Chapter 7
Future Trends and Perspectives

Abstract New methodologies have been proposed to be incorporated in predictive microbiology in foods and quantitative microbial risk assessment (QMRA) to achieve more reliable models and facilitate predictive model applications. The meta-analysis is one of the proposed strategies focused on a systematic analysis of a large collection of data with the intention of generating standardized and summarized information to produce a global estimate. This data analysis approach can be applied to better understand the relationship between environmental factors and kinetic parameters or to input QMRA studies to assess the effect of a particular intervention or treatment concerning food safety. The emergence of systems biology is also affecting predictive microbiology, offering new and more mechanistic approaches to yield more reliable and robust predictive models. The so-called genomic-scale models are built on a molecular and genomic basis supported by experimental data obtained from the genomic, proteomic, and metabolomic research areas. Although the existing gene-scale models are promising regarding prediction capacity, they are still few and limited to specific model microorganisms and situations. Further research is needed, in the coming decades, to complete omics information and thus to produce more suitable models to be applied to real-world situations in food safety and quality.

Keywords Meta-analysis • Stepwise process • Data bases • Systems biology • Genomic-scale models • Flux balance analysis (FBA) • Network analysis • Objective function

7.1 Introduction

7.1.1 Meta-Analysis Approach and Benchmarking Data

As previously explained, Quantitative Microbial Risk Assessment (QMRA) is an iterative process that gives insight into setting microbiological criteria and

identifying the most relevant factors along the food chain. However, it is recognized that the great amount of data required is the most important drawback to be implemented. Also, as a multidisciplinary area, data processing is becoming more difficult as information is reported in a heterogeneous form. The need to account for variability and uncertainty sources together with the characterization of the main statistical distributions to describe the data leads to the creation of alternative tools to integrate these findings and provide a global estimate. A meta-analysis is a systematic analysis of a large collection of data from individual studies aiming to integrate the information generated in a QMRA study and to produce a global estimate of the effect of a particular intervention or treatment (van Besten and Zwietering 2012). This technique has been more extensively used in food microbiology and can give an improved understanding of main and side effects on microbiological kinetics (Ross et al. 2008).

To start the application of a meta-analysis approach, a sufficient number of data should be generated. Gonzales-Barron and Butler (2011) suggested a stepwise procedure to meta-analysis consisting of (1) systematic review; (2) data extraction to collate quantitative and qualitative information from the primary studies; (3) selection of the appropriate effect size parameter to describe, summarize, and compare the data of the primary studies, and when needed, subsequent translation of the reported findings of the individual studies into the parameter; (4) estimation of the overall effect size by combining the primary studies; (5) assessment of heterogeneity among the studies; and, finally, (6) the presentation of the meta-analysis results.

Selection of data coming from primary studies can begin with experimental data from research institutions or extra data available in scientific data bases. However, individual results must be incorporated into the meta-analysis when they are properly defined, structured, and transparently reported.

In the systematic review process the information to be included in the meta-analysis has to be sufficiently accurate to answer the embedded question of a given case study. For instance, in a lettuce disinfection process, one can measure several heads of lettuce to see if there is contamination by *Escherichia. coli*. The data introduced in the meta-analysis approach should justify if the intervention (disinfection) makes a causal inference on the outcome (presence/absence of *E. coli*) and, if so, how large the effect is.

The data extraction from the primary studies should provide the information necessary for summarizing and synthesizing the results and include both numeric and nonnumeric data.

Effect size refers to the degree to which the phenomenon is present in the population (reduction of *E. coli* numbers by disinfection). For the primary studies, meta-analysis converts the effect size into a 'parameter' that allows direct comparison and summation of the primary studies. There are many types of effect size parameters: (1) binary or dichotomous, for example, indicating the presence or absence of the event of interest in each subject, (2) continuous, and (3) ordinal, where the outcome is measured on an ordered categorical scale.

For the estimation of the overall size effect, primary studies may be weighted to reflect sample size, quality of research design, or other factors influencing their

reliability. A relevant factor in precision is the sample size, with larger samples yielding more precise estimates than smaller samples. Another factor affecting precision is the study design, with matched groups yielding more precise estimates (as compared with independent groups) and clustered groups yielding less precise estimates. This consideration can also imply that the obtaining of a lower variance in the primary studies leads to a more accurate global estimate.

On the other hand, it is necessary to make a heterogeneity test among the primary samples to assess the extra-variation in the meta-analysis approach. Generally speaking, individual samples are weighted and statistically compared with aiming at quantifying the variability associated to heterogeneity. In food microbiology, most of the microbial data have been generated in culture media and the effect of environmental factors may not necessarily reflect what might happen in an actual food. Also, results of different studies on factors influencing microbial kinetics are not always similar or may be even contradictory. Variations among microbial strains, individual cell studies, or model estimations contribute positively to increase variability in results. Therefore, quantitative information about the influence of various factors on microbial kinetics is often not adequate under specific conditions, and also often is not available in the published literature.

Finally, results coming from the meta-analysis are presented into several graph types, such as bubble plots, which display point estimates and confidence intervals of each primary study and the overall effects in the global estimate.

The use of data bases in predictive microbiology can provide thousands of records of microbial growth or inactivation kinetics under a wide range of environmental conditions. A systematic and critical analysis of the literature followed by integration of the gathered data results in global estimates of kinetic parameters with their variability, and these can be used to benchmark the latest published data (van Asselt and Zwietering 2006). Meta-analysis has been used in various QMRA studies for relating the microbial concentration of a given hazard to a public health outcome (Pérez-Rodríguez et al. 2007b). However, large variability sources are expected in some cases, mainly because of heterogeneity in primary data. Additionally, overlapping problems are generated when the same information of one variable is obtained from different studies. In spite of these disadvantages, when a large dataset is manipulated, meta-analysis can provide useful links to discern between explanatory variables on the global estimate. The construction of updated data bases on the reviewed question or parameter can also reveal the present knowledge, can highlight default areas where there is a lack of information on factors that might affect the parameter of interest, and can therefore provide direction for future research.

7.2   Mechanistic Predictive Models

Advances in molecular biology, particularly in genome sequencing and high-throughput measurements, enable us to obtain comprehensive data on the cellular system and gain information on the underlying molecules (Kitano 2002).

This genomics revolution has in the past years provided researchers with the option to look genome wide for cellular responses at the level of gene expression (Keijser et al. 2007) and protein presence (Wolff et al. 2006; Hahne et al. 2010). The need of integrating all this complex information has contributed to an emerging scientific field, so-called systems biology, aimed at understanding complex biological systems at the systems level (Kitano 2001). The fundamental idea behind the systems biology approach is that biological systems are hierarchically organized with influences going both up and down through the hierarchy (Brul et al. 2008).

The great avalanche of 'omics' data (i.e., genomic and proteomic data) in systems biology necessitates applying mathematical methods to better understand the interactions and relationships among the different elements within the studied system (Fig. 7.1). Stelling (2004) classified mathematical models applied in systems biology in interaction-based models, constraint-based models, and mechanism-based models. The interaction-based models refer to network topology analysis in which interactions between the different elements in the system, for example, metabolic reactions, protein–protein interactions, and gene regulation, are accounted for by graphical networks. In constraint-based models, physicochemical properties such as reaction stoichiometries and reversibilities impose constraints on network function in addition to network topology. This network reconstruction process ultimately results in the generation of a biochemically, genomically, and genetically (BiGG) structured data base that can be further utilized for both mathematical computation and analysis of high-throughput data sets. The network spans the set of metabolic reactions taking place in a specific biological system, assuming a stationary state (Hertog et al. 2011) in which each reaction is referred to as a flux. The methodologies developed in metabolic engineering such as metabolic control analysis and metabolic flux analysis are applied to analyze steady-state fluxes, although these may also be used to explain oscillatory systems so long as average fluxes are considered (Schuster et al. 2002). More quantitative models can be addressed based on kinetic rates of metabolic reactions included in the biological networks. In this approach, a system of linear differential equations is used to account for reactions rate of the quasi-dynamic or dynamic state fluxes (Hertog et al. 2011). As new genomic data become available, these may aid in the parameterization of metabolic models (Voit 2002). However, one weakness of this approach is that it ignores the variability and noise found in biological networks, which may have important implications in their function (Heath and Kavraki 2009). To overcome this limitation, a stochastic approach has been proposed that basically consists of adding a noise term to the differential equations. Similarly, gene expression regulation (i.e., transcription and translation) and signaling networks have a probabilistic nature that should be accounted for by applying a stochastic approach (Treviño Santa Cruz et al. 2005; McAdams and Arkin 1997).

The latter type of model mentioned by Stelling (2004) refers to mechanism-based models. The author means that with this type of model, models can predict the system dynamics by integrating detailed mechanisms operating in metabolism, signal processing, and gene regulation. The success of this mechanistic approach,
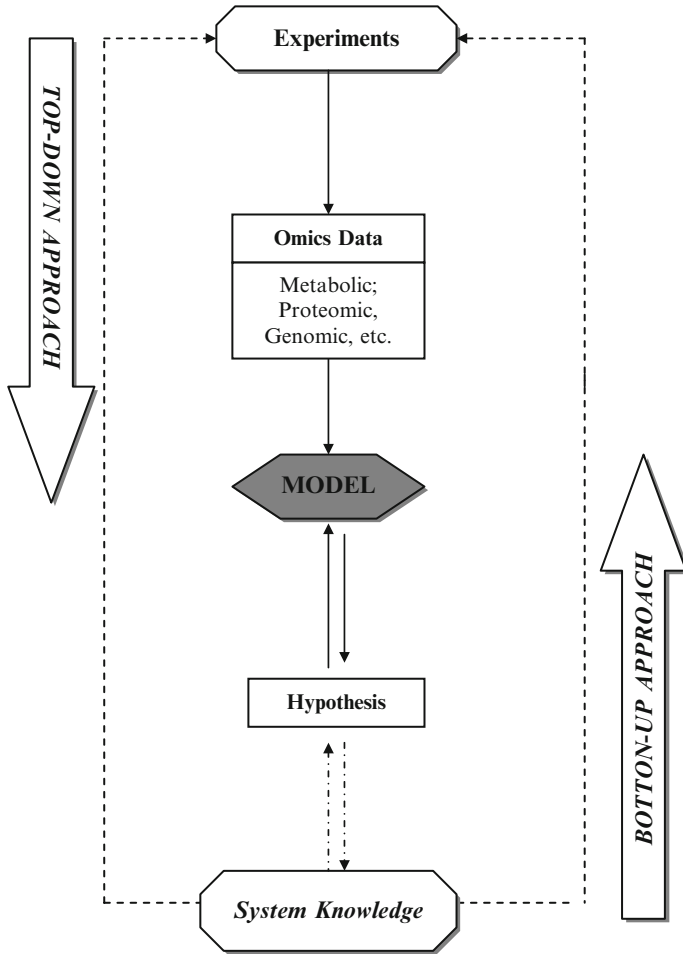
Fig. 7.1 Scheme of the workflow applied to systems biology

that is, integrated modeling, relies largely on the availability of information about the complete mechanism and attendant parameters.

Traditionally, in the field of predictive microbiology applied to foods, the scarce information on the mechanisms involved in the cellular functions has hampered microbiologists from undertaking more mechanistic models, albeit some mechanistic parameters has been introduced in specific cases (Baranyi and Roberts 1995). The emergence of systems biology is creating a new path for microbiologists in predictive microbiology, offering new and more mechanistic approaches to give rise to more reliable and robust models (Brul et al. 2008). In so doing, predictive microbiology will be able to move from the most used empirical modeling, that is, black box models, toward so-called white or gray box models, based on an better

understanding of the biological functions in cells, enabling providing more accurate predictions under specific physical and chemical changes and even extending the model outside the range of space bounded by observations. Such data not only allow for a better fine-tuning of growth/no growth boundaries but will also begin to strengthen die-off/survival models (Brul et al. 2008). Several computer models have been developed on the basis of information derived from systems biology studies and wealthy databases. However, many of the mechanistic studies have been done under conditions and in model microorganisms with relatively low practical relevance (Brul et al. 2008). One of the most studied microorganisms is *E. coli*, as much is known about its metabolism, regulation, and genome, enabling the development of more mechanistic and reliable in silico models for this model microorganism (Reed and Palsson 2003). The experience obtained with *E. coli* has served to be applied to other microorganisms such as *Haemophilus influenzae* (Edward and Palsson 1999), *Helicobacter pylori*, and *Saccharomyces cerevisiae* (Petranovic and Vemuri 2009).

To date, few systems biology-based models have been explored or developed within the area of predictive microbiology in foods (Brul et al. 2008). However, constraint-based models seem to be the first choice by microbiologists to understand the behavior of microorganisms in food-related environments (Métris et al. 2011; Peck et al. 2011). The most significant kinetic reactions constituting the metabolism of the model bacterium are modeled and simulated to know which specific metabolic processes are related to a determined bacterial response (e.g., outgrowth, adaptation, survival). These models consist of describing the fluxes that make up a metabolic network in which each flux accounts for a metabolic reaction as concentration change per time unit for the substrate and product. The reactions can be described by a system of linear differential equations in which stoichiometric coefficient of equations are assumed to be constant because the model represents

$$\frac{dx}{dt} = Sv = 0 \tag{7.1}$$

Here, x defines a vector of the intermediate concentrations of metabolites at a specific time, S is the stoichiometric matrix describing all the metabolic reactions, and S_{ij} corresponds to the ith stoichiometric coefficient in the jth reaction. The thermodynamic constraints and enzyme capacity constraints are represented by vector $v = [v_1, \ldots v_j]$, which includes the reaction rates of each metabolic reaction or flux. Setting Eq. (7.1) to 0 means that conservation laws apply in the production and consumption rates (i.e., $rate_{consumption} = rate_{production}$). A simplified example of the steady-state flux might be the well-known coenzyme nicotinamide adenine dinucleotide (NAD), involved in many metabolic routes as an electron donor. In this case, the reaction would be

$$NAD^+ + H \rightarrow NADH$$
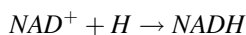
According to the law of conservation, NAD + NADH = a constant, which means that the sum of concentrations of NAD and NADH does not change with time.

The derivation of the reaction rate equations is another important aspect and should be based on an appropriate metabolic network, which should be completely known and closed. The quasi-steady-state and rapid equilibrium approaches can be used to obtain the reaction rate equations. With regard to the kinetic parameters, these might be estimated by using sources such as literature data, electronic data bases, experimental data for dependencies between initial reactions rates and products, inhibitors, substrates, and activators, and finally time-series data for enzyme kinetics and whole pathways (Demin et al. 2005).

Because the system of equations has more fluxes than metabolites, the system is underdetermined (Kauffman et al. 2003), which means that the system has multiple solutions. To reduce the solution space of the system, the model is constrained by imposing different rules, which are often related to thermodynamic feasibility, enzymatic capacity, and mass balance. Model solutions that do not comply with such criteria are excluded from the solution space of the model (Reed and Palsson 2003). Once constraints are defined for the model, the corresponding solution space should be determined. To this end, several mathematical approaches can be taken such as linear optimization, elementary modes and extreme pathways, phenotypic phase plane analysis, gene deletions, or finding objective functions. The linear optimization, which is referred to as flux balance analysis (FBA), is based on an objective function, which is utilized to define the solution space by maximizing or minimizing the defined objective function (Feist and Palsson 2010; Varma and Palsson 1994). The most used objective functions include ATP production, production of a specific by-product, and biomass production (i.e., growth rate) (Van Impe et al. 2011; Reed and Palsson 2003). In this respect, using a biomass production objective function can accurately estimate the growth rate of *E. coli*, as evidenced by the work by Feist et al. (2007).

Métris et al. (2011) performed in silico simulations based on the model of *E. coli* K12 MG1655 previously developed by Feist et al. (2007) considering 1,387 metabolic reactions and 1,260 genes. This study can be considered as one of the first approaches of predictive microbiology in the foods area to systems biology modeling. This model applied the most often used objective function based on optimizing the biomass production, which is associated with growth-associated maintenance (GAM) energy and non-growth-associated maintenance (NGAM) energy. Their values were derived from experiments in a chemostat without added NaCl, which we refer to as the control conditions (Feist et al. 2007). The model was modified to consider exposure to osmotic stress by including changes of concentrations in an osmoprotectant associated with osmolarity changes. The work did not find definitive results relating the changes of these substances with a decrease of the growth rate. Similarly, the model was tested to ascertain if biomass composition derived from osmotic stress might explain the decrease of growth rates observed in experiments; however, again the results were not conclusive. Finally, the authors suggested that more specific objective functions should be developed to explain the chemicophysical limitations of the growth rate. For that, authors suggested including

gene regulation, crowding, and other additional cell resources such as ribosomal content and some tradeoff observed under osmotic stress. This work and its results provide evidence that a new modeling approach is emerging, although still with important gaps and limitations. Nonetheless, it might provide the necessary theoretical basis to develop more mechanistic predictive models in foods (Fig. 7.1).