properties [22, 112, 162], study of phenotypic behavior [51, 148], biological discovery [38, 77, 151], and metabolic engineering [23, 132, 135].

In the future, we can look forward to complete reconstructions of all known biochemical processes in *E. coli* and *Salmonella.* Such models ideally serve as a structured self consistent representations of our knowledge. Inevitably they will grow in scope and more details will need to be added such that model predictions finally reach par with experimental observations. Even now, in certain situations, model predictions not only provide qualitatively and quantitatively accurate predictions of experiments, but can also be used to suggest profitable avenues for experimental confirmation. Even though reconstructions continue to grow in size and scope, fundamentally, they will still be systems of biochemical reactions, just like the core *E. coli* model, but with the rest of the cell built around it. Ultimately, the full utility of computational models will be realized at the fingertips of biological domain specific experts. However, if all other domain specific experts have contributed to the same model, as each expert probes her or his own area of biological expertise *in silico*, they can do so in the knowledge that their *in silico* predictions automatically account for others' expertise, through the biochemically, genetically and genomically structured model. We hope that this introduction to the core *E. coli* model brings this era a little closer.

8 Supplemental Files

The following supplemental files can be found online at *http://systemsbiology.ucsd.edu/*:

ecoli_core_model.xls A Microsoft Excel file that describes the reactions and metabolites in the core *E. coli* model. This file includes the full S matrix.

core_regulatory_rules.xls A Microsoft Excel file that gives the Boolean regulatory rules for every gene in the regulated model.

ecoli_core_model.mat A MATLAB data file that contains the core *E. coli* model (without regulation) in a format that can be used with the COBRA Toolbox (see [169] for details on this toolbox).

modelReg.mat A MATLAB data file that contains the regulated core *E. coli* model for the COBRA Toolbox.

optimizeRegModel.m A new COBRA Toolbox function that is needed to run simulations using the regulated model. This function uses flux balance analysis to determine the state of the network while considering regulatory constraints.

dynamicRFBA.m A new COBRA Toolbox function that performs dynamic flux balance analysis using the regulated model (see [39] for details on dynamic rFBA).

solveBooleanRegModel.m A COBRA Toolbox function that is required by both optimizeRegModel.m and dynamicRFBA.m.

ecoli_core_model.xml An SBML file for the core *E. coli* model without regulation. COBRA Toolbox users with the SBML Toolbox can load this file using the function 'readCbModel'.

Acknowledgment

References

[1] **Alberty, R. A.**, 2003. Thermodynamics of Biochemical Reactions. Wiley.

[2] **Alefounder, P. R., Baldwin, S. A., Perham, R. N., and Short, N. J.**, 1989. Cloning, sequence analysis and over-expression of the gene for the class II fructose 1,6-bisphosphate aldolase of *Escherichia coli*. Biochem J **257**:529–534.

[3] **Almaas, E., Kovács, B., Vicsek, T., Oltvai, Z. N., and Barabási, A.-L.**, 2004. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. Nature **427**:839–843.

[4] **Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G.**, 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet **25**:25–29.

[5] **Bairoch, A.**, 2000. The ENZYME database in 2000. Nucleic Acids Res **28**:304–305.

[6] **Baldwin, S. A. and Perham, R. N.**, 1978. Novel kinetic and structural properties of the class-I D-fructose 1,6-bisphosphate aldolase from *Escherichia coli* (Crookes' strain). Biochem J **169**:643–652.

[7] **Barrett, C. L., Herring, C. D., Reed, J. L., and Palsson, B. Ø.**, 2005. The global transcriptional regulatory network for metabolism in *Escherichia coli* exhibits few dominant functional states. Proc Natl Acad Sci U S A **102**:19103–19108.

[8] **Barrett, C. L., Kim, T. Y., Kim, H. U., Palsson, B. Ø., and Lee, S. Y.**, 2006. Systems biology as a foundation for genome-scale synthetic biology. Curr Opin Biotechnol **17**:488–492.

[9] **Becker, S. A. and Palsson, B. Ø.**, 2008. Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol **4**:e1000082.

[10] **Bell, P. J., Andrews, S. C., Sivak, M. N., and Guest, J. R.**, 1989. Nucleotide sequence of the FNR-regulated fumarase gene (fumB) of *Escherichia coli* K-12. J Bacteriol **171**:3494–3503.

[11] **Berry, M. B., Bae, E., Bilderback, T. R., Glaser, M., and Phillips, G. N.**, 2006. Crystal structure of ADP/AMP complex of *Escherichia coli* adenylate kinase. Proteins **62**:555–556.

[12] **Bizouarn, T., Fjellström, O., Meuller, J., Axelsson, M., Bergkvist, A., Johansson, C., Göran Karlsson, B., and Rydström, J.**, 2000. Proton translocating nicotinamide nucleotide transhydrogenase from *E. coli*. Mechanism of action deduced from its structural and catalytic properties. Biochim Biophys Acta **1457**:211–228.

[13] **Bledig, S. A., Ramseier, T. M., and Saier, M. H.**, 1996. Frur mediates catabolite activation of pyruvate kinase (pykF) gene expression in *Escherichia coli*. J Bacteriol **178**:280–283.

[14] **Bongaerts, J., Zoske, S., Weidner, U., and Unden, G.**, 1995. Transcriptional regulation of the proton translocating NADH dehydrogenase genes (nuoA-N) of *Escherichia coli* by electron acceptors, electron donors and gene regulators. Mol Microbiol **16**:521–534.

[15] **Bonneau, R., Facciotti, M. T., Reiss, D. J., Schmid, A. K., Pan, M., Kaur, A., Thorsson, V., Shannon, P., Johnson, M. H., Bare, J. C., Longabaugh, W., Vuthoori, M., Whitehead, K., Madar, A., Suzuki, L., Mori, T., Chang, D.-E., Diruggiero, J., Johnson, C. H., Hood, L., and Baliga, N. S.**, 2007. A predictive model for transcriptional control of physiology in a free living cell. Cell **131**:1354–1365.

[16] **Botsford, J. L. and Harman, J. G.**, 1992. Cyclic AMP in prokaryotes. Microbiol Rev **56**:100–122.

[17] **Branlant, G. and Branlant, C.**, 1985. Nucleotide sequence of the *Escherichia coli* gap gene. Different evolutionary behavior of the NAD+-binding domain and of the catalytic domain of D-glyceraldehyde-3-phosphate dehydrogenase. Eur J Biochem **150**:61–66.

[18] **Breitling, R., Vitkup, D., and Barrett, M. P.**, 2008. New surveyor tools for charting microbial metabolic maps. Nat Rev Microbiol **6**:156–161.

[19] **Bridger, W. A., Wolodko, W. T., Henning, W., Upton, C., Majumdar, R., and Williams, S. P.**, 1987. The subunits of succinyl-coenzyme A synthetase–function and assembly. Biochem Soc Symp **54**:103–111.

[20] **Brock, M., Maerker, C., Schütz, A., Völker, U., and Buckel, W.**, 2002. Oxidation of propionate to pyruvate in *Escherichia coli*. Involvement of methylcitrate dehydratase and aconitase. Eur J Biochem **269**:6184–6194.

[21] **Brune, M., Schumann, R., and Wittinghofer, F.**, 1985. Cloning and sequencing of the adenylate kinase gene (adk) of *Escherichia coli*. Nucleic Acids Res **13**:7139–7151.

[22] **Burgard, A. P., Nikolaev, E. V., Schilling, C. H., and Maranas, C. D.**, 2004. Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Res **14**:301–312.

[23] **Burgard, A. P., Pharkya, P., and Maranas, C. D.**, 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng **84**:647–657.

[24] **Burke, W. F., Johanson, R. A., and Reeves, H. C.**, 1974. NADP+-specific isocitrate dehydrogenase of *Escherichia coli*. II. Subunit structure. Biochim Biophys Acta **351**:333–340.

[25] **Cain, B. D. and Simoni, R. D.**, 1989. Proton translocation by the F1F0ATPase of *Escherichia coli*. Mutagenic analysis of the a subunit. J Biol Chem **264**:3292–3300.

[26] **Camarena, L., Poggio, S., García, N., and Osorio, A.**, 1998. Transcriptional repression of gdhA in *Escherichia coli* is mediated by the Nac protein. FEMS Microbiol Lett **167**:51–56.

[27] **Chen, L. and Vitkup, D.**, 2006. Predicting genes for orphan metabolic activities using phylogenetic profiles. Genome Biol **7**:R17.

[28] **Cho, B. K., Barrett, C. L., Knight, E. M., Park, Y. S., and Palsson, B. Ø.**, 2008. Genome-scale reconstruction of Lrp-regulatory network in *Escherichia coli*. Proc Natl Acad Sci U S A **accepted**.

[29] **Cho, B.-K., Knight, E. M., Barrett, C. L., and Palsson, B. Ø.**, 2008. Genome-wide analysis of Fis binding in *Escherichia coli* indicates a causative role for A-/AT-tracts. Genome Res **18**:900–910.

[30] **Cho, B. K., Knight, E. M., and Palsson, B. Ø.**, 2006. Transcriptional regulation of the fad regulon genes of *Escherichia coli* by ArcA. Microbiology **152**:2207–2219.

[31] **Cho, B. K., Knight, E. M., and Palsson, B. Ø.**, 2008. Genome-wide Identification of Protein Binding Locations using Chromatin Immunoprecipitation coupled with Microarrays. Methods in Molecular Biology **439**:131–145.

[32] **Christie, K. R., Weng, S., Balakrishnan, R., Costanzo, M. C., Dolinski, K., Dwight, S. S., Engel, S. R., Feierbach, B., Fisk, D. G., Hirschman, J. E., Hong, E. L., Issel-Tarver, L., Nash, R., Sethuraman, A., Starr, B., Theesfeld, C. L., Andrada, R., Binkley, G., Dong, Q., Lane, C., Schroeder, M., Botstein, D., and Cherry, J. M.**, 2004. Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. Nucleic Acids Res **32**:D311–D314.

[33] **Claudel-Renard, C., Chevalet, C., Faraut, T., and Kahn, D.**, 2003. Enzyme-specific profiles for genome annotation: PRIAM. Nucleic Acids Res **31**:6633–6639.

[34] **Cole, S. T., Condon, C., Lemire, B. D., and Weiner, J. H.**, 1985. Molecular biology, biochemistry and bioenergetics of fumarate reductase, a complex membrane-bound iron-sulfur flavoenzyme of *Escherichia coli*. Biochim Biophys Acta **811**:381–403.

[35] **Condon, C., Cammack, R., Patil, D. S., and Owen, P.**, 1985. The succinate dehydrogenase of *Escherichia coli*. Immunochemical resolution and biophysical characterization of a 4-subunit enzyme complex. J Biol Chem **260**:9427–9434.

[36] **Cooper, R. A. and Kornberg, H. L.**, 1965. Net formation of phosphoenolpyruvate from pyruvate by *Escherichia coli*. Biochim Biophys Acta **104**:618–620.

[37] **Cortay, J. C., Nègre, D., Galinier, A., Duclos, B., Perrière, G., and Cozzone, A. J.**, 1991. Regulation of the acetate operon in *Escherichia coli*: purification and functional characterization of the IclR repressor. EMBO J **10**:675–679.

[38] **Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J., and Palsson, B. Ø.**, 2004. Integrating high-throughput and computational data elucidates bacterial networks. Nature **429**:92–96.

[39] **Covert, M. W., Schilling, C. H., and Palsson, B. Ø.**, 2001. Regulation of gene expression in flux balance models of metabolism. J Theor Biol **213**:73–88.

[40] **Csonka, L. N. and Fraenkel, D. G.**, 1977. Pathways of NADPH formation in *Escherichia coli*. J Biol Chem **252**:3382–3391.

[41] **Cunningham, L. and Guest, J. R.**, 1998. Transcription and transcript processing in the sdhCDAB-sucABCD operon of *Escherichia coli*. Microbiology **144 (Pt 8)**:2113–2123.

[42] **Daldal, F.**, 1984. Nucleotide sequence of gene pfkB encoding the minor phosphofructokinase of *Escherichia coli* K-12. Gene **28**:337–342.

[43] **Daldal, F., Babul, J., Guixé, V., and Fraenkel, D. G.**, 1982. An alteration in phosphofructokinase 2 of *Escherichia coli* which impairs gluconeogenic growth and improves growth on sugars. Eur J Biochem **126**:373–379.

[44] **Davies, S. J., Golby, P., Omrani, D., Broad, S. A., Harrington, V. L., Guest, J. R., Kelly, D. J., and Andrews, S. C.**, 1999. Inactivation and regulation of the aerobic C(4)-dicarboxylate transport (dctA) gene of *Escherichia coli*. J Bacteriol **181**:5624–5635.

[45] **DeJongh, M., Formsma, K., Boillot, P., Gould, J., Rycenga, M., and Best, A.**, 2007. Toward the automated generation of genome-scale metabolic networks in the SEED. BMC Bioinformatics **8**:139.

[46] **Delbaere, L. T. J., Sudom, A. M., Prasad, L., Leduc, Y., and Goldie, H.**, 2004. Structure/function studies of phosphoryl transfer by phosphoenolpyruvate carboxykinase. Biochim Biophys Acta **1697**:271–278.

[47] **Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., Srivas, R., , and Palsson, B. Ø.**, 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci U S A **104**:1777–1782.

[48] **Duarte, N. C., Herrgård, M. J., and Palsson, B. Ø.**, 2004. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. Genome Res **14**:1298–1309.

[49] **Dym, O., Pratt, E. A., Ho, C., and Eisenberg, D.**, 2000. The crystal structure of D-lactate dehydrogenase, a peripheral membrane respiratory enzyme. Proc Natl Acad Sci U S A **97**:9413–9418.

[50] **Eberstadt, M., Grdadolnik, S. G., Gemmecker, G., Kessler, H., Buhr, A., and Erni, B.**, 1996. Solution structure of the IIB domain of the glucose transporter of *Escherichia coli*. Biochemistry **35**:11286–11292.

[51] **Edwards, J. S., Ibarra, R. U., and Palsson, B. Ø.**, 2001. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat Biotechnol **19**:125–130.

[52] **Edwards, J. S. and Palsson, B. Ø.**, 2000. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci U S A **97**:5528–5533.

[53] **Essenberg, M. K. and Cooper, R. A.**, 1975. Two ribose-5-phosphate isomerases from *Escherichia coli* K12: partial characterisation of the enzymes and consideration of their possible physiological roles. Eur J Biochem **55**:323–332.

[54] **Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J. J., and Gardner, T. S.**, 2007. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. PLoS Biol **5**:e8.

[55] **Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., and Palsson, B. Ø.**, 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol **3**:121.

[56] **Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., and Palsson, B. Ø.**, 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol **3**:e121.

[57] **Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L., and Palsson, B. Ø.**, 2009. Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol **7**:129–143.

[58] **Feist, A. M. and Palsson, B. Ø.**, 2008. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. Nat Biotechnol **26**:659–667.

[59] **Ferrández, A., Garciá, J. L., and Díaz, E.**, 1997. Genetic characterization and expression in heterologous hosts of the 3-(3-hydroxyphenyl)propionate catabolic pathway of *Escherichia coli* K-12. J Bacteriol **179**:2573–2581.

[60] **Fersht, A.**, 1999. Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding. W. H. Freeman, New York.

[61] **Flint, D. H.**, 1994. Initial kinetic and mechanistic characterization of *Escherichia coli* fumarase A. Arch Biochem Biophys **311**:509–516.

[62] **Fong, S. S., Burgard, A. P., Herring, C. D., Knight, E. M., Blattner, F. R., Maranas, C. D., and Palsson, B. Ø.**, 2005. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. Biotechnol Bioeng **91**:643–648.

[63] **Foster, J. W.**, 2004. *Escherichia coli* acid resistance: tales of an amateur acidophile. Nat Rev Microbiol **2**:898–907.

[64] **Fraser, H. I., Kvaratskhelia, M., and White, M. F.**, 1999. The two analogous phosphoglycerate mutases of *Escherichia coli*. FEBS Lett **455**:344–348.

[65] **Froman, B. E., Tait, R. C., and Gottlieb, L. D.**, 1989. Isolation and characterization of the phosphoglucose isomerase gene from *Escherichia coli*. Mol Gen Genet **217**:126–131.

[66] **Garrido-Pertierra, A. and Cooper, R. A.**, 1983. Evidence for two distinct pyruvate kinase genes in *Escherichia coli* K-12. FEBS Lett **162**:420–422.

[67] **Gianchandani, E. P., Papin, J. A., Price, N. D., Joyce, A. R., and Palsson, B. Ø.**, 2006. Matrix Formalism to Describe Functional States of Transcriptional Regulatory Systems. PLoS Comput Biol **2**.

[68] **Güldener, U., Münsterkötter, M., Kastenmüller, G., Strack, N., van Helden, J., Lemer, C., Richelles, J., Wodak, S. J., García-Martínez, J., Pérez-Ortín, J. E., Michael, H., Kaps, A., Talla, E., Dujon, B., André, B., Souciet, J. L., Montigny, J. D., Bon, E., Gaillardin, C., and Mewes, H. W.**, 2005. CYGD: the Comprehensive Yeast Genome Database. Nucleic Acids Res **33**:D364–D368.

[69] **Golby, P., Davies, S., Kelly, D. J., Guest, J. R., and Andrews, S. C.**, 1999. Identification and characterization of a two-component sensor-kinase and response-regulator system (DcuS-DcuR) controlling gene expression in response to C4-dicarboxylates in *Escherichia coli*. J Bacteriol **181**:1238–1248.

[70] **Grainger, D. C., Aiba, H., Hurd, D., Browning, D. F., and Busby, S. J. W.**, 2007. Transcription factor distribution in *Escherichia coli*: studies with FNR protein. Nucleic Acids Res **35**:269–278.

[71] **Green, J. and Guest, J. R.**, 1994. Regulation of transcription at the ndh promoter of *Escherichia coli* by FNR and novel factors. Mol Microbiol **12**:433–444.

[72] **Guet, C. C., Elowitz, M. B., Hsing, W., and Leibler, S.**, 2002. Combinatorial synthesis of genetic networks. Science **296**:1466–1470.

[73] **Gui, L., Sunnarborg, A., and LaPorte, D. C.**, 1996. Regulated expression of a repressor protein: FadR activates iclR. J Bacteriol **178**:4704–4709.

[74] **Hansen, E. J. and Juni, E.**, 1974. Two routes for synthesis of phosphoenolpyruvate from C4-dicarboxylic acids in *Escherichia coli*. Biochem Biophys Res Commun **59**:1204–1210.

[75] **Hansen, E. J. and Juni, E.**, 1975. Isolation of mutants of *Escherichia coli* lacking NAD- and NADP-linked malic. Biochem Biophys Res Commun **65**:559–566.

[76] **Harris, R. M., Webb, D. C., Howitt, S. M., and Cox, G. B.**, 2001. Characterization of PitA and PitB from *Escherichia coli*. J Bacteriol **183**:5008–5014.

[77] **Herrgård, M. J., Fong, S. S., and Palsson, B. Ø.**, 2006. Identification of genome-scale metabolic network models using experimentally measured flux profiles. PLoS Comput Biol **2**:e72.

[78] **Herrgård, M. J., Lee, B.-S., Portnoy, V., and Palsson, B. Ø.**, 2006. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. Genome Res **16**:627–635.

[79] **Hines, J. K., Fromm, H. J., and Honzatko, R. B.**, 2006. Novel allosteric activation site in *Escherichia coli* fructose-1,6-bisphosphatase. J Biol Chem **281**:18386–18393.

[80] **Hoyt, J. C., Robertson, E. F., Berlyn, K. A., and Reeves, H. C.**, 1988. *Escherichia coli* isocitrate lyase: properties and comparisons. Biochim Biophys Acta **966**:30–35.

[81] **Hu, Z., Killion, P. J., and Iyer, V. R.**, 2007. Genetic reconstruction of a functional transcriptional regulatory network. Nat Genet **39**:683–687.

[82] **Hua, Q., Yang, C., Baba, T., Mori, H., and Shimizu, K.**, 2003. Responses of the central metabolism in *Escherichia coli* to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase knockouts. J Bacteriol **185**:7053–7067.

[83] **Ideker, T. E., Thorsson, V., and Karp, R. M.**, 2000. Discovery of regulatory interactions through perturbation: inference and experimental design. Pacific Symposium on Biocomputing :305–316.

[84] **Iida, A., Teshiba, S., and Mizobuchi, K.**, 1993. Identification and characterization of the tktB gene encoding a second transketolase in *Escherichia coli* K-12. J Bacteriol **175**:5375–5383.

[85] **Iuchi, S. and Lin, E. C.**, 1988. arcA (dye), a global regulatory gene in *Escherichia coli* mediating repression of enzymes in aerobic pathways. Proc Natl Acad Sci U S A **85**:1888–1892.

[86] **Iwakura, M., Hattori, J., Arita, Y., Tokushige, M., and Katsuki, H.**, 1979. Studies on regulatory functions of malic enzymes. VI. Purification and molecular properties of NADP-linked malic enzyme from *Escherichia coli* W. J Biochem **85**:1355–1365.

[87] **Jiang, G. R., Nikolova, S., and Clark, D. P.**, 2001. Regulation of the ldhA gene, encoding the fermentative lactate dehydrogenase of *Escherichia coli*. Microbiology **147**:2437–2446.

[88] **Josephson, B. L. and Fraenkel, D. G.**, 1969. Transketolase mutants of *Escherichia coli*. J Bacteriol **100**:1289–1295.

[89] **Josephson, B. L. and Fraenkel, D. G.**, 1974. Sugar metabolism in transketolase mutants of *Escherichia coli*. J Bacteriol **118**:1082–1089.

[90] **Joyce, A. R., Reed, J. L., White, A., Edwards, R., Osterman, A., Baba, T., Mori, H., Lesely, S. A., Palsson, B. Ø., and Agarwalla, S.**, 2006. Experimental and computational assessment of conditionally essential genes in *Escherichia coli*. J Bacteriol **188**:8259–8271.

[91] **Kai, Y., Matsumura, H., and Izui, K.**, 2003. Phosphoenolpyruvate carboxylase: three-dimensional structure and molecular mechanisms. Arch Biochem Biophys **414**:170–179.

[92] **Kanehisa, M. and Goto, S.**, 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res **28**:27–30.

[93] **Karp, P. D., Keseler, I. M., Shearer, A., Latendresse, M., Krummenacker, M., Paley, S. M., Paulsen, I., Collado-Vides, J., Gama-Castro, S., Peralta-Gil, M., Santos-Zavaleta, A., Peñaloza-Spínola, M. I., Bonavides-Martinez, C., and Ingraham, J.**, 2007. Multidimensional annotation of the *Escherichia coli* K-12 genome. Nucleic Acids Res **35**:7577–7590.

[94] **Kasimoglu, E., Park, S. J., Malek, J., Tseng, C. P., and Gunsalus, R. P.**, 1996. Transcriptional regulation of the proton-translocating ATPase (atpIBEFHAGDC) operon of *Escherichia coli*: control by cell growth rate. J Bacteriol **178**:5563–5567.

[95] **Katz, J. and Rognstad, R.**, 1967. The labeling of pentose phosphate from glucose-14C and estimation of the rates of transaldolase, transketolase, the contribution of the pentose cycle, and ribose phosphate synthesis. Biochemistry **6**:2227–2247.

[96] **Ken A Dill, S. B.**, 2003. Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology. Taylor & Francis, Inc.

[97] **Keseler, I. M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I. T., Peralta-Gil, M., and Karp, P. D.**, 2005. EcoCyc: a comprehensive database resource for *Escherichia coli*. Nucleic Acids Res **33**:D334–D337.

[98] **Kessler, D., Herth, W., and Knappe, J.**, 1992. Ultrastructure and pyruvate formate-lyase radical quenching property of the multienzymic AdhE protein of *Escherichia coli*. J Biol Chem **267**:18073–18079.

[99] **Kharchenko, P., Vitkup, D., and Church, G.**, 2004. Filling gaps in a metabolic network using expression information. Bioinformatics **20 Suppl 1**:i178–i185.

[100] **Kühnel, K. and Luisi, B. F.**, 2001. Crystal structure of the *Escherichia coli* RNA degradosome component enolase. J Mol Biol **313**:583–592.

[101] **Kim, T. H. and Ren, B.**, 2006. Genome-wide analysis of protein-DNA interactions. Annu Rev Genomics Hum Genet **7**:81–102.

[102] **Knappe, J. and Sawers, G.**, 1990. A radical-chemical route to acetyl-CoA: the anaerobically induced pyruvate formate-lyase system of *Escherichia coli*. FEMS Microbiol Rev **6**:383–398.

[103] **Kobayashi, K., Tagawa, S., and Mogi, T.**, 1999. Electron transfer process in cytochrome bd-type ubiquinol oxidase from *Escherichia coli* revealed by pulse radiolysis. Biochemistry **38**:5913–5917.

[104] **Kornberg, H. L.**, 1965. The coordination of metabolic routes. University Press, London.

[105] **Kornberg, H. L.**, 1966. Anaplerotic sequences and their role in metabolism. Essays Biochem **2**:1–31.

[106] **Krieger, C. J., Zhang, P., Mueller, L. A., Wang, A., Paley, S., Arnaud, M., Pick, J., Rhee, S. Y., and Karp, P. D.**, 2004. MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res **32**:D438–D442.

[107] **Kumar, V. S. and Maranas, C. D.**, 2009. GrowMatch: an automated method for reconciling *in silico*/*in vivo* growth predictions. PLoS Comput Biol **5**:e1000308.

[108] **Lee, J., Owens, J. T., Hwang, I., Meares, C., and Kustu, S.**, 2000. Phosphorylation-induced signal propagation in the response regulator ntrC. J Bacteriol **182**:5188–5195.

[109] **Loh, K. D., Gyaneshwar, P., Papadimitriou, E. M., Fong, R., Kim, K.-S., Parales, R., Zhou, Z., Inwood, W., and Kustu, S.**, 2006. A previously undescribed pathway for pyrimidine catabolism. Proc Natl Acad Sci U S A **103**:5114–5119.

[110] **Lyngstadaas, A., Sprenger, G. A., and Boye, E.**, 1998. Impaired growth of an *Escherichia coli* rpe mutant lacking ribulose-5-phosphate epimerase activity. Biochim Biophys Acta **1381**:319–330.

[111] **Maglott, D., Ostell, J., Pruitt, K., and Tatusova, T.**, 2005. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res **33**:D54–D58.

[112] **Mahadevan, R. and Schilling, C. H.**, 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng **5**:264–276.

[113] **Mahajan, S. K., Chu, C. C., Willis, D. K., Templin, A., and Clark, A. J.**, 1990. Physical analysis of spontaneous and mutagen-induced mutants of *Escherichia coli* K-12 expressing DNA exonuclease VIII activity. Genetics **125**:261–273.

[114] **Markowitz, V. M., Korzeniewski, F., Palaniappan, K., Szeto, E., Werner, G., Padki, A., Zhao, X., Dubchak, I., Hugenholtz, P., Anderson, I., Lykidis, A., Mavromatis, K., Ivanova, N., and Kyrpides, N. C.**, 2006. The integrated microbial genomes (IMG) system. Nucleic Acids Res **34**:D344–D348.

[115] **Meléndez-Hevia, E. and Isidoro, A.**, 1985. The game of the pentose phosphate cycle. J Theor Biol **117**:251–263.

[116] **Membrillo-Hernández, J. and Lin, E. C.**, 1999. Regulation of expression of the adhE gene, encoding ethanol oxidoreductase in *Escherichia coli*: transcription from a downstream promoter and regulation by fnr and RpoS. J Bacteriol **181**:7571–7579.

[117] **Mikulskis, A., Aristarkhov, A., and Lin, E. C.**, 1997. Regulation of expression of the ethanol dehydrogenase gene (adhE) in *Escherichia coli* by catabolite repressor activator protein Cra. J Bacteriol **179**:7129–7134.

[118] **Molina, I., Pellicer, M. T., Badia, J., Aguilar, J., and Baldoma, L.**, 1994. Molecular characterization of *Escherichia coli* malate synthase G. Differentiation with the malate synthase A isoenzyme. Eur J Biochem **224**:541–548.

[119] **Muirhead, H.**, 1990. Isoenzymes of pyruvate kinase. Biochem Soc Trans **18**:193–196.

[120] **Narindrasorasak, S. and Bridger, W. A.**, 1977. Phosphoenolypyruvate synthetase of *Escherichia coli*: molecular weight, subunit composition, and identification of phosphohistidine in phosphoenzyme intermediate. J Biol Chem **252**:3121–3127.

[121] **Neidhardt (Ed. in Chief), F. C., Curtis, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger (eds), H. E.**, 1996. *Escherichia coli* and *Salmonella*: Cellular and molecular biology. American Society for Microbiology Press, Washington, DC.

[122] **Nellemann, L. J., Holm, F., Atlung, T., and Hansen, F. G.**, 1989. Cloning and characterization of the *Escherichia coli* phosphoglycerate kinase (pgk) gene. Gene **77**:185–191.

[123] **Nègre, D., Oudot, C., Prost, J. F., Murakami, K., Ishihama, A., Cozzone, A. J., and Cortay, J. C.**, 1998. FruR-mediated transcriptional activation at the ppsA promoter of *Escherichia coli*. J Mol Biol **276**:355–365.

[124] **Nguyen, N. T., Maurus, R., Stokell, D. J., Ayed, A., Duckworth, H. W., and Brayer, G. D.**, 2001. Comparative analysis of folding and substrate binding sites between regulated hexameric type II citrate synthases and unregulated dimeric type I enzymes. Biochemistry **40**:13177–13187.

[125] **Niersbach, M., Kreuzaler, F., Geerse, R. H., Postma, P. W., and Hirsch, H. J.**, 1992. Cloning and nucleotide sequence of the *Escherichia coli* K-12 ppsA gene, encoding PEP synthase. Mol Gen Genet **231**:332–336.

[126] **Nogales, J., Palsson, B. Ø., and Thiele, I.**, 2008. A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. BMC Syst Biol **2**:79.

[127] **Okinaka, R. T. and Dobrogosz, W. J.**, 1967. Catabolite repression and pyruvate metabolism in *Escherichia coli*. J Bacteriol **93**:1644–1650.

[128] **Osterman, A.**, 2006. A hidden metabolic pathway exposed. Proc Natl Acad Sci U S A **103**:5637–5638.

[129] **Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., and Vonstein, V.**, 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res **33**:5691–5702.

[130] **Paley, S. M. and Karp, P. D.**, 2002. Evaluation of computational metabolic-pathway predictions for *Helicobacter pylori*. Bioinformatics **18**:715–724.

[131] **Palsson, B. Ø.**, 2006. Systems biology: Properties of reconstructed networks. Cambridge University Press, Cambridge.

[132] **Park, J. H., Lee, K. H., Kim, T. Y., and Lee, S. Y.**, 2007. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation. Proc Natl Acad Sci U S A **104**:7797–7802.

[133] **Park, S. J., Chao, G., and Gunsalus, R. P.**, 1997. Aerobic regulation of the sucABCD genes of *Escherichia coli*, which encode alpha-ketoglutarate dehydrogenase and succinyl coenzyme A synthetase: roles of ArcA, Fnr, and the upstream sdhCDAB promoter. J Bacteriol **179**:4138–4142.

[134] **Park, S. J. and Gunsalus, R. P.**, 1995. Oxygen, iron, carbon, and superoxide control of the fumarase fumA and fumC genes of *Escherichia coli*: role of the arcA, fnr, and soxR gene products. J Bacteriol **177**:6255–6262.

[135] **Patil, K. R., Rocha, I., Förster, J., and Nielsen, J.**, 2005. Evolutionary programming as a platform for *in silico* metabolic engineering. BMC Bioinformatics **6**:308.

[136] **Pellicer, M. T., Fernandez, C., Badía, J., Aguilar, J., Lin, E. C., and Baldom, L.**, 1999. Cross-induction of glc and ace operons of *Escherichia coli* attributable to pathway intersection. Characterization of the glc promoter. J Biol Chem **274**:1745–1752.

[137] **Perham, R. N. and Packman, L. C.**, 1989. 2-Oxo acid dehydrogenase multienzyme complexes: domains, dynamics, and design. Ann N Y Acad Sci **573**:1–20.

[138] **Peterson, J. D., Umayam, L. A., Dickinson, T., Hickey, E. K., and White, O.**, 2001. The Comprehensive Microbial Resource. Nucleic Acids Res **29**:123–125.

[139] **Peyru, G. and Fraenkel, D. G.**, 1968. Genetic mapping of loci for glucose-6-phosphate dehydrogenase, gluconate-6-phosphate dehydrogenase, and gluconate-6-phosphate dehydrase in *Escherichia coli*. J Bacteriol **95**:1272–1278.

[140] **Pichersky, E., Gottlieb, L. D., and Hess, J. F.**, 1984. Nucleotide sequence of the triose phosphate isomerase gene of *Escherichia coli*. Mol Gen Genet **195**:314–320.

[141] **Pál, C., Papp, B., and Lercher, M. J.**, 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat Genet **37**:1372–1375.

[142] **Pál, C., Papp, B., and Lercher, M. J.**, 2005. Horizontal gene transfer depends on gene content of the host. Bioinformatics **21 Suppl 2**:ii222–ii223.

[143] **Pál, C., Papp, B., Lercher, M. J., Csermely, P., Oliver, S. G., and Hurst, L. D.**, 2006. Chance and necessity in the evolution of minimal metabolic networks. Nature **440**:667–670.

[144] **Plumbridge, J.**, 1998. Control of the expression of the manXYZ operon in *Escherichia coli*: Mlc is a negative regulator of the mannose PTS. Mol Microbiol **27**:369–380.

[145] **Price, N. D., Papin, J. A., Schilling, C. H., and Palsson, B. Ø.**, 2003. Genome-scale microbial *in silico* models: the constraints-based approach. Trends Biotechnol **21**:162–169.

[146] **Prodromou, C., Haynes, M. J., and Guest, J. R.**, 1991. The aconitase of *Escherichia coli*: purification of the enzyme and molecular cloning and map location of the gene (acn). J Gen Microbiol **137**:2505–2515.

[147] **Quail, M. A. and Guest, J. R.**, 1995. Purification, characterization and mode of action of PdhR, the transcriptional repressor of the pdhR-aceEF-lpd operon of *Escherichia coli*. Mol Microbiol **15**:519–529.

[148] **Rafael U Ibarra and Jeremy S Edwards and Bernhard Ø Palsson**, 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. Nature **420**:186–189.

[149] **Reed, J. L., Famili, I., Thiele, I., and Palsson, B. Ø.**, 2006. Towards multidimensional genome annotation. Nat Rev Genet **7**:130–141.

[150] **Reed, J. L., Patel, T. R., Chen, K. H., Joyce, A. R., Applebee, M. K., Herring, C. D., Bui, O. T., Knight, E. M., Fong, S. S., and Palsson, B. Ø.**, 2006. Systems approach to refining genome annotation. Proc Natl Acad Sci U S A **103**:17480–17484.

[151] **Reed, J. L., Patel, T. R., Chen, K. H., Joyce, A. R., Applebee, M. K., Herring, C. D., Bui, O. T., Knight, E. M., Fong, S. S., and Palsson, B. Ø.**, 2006. Systems approach to refining genome annotation. Proc Natl Acad Sci U S A **103**:17480–17484.

[152] **Reed, J. L., Vo, T. D., Schilling, C. H., and Palsson, B. Ø.**, 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol **4**:R54.

[153] **Reed, L. J., Pettit, F. H., Eley, M. H., Hamilton, L., Collins, J. H., and Oliver, R. M.**, 1975. Reconstitution of the *Escherichia coli* pyruvate dehydrogenase complex. Proc Natl Acad Sci U S A **72**:3068–3072.

[154] **Reidl, J. and Boos, W.**, 1991. The malX malY operon of *Escherichia coli* encodes a novel enzyme II of the phosphotransferase system recognizing glucose and maltose and an enzyme abolishing the endogenous induction of the maltose system. J Bacteriol **173**:4862–4876.

[155] **Ren, Q., Chen, K., and Paulsen, I. T.**, 2007. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. Nucleic Acids Res **35**:D274–D279.

[156] **Repaske, R. and Clayton, M. A.**, 1978. Control of *Escherichia coli* growth by CO2. J Bacteriol **135**:1162–1164.

[157] **Resendis-Antonio, O., Reed, J. L., Encarnación, S., Collado-Vides, J., and Palsson, B. Ø.**, 2007. Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. PLoS Comput Biol **3**:1887–1895.

[158] **Rhee, S. G., Ubom, G. A., Hunt, J. B., and Chock, P. B.**, 1982. Catalytic cycle of the biosynthetic reaction catalyzed by adenylylated glutamine synthetase from *Escherichia coli*. J Biol Chem **257**:289–297.

[159] **Riley, M., Abe, T., Arnaud, M. B., Berlyn, M. K. B., Blattner, F. R., Chaudhuri, R. R., Glasner, J. D., Horiuchi, T., Keseler, I. M., Kosuge, T., Mori, H., Perna, N. T., Plunkett, G., Rudd, K. E., Serres, M. H., Thomas, G. H., Thomson, N. R., Wishart, D., and Wanner, B. L.**, 2006. *Escherichia coli* K-12: a cooperatively developed annotation snapshot–2005. Nucleic Acids Res **34**:1–9.

[160] **Rypniewski, W. R. and Evans, P. R.**, 1989. Crystal structure of unliganded phosphofructokinase from *Escherichia coli*. J Mol Biol **207**:805–821.

[161] **Saier, M. H.**, 1998. Multiple mechanisms controlling carbon metabolism in bacteria. Biotechnol Bioeng **58**:170–174.

[162] **Samal, A. and Jain, S.**, 2008. The regulatory network of *E. coli* metabolism as a Boolean dynamical system exhibits both homeostasis and flexibility of response. BMC Syst Biol **2**:21.

[163] **Sauer, U., Canonaco, F., Heri, S., Perrenoud, A., and Fischer, E.**, 2004. The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. J Biol Chem **279**:6613–6619.

[164] **Sawers, G.**, 1993. Specific transcriptional requirements for positive regulation of the anaerobically inducible pfl operon by ArcA and FNR. Mol Microbiol **10**:737–747.

[165] **Sawers, G.**, 2001. A novel mechanism controls anaerobic and catabolite regulation of the *Escherichia coli* tdc operon. Mol Microbiol **39**:1285–1298.

[166] **Sawers, G. and Watson, G.**, 1998. A glycyl radical solution: oxygen-dependent interconversion of pyruvate formate-lyase. Mol Microbiol **29**:945–954.

[167] **Schneider, D., Pohl, T., Walter, J., Dörner, K., Kohlstädt, M., Berger, A., Spehr, V., and Friedrich, T.**, 2008. Assembly of the *Escherichia coli* NADH:ubiquinone oxidoreductase (complex I). Biochim Biophys Acta **1777**:735–739.

[168] **Schomburg, I., Chang, A., and Schomburg, D.**, 2002. BRENDA, enzyme data and metabolic information. Nucleic Acids Res **30**:47–49.

[169] **Scott A Becker and Adam M Feist and Monica L Mo and Gregory Hannum and Bernhard Ø Palsson and Markus J Herrgard**, 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat Protoc **2**:727–738.

[170] **Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., and Friedman, N.**, 2003. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet **34**:166–176.

[171] **Segrè, D., Vitkup, D., and Church, G. M.**, 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci U S A **99**:15112–15117.

[172] **Shimada, T., Ishihama, A., Busby, S. J. W., and Grainger, D. C.**, 2008. The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. Nucleic Acids Res **36**:3950–3955.

[173] **Shlomi, T., Berkman, O., and Ruppin, E.**, 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc Natl Acad Sci U S A **102**:7695–7700.

[174] **Skarstedt, M. T. and Silverstein, E.**, 1976. *Escherichia coli* acetate kinase mechanism studied by net initial rate, equilibrium, and independent isotopic exchange kinetics. J Biol Chem **251**:6775–6783.

[175] **Spehr, V., Schlitt, A., Scheide, D., Guénebaut, V., and Friedrich, T.**, 1999. Overexpression of the *Escherichia coli* nuo-operon and isolation of the overproduced NADH:ubiquinone oxidoreductase (complex I). Biochemistry **38**:16261–16267.

[176] **Sprenger, G. A., Schörken, U., Sprenger, G., and Sahm, H.**, 1995. Transketolase A of *Escherichia coli* K12. Purification and properties of the enzyme from recombinant strains. Eur J Biochem **230**:525–532.

[177] **Stein, L.**, 2001. Genome annotation: from sequence to biology. Nat Rev Genet **2**:493–503.

[178] **Stoesser, G., Tuli, M. A., Lopez, R., and Sterk, P.**, 1999. The EMBL Nucleotide Sequence Database. Nucleic Acids Res **27**:18–24.

[179] **Stolz, B., Huber, M., Markovic-Housley, Z., and Erni, B.**, 1993. The mannose transporter of *Escherichia coli*. Structure and function of the IIABMan subunit. J Biol Chem **268**:27094–27099.

[180] **Stroppolo, M. E., Falconi, M., Caccuri, A. M., and Desideri, A.**, 2001. Superefficient enzymes. Cell Mol Life Sci **58**:1451–1460.

[181] **Sutherland, P. and McAlister-Henn, L.**, 1985. Isolation and expression of the *Escherichia coli* gene encoding malate dehydrogenase. J Bacteriol **163**:1074–1079.

[182] **Suzuki, T.**, 1969. Phosphotransacetylase of *Escherichia coli* B, activation by pyruvate and inhibition by NADH and certain nucleotides. Biochim Biophys Acta **191**:559–569.

[183] **Tanaka, Y., Kimata, K., Inada, T., Tagami, H., and Aiba, H.**, 1999. Negative regulation of the pts operon by Mlc: mechanism underlying glucose induction in *Escherichia coli*. Genes Cells **4**:391–399.

[184] **Thiele, I., Jamshidi, N., Fleming, R. M. T., and Palsson, B. Ø.**, 2009. Genome-scale reconstruction of *Escherichia coli's* transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. PLoS Comput Biol **5**:e1000312.

[185] **Thomason, L. C., Court, D. L., Datta, A. R., Khanna, R., and Rosner, J. L.**, 2004. Identification of the *Escherichia coli* K-12 ybhE gene as pgl, encoding 6-phosphogluconolactonase. J Bacteriol **186**:8248–8253.

[186] **Thomson, G. J., Howlett, G. J., Ashcroft, A. E., and Berry, A.**, 1998. The dhnA gene of *Escherichia coli* encodes a class I fructose bisphosphate aldolase. Biochem J **331 (Pt 2)**:437–445.

[187] **Tseng, C. P.**, 1997. Regulation of fumarase (fumB) gene expression in *Escherichia coli* in response to oxygen, iron and heme availability: role of the arcA, fur, and hemA gene products. FEMS Microbiol Lett **157**:67–72.

[188] **Tseng, C. P., Albrecht, J., and Gunsalus, R. P.**, 1996. Effect of microaerophilic cell growth conditions on expression of the aerobic (cyoABCDE and cydAB) and anaerobic (narGHJI, frdABCD, and dmsABC) respiratory pathway genes in *Escherichia coli*. J Bacteriol **178**:1094–1098.

[189] **Valley, G. and Rettger, L. F.**, 1927. The influence of carbon dioxide on bacteria. J Bacteriol **14**:101–137.

[190] **Varma, A. and Palsson, B. Ø.**, 1993. Metabolic capabilities of *Escherichia coli*: I. Synthesis of biosynthetic precursors and cofactors. J Theor Biol **165**:477–502.

[191] **Varma, A. and Palsson, B. Ø.**, 1993. Metabolic capabilities of *Escherichia coli*: II. Optimal growth patterns. J Theor Biol **165**:503–522.

[192] **Varma, A. and Palsson, B. Ø.**, 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl Environ Microbiol **60**:3724–3731.

[193] **Veronese, F. M., Boccù, E., and Fontana, A.**, 1976. Isolation and properties of 6-phosphogluconate dehydrogenase from *Escherichia coli*. Some comparisons with the thermophilic enzyme from *Bacillus stearothermophilus*. Biochemistry **15**:4026–4033.

[194] **Wade, J. T., Struhl, K., Busby, S. J. W., and Grainger, D. C.**, 2007. Genomic analysis of protein-DNA interactions in bacteria: insights into transcription and chromosome organization. Mol Microbiol **65**:21–26.

[195] **Wallace, B., Yang, Y. J., Hong, J. S., and Lum, D.**, 1990. Cloning and sequencing of a gene encoding a glutamate and aspartate carrier of *Escherichia coli* K-12. J Bacteriol **172**:3214–3220.

[196] **Westerhoff, V. H. and van Dam, J. L.**, 1987. Thermodynamics and Control of Biological Free-energy Transduction. Elsevier, Amsterdam.

[197] **Wilks, J. C. and Slonczewski, J. L.**, 2007. pH of the cytoplasm and periplasm of *Escherichia coli*: rapid measurement by green fluorescent protein fluorimetry. J Bacteriol **189**:5601–5607.

[198] **Woods, S. A., Schwartzbach, S. D., and Guest, J. R.**, 1988. Two biochemically distinct classes of fumarase in *Escherichia coli*. Biochim Biophys Acta **954**:14–26.

[199] **Workman, C. T., Mak, H. C., McCuine, S., Tagne, J.-B., Agarwal, M., Ozier, O., Begley, T. J., Samson, L. D., and Ideker, T.**, 2006. A systems approach to mapping DNA damage response pathways. Science **312**:1054–1059.

[200] **Wu, L. F. and Mandrand-Berthelot, M. A.**, 1995. A family of homologous substrate-binding proteins with a broad range of substrate specificity and dissimilar biological functions. Biochimie **77**:744–750.

[201] **Yankovskaya, V., Horsefield, R., Törnroth, S., Luna-Chavez, C., Miyoshi, H., Léger, C., Byrne, B., Cecchini, G., and Iwata, S.**, 2003. Architecture of succinate dehydrogenase and reactive oxygen species generation. Science **299**:700–704.

[202] **Yuan, J., Fowler, W. U., Kimball, E., Lu, W., and Rabinowitz, J. D.**, 2006. Kinetic flux profiling of nitrogen assimilation in *Escherichia coli*. Nat Chem Biol **2**:529–530.

[203] **Zhou, D. and Yang, R.**, 2006. Global analysis of gene transcription regulation in prokaryotes. Cell Mol Life Sci **63**:2260–2290.