

Chapter 2

A Roadmap Towards a Systems Biology Description of Bacterial Nitrogen Fixation

Marie Lisandra Zepeda-Mendoza and Osbaldo Resendis-Antonio

2.1 Introduction

Nitrogen fixation is a fundamental natural process in which atmospheric nitrogen is reduced to ammonia. Several types of microorganisms that live in a variety of physiological conditions can perform nitrogen fixation. These microorganisms include bacteria such as *Rhizobium etli* (Masson-Boivin et al. 2009), *Klebsiella oxytoca* (Luftu-Cakmakci et al. 1981), *Frankia alni* (Schwmtzer and Tjepkema 1990), and cyanobacteria (Berman-Frank et al. 2003), as well as archaea such as *Methanococcus thermolithotrophicus* (Belay et al. 1984) and *Methanosarcina barkeri* (Bomar et al. 1985). Genome information from these microorganisms, in combination with data from high-throughput technologies, provides valuable material for elucidating the biological principles that characterize nitrogen fixation. Notably, the advent of high-throughput technologies has advanced our global understanding of the way that transcriptional regulatory and metabolic networks work together to support this biological process. While this task may sound easy, success is far from being a direct enterprise. The development of new paradigms is central to understand their basic mechanics and to make practical advancements in crop improvements. Thus, systems biology is a new field that can make notable contributions to these goals.

The central aim of this chapter is to present a conceptual view of how a systems biology description can be useful to construct hypotheses to improve our understanding of nitrogen fixation and to use the *in silico* modeling to perform a systematic and quantitative analysis of this biological phenomenon. We hope that

M.L. Zepeda-Mendoza
Licenciatura en Ciencias Genómicas-UNAM, Av. Universidad s/n, Col. Chamilpa,
Cuernavaca, Morelos C.P. 62210, Mexico

O. Resendis-Antonio (✉)
Instituto Nacional de Medicina Genómica (INMEGEN), Periferico Sur 4809, Arenal Tepepan,
Tlalpan 14610 Mexico City, Mexico
e-mail: oresendis@inmegen.gob.mx

the present study stimulates interest in this new scientific frontier and show that these computational methodologies can be useful to integrate information and generate knowledge for systematically uncovering the underlying metabolic activity during bacterial nitrogen fixation in *R. etli*.

2.1.1 Nitrogen Fixation

Nitrogen fixation can be performed by *rhizobia* soil bacteria in symbiosis with legume plants. This process has been extensively studied, and the entire genome sequences of select *Rhizobiaceae* bacteria (such as *Azorhizobium*, *Allorhizobium*, *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium*, and *Sinorhizobium*) have been reported. During nitrogen fixation, the bacteria use the nitrogenase enzyme to transform atmospheric N_2 into ammonia. In addition to its crucial role in the nitrogen cycle, there are also important agricultural and environmental reasons for studying this process. For example, modern agriculture relies on inefficient industrial fertilizers to maximize crop production. The use of chemical fertilizers severely damages the environment. Large quantities of fossil fuels are needed for nitrogenous fertilizer production and fertilizer decomposition releases highly active greenhouse gases (Crutzen et al. 2007). Furthermore, fertilizer loss due to leaching causes waterway eutrophication (Graham and Vance 2003). However, nitrogen fixation can provide a clean and natural strategy for improving field crops, thereby avoiding or reducing environmental pollution and making strides towards sustainable agriculture. Taken together, these aspects highlight the importance of optimizing nitrogen input through its natural mechanisms. However, nitrogen fixation is a highly complex biochemical process that requires active signaling and metabolic interchanges between the plant and its symbiotic bacteria. This biological process can be divided into three main phases, which can be briefly explained as follows:

- *Bacterial attraction*: Symbiosis starts when the roots of the plant excrete phenolic flavonoid compounds. Bacteria expressing NodD proteins recognize these compounds and are attracted to the roots (Redmond et al. 1986).
- *Nodule formation*: Once bacteria are localized into the root, they produce strain-specific chito-oligosaccharides, known as nod factors, which induce nodule formation. Nodules are special plant structures that house the bacteria while they are in symbiosis with the plant (Caetano-Anolles and Gresshoff 1991; Ferguson et al. 2010). Then, the bacteria enter the plant through its root hairs. This process typically occurs at the root tips, but the bacteria can also enter through cracks in the epidermal tissue of the root. Bacterial entry causes cellular-level ionic changes in the plant (Felle et al. 1999) and the root is deformed to promote cortical cell divisions. Legumes have two types of nodules: determinate and indeterminate. The host plant determines the nodule type; some physical and biological properties are listed for each case in Table 2.1.

Table 2.1 Characteristics of the two types of nodules

Characteristic	Indeterminate nodule	Determinate nodule
Initial cell divisions	First anticlinically in the inner cortex and then periclinically in the endodermis and pericycle	First subepidermically in the outer cortex
Nodule shape	Cylindrical with a persistent meristem	Spherical, lacking a persistent meristem
Nodule vascular system	Less branched compared to determinate nodules	More branched compared to indeterminate nodules
Bacteroid population	Heterogeneous due to continued cell division activity One bacteroid per symbiosome	Homogeneous, because bacterial differentiation into bacteroids occurs synchronously and then the bacteroids undergo senescence Multiple bacteroids per symbiosome
Bacteroid	Enlarged shape because of having multiple genome amplification; branched by having membrane modifications; with low viability (Vasse et al. 1990; Mergaert et al. 2006)	Normal rod size with normal genome content; with high viability (Mergaert et al. 2006)
Plant examples	Alfalfa (<i>Medicago sativa</i>), clover (<i>Trifolium</i>), pea (<i>Pisum sativum</i>), barrel medic (<i>Medicago truncatula</i>) (Bond 1948; Libbenga and Harkes 1973; Newcomb 1976; Newcomb et al. 1979)	Soybean (<i>Glycine max</i>), bean (<i>Phaseolus vulgaris</i>), Pongamia, <i>Lotus japonicus</i> (Bond 1948; Libbenga and Harkes 1973; Newcomb 1976; Newcomb et al. 1979; Turgean and Bauer 1982; Calvert et al. 1984; Gresshoff and Delves 1986; Rolfe and Gresshoff 1988; Mathews et al. 1989)

- *Nitrogen-fixing bacteroids*: The bacteria continue migrating into the plant via infection threads until they reach the inner cortex and the nodule primordium. The bacteria are released into an infection droplet that is excreted near the growing tip of the infection thread. Then, in a process resembling endocytosis, the bacteria are surrounded by a plant-derived membrane, called the peribacteroid membrane, to form a symbiosome (Udvardi and Day 1997). Inside the nodule, the bacteria differentiate into bacteroids. At this stage, the bacteroids are distinctly different from the free-living form of bacteria (see Table 2.1). Once inside the mature nodule, a bacteroid is capable of fixing atmospheric nitrogen by maintaining an ammonia–carbon source exchange with the plant. This nitrogen-fixing capacity is the result of global gene expression changes that give the bacteria highly specialized metabolic activities.

The nitrogenase enzyme is responsible for atmospheric nitrogen reduction. This enzyme is highly oxygen sensitive, but the nodule protects it by providing a microaerobic environment. While nitrogenase performs the nitrogen reduction, the biochemical reaction also depends on the coordinated participation of other

metabolic pathways to produce the required substrates. In other words, the nitrogenase enzyme participates in an essential reaction, but many other metabolic reactions are necessary to maintain the symbiotic nitrogen fixation. Hence, optimal nitrogen fixation also requires active metabolite exchange between the plant and the bacteroid. Amino acid transport between plant and bacteroid is essential to support the nitrogen fixation process (Lodwig et al. 2003). Thus, there is evidence that glutamate, or one of its derivatives, is provided by the plant whereas aspartate and alanine are secreted by the bacteroid (Day et al. 2001).

2.1.2 Symbiotic Relationship Between *R. etli* and *Phaseolus vulgaris*

It is important to clarify that symbiotic nitrogen fixation is a highly selective interaction; i.e., all legumes do not attract all *rhizobia*. Therefore, the bacterium and plant that are used to study and model nitrogen fixation must be carefully selected. For example, several studies have focused on the interaction between alfalfa and *Sinorhizobium meliloti*, which is one of the best-described interactions in symbiotic nitrogen fixation (Jones et al. 2007). Given this plant–bacteria specificity requirement, this chapter will focus on the metabolic analysis of symbiosis between the common bean (*P. vulgaris*) and *R. etli*.

We have chosen to focus on *R. etli* and *P. vulgaris* because of the large amount of physiological, molecular, and genetic information available on this symbiotic relationship. Furthermore, *P. vulgaris* is highly important to the agricultural market. The common bean represents 50 % of worldwide legume consumption (McClean et al. 2004), which is approximately 23 million tons, according to the Food and Agriculture Organization of the United Nations (<http://www.fao.org/corp/statistics/en>). In addition, *P. vulgaris* is a promiscuous legume because it can form nodules with various *rhizobia* (*R. etli*, *R. leguminosarum*, *R. propici*, *R. gallicum*, and *R. giardini*). However, *R. etli* is the most abundant symbiont for the wild-type bean. Taken together, we believe that this specific interaction is an appropriate model for studying legume–*Rhizobium* symbiosis with clear implications in agriculture technology.

2.2 Systems Biology of Bacterial Nitrogen Fixation

Nitrogen fixation involves a variety of signaling and metabolic pathways that coexist in a complex fashion. Consequently, understanding this process requires computational algorithms that can survey its complexity in a systematic and coherent fashion and can also drive additional and improved experimental study designs. While the metabolic and genetic aspects of this symbiosis have been

studied extensively, many intermediate and late-stage events remain poorly understood. Notably, the recent advent of high-throughput technologies has produced extensive data on gene expression patterns and metabolic activities that occur during nitrogen fixation. However, interpreting data from multiple high-throughput technologies (e.g., fluxomics, sequencing, microarrays, protein interaction data, and metabolomics) is not a trivial task due to the varying levels of biological description and heterogeneity of these datasets. In this context, genome-scale computational methods must be developed in a systems biology framework to systematically integrate this myriad of data and construct biological hypotheses. Thus, a systems biology framework can contribute to the achievement of three important goals (1) data integration and interpretation, (2) computational modeling of the metabolic phenotypes associated with nitrogen fixation, and (3) experimental assessment of the *in silico* predictions. A metabolic phenotype refers to the set of metabolic reactions that are required to support nitrogen fixation.

Systems biology has two main schemes: top-down and bottom-up. The top-down scheme provides a descriptive data integration approach, enables genome-scale monitoring of cellular activity, and allows the user to focus on specific areas of interest in terms of cellular activity. The top-down approach also allows researchers to discover patterns, referred to as emergent properties that can only be observed when looking at the system as a whole (Bhalla and Iyengar 1999; Rodriguez-Plaza et al. 2012). In comparison, a bottom-up scheme provides a more systemic and quantitative analysis of the ways that specific external perturbations can affect biological networks such as transcriptional, signaling, and metabolic networks (Fig. 2.1).

Although these two approaches can be viewed as separate strategies, a combination of the bottom-up and top-down schemes is required to completely study a biological system. A computational algorithm known as constraint-based modeling has been developed for this purpose and has been successfully applied to a variety of biological systems (Palsson 2006; Larhlimi and Bockmayr 2007, 2009; Resendis-Antonio et al. 2007). This framework enables the observation of genotype–phenotype relationships for a selected organism. The steps to form a constraint-based model are as follows (see Fig. 2.2). First, an organism’s genomic data are obtained. Next, a metabolic network is reconstructed with the use of bioinformatic tools and literature reviews. Then, computational simulations are conducted to predict the organism’s potential response to external perturbations. A more detailed description of this process is provided in the following sections.

2.2.1 Metabolic Network Reconstruction

The first step in reconstructing a metabolic network is data collection. All of the relevant information regarding the process of interest must be collected. There are many different sources of biological information, ranging from laboratory experiments and literature research to new high-throughput technologies. The

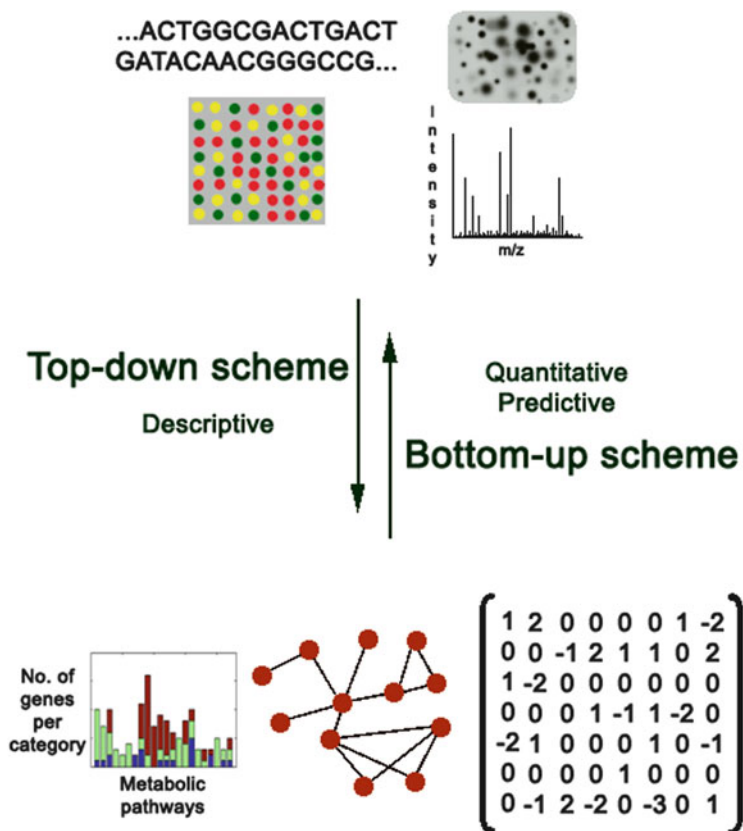


Fig. 2.1 The two schemes in systems biology. The top-down scheme starts with data coming from different sources (microarrays, DNA or RNA sequencing, mass spectrometry, etc.) to provide a description of the cellular activity in a genomic scale. The bottom-up scheme works with integrated data to build models and carry out *in silico* simulations to do predictions and guide hypothesis testing

amount of physiological data has grown exponentially in recent years in a variety of Rhizobiaceae; in particular high-throughput technologies have provided information across a wide range of biological levels such as transcriptome, proteome, and metabolome (see, for instance, Sarma and Emerich 2005, 2006; Resendis-Antonio et al. 2011, 2012). All of this physiological information needs to be collected to warrant a high-quality microorganism's metabolic network reconstruction.

Once the necessary information has been collected, we can systematically integrate the data by reconstructing a metabolic network. At this stage, we have a set of metabolic interactions that are supported by experimental evidence, bioinformatic predictions, or both. In turn, each reaction is associated with basic genome information such as which enzyme carries out the reaction, which genes encode the enzyme, the processed mRNA sequence, and the maximum reaction flux. However,

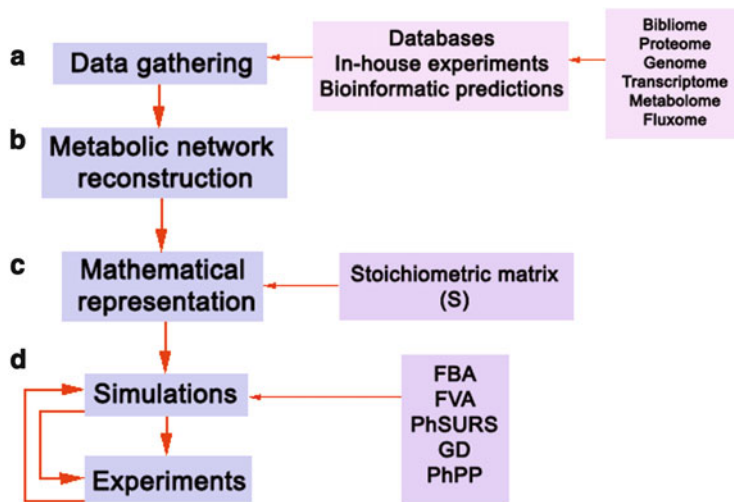


Fig. 2.2 Steps of the constraint-based modeling. **(a) Data gathering.** Data can come from many different sources, such as metabolomic, transcriptomic, proteomic, fuxomic, and genomic experiments carried out in-house or stored in public databases, as well as from bioinformatic predictions and from the literature. **(b) Metabolic network reconstruction.** The gathered data has to be integrated in a coherent fashion by taking into account all the available information related to every reaction. This integrated data constitutes the metabolic reconstruction. **(c) Mathematical representation of the network.** Once it has been reconstructed, a metabolic network is mathematically represented through an $m \times n$ matrix, called stoichiometric matrix (S), in which the m rows are the metabolites and the n columns are the reactions. The entry value in (m,n) is the stoichiometric coefficient of the metabolite m in the reaction n , and the value is positive if the metabolite is produced in the reaction and negative if it is consumed. **(d) Simulations and experimental cross talk.** In silico simulations are made based on the model by changing some parameters of the components to make predictions and lead future hypothesis testing experiments. Thus a retroactive process between laboratory experiments and computational analysis is established. Some of the most used and important simulations are flux balance analysis (FBA), flux variability analysis (FVA), phenotypic space uniform random sampling (PhSURS), gene deletion (GD), and phenotypic phase-plane (PhPP)

given that our knowledge is incomplete, this metabolic reconstruction frequently contains gaps that can limit the in silico phenotype analysis. In order to fill these missing metabolic reactions it is also crucial to include bioinformatic predictions or alternative computational methods (Orth and Palsson 2010). Two types of missing metabolic reactions can be distinguished: gap and orphan reactions. The first ones are those without experimental results confirming its presence but with the metabolic context suggesting it. The second ones include all the reactions that we expect to exist, preferably with experimental evidence suggesting it, but there is no clear evidence to which genes they are associated (Orth and Palsson 2010). For example, if experimental data were not available for all the enzymes in the glycolysis pathway, the missing enzymes could be predicted based on a variety of techniques that include the completely annotated pathway in a closely related species with

some bioinformatics and experimental methods. Thus, bioinformatic predictions can fill any gaps in a pathway. This data collection and gap-filling process are fundamental to ensure a proper metabolic reconstruction that will increase the predictive capabilities of the model.

2.2.2 *Mathematical Representation of the Metabolic Reconstruction*

Once the data have been integrated into a metabolic reaction set, the reconstruction can be mathematically represented as a $m \times n$ matrix, called the stoichiometric matrix (S), in which each of the rows represents a metabolite and each of the columns represents a reaction. If the metabolite m is produced in the reaction n , the entry value of (m,n) in S is the metabolite's stoichiometric coefficient in that reaction. If a metabolite is consumed, its entry (m,n) is the negative of its stoichiometric coefficient. As an example, we present the metabolic network reconstruction of the glycolysis pathway. The glycolytic reactions and their corresponding enzymes are shown in Table 2.2. The corresponding S is shown in Fig. 2.3.

The organization of the reconstructed network can also be represented graphically. Cellular metabolism involves two sets of objects: reactions and metabolites. Given this bipartite nature, metabolic networks can be separated into their two components and represented as a reaction or as a metabolite graph (Montañez et al. 2010). In a reaction graph, each node is an enzyme that carries out a given reaction where links connect any two nodes that share a common metabolite. In a metabolite graph, the compounds are the nodes and any two nodes are linked if they participate in the same reaction. An example of these complementary representations is shown in Fig. 2.4.

The type of question asked in a study determines which graphical representation is the most appropriate. For example, if the question is how the network metabolites work together to optimize a biological process, then the metabolite graph should be used (Resendis-Antonio et al. 2012). However, if the question is how the metabolic reactions work together to optimize a biological process, then the reaction graph should be used. Although the questions may sound similar, they have important intrinsic differences. In the first question, the focus is solely on the metabolite and does not involve which reaction it belongs to, i.e., the metabolite can be linked to many different reactions. In the second scenario, the focus is on the enzyme and not its associated metabolites.

Table 2.2 Glycolysis metabolic pathway reconstruction

Reaction	Enzyme
$\text{Glucose} + \text{ATP}^{4-} \rightarrow \text{G6P}^{2-} + \text{ADP}^{3-} + \text{H}^+$	Hexokinase (HK)
$\text{G6P}^{2-} \leftrightarrow \text{F6P}^{2-}$	Phosphoglucose isomerase (PGI)
$\text{F6P}^{2-} + \text{ATP}^{4-} \rightarrow \text{F1,6BP}^{4-} + \text{ADP}^{3-} + \text{H}^+$	Phosphofructokinase (PFK-1)
$\text{F1,6BP}^{4-} \leftrightarrow \text{DHAP}^{2-} + \text{G3P}^{2-}$	Fructose-bisphosphate aldolase (ALDO)
$\text{DHAP}^{2-} \leftrightarrow \text{G3P}^{2-}$	Triosephosphate isomerase (TPI)
$\text{G3P}^{2-} + \text{P}_i^{2-} + \text{NAD}^+ \leftrightarrow \text{1,3BPG}^{4-} + \text{NADH} + \text{H}^+$	Glyceraldehyde phosphate dehydrogenase (GAPDH)
$\text{1,3BPG}^{4-} + \text{ADP}^{3-} \leftrightarrow \text{3PG}^{3-} + \text{ATP}^{4-}$	Phosphoglycerate kinase (PGK)
$\text{3PG}^{3-} \leftrightarrow \text{2PG}^{3-}$	Phosphoglycerate mutase (PGM)
$\text{2PG}^{3-} \leftrightarrow \text{PEP}^{3-} + \text{H}_2\text{O}$	Enolase (ENO)
$\text{PEP}^{3-} + \text{ADP}^{3-} + \text{H}^+ \rightarrow \text{Pyr}^- + \text{ATP}^{4-}$	Pyruvate kinase (PK)

Glucose	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ATP ⁴⁻	-1	0	0	-1	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	1
G6P ²⁻	1	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ADP ³⁻	1	0	0	1	0	0	0	0	0	0	0	0	-1	1	0	0	0	0	-1
H ⁺	1	0	0	1	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	-1
F6P ²⁻	0	1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F1,6BP ⁴⁻	0	0	0	1	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
DHAP ²⁻	0	0	0	0	1	-1	-1	1	0	0	0	0	0	0	0	0	0	0	0
G3P ²⁻	0	0	0	0	1	-1	1	-1	-1	1	0	0	0	0	0	0	0	0	0
P _i ²⁻	0	0	0	0	0	0	0	0	0	-1	1	0	0	0	0	0	0	0	0
NAD ⁺	0	0	0	0	0	0	0	0	0	-1	1	0	0	0	0	0	0	0	0
1,3BPG ⁴⁻	0	0	0	0	0	0	0	0	0	1	-1	-1	1	0	0	0	0	0	0
NADH	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0
3PG ³⁻	0	0	0	0	0	0	0	0	0	0	0	1	-1	-1	1	0	0	0	0
2PG ³⁻	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	-1	1	0	0
PEP ³⁻	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	-1	0
H ₂ O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0
Pyr ⁻	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Fig. 2.3 Representation of the glycolysis pathway through a stoichiometric matrix. In this figure we have used the following notations: glucose-6-phosphate (G6P), fructose-6-phosphate (F6P), fructose-1,6-bisphosphate (F1,6BP), dihydroxyacetone phosphate (DHAP), glyceraldehyde-3-phosphate (G3P), 1,3-bisphosphoglycerate (1,3BPG), 3-phosphoglycerate (3PG), 2-phosphoglycerate (2PG), phosphoenolpyruvate (PEP), and pyruvate (Pyr). From left to right the reactions are in the same order as shown in Table 2.2

2.2.2.1 Topological Analysis of the Genome-Scale Metabolic Reconstruction

Graph theory can be used to perform topological studies on metabolic networks. Thus, topological analysis is important for uncovering cellular organizational principles and understanding its mechanisms for coordinating biological functions. Several classic topological properties are explained below.

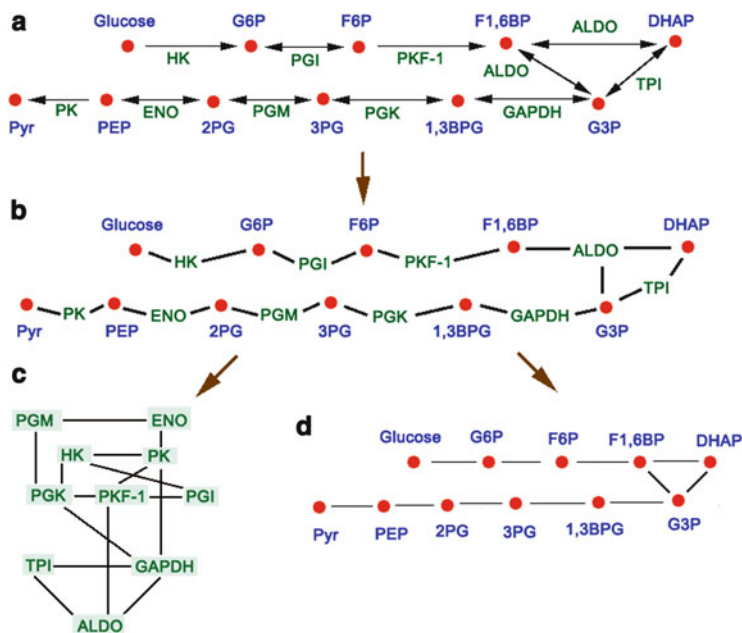


Fig. 2.4 Graphical representations of a metabolic network. (a) Conventional biochemical representation of the glycolysis pathway. This pathway can also be represented as a bipartite graph. (b) The bipartite graph of the glycolysis pathway consists of two different sets, enzymes and metabolites, represented by two different types of nodes. Nodes of one type can only be linked to nodes of the other type, so here a metabolite node is linked to an enzyme node if it participates in the reaction carried out by that enzyme. This multipartite graph can be separated into two graphs, each one with one set of data. (c) The reaction graph is built with the enzymes set. In this graph each node is an enzyme and a link connects two nodes if they share any metabolite. (d) The metabolite graph is built with the metabolites set. Here each node represents a metabolite and the links connecting them are the metabolic reactions that transform one metabolite to another. Here graphs (b), (c), and (d) are undirected because they are only for the purpose of representing general relationships, not taking into account the causality as directed networks do when the question of the study has been defined

- *Degree distribution:* The degree or connectivity of a node describes its number of nearest neighbors, i.e., its number of links. The degree values of all of the nodes in the network form a distribution; this distribution describes the number of nodes per degree (Newman 2003). The probability distribution of metabolic networks can usually fit a power law, reflecting a scale-free topology (Barabasi 2009). Scale-free networks are characterized by a few components with high connectivity (called hubs, e.g., ATP) and numerous components with low connectivity. There is evidence that this characteristic yields network robustness against the random loss of components (Han et al. 2004; Barabasi and Oltvai 2004; Jeong et al. 2001). However, other studies have suggested that this finding should be carefully reviewed in biological systems (Stumpf and Porter 2012).

- *Clustering coefficient (C)*: The clustering coefficient describes the degree to which the neighbors of a given node are connected to one another. In a social context, this is similar to asking “how many of my friends are also friends?” The clustering coefficient has a value between 0 and 1, with 1 indicating that all of the neighbors are connected to one another (Watts and Strogatz 1998).
- *Shortest path length*: The shortest path length is defined as the average of the minimum number of links (the shortest path) between every two nodes in the network. This measure is typically very small in metabolic networks (Jeong et al. 2000; Wagner and Fell 2001), although not all studies are in agreement (Arita 2004).
- *Module*: A module is a subnetwork of nodes with high clustering coefficients between one another. In other words, it is a set of nodes that are more closely related to one another than to the rest of the network. Nodes that share a common biological function are called functional modules. If there is no implicated biological function among the nodes, they are referred to as structural modules (Barabasi et al. 2011). However, there is evidence that functional modules can be associated to topological modules (Resendis-Antonio et al. 2012).

2.2.3 Computational Simulations

After the metabolic network reconstruction has been mathematically represented as S , the metabolic capabilities of the organism can be evaluated. However, before continuing with this aim, several issues must be solved. Dynamic modeling of a genome-scale metabolic reconstruction using ordinary differential equations (ODE) or partial differential equations (PDE) requires specific knowledge of the associated enzyme kinetics. The lack of specific information on metabolic reactions is a central problem in systems biology. Thus, there is a need to develop alternative methods that can overcome these limitations (Orth et al. 2011; Palsson 2011). In order to overcome such limitations, some frameworks have been developed that contribute to the exploration of the metabolic responses of a microorganism at steady-state behavior without an exhaustive number of kinetic parameters. Constraint-based modeling is a systemic and comprehensive systems biology approach that addresses these issues by imposing physical, biological, and thermodynamic constraints on the entire set of metabolic reactions under steady-state conditions (Fig. 2.5a) (Palsson 2006; Price et al. 2004). In this approach, an organism’s metabolic capacities are entirely defined by the properties of the stoichiometric matrix S and the enzymatic constraints of each metabolic reaction. Thus, it is possible to analyze the phenotypic space of an organism from its genome-scale metabolic reconstruction.

While the constraints define and limit the metabolic responses of a microorganism, a large number of possible metabolic states still exist. Therefore, developing computational methods to survey the properties of this space is a priority. To this end, a variety of methods that link the metabolic activity of a network to the

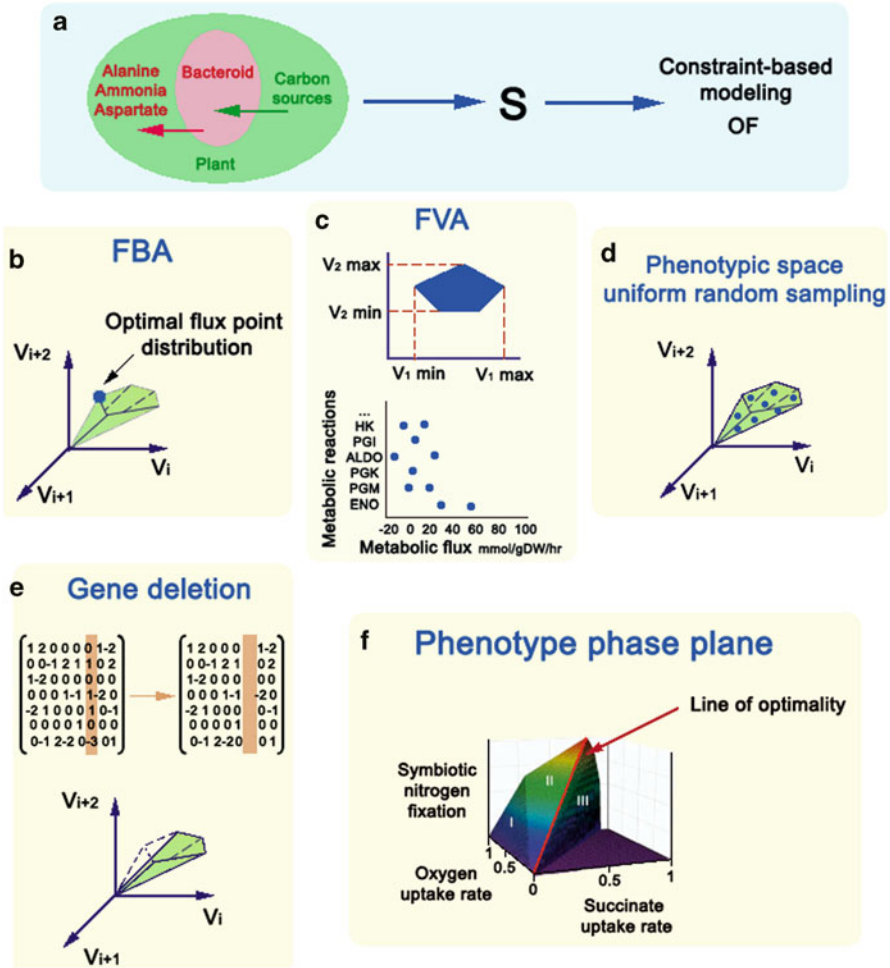


Fig. 2.5 Constraint-based modeling and in silico simulations. **(a)** *Constraint-based modeling*. Once the S is constructed and the objective function (OF) selected, an approach to analyze the physiological capabilities of the bacteroid during nitrogen fixation is constraint-based modeling. It works by studying the results of in silico simulations that impose a variety of biological and thermodynamic constraints subject to a steady-state condition. **(b)** *Flux balance analysis (FBA)*. From all the space of possible flux values (the solution space), FBA finds the steady-state flux distribution subject to mass balance and thermodynamics constraints that maximizes an OF, in our case the bacterial nitrogen fixation. To meet the steady-state assumption, the equality $Sv = 0$ is set, with v containing the flux values of all the reactions in the network. **(c)** *Flux variability analysis (FVA)*. FVA is used when there is more than one optimal flux distribution. It computes the minimal and maximal flux of each reaction such that a fixed OF is produced. **(d)** *Phenotypic space uniform random sampling*. This method is used to characterize the solution space. It samples some points of the space in a uniform random way, thus allowing an unbiased characterization of the solution space. Some aspects that can be studied with this unbiased method are the flux correlation between reactions and the characterization of the size and shape of the solution space. **(e)** *Gene deletions*. Gene deletions can be simulated by deleting from the S the reactions associated to the deleted gene product. Theoretically speaking, it results in a reduction of the solution space. It can be used to investigate gene dispensability and mutant phenotypes. **(f)** *Phenotypic phase plane (PhPP)*.

phenotype of a microorganism have been developed. These methods address questions aimed at understanding genotype–phenotype relationships for a biological process of interest. Before describing examples of these methods, two key concepts are explained: solution space and objective function.

- *Solution space*: The metabolic capacities of the model organism during the process of interest together with the enzymatic constraints of each metabolic reaction constitute a phenotypic space called the solution space. The solution space contains all of the possible steady-state metabolite fluxes in the reconstructed metabolic network. Consequently, the metabolic activities identified in different regions of the solution space can be associated with a wide range of phenotypes proper to diverse physiological conditions, including growth rates, metabolite production rates, and the rate of bacterial nitrogen fixation. Mathematically speaking, to represent a cellular metabolism at a steady state, the matrix S must be multiplied by the vector x such that the product equals 0, in other words

$$S \cdot x = 0$$

Because S is a matrix with m rows and n columns, the set of all vectors x that satisfy this equation forms an n -dimensional solution space. It is important to mention that the size of the solution space is *constrained* by the minimum and maximum flux values assigned to each reaction, see Fig. 2.5b.

- *Objective function (OF)*: The objective function represents a linear combination of the metabolite fluxes that are needed for a metabolic process to occur. In other words, the OF is the sum of the fluxes of metabolites, each of which is multiplied by a coefficient that indicates their weighted contribution to the metabolic process. Mathematically speaking, this function is represented as

$$OF = c^T \cdot v$$

where c and v are a weight and metabolic flux vector respectively. The flux vector contains the flux information of all the metabolites included in the reconstruction, it indicates the flux of the metabolites participating in the OF. It is important to select an OF that represents the specific physiological conditions in the organism of interest. In the particular case of nitrogen fixation for *R. etli*, the OF can be defined by taking into account various information sources, such as literature

Fig. 2.5 (continued) This method characterizes the steady-state solution space projected in two or three dimensions to divide the steady-state flux distributions into regions with similar metabolic flux patterns and a similar OF response if certain metabolites were additionally supplied to the network. The line of optimality corresponds to the conditions where metabolic activity is organized in the most efficient way to maximize the OF

reviews and metabolome data generated during the nitrogen fixation stage of *R. etli* (Resendis-Antonio et al. 2012).

Now that we have described these key concepts, we explain below some of the methods used in constraint-based modeling for exploring the phenotype capacities of genome-scale metabolic reconstructions.

2.2.3.1 Flux Balance Analysis

Flux balance analysis (FBA) uses optimization principles to identify the metabolic state of a microorganism and works as explained next. First, an objective function (OF) is defined to simulate the microorganism's physiological state in the process of interest. In order to reduce the number of kinetic parameters required to simulate the metabolic activity, the biological process is assumed to occur at a steady state while maximizing the OF. Finally, the metabolic flux distribution that maximizes the OF in the solution space is computed by linear programming (Fig. 2.5b). Thus, by maximizing the OF, one can identify the metabolic activity along the metabolic pathways required to produce an efficient biological process which, in our case, is nitrogen fixation.

FBA has implicit advantages and disadvantages. On one hand, it requires a limited number of parameters to predict the metabolic state associated with a phenotype. Furthermore, it represents an elegant formalism for analyzing genome-scale metabolic reconstructions while encompassing cross talk from high-throughput technologies (Orth et al. 2010; Resendis-Antonio et al. 2012). However, these advantages can also be seen as a drawback. Due to a lack of kinetic parameters, FBA cannot predict metabolic concentrations and its outputs are only valid under a steady-state assumption (Orth et al. 2010). The temporal behavior of metabolic concentrations and fluxes can be analyzed by dynamic flux balance analysis or genome-scale linear perturbation models (Varma and Palsson 1994; Resendis-Antonio 2009; Jamshidi and Palsson 2008). However, these issues remain open to further investigation in systems biology.

2.2.3.2 Flux Variability Analysis

Given the high dimensionality of the optimization problem described in Sect. 2.2.3.1, the metabolic phenotype that maximizes the OF is usually not unique. This result is due to redundant mechanisms underlying the metabolic network. Flux variability analysis (FVA) is a computational tool that is useful for quantifying this redundancy by identifying the set of metabolic phenotypes that result in an equivalent maximal (or minimal) OF. This analysis is extremely important for identifying biochemical pathways that can potentially generate the same phenotype. The results are fundamental for a proper interpretation of the *in silico* outputs. From a computational point of view, FVA computes the minimal and maximal fluxes of each

reaction for a fixed OF (Fig. 2.5c). This analysis identifies the feasible range of flux values for each reaction, thus representing the different metabolic phenotypes that a microorganism can use to adapt to its environment. In other words, this method does not identify all the optimal solutions but rather identifies the range of flux variability that is possible within any given solution (Mahadevan and Schilling 2003).

2.2.3.3 Phenotypic Space Uniform Random Sampling

A method called phenotypic space uniform random sampling can be used to survey the range of possible biochemical states without the bias of an OF. This analysis enables characterization of the size and shape of the solution space and provides information about the variety of a microorganism's metabolic phenotype states. The method consists of sampling the solution space in a random and uniform manner to extract interesting information about its characteristics. For example, pairwise correlation coefficients can be calculated between all of the metabolic fluxes to identify the degree of correlation between each pair of reactions (Fig. 2.5d) (Price et al. 2004).

2.2.3.4 Gene Deletions

Gene deletions can be simulated *in silico* by deleting the columns from S associated with the gene or genes product of interest or, alternatively, by constraining its corresponding upper and lower bound in the flux vector. Hence, these *in silico* gene deletions result in a reduction of the solution space (Fig. 2.5e). This type of study can be used to investigate gene dispensability (Duarte et al. 2004), to examine the evolution of wild-type strains towards new optimal states under selection pressures (Fong et al. 2003), and to predict the metabolic phenotypes of mutant strains (Resendis-Antonio et al. 2007).

2.2.3.5 Phenotypic Phase Plane

The phenotypic phase plane is a method for characterizing a projection of the steady-state solution space in two or three dimensions. In this analysis, steady-state flux distributions can be divided into a finite number of regions based on similar metabolic flux patterns and shadow prices. A shadow price describes how the OF would change if additional metabolites were supplied to the network (Edwards et al. 2001). Thus, the regions of the plane are classified based on the extent to which metabolite availability limits the process of interest and which biochemical activity is the optimal to induce a specific phenotype, in this case nitrogen fixation (Fig. 2.5f).

2.2.4 *Experimental Cross Talk*

The described methods above are useful for examining a microorganism's metabolic capabilities in a given physiological state. These methods allow us to link a cell's metabolic activity to a specific phenotype. However, in silico predictions can differ from experimental results due to network incompleteness introduced during the metabolic reconstruction, an improperly defined OF, or an inaccurate description of the cell's environmental conditions. Thus, an iterative, retroactive process is needed to reconcile laboratory data and computational predictions.

Simulation results can lead to hypothesis that in order to be verified require the design of new experiments. These experiments can provide additional data, resulting in information that contributes to construct a more realistic computational representation of nitrogen fixation. For instance, this approach could be used to refine the OF definition. The corrected OF could then be used as input for computational methods, the results of which will either further confirm the metabolic activity of previous simulations or lead to the identification of other pathways for future research.

2.3 Systems Biology Description of *R. etli* and *P. vulgaris* Symbiosis

A systems biology analysis of the symbiotic interactions between *R. etli* and *P. vulgaris* was conducted for nitrogen fixation. Top-down and bottom-up strategies were used to achieve the following goals (1) genome-scale metabolic reconstruction and topological analysis of the *R. etli*, (2) constraint-based metabolic modeling of the nitrogen-fixing bacteroid, and (3) construction of a computational platform capable to predict the metabolic phenotype in *R. etli* and compare between our computational predictions and the experimental results.

2.3.1 *Metabolic Reconstruction of R. etli*

The metabolic reconstruction of *R. etli* for simulating the symbiotic interaction with *P. vulgaris* was based on the integration of several levels of biological data. This information came from various papers reported in the literature and data obtained from high-throughput technologies. This latter source including proteomic, transcriptomic, and metabolomic experiments (Resendis-Antonio et al. 2007, 2011, 2012). In addition other information sources were used, including the *R. etli* genome annotation from KEGG (Kyoto Encyclopedia of Genes and Genomes) and available scientific literature (e.g., biochemical textbooks and previous reports). In total, the current version of the *R. etli* metabolic reconstruction,

identified as iOR450, consists of 450 genes, 405 reactions, and 377 metabolites. Although *R. etli* contains approximately 7,000 genes and thus 450 genes may sound too small to be representative of its genome, this reconstruction is based on well-curated and reliable experimental data.

Functional classification of these 450 genes revealed metabolic pathways with an important role in nitrogen fixation, including glycolysis, the TCA cycle, the pentose phosphate pathway, oxidative phosphorylation, amino acid production, glycogen and poly- β -hydroxybutyrate (PHB) biosynthesis, nitrogen reduction, secretion systems, and fatty acid metabolism. This version of the reconstruction is the basic platform for expansion and improvement in future versions. For instance, a significant number of the identified proteins are related to transport (e.g., the transport of small molecules), thus reflecting extensive metabolic cross talk between the plant and the bacteroid (Resendis-Antonio et al. 2007, 2011, 2012). Our metabolic reconstruction will include this physiological information and these processes will be experimentally verified. Currently, this reconstruction is a cornerstone in exploring the metabolic capacities of *R. etli* when facing different environmental conditions.

The graphical representation of the network provides information to elucidate how the metabolite organization within the network supports bacterial nitrogen fixation. For instance, structural modules were identified based on a purely topological criterion to elucidate their relationship with *R. etli*'s biological functions (Ravasz et al. 2002; Resendis-Antonio et al. 2005). Hence, based on a topological criterion (defined as the inverse square of the minimal path length between every pair of nodes), there were reported nine topological modules whose metabolic composition fell into three main groups: nucleic acids, peptides, and lipids. Notably, the functional composition of these modules and the quantitative metabolome data have contributed to establish hypotheses of how the production of metabolites is required for sustaining an optimal nitrogen fixation (Resendis-Antonio et al. 2012). This kind of analysis is an adequate scheme to survey the basic organizational principles by which a biological process can happen in nature. Thus, by correlating the metabolites of modules with the high-throughput data, it was observed that the concentrations of most of the metabolites in each nodule upregulate their concentration during nitrogen fixation compared to the free-living bacteria. This finding supports the idea that a coherent functional biological activity is required at a functional level (Hartwell et al. 1999; Resendis-Antonio et al. 2005, 2012).

2.3.2 Constraint-Based Modeling and Simulation Results

Once the metabolic reconstruction was complete, a constraint-based model allowed us to explore the metabolic capacities of the network for *R. etli*. To this end, first we have constructed an appropriate OF for mimicking the metabolic activity during nitrogen fixation. This function was defined based on data from literature review

and high-throughput experiments, specially taking into account the following information:

- Plant–bacteroid exchange of certain amino acids may be a general mechanism in *rhizobia* (Prell and Poole 2006).
- Sixteen ATP molecules are required to reduce one N_2 molecule into two ammonium molecules, which are subsequently exported to the plant (Patriarca et al. 2002; Lodwig and Poole 2003).
- Glycogen and PHB accumulate during nitrogen fixation and serve as carbon storage (Bergersen and Turner 1990; Lodwig et al. 2005; Sarma and Emerich 2006; Trainer and Charles 2006).
- Mutations in the biosynthesis of branched chain amino acids, such as L-valine, produce defective nodule formations (De las Nieves Peltzer et al. 2008), which indicates that this metabolite is an essential component. At a similar level, we considered that L-histidine is a central compound in nitrogen fixation (Dixon and Kahn 2004). In agreement with this experimental findings, both amino acids were included in the OF.
- Metabolome experiments have identified new metabolites with statistically significant changes during bacteroid activity compared to free-living bacteria. In order to give a better constraint-based analysis, there has been suggested that these metabolites can be used to define a more complete OF. These metabolites include CMP, 3-phospho-D-glycerate, and 2-oxoglutarate (Resendis-Antonio et al. 2012).

By integrating these findings, an OF representing the metabolic flux of key metabolites in nitrogen fixation was defined as follows:

$$Z^{\text{Fix}} = \text{glycogen} + \text{hist}[c] + \text{lys}[c] + \text{phb}[c] + \text{val}[c] + \text{ala}[e] + \text{asp}[e] + \text{nh4}[e] \\ + \text{mal}[c] + \text{trp}[c] + \text{arg}[c] + \text{cit}[c] + \text{cmp}[c] + \text{fum}[c] + 3\text{pg}[c] + \text{akg}[c]$$

where glycogen, histidine, lysine, polyhydroxybutyrate, valine, alanine, aspartate, and ammonium are denoted as glycogen, hist[c], lys, phb[c], val[c], ala[e], asp[e], and nh4[e], respectively. Similarly, mal, trp, arg, cit, cmp, fum, 3 pg, and ak denote malate, tryptophan, arginine, citrate, CMP, fumarate, 3-phospho-D-glycerate, and 2-oxoglutarate, respectively. All these metabolites are required to support an effective symbiotic nitrogen fixation, and the location of the metabolites is represented by [c] or [e] for cytoplasmic or external compounds, respectively. Once the OF was defined, a series of studies and simulations were performed to explore, characterize, and predict bacteroid metabolic phenotypes.

Thus, FBA was performed to examine the metabolic fluxes associated with nitrogen fixation. With the FBA results and based on the gene products associated with these reactions, one can predict the set of essential genes for carrying out nitrogen fixation. Furthermore, FVA was used to characterize the core metabolic activity within the set of alternative solutions. The reactions with zero range in flux variability were selected as part of central metabolism in bacterial nitrogen fixation.

The experimental data had to be related to the computational results to biologically support the model; consequently the agreement between the computational interpretations and the experimental data must be quantified to make this comparison. To this end, a consistency coefficient has been defined as the fraction of genes and enzymes that FBA predicted as active compared to those that were detected by transcriptome and proteome technologies. Based on this parameter, the consistency coefficients for the simulation of nitrogen fixation were reported to be 0.61 for genes and 0.71 for enzymes. To ensure the quality of the reconstruction and evaluate the coherence of the computational simulations, we proceed to assess the capacity of the model to predict physiological knowledge when gene deletion occurs (see Fig. 2.5e). To evaluate capacity of the model to predict physiological behaviors after gene silencing, *in silico* gene deletion analysis was performed on the genes encoding PHB synthase, glycogen synthase, arginine deiminase, myo-inositol dehydrogenase, pyruvate carboxylase, citric acid cycle enzymes, PEP carboxykinase, bisphosphate aldolase, and nitrogenase. These simulations were in qualitative agreement with the experimental counterpart in a variety of Rhizobiaceae (Resendis-Antonio et al. 2007, 2010, 2012).

On the other hand, the robustness of the topological structure of metabolic network (e.g., its functional modules) was explored by changing its external environment, specifically the uptake rates of succinate and inositol. The effects of these changes were evaluated by constructing a phenotypic phase plane. Then, a subset of 20 points, corresponding to select uptake rate conditions within the metabolic phase plane, was subjected to FBA. The metabolic subnetworks were graphically represented based on the FBA results. The topological variations between these subnetworks were defined by the fraction of overlapping metabolites. This study concluded that the metabolic profile required for optimizing nitrogen fixation does not significantly change for a wide range of succinate or inositol uptake rates. This finding suggests that the metabolic network supporting nitrogen fixation is robust to environmental changes (Resendis-Antonio et al. 2012).

2.3.3 Comparative Analysis of Simulations and Experimental Data

A detailed comparison between the *in silico* simulation results and the high-throughput data highlighted several important aspects of bacterial nitrogen fixation; such processes are described in Table 2.3. Overall, in this chapter we support that a systems biology approach enable the reconstruction and comprehensive analysis of *R. etli* metabolism during symbiosis with *P. vulgaris*, providing a systematic framework to reach a broader view of this important interaction. The current metabolic description can be improved by continually assessing the discrepancies between the computational results and the experimental data, whereas its

Table 2.3 Comparison of the computational and experimental results

Biological process	Computational results	Experimental results
Entner–Doudoroff and pentose phosphate pathways	Besides gluconeogenesis, the existence of a fueling pathway based on pentoses is predicted in the bacteroid	Metabolome experiments detected metabolites of the pentose phosphate and the Entner–Doudoroff pathways (Resendis-Antonio et al. 2012)
Nitrogen fixation enzymes	As expected, if viewing this finding as a positive control for the OF, these enzymes were predicted by the model to be central in the nitrogen fixation	As expected, the <i>nif</i> and <i>nifx</i> genes and coded proteins that are involved in nitrogen fixation were identified by the microarrays and the proteome experiments
Nucleotides metabolism	Some enzymes of the purine and the pyrimidine pathways actively participate for an optimal nitrogen fixation	Several of the enzymes that participate in the purine and pyrimidine pathways were identified in the proteome data (Resendis-Antonio et al. 2011)
Oxidative phosphorylation	Nonzero fluxes through oxidative phosphorylation and removal of all cytochrome oxidase reactions result in total loss of nitrogen fixation	The essentiality of respiration for nitrogen fixation has been previously reported (Lodwig and Poole 2003; Batut and Boistard 1994)
TCA cycle	The model predicts incomplete use of the TCA cycle	Experiments are not always able to detect all the TCA cycle enzymes. Besides, mutants in the TCA cycle in <i>B. japonicus</i> are still able to fix nitrogen, suggesting that a complete set of TCA cycle enzymes is not required for fixation (Lodwig and Poole 2003; Green and Emerich 1997)
Gluconeogenesis pathway	The model concluded that the gluconeogenesis pathway is active in nitrogen fixation	It is known for <i>R. etli</i> that the gluconeogenesis pathway is active in nitrogen fixation (Lodwig and Poole 2003). This finding was also identified in the proteome experiment (Resendis-Antonio et al. 2011)
Ammonium assimilation	FBA predicted that there is no activity in the ammonium assimilation pathway during symbiosis	Ammonium assimilation is not observed during nitrogen fixation and an increase in ammonium assimilation negatively affects nodulation (Mendoza et al. 1995)
PHB and glycogen accumulation	Simulations of deletion of PHB synthase predict that symbiotic nitrogen fixation increases	In <i>R. etli</i> PHB synthase deletion causes increase in nitrogen fixation (Cevallos et al. 1996). A similar result is observed upon glycogen synthase deletion in <i>R. tropici</i> (Marroqui

(continued)

Table 2.3 (continued)

Biological process	Computational results	Experimental results
		et al. 2001). Furthermore, it has been suggested that inhibition of one of the polymers results in accumulation of the other (Cevallos et al. 1996)
Arginine deiminase pathway	Arginine deiminase deletion predicts a decrease in symbiotic nitrogen fixation	Nitrogen fixation is reduced upon arginine deiminase deletion (D’Hooghe et al. 1997)
Myo-inositol catabolic pathway	The model predicts that the activity of the enzyme increases nitrogen fixation and that its deletion decreases the fixation activity	Mutation of myo-inositol dehydrogenase in <i>Sinorhizobium fredii</i> increases nitrogen fixation (Jiang et al. 2001). Besides, myo-inositol, 2-dehydrogenase proteins were detected in <i>R. etli</i> (Resendis-Antonio et al. 2011)
Fatty acids metabolism	Analysis of the functional modules suggests that the fatty acids are important for nitrogen fixation (Resendis-Antonio et al. 2012)	Fatty acid metabolism can play a significant role in nitrogen fixation in <i>R. etli</i> (Resendis-Antonio et al. 2011). An explanation could be that it can supply a variety of precursors to the bacteroid, such as components of the rhizobial membrane, lipopolysaccharides, and coenzymes required in signal transduction

agreements can lead to the design of new experiments (Resendis-Antonio et al. 2012).

2.4 Conclusion

High-throughput technologies provide valuable data for describing the global landscape of cellular activity. However, these techniques do not have the full capacity for describing and predicting the integrated functions of biological processes. Using a systems biology approach, we were able to construct a proper computational framework that serves as a guide for integrating various types of “-omics” data (Palsson 2006, 2011). With this computational framework, we could describe and predict metabolic activities, as well as design experiments that explore genotype–phenotype relationships involved in nitrogen fixation. The combination of experimental data and computational modeling advances our understanding of the main metabolic mechanisms that support bacterial nitrogen fixation. This achievement will undoubtedly have important effects in developing sustainable

agricultural programs by optimizing cost-effective crop improvements and ultimately diminishing the pollution effects of chemical fertilizers.

Acknowledgments Osbaldo Resendis-Antonio thanks the financial support of the Research Chair on Systems Biology-INMEGEN, FUNTEL-Mexico. The authors are grateful for comments and suggestions from a referee during the review process of this chapter.

References

- Arita M (2004) The metabolic world of *Escherichia coli* is not small. *Proc Natl Acad Sci USA* 101 (6):1543–1547
- Barabasi AL (2009) Scale-free networks: a decade and beyond. *Science* 325:412–413
- Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5:101–113
- Barabasi AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. *Nat Rev Genet* 12(1):56–68
- Batut J, Boistard P (1994) Oxygen control in *Rhizobium*. *Antonie Van Leeuwenhoek* 66:129–150
- Belay N, Sparling R, Daniels L (1984) Dinitrogen fixation by a thermophilic methanogenic bacterium. *Nature* 312:286–288
- Bergersen FJ, Turner GL (1990) Bacteroids from soybean root nodules: accumulation of poly- β -hydroxybutyrate during supply of malate and succinate in relation to N_2 fixation in flow-chamber reactions. *Proc R Soc Lond B Biol Sci* 240:39–59
- Berman-Frank I, Lundgren P, Falkowski P (2003) Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Res Microbiol* 154:157–164
- Bhalla US, Iyengar R (1999) Emergent properties of biological signaling pathways. *Science* 283 (5400):381–387
- Bomar M, Knoll K, Widdel F (1985) Fixation of molecular nitrogen by *Methanosarcina barkeri*. *FEMS Microbiol Lett* 31:47–55
- Bond L (1948) Origin and developmental morphology of root nodules of *Pisum sativum*. *Bot Gaz* 109:411–434
- Caetano-Anolles G, Gresshoff PM (1991) Plant genetic control of nodulation. *Annu Rev Microbiol* 45:345–382
- Calvert HE, Pence MK, Pierce M, Malik NSA, Bauer WD (1984) Anatomical analysis of the development and distribution of *Rhizobium* infection in soybean roots. *Can J Bot* 62:2375–2384
- Cevallos MA, Encarnacion S, Leija A, Mora Y, Mora J (1996) Genetic and physiological characterization of a *Rhizobium etli* mutant strain unable to synthesize poly- β -hydroxybutyrate. *J Bacteriol* 178:1646–1654
- Crutzen P, Mosier AR, Smith KA, Winiwarter W (2007) NO release from agro-fuel production negates global warming reduction by replacing fossil fuels. *Atmos Chem Phys Discuss* 7:11191–11205
- D'Hooghe I, Vander Wauven C, Michiels J, Tricot C, de Wilde P et al (1997) The arginine deiminase pathway in *Rhizobium etli*: DNA sequence analysis and functional study of the arcABC genes. *J Bacteriol* 179:7403–7409
- Day DA, Poole PS, Tyerman SD, Rosendahl L (2001) Ammonia and amino acid transport across symbiotic membranes in nitrogen-fixing legume nodules. *Cell Mol Life Sci* 58:61–71
- De las Nieves Peltzer M, Roques N, Poinso V, Aguilar OM, Batut J, Capela D (2008) Auxotrophy accounts for nodulation defect of most *Sinorhizobium meliloti* mutants in the branched-chain amino acid biosynthesis. *Mol Plant Microbe Interact* 21(9):1232–1241

- Dixon R, Kahn D (2004) Genetic regulation of biological nitrogen fixation. *Nat Rev Microbiol* 2:621–631
- Duarte NC, Herrgard MJ, Palsson BØ (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* 14:1298–1309
- Edwards JS, Ramakrishna R, Palsson BØ (2001) Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol Bioeng* 77:27–36
- Felle HH, Kondorosi E, Kondorosi A, Schultze M (1999) Elevation of the cytosolic free $[Ca^{2+}]$ is indispensable for the transduction of the nod factor signal in alfalfa. *Plant Physiol* 121:273–279
- Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE, Gresshoff PM (2010) Molecular analysis of legume nodule development and autoregulation. *J Integr Plant Biol* 52 (1):61–76
- Fong SS, Marciniak JY, Palsson BØ (2003) Description and interpretation of adaptive evolution of *Escherichia coli* K-12 MG1655 by using a genome-scale in silico metabolic model. *J Bacteriol* 185:6400–6408
- Graham PH, Vance CP (2003) Legumes: importance and constraints to greater use. *Plant Physiol* 131:872–877
- Green LS, Emerich DW (1997) The formation of nitrogen-fixing bacteroids is delayed but not abolished in soybean infected by an alpha-ketoglutarate dehydrogenase-deficient mutant of *Bradyrhizobium japonicum*. *Plant Physiol* 114:1359–1368
- Gresshoff PM, Delves AC (1986) Plant genetic approaches to symbiotic nodulation and nitrogen fixation in legumes. In: Blonstein AD, King PJ (eds) *Plant gene research III. A genetical approach to plant biochemistry*. Springer, Wien, pp 159–206
- Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, Vidal M (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 430(6995):88–93
- Hartwell H, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. *Nature* 402:C47–C52
- Jamshidi N, Palsson BØ (2008) Formulating genome-scale kinetic models in the post-genome era. *Mol Syst Biol* 4:171
- Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL (2000) The large-scale organization of metabolic networks. *Nature* 407:651–654
- Jeong H, Mason SP, Barabási AL, Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* 411(6833):41–42
- Jiang G, Krishnan AH, Kim YW, Wacek TJ, Krishnan HB (2001) A functional myo-inositol dehydrogenase gene is required for efficient nitrogen fixation and competitiveness of *Sinorhizobium fredii* USDA191 to nodulate soybean (*Glycine max* [L.] Merr.). *J Bacteriol* 183:2595–2604
- Jones KM, Kobayashi H, Davies BW, Taga ME, Walker GC (2007) How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. *Nat Rev Microbiol* 5(8):619–633
- Larhlimi A, Bockmayr A (2007) Constraint-based analysis of gene deletion in a metabolic network. Workshop on constraint based methods for bioinformatics, WCB'07, Porto, pp 48–55
- Larhlimi A, Bockmayr A (2009) A new constraint-based description of the steady-state flux cone of metabolic networks. *Discrete Appl Math* 157:2257–2266
- Libbenga KR, Harkes PAA (1973) Initial proliferation of cortical cells in the formation of root nodules in *Pisum sativum* L. *Planta* 114:17–28
- Lodwig E, Poole P (2003) Metabolism of *Rhizobium* bacteroid. *Crit Rev Plant Sci* 22:37–78
- Lodwig EM, Hosie AHF, Bourdès A, Findlay K, Allaway D, Karunakaran R, Downie JA, Poole PS (2003) Amino-acid cycling drives nitrogen fixation in the legume-*Rhizobium* symbiosis. *Nature* 422:722–726
- Lodwig EM, Leonard M, Marroqui S, Wheeler TR, Findlay K et al (2005) Role of polyhydroxybutyrate and glycogen as carbon storage compounds in pea and bean bacteroids. *Mol Plant Microbe Interact* 18:67–74

- Luftu-Cakmakci M, Evans HJ, Seidler RJ (1981) Characteristics of nitrogen-fixing *Klebsiella oxytoca* isolated from wheat roots. *Plant Soil* 61:53–63
- Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* 5:264–276
- Marroqui S, Zorreguieta A, Santamaria C, Temprano F, Soberon M, Megias M, Downie JA (2001) Enhanced symbiotic performance by *Rhizobium tropici* glycogen synthase mutants. *J Bacteriol* 183:854–864
- Masson-Boivin C, Giraud E, Perret X, Batut J (2009) Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends Microbiol* 17(10):458–466
- Mathews A, Carroll BJ, Gresshoff PM (1989) Development of Bradyrhizobium infection in supernodulating and non-nodulating mutants of soybean (*Glycine max* [L.] Merrill). *Protoplasma* 150:40–47
- McClean P, Kami J, Gepts P (2004) Genomic and genetic diversity in common bean. In: Wilson RF, Stalker HT, Brummer EC (eds) *Legume crop genomics*. AOCS, Champaign, IL, pp 60–82
- Mendoza A, Leija A, Martinez-Romero E, Hernandez G, Mora J (1995) The enhancement of ammonium assimilation in *Rhizobium etli* prevents nodulation of *Phaseolus vulgaris*. *Mol Plant Microbe Interact* 8:584–592
- Mergaert P, Uchiumi T, Alunni B, Evanno G et al (2006) Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. *Proc Natl Acad Sci USA* 103 (13):5230–5235
- Montañez R, Medina MA, Sole RV, Rodriguez-Caso C (2010) When metabolism meets topology: reconciling metabolite and reaction networks. *Bioessays* 32:246–256
- Newcomb W (1976) A correlated light and electron microscopic study of symbiotic growth and differentiation in *Pisum sativum* root nodules. *Can J Bot* 54:2163–2186
- Newcomb W, Sippel D, Peterson RL (1979) The early morphogenesis of *Glycine max* and *Pisum sativum* root nodules. *Can J Bot* 57:2603–2616
- Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45(2):167–256
- Orth JD, Palsson BØ (2010) Systematizing the generation of missing metabolic knowledge. *Biotechnol Bioeng* 107(3):403–412
- Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? *Nat Biotechnol* 28 (3):245–248
- Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BØ (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism. *Mol Syst Biol* 7:535
- Palsson BØ (2006) *Systems biology: properties of reconstructed networks*. Cambridge University Press, Cambridge
- Palsson BØ (2011) *Systems biology: simulation of dynamic network states*. Cambridge University Press, Cambridge
- Patriarca EJ, Tate R, Iaccarino M (2002) Key role of bacterial NH₄⁺ metabolism in Rhizobium-plant symbiosis. *Microbiol Mol Biol Rev* 66:203–222
- Prell J, Poole P (2006) Metabolic changes of rhizobia in legume nodules. *Trends Microbiol* 14:161–168
- Price ND, Reed JL, Palsson BØ (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature* 2(11):886–897
- Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL (2002) Hierarchical organization of modularity in metabolic networks. *Science* 297(5586):1551–1555
- Redmond JW, Batley M, Djordjevic MA, Innes RW, Kuempel PL, Rolfe BG (1986) Flavones induce expression of nodulation genes in Rhizobium. *Nature* 323:632–635
- Resendis-Antonio O (2009) Filling kinetic gaps: dynamic modeling of metabolism where detailed kinetic information is lacking. *PLoS One* 4(3):e4967
- Resendis-Antonio O, Freyre-Gonzalez JA, Menchaca-Mendez R, Gutierrez-Rios RM, Martinez-Antonio A, Avila-Sanchez C, Collado-Vides J (2005) Modular analysis of the transcriptional regulatory network of *E. coli*. *Trends Genet* 21(1):16–20

- Resendis-Antonio O, Reed JL, Encarnacion S, Collado-Vides J, Palsson BØ (2007) Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. PLoS Comput Biol 3:1887–1895
- Resendis-Antonio O, Hernandez M, Salazar E, Contreras S, Martinez-Batallar G, Mora Y, Encarnacion S (2011) Systems biology of bacterial nitrogen fixation: high-throughput technology and its integrative description with constraint-based modeling. BMC Syst Biol 5:120
- Resendis-Antonio O, Hernandez M, Mora Y, Encarnacion S (2012) Functional modules, structural topology, and optimal activity in metabolic networks. PLoS Comput Biol 8(10):e1002720
- Rodríguez-Plaza JG, Villalon-Rojas A, Herrera S, Garza-Ramos G, Torres Larios A, Amero C, Zarraga Granados G, Gutierrez Aguilar M, Lara Ortiz MT, Polanco Gonzalez C, Uribe Carvajal S, Coria R, Peña Diaz A, Bredesen DE, Castro-Obregon S, del Rio G (2012) Moonlighting peptides with emergent function. PLoS One 7(7):e40125
- Rolfe BG, Gresshoff PM (1988) Genetic analysis of legume nodule initiation. Annu Rev Plant Physiol Plant Mol Biol 39:297–319
- Sarma AD, Emerich DW (2005) Global protein expression pattern of *Bradyrhizobium japonicum* bacteroids: a prelude to functional proteomics. Proteomics 5:4170–4184
- Sarma AD, Emerich DW (2006) A comparative proteomic evaluation of culture grown versus nodule isolated *Bradyrhizobium japonicum*. Proteomics 6:3008–3028
- Schwmtzer CR, Tjepkema JD (1990) The biology of Frankia and actinorhizal plants. Academic, San Diego, CA
- Stumpf MPH, Porter MA (2012) Critical truths about power laws. Science 335:665
- Trainer MA, Charles TC (2006) The role of PHB metabolism in the symbiosis of rhizobia with legumes. Appl Microbiol Biotechnol 71:377–386
- Turgeon BG, Bauer WD (1982) Early events in the infection of soybean by *Rhizobium japonicum*. Time course and cytology of the initial infection process. Can J Bot 60:152–161
- Udvardi M, Day D (1997) Metabolite transport across symbiotic membranes of legume nodules. Annu Rev Plant Physiol Plant Mol Biol 48:493–523
- Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic byproduct secretion in wild-type *Escherichia coli* W3110. Appl Environ Microbiol 60:3724–3731
- Vasse J, de Billy F, Camut S, Truchet G (1990) Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. J Bacteriol 172(8):4295–4306
- Wagner A, Fell DA (2001) The small world inside large metabolic networks. Proc R Soc Lond B Biol Sci 268:1803–1810
- Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. Nature 393 (6684):409–410

Symbiotic Endophytes

Aroca, R. (Ed.)

2013, VIII, 348 p. 36 illus., 16 illus. in color., Hardcover

ISBN: 978-3-642-39316-7