



## Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks

**Saa, Pedro A.; Nielsen, Lars Keld**

*Published in:*  
Biotechnology Advances

*Link to article, DOI:*  
[10.1016/j.biotechadv.2017.09.005](https://doi.org/10.1016/j.biotechadv.2017.09.005)

*Publication date:*  
2017

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*

Saa, P. A., & Nielsen, L. K. (2017). Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. *Biotechnology Advances*, 35(8), 981-1003. DOI: 10.1016/j.biotechadv.2017.09.005

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

ACCEPTED MANUSCRIPT

**Abstract**

Kinetic models are critical to predict the dynamic behaviour of metabolic networks. Mechanistic kinetic models for large networks remain uncommon due to the difficulty of fitting their parameters. Recent modelling frameworks promise new ways to overcome this obstacle while retaining predictive capabilities. In this review, we present an overview of the relevant mathematical frameworks for kinetic formulation, construction and analysis. Starting with kinetic formalisms, we next review statistical methods for parameter inference, as well as recent computational frameworks applied to the construction and analysis of kinetic models. Finally, we discuss opportunities and limitations hindering the development of larger kinetic reconstructions.

**Keywords**

Kinetic models, metabolic modelling, enzyme kinetics, parameter estimation, metabolic control analysis, Monte Carlo simulation.

ACCEPTED MANUSCRIPT

## 1. Introduction

Mathematical models are essential to broaden our knowledge of metabolic networks. They provide a rational and systematic framework for integrating existing biological knowledge with experimental data, thus enabling appraisal of the complex regulation underpinning the operation of metabolism. During the past decades, several modelling frameworks have been developed for predicting the dynamic behaviour of cellular metabolism supported by the rapid progress in high-throughput omics data generation and advanced metabolic engineering techniques (Chowdhury et al., 2015). The ultimate goal is to integrate these data with mechanistic models to increase our understanding about metabolic networks as well as the information content of the models.

Metabolic network models are described by the set of biochemical reactions mediated by enzymes. Enzymes are proteins whose expression is determined by the genetic program of the cell under specific environmental conditions. The presence of a specific enzyme in the genome implies that a cell has the metabolic capability of the corresponding biotransformation. Annotation and assembly of the repertoire of Gene-Protein-Reactions (GPR) associations from genome sequences and multiple data sources (Fig. 1A) constitutes then a formal representation of the metabolic potential of the cell (Fig. 1B). Subsequent integration of different omics data onto the reconstructed network produces a metabolic model amenable to structural and dynamic analyses (Fig. 1C). Structural analysis relies solely on reaction stoichiometries under steady-state and constitutes the basis of parameter-free Constrained-Based Modelling (CBM) methods (Fig. 1D) (for a detailed review refer to Lewis et al. (2012)). These methods have generated fundamental biological insights into the operation of metabolic networks (Ibarra et al., 2002; Schuetz et al., 2012) as well as great advances in biotechnological applications (Lee and Kim, 2015; Yim et al., 2011). By applying various optimization (Burgard et al., 2003; Edwards et al., 2001; Mahadevan and Schilling, 2003) and sampling methods (Almaas et al., 2004; Saa and Nielsen, 2016b), CBM methods enable exploration of the space of feasible metabolic states allowed by the network structure and physiological constraints. The latter methods are however of limited use for the prediction of how metabolic states are achieved, as they lack kinetic information. In contrast, kinetic models of metabolism explicitly describe reaction fluxes as a function of metabolite and enzyme concentrations, enabling dynamic interrogation and quantitative integration of metabolomic, proteomic and transcriptomic data (Fig. 1E).

**Fig. 1.** Model-centric workflow for metabolic networks reconstruction and analysis. **A** Network reconstruction starts with the annotation of the genome sequence with the encoded metabolic enzymes. Relationships between gene, proteins and reactions are stored in GPR (Gene-Protein-Reaction) associations, enabling rational representation of the biological information flow. Discrepancies and missing information are resolved (where possible) with the support of data from the literature and comprehensive databases (e.g., BRENDA (Placzek et al., 2016), MetaCyc (Caspi et al., 2016), KEGG (Kanehisa et al., 2016)) in a thorough process of manual curation. **B** The set of curated enzymatic reactions determines the spectrum of metabolic capabilities of the cell. This spectrum is mathematically described by the stoichiometry of the biochemical reactions and defines the topology of the reconstructed network. **C** Integration of diverse ‘omics’ data with the metabolic reconstruction enables construction of a metabolic model amenable for rational interrogation and biological discovery. **D** Structural analysis of the metabolic model is readily achieved using constrained-based modelling methods. Stoichiometric, thermodynamic and kinetic (capacity) constraints defines the space of possible network states which can be readily explored using state-of-the-art optimization and/or sampling methods. This structural analysis is however static, and, it does not quantitatively explain how fluxes are achieved. **E** Inclusion of kinetic descriptions for the all the

enzymes in the model enables prediction of metabolic states as well as dynamic interrogation of the system. The resulting kinetic model can reconcile higher amounts of data; however, it requires substantially more information for its construction.

Metabolic reactions in kinetic models are described by disparate non-linear rate laws, typically involving highly parameterized mathematical expressions. Early modelling efforts proposed different simplified kinetic formalisms to simplify their structure and ease parameter fitting from *in vivo* data (Hatzimanikatis and Bailey, 1997; Savageau, 1969; Visser and Heijnen, 2003). Although these efforts have yielded valuable insights about the design principles (Savageau et al., 2009) and the dynamic behaviour of different metabolic (Alvarez-Vasquez et al., 2005; Visser et al., 2004), signalling (Vera et al., 2007) and even genetic systems (Atkinson et al., 2003), most of their predictions are inherently limited to the proximity of the chosen operation point. Furthermore, these models ignore thermodynamic relationships between parameters, often rendering unrealistic behaviours. Conversely, mechanistic-based rate laws are thermodynamically consistent and hold greater prediction power; however, they require a substantial volume of data to fit a multitude of parameters. Approximate mechanistically-inspired formalisms have alleviated these issues to some extent (Ederer and Gilles, 2007; Hofmeyr and Cornish-Bowden, 1997; Liebermeister and Klipp, 2006a; Liebermeister et al., 2010), however they still inevitably suffer from parameter identifiability issues (Heijnen and Verheijen, 2013). Fitting mechanistically-grounded kinetic models using conventional methods has previously been deemed impracticable, as homeostatic control often renders several parameters highly correlated or even outright unidentifiable even in the presence of large amounts of data (Degenring et al., 2004; Hadlich et al., 2009). However, recent Monte Carlo and other simulation-based strategies for kinetic model construction and analysis (Bordbar et al., 2015; Chakrabarti et al., 2013; Saa and Nielsen, 2016a; Steuer et al., 2006; Tran et al., 2008) have shown that satisfactory predictions can be achieved even when many parameters are poorly resolved and/or uncertain. Indeed, scrutiny of most kinetic models has revealed that good predictions do not necessarily require precise parameters (Gutenkunst et al., 2007). As such, modelling frameworks are moving away from precisely fitting coarse grained models, towards building mechanistic models capable of identifying emergent regulatory and dynamic behaviours (Link et al., 2014).

The potential key role of kinetic models in the field of systems biotechnology is certainly undeniable. These models are the only capable of reconciling the multiple layers of omics data, i.e., transcriptomics/proteomics, metabolomics and fluxomics, within a common and coherent mathematical framework. Recent examples of the application of these models include strain design and optimization (Andreozzi et al., 2016a; Khodayari and Maranas, 2016; Savoglidis et al., 2016), identification of drug targets and side effects (Bordbar et al., 2015; Haanstra et al., 2017; Murabito et al., 2011), unravelling key regulatory interactions (Link et al., 2013; Saa and Nielsen, 2016a), to name a few. These case studies constitute a first glance of the potential applications of kinetic models, which justifies the renewed interest of the scientific community in these models. Supported by advanced frameworks for kinetic modelling, kinetic models are increasingly deployed to understand complex metabolic phenotypes.

This review presents a comprehensive overview of mathematical frameworks for kinetic modelling, starting with the relevant formalisms used to describe enzyme-catalysed reactions. We next review relevant classical statistical methods for parameter inference, as well as more recent computational frameworks specific for the analysis of kinetic models.

Considering the importance and potential applications of the latter, we focused our attention on these and critically reviewed their main features and capabilities. Finally, we discuss current limitations hindering the development of larger and more detailed kinetic models. From incomplete *a priori* knowledge of regulatory interactions, to the integration of diverse regulatory events at transcriptional, translational, post-translational, and metabolic (i.e., allosteric) level, we have highlighted the main theoretical as well as practical challenges limiting further progress. Although systematic construction of genome-scale kinetic models has not been achieved yet – despite previous speculations (Stanford et al., 2013) –, recent progress in computational frameworks added to the increasing availability of comprehensive multi-omics datasets (e.g., Buescher et al. (2012), Hackett et al. (2016) and Ishii et al. (2007)) suggest that attainment of this goal is close.

## 2. Mathematical formalisms for describing enzymatic reactions

The dynamic behaviour of metabolic networks is represented by the set of Ordinary Differential Equations (ODEs) describing the mass balances for the reacting species in the system,

$$\frac{d\mathbf{x}}{dt} = \mathbf{S} \cdot \mathbf{v}(\mathbf{E}; \mathbf{x}; \mathbf{k}) \quad , \quad \mathbf{x}(0) = \mathbf{x}_0 \quad (1)$$

where  $\mathbf{S}$  and  $\mathbf{v}$  denote the stoichiometric matrix and the vector of metabolic reactions or fluxes, respectively. While  $\mathbf{S}$  is readily reconstructed from genomic information supported by appropriate literature data,  $\mathbf{v}$  depends on the metabolite concentrations ( $\mathbf{x}$ ) as well as kinetic parameters ( $\mathbf{k}$ ) and enzyme concentrations ( $\mathbf{E}$ ). In Eq. 1,  $\mathbf{x}_0$  describes the initial conditions for the metabolite concentrations. A common assumption for Eq. 1 is the steady state condition, i.e.,  $\mathbf{S} \cdot \mathbf{v} = \mathbf{0}$ , whose solution yields the metabolic state of the cell.

The rate of an enzyme-catalysed reaction ( $v_i$ ) depends proportionally on the enzyme concentration (or capacity)  $E_i$  as well as on a complex catalytic term  $f_i$  function of metabolic reactants  $\mathbf{x}_M$ , effectors  $\mathbf{x}_E$  and kinetic parameters  $\mathbf{k}_i$ . The product of the above factors describes the enzyme activity responsible for the metabolic flux, i.e.,  $v_i = E_i \cdot f_i(\mathbf{x}_M; \mathbf{x}_E; \mathbf{k}_i)$ . Since the development of the Law of Mass Action in the middle of the 19th century, many different mathematical formalisms have been proposed and used to describe enzymatic reactions (Fig. 2A). Over time, formalisms have favoured either mechanistic or practical considerations. For instance, initial rate laws were mainly based on a defined mechanism of action grounded on first principles and their purpose was to rationally explain fundamental kinetic behaviours and observations. These formalisms typically employed several parameters, and thus, their application was limited to those sufficiently well-studied systems where enough data was available. In contrast, recent efforts have focussed on the development of more compact (approximate) kinetic formulae that ease parameter fitting while retaining reasonable predictive power. If such approximate rate law is purely based on the metabolic interaction network, then the latter can be formulated in a common or ‘canonical’ format. On the other hand, if the simplified rate law is the result of simplifying assumptions of a mechanistic rate law, then the formalism is considered to be approximate. The selection of a canonical, approximate or mechanistic kinetic formalism depends on the

specific application, modelling assumptions and availability of data (Fig. 2B). For instance, while large kinetic models can be readily constructed and fitted using limited data employing canonical formalisms, their predictive power is typically restricted to the neighbourhood around an ‘operating point’. In contrast, mechanistic and approximate formalisms enjoy a wider applicability range, but they require substantial data to fit their kinetic parameters (particularly the former). In the following, we reviewed prevailing kinetic formalisms used to describe the range of kinetics encountered in nature. In particular, we focused our attention to the formalisms most often used to model enzyme-catalysed reactions in metabolic models.

**Fig. 2.** Historical development and key features of popular mathematical formalisms for describing enzyme kinetics. **A** While during the late 40s to the early 70s considerable efforts were devoted to formulating more general mechanistic-based formalisms to explain fundamental kinetic behaviours, since the late 70s efforts have been focused on developing more compact approximate formulations to ease the construction of useful kinetic models. **B** Depending on the degree of detail, kinetics supported and modelling assumptions, different formalisms can be more appropriate for a given task provided a certain amount of data for fitting a multitude of parameters. The validity of the selected formalism depends largely on the intended use of the modeller. Abbreviations: MWC, Monod-Wyman-Changoux; GMA, Generalized Mass Action; KFN, Koshland-Némethy-Filmer; TKM, Thermodynamic-Kinetic Modelling.

## 2.1. Canonical rate laws

Canonical rate laws rely exclusively on the structure of the interaction network, and thus, they are readily written in a standard form for any enzyme-catalysed reaction. These rate laws are typically formulated as power-law or log-linear functions based on linear Taylor approximation. In the following, we revised the most relevant canonical formats used for modelling enzymatic reactions.

### 2.1.1. Mass Action

The Law of Mass Action (LMA) dates back to 1864 and it was first proposed by Cato M. Gulberg and Peter Wagee (Guldberg and Waage, 1864). Its postulates were based on the previous ideas of Claude L. Berthollet, and it was later re-discovered by Henricus van 't Hoff in 1877 (Voit et al., 2015). This law states that the rate of chemical conversion is proportional to the product of the reactants concentration each raised to the power of the (integer) stoichiometric coefficient of the chemical equation. For example, let us consider the reversible reaction  $\alpha_S S \leftrightarrow \alpha_P P$ , then the rate of product formation and consumption of reactant,

$$\begin{aligned} v_i^+ &= k_i^+ \cdot x_S^{\alpha_S} \\ v_i^- &= k_i^- \cdot x_P^{\alpha_P} \end{aligned} \tag{2}$$

where  $k_i^+$  and  $k_i^-$  denote respectively the forward and reverse rate constants, and  $\alpha_S$  and  $\alpha_P$  are the stoichiometric coefficients of the chemical equation (also known as reaction orders).



In the context of enzymatic reactions, both rate constants are proportional to the enzyme concentration  $E_i$ . The net (apparent) rate of reaction is given by  $v_i = v_i^+ - v_i^-$ , and at equilibrium, i.e.,  $v_i = 0$ ,  $K^{\text{eq}} = P^{\text{eq}}/S^{\text{eq}} = k_i^+/k_i^-$ , where  $K^{\text{eq}}$  is the equilibrium constant. The existence of a positive  $K^{\text{eq}}$  implies  $k_i^+ > 0$  and  $k_i^- > 0$  (reversibility), such that the principle of detailed balance is obeyed (Tolman, 1938), i.e., all fluxes must vanish at equilibrium. Because equilibrium depends exclusively on the reaction thermodynamics, i.e.,  $\Delta G_r^\circ = -\bar{R}T \ln(K^{\text{eq}})$  where  $T$  is the temperature in Kelvin and  $\bar{R}$  is the universal gas constant, the LMA requires just one kinetic parameter (e.g., forward rate constant) to describe the velocity rate of any reaction with known thermodynamics. This convenient feature has been exploited to construct large-scale kinetic models for dynamic analysis (Jamshidi and Palsson, 2008, 2010) and unravelling drug side effects (Bordbar et al., 2015) (see Section 4.1.3. MASS framework for further details).

Although mass action kinetics is thermodynamically feasible and coherent, it ignores basic kinetic behaviours – e.g., enzyme saturation, activation and inhibition – when formulated solely based on the chemical reaction stoichiometry. In contrast, application of this law to the enzyme-catalysed elementary steps or elementary reactions provides a correct description of the reaction rate as well as appropriate constraints on the rate constants (Horn and Jackson, 1972). In the context of biochemical reactions, however, ‘mass action kinetics’ have historically been formulated based on the reaction stoichiometry rather than on the reaction mechanism (Bailey and Ollis, 1977). As the more general formalism is based on modelling elementary reactions describing events at molecular level, we have designated this formalism Elementary Reaction Mass Action (ER-MA). Owing to the mechanistic character of the latter formalism, its features are described in the Mechanistic rate laws Section 2.2.1.

### 2.1.2. Generalized Mass Action and S-system models

Generalized Mass Action (GMA) and S-system models are two very similar formalisms based on power law approximations for describing non-linear reacting systems. Both approaches are part of the Biochemical System Theory (BST) of Savageau (1969), and substantially expand the range of metabolic regulations supported in LMA. The canonical reaction rate for these two formalisms is given by,

$$v_i = E_i \cdot k_i^+ \cdot \prod_{j=1}^{m_S} x_{S,j}^{h_{i,j}} \cdot \prod_{l=1}^{m_E} x_{E,l}^{h_{i,l}} \quad (3)$$

where  $h_{i,j}$  and  $h_{i,l}$  are the kinetic orders of metabolites  $j$  and  $l$ , and  $m_S$  and  $m_E$  represent respectively the number of substrates and effector (regulatory) metabolites acting on the reaction. The kinetic orders are analogous to reaction orders; however they do not necessarily have to be integer numbers. Positive values for  $h_{i,j}$  and  $h_{i,l}$  describe an activating effect by substrates or allosteric activators, whereas negative or zero values describe inhibition (e.g., inhibiting product or allosteric inhibitor) or no effect on the reaction, respectively. As in LMA, the reaction rate is proportional to the enzyme concentration, which is made explicit in Eq. 3. For reversible reactions, linear combination of two power expressions may be used for formation and consumption.

Although both GMA and S-system models are formulated based on the known structure of the network and share a common mathematical expression for the reaction rates, they differ in how reaction fluxes are aggregated at branch points (Savageau, 1988). While in S-system models the rate of accumulation of the split metabolite is computed as the difference between all producing and consuming reactions – where each term is written as a single product of power law functions – in the case of GMA, the respective accumulation is expressed as the signed sum of the individual producing and consuming reactions each in power law format (Savageau, 1988). The S-system aggregation approximation introduces only small errors even when metabolite concentrations differ substantially from the reference state (Voit, 2000). The latter does not typically hold when enzyme activities are greatly varied (Heijnen, 2005). On a positive note, S-system models substantially simplifies the right-hand side of Eq. 1 to the point where analytical solutions can be obtained for simple networks, and most importantly, reliable parameter estimation methods are applicable (Chou et al., 2006). Owing to their simplicity, these formalisms have been readily applied to study the dynamic behaviour of metabolic networks (Alvarez-Vasquez et al., 2005) and genetic circuits (Atkinson et al., 2003). For an exhaustive list of applications the reader is referred to Chou et al. (2009).

A note of concern of these parameterizations is related to their kinetic and thermodynamic plausibility. Normalization of Eq. 3 to a measured operation point or reference state  $\mathbf{y}_i^{\text{ref}} = (E_i^{\text{ref}}, v_i^{\text{ref}}, \mathbf{x}_M^{\text{ref}}, \mathbf{x}_E^{\text{ref}}, \boldsymbol{\varepsilon}_i^{\text{ref}})$  followed by log-transformation yields the following expression,

$$\ln \frac{v_i}{v_i^{\text{ref}}} = \ln \frac{E_i}{E_i^{\text{ref}}} + \sum_{j=1}^{m_M} \varepsilon_{M,i,j}^{\text{ref}} \ln \frac{x_{M,j}}{x_{M,j}^{\text{ref}}} + \sum_{l=1}^{m_E} \varepsilon_{E,i,l}^{\text{ref}} \ln \frac{x_{E,l}}{x_{E,l}^{\text{ref}}} \quad (4)$$

where  $v_i^{\text{ref}}$  denotes the reference reaction flux and,  $\varepsilon_{M,i,j}^{\text{ref}}$  and  $\varepsilon_{E,i,l}^{\text{ref}}$  represent metabolic reactants (i.e., substrates and products) and effector elasticities for reaction  $i$  at the reference point. In Eq. 4, kinetic orders have been replaced by elasticity parameters that represent the (normalized) sensitivity of the reaction rate to changes in the concentration of a metabolite holding all else constant. A general property of enzyme kinetics is the decreasing concentration elasticity as metabolite concentration increases (Cornish-Bowden, 2012). Given that the elasticities are independent of metabolite concentrations and equal to the reference state, this basic kinetic behaviour is not supported by these rate laws. Additionally, the lack of thermodynamic constraints on their parameters allows the possibility of a chemical perpetuum-mobile in violation of the second law thermodynamics. In summary, these formalisms provide simple expressions suitable for rapid fitting methods, but their prediction power and validity range are inherently narrow.

### 2.1.3. Log-lin and Lin-log kinetics

The first logarithmic format for modelling rate laws was proposed by Hatzimanikatis and Bailey (1997) based on a (log)-linear approximation of the dynamic mass balance in Eq. 1. By using a Taylor approximation, the authors derived a linear (in log-quantities) system of equations that is analytically solvable and amenable for stability and steady-state analysis.

Due to its mathematical features, the formalism was denominated Log-lin (Eq. 5). Reaction rates can be formulated directly from the topology and regulation of the metabolic network, displaying a simple canonical form,

$$\frac{v_i}{v^{\text{ref}}} = 1 + \ln \frac{E_i}{E_i^{\text{ref}}} + \sum_{j=1}^{m_M} \varepsilon_{M,i,j}^{\text{ref}} \ln \frac{x_{M,j}}{x_{M,j}^{\text{ref}}} + \sum_{l=1}^{m_E} \varepsilon_{E,i,l}^{\text{ref}} \ln \frac{x_{E,l}}{x_{E,l}^{\text{ref}}} \quad (5)$$

Equation 5 accommodates large changes in relative metabolite concentrations well, but fails to accurately capture the dynamics upon moderate enzyme concentration changes ( $> \pm 35\%$ ) as a consequence of its log-format (Heijnen, 2005). Notably, the reaction rate is only approximately proportional to relative changes in enzyme concentration when metabolite concentrations are held constant, which goes against fundamental kinetic evidence. On the other hand, thanks to its convenient mathematical form, the framework can be readily extended to the calculation of control properties of linear and branched metabolic pathways (Hatzimanikatis and Bailey, 1997), as well as to the exploration of regulatory interactions capable of improving metabolic performance using elegant optimization formulations (Hatzimanikatis et al., 1996).

The later log-linear format known as Lin-log and proposed by Visser and Heijnen (2003) addressed some of the issues of the Log-formalism using a similar mathematical structure (Eq. 6). Although fairly similar, the Lin-log mathematical form has its basis in previous results on the behaviour of the reaction rate as a linear function of reaction affinity (Rottenberg, 1973; van der Meer et al., 1980), and on the reaction rate dependency on possibly additional metabolic effectors (Nielsen, 1997; Westerhoff and Van Dam, 1987).

$$\frac{v_i}{v^{\text{ref}}} = \frac{E_i}{E_i^{\text{ref}}} \left( 1 + \sum_{j=1}^{m_M} \varepsilon_{M,i,j}^{\text{ref}} \ln \frac{x_{M,j}}{x_{M,j}^{\text{ref}}} + \sum_{l=1}^{m_E} \varepsilon_{E,i,l}^{\text{ref}} \ln \frac{x_{E,l}}{x_{E,l}^{\text{ref}}} \right) \quad (6)$$

As for Log-lin kinetics, the Lin-log approach provides analytical steady-state solutions for Eq. 1, enabling easy computation of control properties (Visser and Heijnen, 2003). In addition, the Lin-log approach has been shown to hold a wider validity range for relative changes in enzyme and metabolite concentrations than Log-lin (Heijnen, 2005), as well as a greater theoretical agreement with known kinetic behaviours, e.g., reaction rate proportional to enzyme concentration for constant metabolite concentrations. This formalism has been applied for fitting metabolite elasticities *in vivo* (Visser et al., 2004; Wu et al., 2006) and for revealing dynamic and control properties for different network configurations *in silico* (Kresnowati et al., 2005; Visser and Heijnen, 2003; Wu et al., 2004).

While both the Log-lin and Lin-log approaches show wider applicability and greater kinetic consistency than GMA and S-system models, e.g., concentration elasticities are no longer constant for different metabolite concentrations; they remain local approximations of the actual kinetics. Fitting of their parameters is thus highly sensitive to the definition of a ‘representative’ reference state that enables reasonable prediction of subsequent metabolic

perturbations. For example, Voit et al. (2007) showed that S-system models could represent better than log-format approaches the situations when variable concentrations are moderately small. As the access to convenient reference states *in vivo* is ultimately dictated by the cell physiology, the predictive power of these formalisms is limited in practice.

## 2.2. Mechanistic rate laws

Mechanistic rate laws depend on a reaction mechanism that is specific for the enzyme. The mathematical description of such mechanism is rooted in first principles underpinned by mass conservation and thermodynamic laws. In the following, the relevant mechanistic formalisms used for describing the kinetics of both catalytic and allosteric enzymes are presented.

### 2.2.1. Elementary Reaction Mass Action

The kinetics of enzyme-catalysed reactions can be accurately captured using the Law of Mass Action at the molecular level. These events are described by uni- or bi-molecular reactions – known as elementary reactions – and represent the most basic steps in biochemical interconversion, namely: binding, dissociation and catalysis (or conversion) (Turanyi and Tomlin, 2014). Although, under some particular conditions (e.g., high pressure) ter-molecular events are not only likely but also important (Bernshtein and Oref, 2004), the chances of collisions between three entities is highly improbable under standard conditions, i.e., moderate temperature and pressure. Furthermore, even if ter-molecular events were present, they often involve the formation of a reaction intermediate through a bimolecular reaction followed by a collision with the third reactant, e.g., free radical (Turanyi and Tomlin, 2014). For practical purposes and in the context of enzymatic reactions, elementary reactions are thus regarded either as uni- or bi-molecular steps governed by the law of mass action.

Let us again revisit the reversible reaction  $\alpha_S S \leftrightarrow \alpha_P P$ , this time considering the mechanism of enzymatic action for this conversion. For illustration purposes, let us also assume that the stoichiometric coefficients are both equal to one, i.e.,  $\alpha_S = \alpha_P = 1$ , and that the conversion and release of the product P happens simultaneously. The reason for the latter arbitrary assumption is simply for ease of subsequent comparison with the Michaelis-Menten model (see Section 2.3.1), and will be apparent soon. The simplest reaction mechanism for this conversion is given by,



Here, we have considered that the enzyme can exist in two intermediate forms, either free ( $E_{\text{free}}$ ) or forming a complex with the substrate ( $ES$ ). Note that the conservation law

$\sum_{i \in \{E_{\text{free}}, ES\}} E_i = E_{\text{total}}$  holds true for the enzyme intermediates. The reaction rate for this

reversible conversion can be written as  $v = k_{+2}ES(t) - k_{-2}P \cdot E_{\text{free}}(t)$ , where  $E_{\text{free}}$  and  $ES$  are solutions of the following dynamic system,

$$\begin{aligned} \frac{dE_{\text{free}}}{dt} &= -E_{\text{free}}(k_{+1}x_S + k_{-2}x_P) + ES(k_{-1} + k_{+2}) \\ \frac{dES}{dt} &= -\frac{dE_{\text{free}}}{dt}, \quad \mathbf{E}(0) = (E_{\text{free},0}; ES_0) \end{aligned} \quad (8)$$

An interesting feature of Eq. 8 arises from the natural separation of characteristic time scales between enzyme intermediates and reactants. Considering that  $E_{\text{total}} \ll S(0)$  is almost always true in biochemical systems, the relaxation time for the intermediates should be substantially shorter than for the reactants. As such, if enzyme intermediates reach their steady-state considerably faster, then their concentrations can be easily computed assuming a pseudo-steady-state (Bodenstein, 1913). Conveniently, this assumption yields the reaction rate as a function of total enzyme and reactant concentrations only,

$$v = E_{\text{total}} \cdot \frac{k_{+1}k_{+2}x_S - k_{-1}k_{-2}x_P}{k_{-1} + k_{+2} + k_{+1}x_S + k_{-2}x_P} \quad (9)$$

Equation 9 can be expressed in a more familiar form defining  $k_{\text{cat}+} = k_{+2}$ ,  $k_{\text{cat}-} = k_{-1}$ ,  $K_S = k_{+1}/(k_{-1} + k_{+2})$  and  $K_P = k_{-2}/(k_{-1} + k_{+2})$  as follows,

$$v = E_{\text{total}} \cdot \frac{k_{\text{cat}+}x_S/K_S - k_{\text{cat}-}x_P/K_P}{1 + x_S/K_S + x_P/K_P} \quad (10)$$

where  $k_{\text{cat}+}$  and  $k_{\text{cat}-}$  denotes the forward and reverse catalytic constants, respectively, and  $K_S$  and  $K_P$  are the respective substrate and product dissociation constants or Michaelis-Menten constant. Proper conditions for valid use of the pseudo-steady-state assumption are revised elsewhere (Walter, 1974). The parameters in Eq. 10 are known as macroscopic or kinetic parameters, and as opposed to the rate constants in Eq. 9, they can be readily fitted using enzymatic assay data under initial velocity conditions, i.e., high substrate concentration and products absent (Cornish-Bowden, 2012). Appropriate definition of these parameters for more complex mechanisms is achieved using Cleland's rules (Cleland, 1963).

Analysis of Eqs. 9-10 at equilibrium reveals the existence of a relationship between the equilibrium constant and the rate (Eq. 11) and kinetic parameters (Eq. 12). Eq. 11 corresponds to a particular case of the Wegscheider cyclicity condition (Wegscheider, 1901), which is a consequence of microscopic reversibility and leads to detailed balance (Tolman, 1925). Eq. 12 is commonly known as Haldane relationship (Haldane, 1930), and it explicitly links observable kinetic parameters to the thermodynamic equilibrium constant. Importantly, both expressions entail relationships between the rate and kinetic parameters that 1) are valid

for all states near and far from equilibrium, and 2) ensure thermodynamic plausibility of the process.

$$K^{\text{eq}} = \frac{k_{+1}k_{+2}}{k_{-1}k_{-2}} \quad (11)$$

$$K^{\text{eq}} = \frac{k_{\text{cat+}}K_{\text{P}}}{k_{\text{cat-}}K_{\text{S}}} \quad (12)$$

Rate parameters derived using the ER-MA formalism are mechanistically sound, i.e., obey mass conservation and are constrained by thermodynamic laws. However, they often require many parameters (two per elementary step) to describe realistic mechanisms, rendering cumbersome rate laws (Leskovac, 2003). Different approximate kinetic expressions have been proposed to overcome this difficulty (reviewed in Section 2.3). Moreover, cooperative and allosteric behaviours – i.e., binding of reactants and/or effectors to sites other than the active site affecting enzyme activity – are not immediately covered by ER-MA. Additional assumptions and/or hypothetical mechanisms must be formulated to arrive at a meaningful description of allosteric effects. In the following, the relevant allosteric formalisms are revised.

### 2.2.2. Allosteric kinetics and the Monod-Wyman-Changeux model

Since the early haemoglobin binding experiments of Christian Bohr in the beginning of the 20<sup>th</sup> century (Bohr, 1904), the increase in binding affinity as ligand (in this case oxygen) concentration increases was recognized as a natural occurring behaviour in oligomeric proteins with multiple ligand binding sites. Due to the ‘cooperative’ character of subsequent bindings, this phenomenon was termed cooperative binding. The first mathematical description of this behaviour was suggested by Archibald Hill (Hill, 1910) and was phenomenologically modelled as follows,

$$\bar{y} = \frac{x_{\text{S}}^{n_{\text{H}}}}{K_{\text{S}}^{n_{\text{H}}} + x_{\text{S}}^{n_{\text{H}}}} \quad (13)$$

where  $\bar{y}$  represents the fraction of bound enzyme (or fractional occupancy),  $x_{\text{S}}$  denotes the ligand concentration, and  $n_{\text{H}}$  is the Hill coefficient (positive value). The latter coefficient quantifies the type and strength of cooperativity. For  $n_{\text{H}} > 1$ , the enzyme is said to undergo positivity cooperativity, whereas for  $n_{\text{H}} < 1$  the cooperativity is negative. The particular case of  $n_{\text{H}} = 1$  implies non-cooperative behaviour. Later experiments from Gilbert S. Adair in haemoglobin showed that the binding affinity is not constant but oxygen-dependent (Adair, 1925). By assuming binding saturation in stages and mass action law, Adair proposed the following expression,

$$\bar{y} = \frac{1}{n} \left( \frac{\sum_{j=1}^n j \cdot K_{S,j} x_S^j}{1 + \sum_{j=1}^n K_{S,j} x_S^j} \right) \quad (14)$$

where  $K_{S,j}$  represents apparent macroscopic association constants and  $n$  denotes the number of ligand binding sites in the protein. Although this mathematical description draws from the physical system by explicitly considering the number of binding sites, it does not offer a mechanistic explanation as to why various dissociation constants differ from one another. Successive formulations by Klotz (1946) and Pauling (1935) refined the calculation and interpretation of the dissociation constants of Adair's model. Particularly, the Klotz model (popularly referred to as the Adair-Klotz model) has found useful applications for describing calmodulin (Dagher et al., 2011) and haemoglobin (Yonetani et al., 2002) binding data.

A major breakthrough came with postulation of the mechanistic concerted model of Monod, Wyman and Changeux (MWC) (Monod et al., 1965) and the sequential model of Koshland, Némethy and Filmer (KNF) (Koshland et al., 1966). While both models rest on the assumption of an equilibrium between a relaxed (R) and tense (T) conformational states, transitions between states are treated differently. The MWC model requires maintenance of conformational symmetry (i.e., concerted transition) as ligands bind to succeeding binding sites (here assumed to be protein subunits). In contrast, the KNF model does not require the latter but rather strict induced fit. Notably, the MWC goes one step further than the KNF by offering thermodynamic and structural support to its main assumptions. Of all allosteric models, the MWC model has been the preferred choice for mechanistically describing allosteric and cooperative interactions (Cornish-Bowden, 2014).

In the MWC model, two functions are required to describe the kinetic behaviour of an oligomeric enzyme, namely the fraction of protein in the R state ( $\bar{R}$ ) and  $\bar{y}$  the fractional occupancy or 'saturation function',

$$\begin{aligned} \bar{R} &= \frac{(1 + x_S/K_S^R)^n}{L(1 + x_S/K_S^T)^n + (1 + x_S/K_S^R)^n} \\ \bar{y} &= \frac{L \cdot x_S/K_S^T (1 + x_S/K_S^T)^{n-1} + x_S/K_S^R (1 + x_S/K_S^R)^{n-1}}{L(1 + x_S/K_S^T)^n + (1 + x_S/K_S^R)^n} \end{aligned} \quad (15)$$

where  $L$  is the equilibrium allosteric constant, and  $K_S^R$  and  $K_S^T$  denote the ligand (or substrate) dissociation constants for the R and T states, respectively. Here, the allosteric constant is thermodynamically and structurally related to kinetic properties of the enzyme through  $L = \exp(-\Delta\Delta G/\bar{R}T) = T_0/R_0$ , where  $\Delta\Delta G$  denotes the Gibbs free energy difference of conformations between the R and T states, and  $T_0$  and  $R_0$  are the enzyme conformation

abundances in the absence of ligands. In regards to cooperativity, the MWC model naturally supports this kinetic behaviour as the result of higher abundance of the R state due to increasing ligand concentration. When the bound ligand affects its own binding, the allosteric effect is called homotropic, otherwise it is called as heterotropic. Importantly, most allosteric effects can be readily accommodated within this framework with the exception of negative cooperativity (Monod et al., 1965). Applications, limitations and further details about the MWC formalism can be found elsewhere (Changeux, 2012).

In the context of biochemical reactions mediated by allosteric enzymes, Eq. 15 is somehow difficult to interpret in catalytic terms. A mathematically convenient expression based on the MWC was developed by Popova and Sel'kov (Popova and Sel'kov, 1975, 1976) – termed Generalized MWC – which provided an elegant definition of the reaction rate for any of oligomeric enzyme. The proposed formula recast the reaction rate as the product of a catalytic and regulatory term. The catalytic term describes the catalytic rate of the subunits (protomers) in the R state, whereas the regulatory term describes the conformational transitions from the T to the R conformation.

$$v_i = \underbrace{E_i \cdot n f_R(\mathbf{x}_M; \mathbf{k}_R)}_{\text{catalytic}} \cdot \underbrace{\frac{1 + (f_T(\mathbf{x}_M; \mathbf{k}_T)/f_R(\mathbf{x}_M; \mathbf{k}_R)) \cdot Q(L; \mathbf{x}_M; \mathbf{x}_E; \mathbf{k}_R; \mathbf{k}_T; \mathbf{k}_E)}{1 + Q(L; \mathbf{x}_M; \mathbf{x}_E; \mathbf{k}_R; \mathbf{k}_T; \mathbf{k}_E)}}_{\text{regulatory}} \quad (16)$$

In Eq. 16,  $f$  represents a catalytic function that models the reaction mechanism of a single protomer. As per the MWC conformational symmetry, protomers are assumed to be equal, and thus, they exhibit the same reaction mechanism with parameters differing only based on the protomer state, i.e.,  $f_R$  and  $f_T$  have the same mathematical form but with different parameterizations for the R ( $\mathbf{k}_R$ ) and T ( $\mathbf{k}_T$ ) states. Derivation of the catalytic function can be achieved using thermodynamically feasible kinetic frameworks like the ER-MA (Saa and Nielsen, 2015). On the other hand, the regulatory term in Eq. 16 determines the conformational transitions and includes information about allosteric sites and effectors ( $\mathbf{x}_E$ ). Importantly, this mechanism is invariant with respect to the catalytic mechanism and rests solely on the mechanism of conformational transitions described by the  $Q$  function. This general framework has found broad application in the study of key regulatory enzymes (Peskov et al., 2008; Saa and Nielsen, 2015, 2016c), as well as in the construction of large-scale kinetic models of central carbon metabolism (Peskov et al., 2012). A major drawback of this framework is the excessive number of parameters needed – twice the amount of ER-MA plus at least one additional allosteric parameter –, hindering its fitting and widespread use.

### 2.3. Approximate kinetics

Approximate formalisms typically display a mechanistic inspiration; however, they propose a series of simplifying assumptions or suitable analogies that yield more compact rate laws. Depending on the degree of simplification, these models can produce more interpretable, easier-to-fit and often standardizable rate law formats. In the following, the most representative approximate kinetic formats are reviewed.



### 2.3.1. Michaelis-Menten

The Michaelis-Menten rate law (Michaelis and Menten, 1913) is the most prevalent rate model in biochemistry. Let us recall the Uni-Uni conversion  $S \leftrightarrow P$ , and let us this time assume that there is no reverse catalytic step, i.e.,  $k_{-2} = 0$  in Eq. 7. The corresponding mechanism describing this situation is then the following,



If we assume that  $S$  reaches instantaneously chemical equilibrium (rapid equilibrium assumption), i.e., the first reversible step is much faster than the second irreversible step, the reaction rate can be written in the popular form,

$$v_i = \frac{V_{\text{max},i} x_S}{K_d + x_S} \quad (18)$$

where  $V_{\text{max},i} = E_i \cdot k_{\text{cat}}$  denotes the maximum reaction velocity and  $K_d = k_{+1}/k_{-1}$  is the enzyme-substrate dissociation constant. One can arrive to an almost indistinguishable expression for Eq. 18 using the Quasi-Steady-State Assumption (QSSA) on the enzyme intermediates (Briggs and Haldane, 1925), with the only difference being the definition of the dissociation constant  $K_S = (k_{+1} + k_{\text{cat}})/k_{-1}$ . The use of either assumption demands different experimental conditions for its safe use (Keleti, 1986), albeit they yield identical parametric expressions. Importantly, multi-substrate reversible rate laws based on the rapid equilibrium assumption (as in Michaelis-Menten kinetics) have been developed (Alberty, 1953) and employed extensively for modelling enzymatic reactions (Cornish-Bowden, 2015). In spite of its broad application and accepted universality, early work showed that several enzymes do not precisely follow the Michaelis-Menten model due to inherent complexities in their kinetics (Hill et al., 1977). Although amenable and elegant, the reversible Michaelis-Menten equation inevitably incurs a loss of kinetic detail owing to its simplifying assumptions.

### 2.3.2. Generalized reversible Hill equation

Mechanistic equations for cooperative kinetics are commonly very complicated and necessitate the definition of several parameters. Determination of these parameters is often not possible as complete kinetic studies are not available. In order to overcome this limitation, Hofmeyr and Cornish-Bowden (1997) proposed an approximate thermodynamically consistent rate law for cooperative kinetics. The suggested parameterization is based on the functional form of the Hill equation (refer to Eq. 13), but it was made reversible, capable of including known modifiers and applicable to multimeric enzymes. The reversible Hill Equation for a Uni-Uni reaction incorporating the allosteric effect of one effector  $x_E$  reads,

$$v_i = E_i \cdot \frac{k_{\text{cat}+} x_S / K_S (1 - x_P / (x_S \cdot K^{\text{eq}})) \cdot (x_S / K_S + x_P / K_P)^{n_H - 1}}{1 + (x_E / K_E)^{n_H} + \eta (x_E / K_E)^{n_H} + (x_S / K_S + x_P / K_P)^{n_H}} \quad (19)$$

where  $\eta$  denotes a modifier interaction factor and  $K_E$  represents the effector dissociation constant. In Eq. 19, the Hill coefficient ( $n_H$ ) carries the same meaning as in Eq. 13 in that it quantifies the degree of cooperativity of the enzyme. The maximum value for this coefficient is bounded by the number of interacting sites of the enzyme (limiting cooperativity case), and in general, it is not an integer number. Inspection of the reversible Hill Equation shows that the model is thermodynamically feasible as it complies with the appropriate Haldane relationship (Eq. 12). Furthermore, in the absence of the effector ( $x_E = 0$ ) and imposing non-cooperative behaviour ( $n_H = 1$ ), the kinetic expression results in the mechanistic rate law defined in Eq. 10.

Subsequent work from Rohwer et al. (2006) extended this rate law to multi-substrate reactions in the presence and absence of modifiers. For obvious reasons, the extended version was called generalized reversible Hill Equation. The same authors showed that the proposed rate law displayed better agreement with experimental data for the pyruvate kinase of *Bacillus stearothermophilus* than the MWC (Rohwer et al., 2006). A follow-up study suggested a more realistic behaviour of the generalized reversible Hill equation than a variant of the MWC known as exclusive MWC (Olivier et al., 2006). In spite of the reported – albeit isolated – data, this rate law has not found wide application in the field.

### 2.3.3. Convenience and Modular kinetics

To address the excessive number of parameters of fully mechanistic rate laws, the convenience kinetics formalism was proposed as a flexible framework for describing enzyme kinetics consistent with thermodynamic constraints (Liebermeister and Klipp, 2006a). Convenience kinetics are a special form of the ER-MA kinetics for enzyme-catalysed reactions with a random-order mechanism. The main assumptions underpinning this formalism are: 1) random-order binding and unbinding of reactants, 2) binding of reactants is reversible and much faster than the catalytic step (i.e., rapid equilibrium assumption), and 3) binding energies of individual reactants do not depend on other reactants already bound to the enzyme (Liebermeister and Klipp, 2006a). The general form of the convenience kinetics reads as follows,

$$v_i = E_i \cdot h(\mathbf{x}_E; \mathbf{k}_E) \cdot \frac{k_{\text{cat}+} \prod_{j \in m_S} (x_j / K_{S,j})^{|\alpha_j|} - k_{\text{cat}-} \prod_{j \in m_P} (x_j / K_{P,j})^{|\alpha_j|}}{\prod_{j \in m_S} \sum_{r=0}^{|\alpha_j|} (x_j / K_{S,j})^r + \prod_{j \in m_P} \sum_{r=0}^{|\alpha_j|} (x_j / K_{P,j})^r - 1} \quad (20)$$

where  $\alpha_j$  denotes the stoichiometric coefficient of species  $j$  in the catalysed reaction and  $h(\mathbf{x}_E; \mathbf{k}_E)$  is a suitable function describing the kinetic modulation of the enzyme by additional metabolic effectors  $\mathbf{x}_E$ . In Eq. 20, the dissociation or Michaelis parameters  $K_{S,j}$  and  $K_{P,j}$  coincide with half saturation constants which can be readily obtained from enzymatic essays (Cornish-Bowden, 2012). Although in principle simple, convenience rate laws are complex enough to represent reasonable well mechanistically derived rate laws for common mechanisms such as ordered and ping-pong (Liebermeister and Klipp, 2006a).

A central claim of the convenience kinetics is related to their thermodynamic feasibility. In order to ensure thermodynamic plausibility, the authors introduced thermodynamically independent system parameters whose combination yields kinetic parameters consistent with the Haldane relationships (Liebermeister and Klipp, 2005, 2006a). Furthermore, definition of probability distributions for this set (e.g., multivariate Gaussian distributions) accompanied with the integration of metabolic data showed interesting potential for exploring the dynamic behaviour of the threonine pathway using Bayesian inference (Liebermeister and Klipp, 2006b).

More recently, the convenience rate laws have been extended to describe additional kinetic behaviours using Modular rate laws (Eq. 21) (Liebermeister et al., 2010).

$$v_i = E_i \cdot \frac{k_{\text{cat}+} \prod_{j \in m_S} (x_j / K_{S,j})^{|\alpha_j|} - k_{\text{cat}-} \prod_{j \in m_P} (x_j / K_{S,j})^{|\alpha_j|}}{D_r + D_r^{\text{reg}}} \quad (21)$$

In Eq. 21,  $D$  represent denominator terms representing enzyme regulation through the availability of binding sites ( $D_r$ ) and/or specific activity modulation ( $D_r^{\text{reg}}$ ). Depending on the reaction mechanism and type of enzyme regulation, several mathematical forms are available for describing different situations. More precisely, five generic and thermodynamically safe rate laws are introduced, namely: ‘common’ (equal to the original convenience rate law), ‘direct binding’, ‘simultaneous binding’, ‘power-law’, and ‘force-dependent’, along with a description of their properties (Liebermeister et al., 2010). The most attractive feature of the modular and convenience rate laws is the compact and flexible format. Lately, this feature has been exploited for the (semi-)automatic construction and analysis of large-scale kinetic models (Andreozzi et al., 2016a; Hackett et al., 2016; Stanford et al., 2013).

#### 2.3.4. Thermodynamic-Kinetic Modelling (TKM) formalism

The idea of borrowing concepts from electric potentials and using them to draw a parallel in the setting of irreversible thermodynamics has been explored for describing enzyme kinetics (Ederer and Gilles, 2007; Yang et al., 2006). Out of the reported formalisms, the TKM formalism (Ederer and Gilles, 2007) stands out as a comprehensive framework for building thermodynamically feasible kinetics in observance of the principle of detailed balance as imposed by the Wegscheider conditions (Wegscheider, 1901). Specifically, TKM models are parameterized based on the definition of a thermokinetic potential ( $\xi_j$ ) proportional to the metabolite concentration. The proportionality constant is known as capacity ( $C_j$ ) and depends

on the specific metabolite. Following the flux-force relationship  $F = R \cdot v$  (Westerhoff and Van Dam, 1987), the reaction rate is proportional to a thermokinetic force ( $F$ ) function of thermokinetic potentials up to a resistance parameter ( $R$ ). The general form of the TKM rate law is given in Eq. 22.

$$v = \frac{F(\xi)}{R(\xi)} = E_i \cdot \frac{\prod_{j \in m_s} \xi_j^{|\alpha_j|} - \prod_{j \in m_p} \xi_j^{|\alpha_j|}}{R(\xi)} \quad (22)$$

For pure mass action kinetics, the resistance is constant and readily derived. For non-mass action kinetics, the resistance can be either a rational or polynomial function of the thermokinetic potentials. For instance, the resistance for the rate law defined by Eq. 10 is given by  $R = \rho_0 + \rho_s \xi_s + \rho_p \xi_p$ . As non-mass action kinetics is generally the case of metabolic reactions, more complex and often approximate resistance parametric forms are employed. Further suggested applications of the TKM formalism include analysis of futile cycles in complex reaction and signalling networks, kinetic model reduction and sensitivity analysis (Ederer and Gilles, 2007). Comparison of TKM with other approaches as well as analysis of more complex reaction networks (treated as electric circuits in TKM) can be found elsewhere (Ederer and Gilles, 2008). Notably, the parameterization proposed in TKM has recently been shown to be almost equivalent to the modular rate laws parameterization with concentration-dependent conductivities (Liebermeister et al., 2010).

### 3. Kinetic model development

Kinetic model development requires integration of stoichiometric, thermodynamic, kinetic and regulatory information to conform a coherent representation of the system. Once the parametric structure of the model is defined, the next task is the construction of the model using available data from steady-state and/or dynamic time-course metabolic profiles. Broadly, kinetic model construction can be performed in two ways, namely: 1) bottom-up (forward) reconstruction, and 2) top-down (inverse modelling) construction (Fig. 3A). The key difference between the approaches is whether constitutive parts or sub-models are built individually and then merged forming the final model (bottom-up), or whether all parts are simultaneously fitted during model construction (top-down). The global fit of the top-down approach typically offers better prediction fidelity compared to a bottom-up, but it does require larger datasets as well as more sophisticated mathematical methods for attaining satisfactory fits.

Regardless of the strategy chosen for model construction, model parameters need to be estimated. Assignment of numerical values to model parameters enables quantitative interrogation and description of the dynamic behaviour of the system. This task is known as parameter estimation or fitting and it is probably the hardest step in the model development process (Jaqaman and Danuser, 2006). In the statistical field, this task is referred to as parameter inference. Despite the ubiquity of this problem in science, there is no single best method for all modelling applications. Indeed, depending on the parametric structure, nature and amount of data, and underlying modelling assumptions, different parameter inference strategies may be more convenient or attractive than others (Fig. 3B). In the case of kinetic

models, given their deterministic nature and simple formulation, there are several parameter fitting methods available and suitable for application (Bock et al., 2013). In addition, there is a great body of literature for asserting the statistical quality measured by the uncertainty of parameter fits and their predictions (Vanlier et al., 2013). In this section, we review approaches for constructing kinetic models and formal methods for performing parameter inference.

**Fig. 3.** Workflows for kinetic model development and statistical approaches for parameter estimation. **A** Depending on the availability of experimental data and kinetic information, kinetic models can be developed in a bottom-up or top-down fashion. The key difference between these approaches lies on whether the model is constructed reaction-by-reaction using *in vitro* and data from different sources, or whether it is built globally from *in vivo* measurements. **B** Parameter estimation of kinetic models can be achieved using either frequentist or Bayesian inference approaches. While both approaches employ the likelihood function to weight how likely a parameter is, the use of this function differs greatly in each approach. Frequentist approaches compute a ‘best’ point estimate of the parameter known as Maximum Likelihood Estimator (MLE) using optimization, and then, use different approaches to quantify the uncertainty in this estimate. In contrast, Bayesian approaches treat parameters as truly random variables defined by probability density functions known as *priors*. Combination of prior distribution (subjective) with the information contained in the data through the likelihood, yields the parameter’s *posterior* distribution which represents the uncertainty in the parameter. Computation of this *posterior* is typically achieved using sampling techniques.

### 3.1. Approaches for kinetic model construction

Construction of kinetic models can follow a bottom-up, top-down or a combination of both workflows. Bottom-up approaches seek to infer the behaviour of the system by understanding the operation of its individual constituents, whereas as top-down approaches seek to holistically describe the operation of the system. Although both approaches have associated with them difficulties and challenges, they also provide particular opportunities for kinetic model development in specific situations. The main features of each approach as well as illustrative examples are revised next.

#### 3.1.1. Bottom-up reconstruction

Starting from individual reactions or smaller networks of the model, bottom-up approaches formulate mechanistic-based rate expressions to describe the kinetic behaviour of the reaction(s) in each sub-module. In order to fit these parameters, these approaches require deep datasets covering hopefully the full dynamic range of each reaction including in the modules. As this cannot be achieved *in vivo* due to homeostatic control, bottom-up approaches generally rely on data obtained from *in vitro* experiments to fit their parameters. Even then, technical issues and excessive resource demands may preclude the gathering of the required data. Repositories of kinetic data such as BRENDA (Placzek et al., 2016), SABIO-RK (Wittig et al., 2012) and KiMoSys (Costa et al., 2014), as well as kinetic model databases such as BioModels (Chelliah et al., 2015) and JWS (Olivier and Snoep, 2004) have alleviated this need by offering streamlined access to published kinetic information. Additional efforts from different consortia (Tipton et al., 2014; van Eunen et al., 2010) are further advancing enzyme data standardization by defining guidelines for reporting consistent

data. In this way, by increasing the access to quality kinetic information and models, bottom-up modelling efforts can leverage existing information to build more complex models.

While in theory convenient, the actual implementation of bottom-up approaches is not devoid of difficulties. Naïve integration of kinetic information from different sources often leads to biologically unrealistic model behaviours. For example, the use of different *in vitro* experimental conditions for generating the kinetic data added to simplistic kinetic parameterizations can generate erroneous results. An example of the above was observed in an early effort trying to understand glycolysis' *in vivo* behaviour from measured *in vitro* kinetic properties (Teusink et al., 2000). Overcoming these limitations was more recently achieved by including allosteric regulation in the original model, i.e., expanding the kinetic capabilities of the model, and re-fitting kinetic parameters under conditions resembling the *in vivo* environment (van Eunen et al., 2012). Indeed, having a sufficiently complex model representation made up from reactions fitted individually in appropriate experimental conditions can yield biologically meaningful kinetic models (Chassagnole et al., 2002; Cintolesi et al., 2012; Curien et al., 2009; Peskov et al., 2012; Smallbone et al., 2013). Naturally, this type of approach is limited to those systems sufficiently characterized, where both the regulation and kinetic information are known. As this is not the rule but typically the exception, reconstructed models following this workflow generally require several rounds of global tuning using *in vivo* data.

### 3.1.2. Top-down construction

In systems biology, top-down approaches are normally defined as data-driven strategies that integrate various sources of experimental (omics) data. In the context of this review, however, we refer to top-down as the equivalent of inverse modelling approaches, where the complete model is globally inferred from the available *in vivo* data. Given the mathematical complexity of most kinetic models, top-down kinetic model construction is, in principle, more challenging than bottom-up reconstruction. In order to ease model development, early construction efforts have employed approximate kinetic formalisms to enable parameter fitting in several biological systems (Curto et al., 1997; Neves et al., 2002; Visser et al., 2000). The use of simplified expressions, however, inevitably incurs a loss of kinetic detail and frequently also a lack of thermodynamic consistency. Moreover, the latter typically precludes the identification of novel regulatory mechanisms.

During the past decade, different modelling approaches have been proposed to overcome the issue of detailed kinetic model identification. Realization that 'sloppy' parameters can still yield a predictive kinetic model (Gutenkunst et al., 2007), has opened the door for 'unusual' modelling strategies in the field. Where conventional statistical inference focuses on estimating each parameter with high accuracy, new approaches rely on the fact that in highly nonlinear, regulated systems few parameters dictate the behaviour of the system under the studied experimental conditions, while the rest can remain reasonably uncertain (Zamora-Sillero et al., 2011). Sampling-based approaches have gained popularity due to their versatility and effectivity, and different modelling strategies have been implemented for exploring the feasible kinetic space consistent with experimental observations (Tran et al., 2008) as well as unravelling key regulatory mechanisms (Kuepfer et al., 2007; Schaber et al., 2011). In our view, the greatest challenges of these approaches are their computational scalability and appropriate statistical description of the parameters' uncertainties.

### 3.2. Parameter estimation

Irrespective of the approach for constructing the kinetic model (e.g., bottom-up, top-down or a combination of both), parameters must be defined before using the model. Once the parametric form of the kinetic model is defined and the required data is collected, the next step is to estimate the model parameters. To this end, let us consider the probability of observing data  $\mathbf{y}_{\text{obs}}$  given parameter vector  $\theta$ ,  $p(\mathbf{y}_{\text{obs}}|\theta)$ , where  $\theta$  includes all the kinetic parameters of the model as well as possibly additional measurement noise, i.e.,  $\theta = (\mathbf{k}; \xi)$ . In other words,  $\theta$  lists all the parameters required to simulate from the model. A common formulation for  $p(\mathbf{y}_{\text{obs}}|\theta)$  assumes independent additive Gaussian noise with constant variance for each measurement yields the following statistical model,

$$p(\theta|\mathbf{y}_{\text{obs}}) = (2\pi)^{-N_{\text{meas}}/2} \cdot \det(\Sigma_{\text{meas}})^{-1/2} \cdot \exp\left\{-\frac{1}{2} \cdot (\mathbf{y}_{\text{sim}} - \mathbf{y}_{\text{obs}})^T \Sigma_{\text{meas}}^{-1} (\mathbf{y}_{\text{sim}} - \mathbf{y}_{\text{obs}})\right\} \quad (23)$$

where  $\Sigma_{\text{meas}}$  is a diagonal matrix representing the error covariance matrix of the measurements and  $N_{\text{meas}}$  denotes the number of data points. In the statistical frequentist paradigm, one seeks to determine a parameter set  $\hat{\theta}$  such that Eq. 23 is maximum. This procedure yields a Maximum Likelihood Estimator (MLE) ( $\hat{\theta}_{\text{MLE}}$ ) and it is cast as a nonlinear optimization problem. A convenient property of optimization problems is that the optimum remains invariant with respect to monotone transformations. Thus, log-transformation of the above equation (denoted commonly by  $L(\theta)$ ) yields the same solution but with a far simpler quadratic form for which efficient optimization algorithms are available (Hendrix and Tóth, 2010). A further simplification to this problem can be obtained assuming  $\Sigma_{\text{meas}}$  is constant and independent of  $\theta$ . By log-transforming Eq. 23 and imposing the latter assumption, the quantity to be optimized displays the familiar form of a Residual Sum of Squares (RSS) following a  $\chi^2$ -distribution,

$$\chi^2 = (\mathbf{y}_{\text{sim}} - \mathbf{y}_{\text{obs}})^T \Sigma_{\text{meas}}^{-1} (\mathbf{y}_{\text{sim}} - \mathbf{y}_{\text{obs}}) \quad (24)$$

Although simple in principle, finding  $\theta$  that maximizes the above quantity is non-trivial due to the existence of multiple local optima, discrepancies in the scale of parameters and typically poor model identifiability. To address these problems, sophisticated optimization algorithms have been developed that are capable of reaching global optimal solutions (Hendrix and Tóth, 2010). Once a satisfactory parameter set has been found, well-known *a posteriori* tests can be performed to assess the uncertainty and quality of the fitted parameters, e.g., sensitivity analysis, confidence intervals, and profile likelihood. A concise description of the latter tools is shown in the next section.

A very different way of addressing the parameter inference problem is based on Bayesian statistics. While frequentists methods seek to fit parameters to the data following a purely data-driven approach, Bayesian methods adopt a probabilistic approach where

parameters are treated as truly random variables, i.e., they are defined by probability density functions. These (subjective) probability distributions are known as *priors*, and they commonly describe either non-informative (objective) distributions, or, distributions that reflect the current state of belief (Mukherjee and Speed, 2008). Application of Bayes theorem then enables the updating of prior knowledge with new observations yielding an updated probability distribution referred to as the *posterior*. Let us define the parameter vector  $\theta$ , data  $\mathbf{y}_{\text{obs}}$ , prior distribution  $p(\theta)$ , and likelihood function  $p(\mathbf{y}_{\text{obs}}|\theta)$  (not necessarily given by Eq. 23), then Bayesian methods aim to sample from the posterior distribution  $p(\theta|\mathbf{y}_{\text{obs}})$  using Bayes' rule,

$$p(\theta|\mathbf{y}_{\text{obs}}) = \frac{p(\mathbf{y}_{\text{obs}}|\theta) \cdot p(\theta)}{p(\mathbf{y}_{\text{obs}})} \quad (25)$$

where  $p(\mathbf{y}_{\text{obs}})$  denotes the marginal likelihood or evidence. As this quantity does not depend on the parameters, it effectively acts as a normalizing constant such that  $p(\theta|\mathbf{y}_{\text{obs}})$  integrates to one. Bayesian methods provide a statistically sound way for describing parameter uncertainty and systematically advancing knowledge, e.g., posteriors from previous experiments can serve as priors for future experiments. Moreover, determination of bounds on  $p(\theta|\mathbf{y}_{\text{obs}})$  defines credible intervals – analogous to confidence intervals in the frequentist setting – that formally reflect current beliefs about the model. These powerful capabilities, however, incur in the expensive evaluation of high-dimensional integrals for which Monte Carlo sampling techniques are not only suitable, but indispensable. Although computationally more expensive, Bayesian methods are better suited for large models with complex and often unknown parameter correlations (Congdon, 2006).

In the following, we review relevant methods for parameter inference using MLE optimization-based and Monte Carlo (sampling-based) approaches. The key distinction between the classes is whether a point estimate or a parameter sample (i.e., population) is used to describe the model and perform predictions. We have decided against using a statistical division criterion (i.e., frequentist and Bayesian), as we believe it is more practical – from a modelling point of view – to use an operational distinction. Indeed, we note that optimization strategies can be employed both in frequentist and Bayesian settings (to a much lesser extent in the latter case). For example, in the frequentist case the objective function is typically given by Eq. 24 and yields an MLE upon minimization. Similarly, Eq. 25 can be maximized provided one defines the likelihood and priors for the inferred parameters. Such estimate is known as Maximum A Posteriori (MAP) in the Bayesian setting, and it can be considered a regularized version of the MLE. Importantly, in both cases an optimization strategy is used to infer the model parameters. A similar argument about sampling-based strategies can be made as shown in Section 3.2.2.

### 3.2.1. MLE optimization for parameter fitting

Optimization algorithms for parameter estimation can be either local or global. While the former converges substantially faster on an optimum of the objective function, the latter has a lower risk on converging to a local sub-optimal solution. Regardless of the applied algorithm,



knowledge about the topology of the problem can greatly improve convergence rate (Ashyraliyev et al., 2009). Once an optimal solution is found, analysis of the (weighted) model residuals can reveal whether the model describes the data satisfactorily (Cedersund and Roll, 2009). If the fit is acceptable, a range of *a posteriori* tools can be employed to quantify parameter uncertainty and grasp the predictive power of the model. Next, we discuss relevant local and global optimization methods, followed by useful tools for *a posteriori* identifiability analysis of the model. For a more comprehensive review of the latter the reader is referred to Vanlier et al. (2013).

### 3.2.1.1. Local optimization

Local optimization methods use either gradient-free or gradient-based search methods. The former type of algorithms does not make use of the objective function gradient, but instead use different heuristics to arrive to an optimum. The most popular gradient-free methods are direct-search methods, of which the Nelder–Mead (Nelder and Mead, 1965) and Hooke–Jeeves (Hooke and Jeeves, 1961) methods are the most employed. Both methods employ pattern heuristics either to determine convenient directions of movement in the parameter space (Hooke–Jeeves direct search) or to smartly shrink the feasible space until an optimal solution is found (Nelder–Mead adaptive simplex). Despite lacking some of the convergence guarantees of gradient-based methods and often displaying lower efficiencies for multi-dimensional problems (Lagarias et al., 1998), these methods can be more attractive for some applications as they do not require explicit computation of gradients. This may be particularly convenient if the objective function is suspected to exhibit discontinuities or to be non-differentiable in some regions.

Gradient-based optimization methods require a starting point and the gradient of the objective function with respect to the parameters represented by the Jacobian matrix  $\mathbf{J}_\theta$ . In addition, some methods also require a Hessian matrix  $\mathbf{H}_\theta$  containing second derivatives. As analytical closed forms for these matrices are rare and difficult to derive for nonlinear problems, numerical approximations are applied based on, for example, finite differences. Popular gradient-based methods are *steepest descent* (Cauchy, 1847), Newton-Raphson (Raphson, 1690) and *trust region* (Sorensen, 1982). The *steepest descent* and Newton-Raphson methods only differ on how the directions of movement are defined ( $\mathbf{d}_{\text{steep}} = -\mathbf{J}_\theta$  and  $\mathbf{d}_{\text{newton}} = -\mathbf{H}_\theta^{-1} \cdot \mathbf{J}_\theta$ ), while *trust region* methods use a simpler – typically quadratic – objective surrogate which enables efficient computation of the objective function surface. By comparing the expected improvement of the approximation with the original objective, a ‘trust’ region can be build and searched for optima. Out of the three methods, the Newton-Raphson method is the fastest, displaying a quadratic order of convergence (Gerlach, 1994). It does, however, assume non-singular  $\mathbf{H}_\theta$  throughout the iterations and it can potentially fail if started too far from the optimum. Another family of popular local optimization methods restricted to quadratic programming problems, i.e., least-square fitting, are the Gauss-Newton and Levenberg-Marquardt methods (Levenberg, 1944; Marquardt, 1963). These two methods exhibit similarities with the Newton-Raphson algorithm, however they avoid evaluation of the Hessian and employ linear approximations based on different steep descent directions with accelerated convergence.

Gradient-based optimization algorithms are by far the most employed for nonlinear optimization due to their speed and (local) convergence properties. These methods are suitable for fitting kinetic parameters in a bottom-up fashion – i.e., one-reaction-at-a-time –

using, for example, data from *in vitro* enzyme assay experiments. These methods should be avoided when globally fitting kinetic models – i.e., top-down approach – as they fail to thoroughly explore the objective function landscape and arrive to a global optimum. For this task, global optimization methods are preferred.

### 3.2.1.2. Global optimization

Global optimization methods are either deterministic or stochastic. Stochastic methods only enjoy weak theoretical guarantees of converging on the global optimum, whereas deterministic methods can provide higher assurance level for certain nonlinear optimization problems. Even then, nonlinear problems cannot in general be solved with certainty in finite time (Guus et al., 1995).

As global convergence cannot be assured for either type, global optimizers of stochastic nature are preferred for dynamical modelling due to their comparatively lower computational cost for generating ‘good enough’ solutions (Moles et al., 2003). While deterministic methods usually approach the optimization problem using a deterministic (local) optimizer with dispersed initial values, stochastic methods accept or reject parameter sets in a probabilistic manner thereby avoiding getting trapped in local optima (Hendrix and Tóth, 2010). Popular stochastic algorithms for global nonlinear optimization include evolutionary algorithms (Goldberg, 1989) and simulated annealing (Kirkpatrick et al., 1983). Other popular methods known as meta-heuristics such as Taboo Search, Ant Colony Optimization and Particle Swarm methods have also found use for nonlinear optimization (Corne et al., 1999). For a rigorous comparison of the performance of different stochastic optimization methods in biochemical modelling see Moles et al. (2003).

The improved performance of global optimizers has boosted their use for fitting large-scale kinetic models. For instance, a comprehensive kinetic model of the *Escherichia coli* central carbon metabolism with 351 parameters was recently fitted using a genetic algorithm, yielding reasonably accurate description of the fermentation dynamics of wild-type and few genetic mutants (Jahan et al., 2016). Even larger models ( $> 10^3$  parameters) of the *E. coli* central carbon metabolism (Khodayari and Maranas, 2016; Khodayari et al., 2014) have been recently constructed using genetic algorithms in combination with sampling-based frameworks (see Section 4.1.4. Ensemble Modelling for details).

### 3.2.1.3. *A posteriori* analysis

A diverse range of useful *a posteriori* diagnostics exists for statistically assessing the quality of the fitted model parameters. In the frequentist statistical setting, one can evaluate the discrepancies between the model and data using a  $\chi^2$ -test to quantify goodness-of-fit (Franceschini and Macchietto, 2008). If the results are not fully satisfactory, alternative nested models, i.e., models that can be transformed into others by imposing linear constraints, can be fitted and compared with the initial hypothetical model using likelihood ratio tests (Lehmann and Romano, 2005), or other suitable criteria, e.g., Akaike Information Criterion (Akaike, 1973) or Bayesian Information Criterion (Schwarz, 1978). Sensitivity analysis can aid in the task of proposing alternative and simpler model structures by indicating parameters with ‘sloppy’ sensitivities that can be fixed *a priori* in the model (Sacher et al., 2011), as well as more informative experiments (Franceschini and Macchietto, 2008). For an efficient

algorithm for the latter task with applications in biochemical models, the reader is referred to Sánchez et al. (2014).

Once the quality of the fit has been checked, the remaining task is to determine the MLE uncertainty. Under weak conditions and assuming a large number of samples  $N$ , the MLE can be represented by a normal distribution  $N(\hat{\Theta}_{\text{MLE}}, \Sigma_{\text{MLE}})$  with covariance matrix  $\Sigma_{\text{MLE}} = \text{FIM}^{-1}$ , where the FIM denotes the Fisher Information Matrix computed as the negative Hessian of the log-likelihood, i.e.,  $-\mathbf{H}_0^{L(\theta)}$  (Cramér, 1946). If the FIM is invertible, its inverse provides a lower bound for  $\Sigma_{\text{MLE}}$  as indicated by the Cramér-Rao inequality (Cramér, 1946; Rao, 1945). We note that the assumption of  $\hat{\Theta}_{\text{MLE}}$  normality does not always hold, for example, when the parameters are constrained and  $\hat{\Theta}_{\text{MLE}}$  is close to the boundary, or more typically, when the parameters are simply non-identifiable because the FIM is non-invertible. However, if the above regularity condition holds true, then an approximate 95%-confidence interval can be then determined using the appropriate Student's  $t$ -distribution (Vanlier et al., 2013). Similarly, confidence intervals for the model predictions can be constructed using linear approximations based on the Jacobian  $\mathbf{J}_{\hat{\Theta}_{\text{MLE}}}$  of the model at  $\hat{\Theta}_{\text{MLE}}$  (Franceschini and Macchietto, 2008).

As previously mentioned, asymptotic statistics requires reasonably good model identifiability and significant data, which are not commonplace in biochemical models. An alternative for asymptotic methods is the Likelihood Profile (LP) method, which provides more reliable confidence intervals by mapping the optimal likelihood path upon perturbations in individual parameters (Kreutz et al., 2013). Inspection of the likelihood ‘profiles’ can reveal non-intuitive structural and practical identifiability issues in biochemical kinetic models (Raue et al., 2009). In fact, the LP method has been recently identified as the superior method for uncertainty assessment in kinetic models (Fröhlich et al., 2014). For an illustrative review of this method and its application the reader is referred to Kreutz et al. (2013).

### 3.2.2. Monte Carlo-based methods for parameter inference

Conditions for global optimality are rarely met in systems biology models in general (Gutenkunst et al., 2007), and particularly, in detailed kinetic models of metabolism (Heijnen and Verheijen, 2013). Due to the paucity of kinetic data and the highly nonlinear structure of kinetic models, MLE-based approaches are unable to yield a representative best parameter fit given the extremely rugged (i.e., non-convex) likelihood surface. Consequently, ‘best-fits’ obtained from MLE and other similar methods are very likely to represent spurious global optima, often overfitting the data, and more importantly, leading to premature discarding of alternative equally likely hypotheses (Vanlier et al., 2013). In practice, the latter implies that distinct parameter sets can have almost equal likelihood values, rendering parameters structurally and/or practically non-identifiable (Raue et al., 2009). To overcome this issue, sampling-based approaches for parameter inference resort to probability distributions for scanning the likelihood surface (or another functional) conditional on the chosen parameters. In addition, by sampling from defined probability distributions, parameter uncertainty can be properly quantified and accounted for in model predictions. As sampling-based methods require many samples to confidently represent a multi-dimensional space, implementation of these approaches incurs substantially greater computation times compared to MLE optimization-based approaches. However, considering the alternative of just having a (local)

'best' point estimate with poor representativeness and predictive power, the use of sampling-based strategies is not only justified, but also encouraged.

There are a multitude of sampling-based methods for uncertainty quantification in dynamic models and most rely on Monte Carlo simulation. Monte Carlo methods generate random samples from defined probability distributions to approximate a quantity of interest  $g$  in a not necessarily regular multi-dimensional region  $\Omega_g$ . By using an appropriate sampling function on  $\Omega_g$  and averaging the outcomes of random experiments, a reasonably good approximation  $\hat{g}$  of  $g$  can be achieved, provided that a sufficiently large sample is collected. The theoretical basis for this approximation rests on the Law of Large Numbers – for a large sample of size  $N$ ,  $\hat{g}$  converges almost surely on  $g$  – and the Central Limit Theorem – the convergence rate of  $\hat{g}$  on  $g$  is of the order of  $N^{-1/2}$  (Mackay, 1998). A critical and maybe overlooked implication derived from the convergence property is that the Monte Carlo approximation does not depend on the size of  $\Omega_g$ , but only on the number of samples  $N$ . Although impressive at first sight, the latter assumes that samples are *independent*, which it is hard to achieve as directly sampling  $\hat{g}$  is often impossible without the use of elaborated sampling schemes that cannot guarantee complete independence (Rubinstein and Kroese, 2011). Despite these difficulties, Monte Carlo methods enjoy widespread use due to the ease of implementation, flexibility, and more importantly, efficacy for problems that are simply intractable using alternative methods. Amongst the latter, quantification of the uncertainty in fitted parameters of kinetic models of metabolism has been of particular focus in recent years (Achcar et al., 2012; Kerkhoven et al., 2013; Murabito et al., 2014). Both frequentist and Bayesian Monte Carlo methods exist for parameter uncertainty analysis.

#### 3.2.2.1. Frequentist parameter uncertainty modelling

Frequentist Monte Carlo methods rely on parametric or non-parametric bootstrapping to describe parametric uncertainties (Efron, 1979). In parametric bootstrapping data replicates are constructed using the original data and a parameterized error distribution and each replicate is used to re-fit the model to quantify parameter uncertainty. Non-parametric bootstrapping differs from the parametric approach in that data replicates are constructed by resampling with replacement from available experimental replicates, i.e., no probabilistic parameterization is required (Efron, 1981). Although non-parametric techniques are more robust than parametric methods as they do not make assumptions about the underlying error distribution, they require a greater number of bootstrap samples to properly describe the sample variability. Importantly, in both cases, the re-fitting task should be performed using robust global optimizers (e.g., multi-start algorithms (Hendrix and Tóth, 2010)) capable of exhaustively exploring the likelihood or the functional of interest, so that searches do not get trapped in (recurrent) local modes. Using both approaches, confidence intervals for parameter estimates can be obtained based on defined significance levels, and further corrected for bias and skewness (Diciccio and Romano, 1988). For an illustrative biochemical model example of bootstrapping refer to Joshi et al. (2006). Additional bootstrapping applications include model selection (only nested models) of signalling pathways using likelihood ratio tests (Muller et al., 2004). Further details about this test and applications in general can be found elsewhere (Lewis et al., 2011).

#### 3.2.2.2. Bayesian parameter uncertainty modelling

Bayesian statistics provides a comprehensive and mature framework for uncertainty analysis. As previously mentioned, Bayesian methods seek to sample from the posterior distribution  $p(\theta|y_{\text{obs}})$  to perform statistical inference. Hence, instead of restricting the parameter inference problem to the estimation of a single ‘best’ point estimate (e.g., MLE), Bayesian methods aim to compute the full parameter distribution conditional on the gathered data. Assembly of the posterior sample thus enables straightforward analysis of parameter uncertainties. For instance, determination of the 95%-percentile of  $p(\theta|y_{\text{obs}})$  defines credible intervals – analogous to confidence intervals in the frequentist setting – that formally reflect current beliefs about the model (Congdon, 2006). The same procedure can readily be applied to model predictions (forecast) as well as to model selection of any type (not necessarily nested) (Congdon, 2006).

These powerful capabilities do come at the cost of expensive evaluation of multi-dimensional integrals where Monte Carlo methods are not only suitable, but indispensable. Several efficient Monte Carlo methods have been developed to address this challenge (Chen et al., 2000), enabling the implementation of Bayesian methods in biochemical modelling applications for parameter estimation and model selection of signalling pathway models (Eydgahi et al., 2013; Hug et al., 2013; Koutroumpas et al., 2016; Sunnåker et al., 2013b), metabolic models (Liepe et al., 2015; Saa and Nielsen, 2016a, c), as well as for experimental design (Busetto et al., 2013). In particular, the opportunity of defining prior knowledge combined with recent progress in advanced sampling schemes are increasing the use of Bayesian approaches for dynamic metabolic modelling (Vasilakou et al., 2016).

#### 4. Advanced frameworks for kinetic modelling and analysis

Experience shows that the most challenging task in the construction of kinetic models of metabolism is the fitting of their parameters. Conventional fitting strategies are difficult to implement because of structural and practical identifiability issues in complex kinetic models. In spite the above difficulties, the past decade has seen an increase of advanced frameworks for tackling these challenges, offering novel capabilities for kinetic model analysis. Commonly based on Monte Carlo simulation, these advanced frameworks have shown promising results for the interrogation of dynamic properties, prediction of metabolic states and identification of key regulatory mechanisms in metabolic networks.

In the following, we review recently developed modelling frameworks tailored for the analysis of kinetic models of metabolism. Reviewed frameworks include: Structural Kinetic Modelling (SKM) (Steuer et al., 2006), Optimization and Risk Analysis of Complex Living Entities (ORACLE) (Mišković and Hatzimanikatis, 2010), MASS framework (Jamshidi and Palsson, 2008), Ensemble Modelling (EM) (Tran et al., 2008), and General Reaction and Assembly Platform (GRASP) (Saa and Nielsen, 2015). The selection of these frameworks is underpinned by promising recent studies, and it is by no means exhaustive of all modelling approaches for kinetic modelling. However, we have focussed on these approaches as they constitute complete workflows for kinetic model analysis, i.e., from the formulation and construction of the kinetic representation, to its quantitative interrogation. Next, we revise in detail the features, capabilities and applications of each framework.

##### 4.1. Structural Kinetic Modelling (SKM)

SKM is a modelling framework that enables dynamic analysis of a metabolic system with minimal information (Steuer et al., 2006). The key idea underlying this approach is that, while state predictions of the system in Eq. 1 require explicit kinetic parameterizations, dynamic analysis of its behaviour does not. By relying on the structure of the system and building a local linear approximation of its dynamic behaviour, SKM provides a versatile framework for exploring possible systems dynamics and predicting stability and robustness of metabolic states (Steuer et al., 2006), as well as unravelling relevant parameters and interactions underpinning the above (Grimbs et al., 2007).

Central to SKM is the Jacobian matrix  $\mathbf{J}_x^{\text{ref}}$  – a matrix with partial derivatives of the system states –, which captures the dynamic response of the system in the vicinity of a (not necessarily unique or stable) reference metabolic state. The Jacobian represents a linear approximation of system dynamics that can be constructed even if detailed knowledge of the rate equations is lacking. To do so, SKM formulates the above matrix as the product of two matrices  $\Lambda$  and  $\theta_x^v$ , which respectively describe the structure and kinetic-order of the system,

$$\frac{d\bar{\mathbf{x}}}{dt} \approx \mathbf{S} \cdot \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \Big|_{\mathbf{x}^{\text{ref}}} \cdot \bar{\mathbf{x}} = \mathbf{J}_x^{\text{ref}} \cdot \bar{\mathbf{x}} \quad (26)$$

$$\mathbf{J}_x^{\text{ref}} = \Lambda(\mathbf{S}; \mathbf{x}^{\text{ref}}; \mathbf{v}^{\text{ref}}) \cdot \theta_x^v(\mathbf{x}; \mathbf{x}^{\text{ref}}; \mathbf{v}; \mathbf{v}^{\text{ref}})$$

where  $\bar{\mathbf{x}}$  represents deviation state variables defined as  $\bar{\mathbf{x}} = \mathbf{x} - \mathbf{x}^{\text{ref}}$ . Notably, in the above equation  $\Lambda$  is solely function of known structural and constant quantities, whereas  $\theta_x^v$  depends on the degree of saturation of the reactions at the operation point  $(\mathbf{x}^{\text{ref}}; \mathbf{v}^{\text{ref}})$ . Given the general functional form of  $\theta_x^v$ , this matrix is also called the ‘normalized saturation’ or elasticity matrix and possesses a known sparsity pattern. Importantly, the nonzero entries of  $\theta_x^v$  have well-defined intervals that can be uniformly sampled, yielding an ensemble of possible Jacobians (for a detailed overview of the approach, see Fig. 4A). In this way, starting from the structure of the network and with very limited kinetic information, the SKM workflow enables statistical analysis of the characteristic properties of the Jacobian ensemble (i.e., eigenvalues) allowing appraisal of the dynamic properties of the system such as stability and oscillations, as well as identification of transition phases (bifurcations) (Steuer et al., 2006). Main applications of this approach have revealed non-intuitive insights into the influence of allosteric regulations in the dynamic operation of human erythrocytes (Grimbs et al., 2007), as well as specific metabolite levels (specifically low pyruvate and oxaloacetate) required to attain maximally stable networks in the plant mitochondrial TCA cycle (Steuer et al., 2007) (Fig. 4B).

**Fig. 4.** Overview of the Structural Kinetic Modelling (SKM) framework. **A** Starting from the network structure and with minimal experimental and kinetic information, the SKM workflow generates a population of Jacobians that enable exploration of the dynamic properties of a metabolic. **B** Main applications of the SKM framework include the identification of key regulatory (allosteric) checkpoints responsible for an increased stability of the central carbon metabolism of the RBC (Grimbs et al., 2007), as well as identification and determination of metabolite levels required to

attain a more robust operation of the TCA cycle in plants; in this case, low relative levels of pyruvate and oxaloacetate (Steuer et al., 2007).

From a practical standpoint, however, this framework is of rather limited use for kinetic interrogation. As the capabilities of this framework are solely focussed on the dynamic analysis of metabolic networks, SKM is not suitable as a predictor of metabolic states or missing regulatory interactions, i.e., metabolic states and regulatory interactions need to be known *a priori* for their analysis. In addition, Jacobians built in the early SKM framework have not explicitly considered thermodynamic constraints (i.e., Haldane relationships) when sampling saturation levels in the elasticities, rising concerns about the thermodynamic plausibility of the inferred Jacobians. Thermodynamically safe expressions (Liebermeister et al., 2010) and novel sampling methods (Childs et al., 2015) have been proposed to address and remedy this issue, however SKM results still need to be interpreted with caution. The fact that saturation levels are uniformly sampled does not guarantee ‘representativeness’ of the system, and hence, conclusions drawn from this approach should ideally be contrasted against experimental data.

#### 4.2.MASS framework

Construction of genome-scale kinetic models can be readily achieved using the LMA for each metabolic reaction. As only a single rate parameter (forward rate constant) is required to describe the reaction rate – provided that the thermodynamic equilibrium constant is known (refer to Section 2.1.1. Mass Action) – formulation of large-scale kinetic models can be easily realized solely based on stoichiometry. Following this rationale, the MASS framework proposes a scalable framework for constructing genome-scale kinetic models (Jamshidi and Palsson, 2008). Furthermore, the latter derives fundamental dynamic properties of the mass action formulation. Starting from a local linearization around a reference state in Eq. 1, the dynamic mass balance adopts the same form of Eq. 26, i.e.,  $d\bar{x}/dt = \mathbf{J}_x^{\text{ref}} \cdot \bar{x}$ ,  $\bar{x}(0) = \mathbf{0}$ . Application of the LMA for each reaction then yields the following decomposition of the Jacobian of metabolic states,

$$\mathbf{J}_x^{\text{ref}} = \mathbf{S} \cdot \text{diag}(\mathbf{k}_{\text{PERC}}) \cdot \mathbf{\Gamma}(\mathbf{x}; \mathbf{K}^{\text{eq}}) \quad (27)$$

where  $\text{diag}(\mathbf{k}_{\text{PERC}})$  is a diagonal matrix including forward rate or pseudo-elementary rate constants (kinetic information), and  $\mathbf{\Gamma}$  denotes a gradient matrix that incorporates the metabolite concentrations and equilibrium constants (thermodynamic information). Fundamental properties of this decomposition include: structural similarity ( $\mathbf{S} \sim -\mathbf{G}^T$ ), duality ( $\bar{v} = \mathbf{G} \cdot \bar{x}$  and  $\mathbf{J}_v^{\text{ref}} = \mathbf{G} \cdot \mathbf{S}$ ), and timescale decomposition based on the eigenvalues distribution of  $\mathbf{J}_x^{\text{ref}}$  or  $\mathbf{J}_v^{\text{ref}}$ . In terms of model formulation, the workflow is straightforward and comprises the following steps: 1) selection of a steady-state flux solution, 2) definition of concentrations for the each metabolite in the network (here approximations are typically used), 3) determination of equilibrium constants for each reaction (e.g., using Group Contribution methods (Jankowski et al., 2008; Noor et al., 2013)), and 4) calculation of

forward rate constants at the reference state. A complete overview of the MASS framework is depicted in Fig. 5A.

Thanks to the minimal information requirement, application of the MASS framework has been possible to very large metabolic models for exploration of their dynamic properties. Examples of the latter are the dynamic analysis of the RBC metabolic network for time-scale separation (modularization) (Jamshidi and Palsson, 2008), and for studying the metabolic response upon environmental perturbations in regulated and non-regulated model versions (Jamshidi and Palsson, 2010). In a more recent and notable application, the MASS framework was employed to construct personalized whole-cell kinetic models of RBCs (Bordbar et al., 2015) (Fig. 5B). As the metabolic state of the network (fluxes) cannot be uniquely determined in this system, the authors employed a Monte Carlo technique for sampling the flux space and constructing a representative ensemble of baseline kinetic models. Integration of the latter model with high-throughput metabolomic information and dynamic profiling, enabled unprecedented study of variations (at the kinetic level) between different individuals. Most remarkably, the authors showed that personalized kinetic parameters were better predictors of the off-target effects of drugs in susceptible genotypes – in this case ribavirin for haemolytic anaemia – than metabolite concentrations. In addition, given the breadth of the model, the authors could identify reaction modules at genome-scale that emerged using time decomposition based on their characteristic time-scales (Fig. 5B).

**Fig. 5.** Overview of the MASS framework. **A** The MASS framework is likely the most scalable framework for large-scale kinetic model construction. By using the network topology and thermodynamics combined with available omics data at a reference state, this framework enables appraisal of the kinetic ( $k_{\text{PERC}}$ ) and dynamic (Jacobians) properties of large metabolic systems. **B** Application of the MASS framework for the construction of personalized whole-cell models of RBCs from different individuals revealed kinetic differences and network susceptibilities to metabolic drugs, as well as organization of the network in reaction modules operating at distinct time scales (Bordbar et al., 2015).

Of all the kinetic modelling and analysis frameworks, the MASS framework is probably the most scalable due to its data-driven nature and dependency on fundamental stoichiometric and thermodynamic information. Despite the challenge of data completeness and error, sampling-based techniques can be implemented to handle such situations. As such, construction of MASS-derived kinetic models can be very useful for studying kinetic differences between individuals at a defined reference state, e.g., healthy or disease. MASS models, however, hold limited capability for metabolic state prediction, as they rely on a linearization around an operation point parameterized using simple mass action kinetics. A recent comparison between different kinetic formalisms showed that mass action kinetics along with other simplified kinetic expressions can display substantial – and often unexpected – discrepancies when compared against mechanistic expressions in small metabolic models (Du et al., 2016). Thus, reliable predictions of the impact of large metabolic perturbations and/or system properties like metabolic control coefficients cannot be achieved within this framework.

#### 4.3. Optimization and Risk Analysis of Complex Living Entities (ORACLE)



The ORACLE framework originates in the Log-Lin formalism and its extension to the assessment of dynamic uncertainty within the framework of Metabolic Control Analysis (MCA) (Wang et al., 2004). MCA is one of the first formal frameworks for studying the behaviour of metabolic networks and seeks to quantify the sensitivity of the network upon genetic and environmental perturbations (Heinrich and Rapoport, 1974; Kacser and Burns, 1973). More precisely, MCA is a type of sensitivity analysis specific for metabolic networks that yields flux ( $C_E^v$ ) and metabolite ( $C_x^v$ ) control coefficients. Together, these coefficients quantify the impact induced by changes in enzyme activities and metabolite concentrations on reaction fluxes at a reference point. Most remarkably, these coefficients represent system or global properties of the network, which are underpinned by local kinetic properties captured in the reaction elasticities.

MCA was previously limited to well-studied systems where kinetic information and metabolite concentrations were available (Nielsen and Jørgensen, 1995; Poolman et al., 2000; Rossignol et al., 2000). To overcome these limitations, Wang et al. (2004) developed a Monte Carlo-based framework, where scaled metabolite concentrations are uniformly sampled and used to generate a population of elasticities and corresponding control coefficients. Evaluation of the (local) stability of each member in the population is then performed using the MCA framework developed for the Log-lin formalism (Hatzimanikatis and Bailey, 1997), and used to reject unstable (deemed as infeasible) population members. This consistency check also requires metabolite concentrations, which can be sampled within physiological ranges. Analysis of the population of control coefficients can then reveal the control structure of the network under a defined condition (given by a reference flux distribution), as well as possibly non-trivial couplings between reactions (refer to Wang and Hatzimanikatis (2006) for an illustrative application in *Saccharomyces cerevisiae*).

A second generation of the MCA framework under uncertainty came with the introduction of the ORACLE framework as it is known today (Mišković and Hatzimanikatis, 2010). The latter extended the capabilities of the original framework by introducing a convenient parameterization for the elasticities as functions of auxiliary parameters  $\varepsilon = u(\mathbf{e}; \boldsymbol{\gamma})$ , where  $\mathbf{e}$  denotes enzymatic state abundances (e.g., free enzyme, enzyme-substrate complex, etc.) and  $\boldsymbol{\gamma}$  represents coefficients of displacement from equilibrium. The latter coefficients ensure that the choice of the sampled elasticities is consistent with the reaction thermodynamics ( $\Delta G_r$ ), increasing the reliability of the elasticity population. The mathematical form of the  $u$  function is dictated by the rate law used to parameterize each reaction and can be derived directly from the employed approximate (commonly the case) or mechanistic rate law. By integrating these features and proceeding in a similar fashion to the initial framework, a representative population of control coefficients can be obtained that accurately represents the dynamic state of the system. Additional experimental information about control coefficients and/or enzyme states can readily be included if available (Mišković and Hatzimanikatis, 2011). A complete overview of the most recent ORACLE framework is depicted in Fig. 6A.

**Fig. 6.** Overview of the Optimization and Risk Analysis of Complex Living Entities (ORACLE) framework. **A** Formulation of the network structure integrated with fluxomic data supported by directionalities based on thermodynamic and metabolomic data, enables definition of a reference state for the ORACLE application. This framework generates a population of control coefficients, Jacobians and elasticities that together fully characterize the dynamic state of the system in the neighbourhood of the reference state. **B** Recent application of the ORACLE to a BDO-overproducing

*E. coli* strain showed excellent agreement with reported data for the identification of non-intuitive pathways and reactions controlling BDO production and yield in a large-scale kinetic model (Andreozzi et al., 2016a).

Since the appearance of the ORACLE framework, there have been several applications devoted to better understand metabolic robustness and flexibility (Chakrabarti et al., 2013; Soh et al., 2012), and more recently, to rationally engineer microbial strains (Andreozzi et al., 2016a; Savoglidis et al., 2016). Remarkably in Andreozzi et al. (2016a), a comprehensive control analysis of a recombinant *E. coli* producing 1,4-butanediol (BDO) using a large-scale kinetic model revealed metabolic engineering targets for improving BDO production, showing high agreement between model predictions and confirmatory experiments (Fig. 6B). In the case of Savoglidis et al. (2016), a novel mathematical framework called Inverse Metabolic Control Analysis (IMCA) was introduced to study the complex sphingolipid biosynthesis in *S. cerevisiae*. The key novelty of this approach is that it enables determination of the (log)changes in enzyme activities, as function of measured (log)changes in metabolite concentrations and concentration control coefficients derived from the conventional ORACLE. As the enzymes of this pathway catalyse several reactions, IMCA helps reveal the most influential enzymes in the complex network. Another recent extension of the ORACLE approach relates to its application for studying and reducing the uncertainty in large-scale kinetic models using a classification algorithm (iSCHRUNK) (Andreozzi et al., 2016b). Rejection of kinetic parameter samples by imposition of stability and consistency criteria in the ORACLE, implies that there are parameter patterns (rules) that render models infeasible. By learning these patterns, the frequency of feasible parameters can be improved. Furthermore, characterization of the feasible parameter space confirmed previous results on ‘sloppy’ kinetic models, i.e., few parameters are constrained to narrow regions of the space whereas most of the parameters can be largely varied (Daniels et al., 2008; Gutenkunst et al., 2007). It is also suggested that iSCHRUNK can be used as an alternative uniform, non-asymptotic, solution space sampler, although no rigorous proofs of the latter are offered.

The ORACLE framework represents a complete framework for studying dynamic and control properties of metabolic systems at well-defined steady state. As such, ORACLE can suggest metabolic engineering strategies to optimize the performance of the system around the reference phenotype. The latter implies that the predictive power of ORACLE is inherently limited to the proximity of the studied state. Also, the fact that the central quantities generated are elasticities – which can be in turn derived from disparate kinetic formats – somewhat complicates their deconvolution, precluding execution of simple tasks like time course simulation (Srinivasan et al., 2015).

#### 4.4. Ensemble Modelling

The concept of ‘ensemble modelling’ has its roots in the field of statistical mechanics (Gibbs, 1902), and it was not until the last decade that was introduced in the field of systems biology (Alves and Savageau, 2000; Battogtokh et al., 2002). Simply put, an ensemble of models is a collection of different models describing competing hypothesis about a determined phenomenon. As the amount of data increases, the confidence and predictive power of the ensemble increases and converges to a single explanatory model with most of the weight. In the context of systems biology, however, there have been two uses and/or interpretations for

ensemble modelling. The first interpretation follows the above definition and seeks to formulate different biochemical models describing alternative regulatory mechanisms (Battogtokh et al., 2002; Kuepfer et al., 2007; Schaber et al., 2011). The second interpretation is more similar to a relaxation approach, where a single model structure is parametrized by different parameter values yielding quantitatively distinct models (Alves and Savageau, 2000; Tran et al., 2008).

Following the latter interpretation, the Ensemble Modelling (EM) approach was introduced by Tran et al. (2008) for describing feasible kinetic models consistent with experimental data. One of the most remarkable achievements of this approach was the derivation of an elegant procedure for parameterizing catalytic reactions around a reference state using the ER-MA formalism. The key idea behind the EM approach is that rate constants can be normalized ( $\tilde{\mathbf{k}}$ ) and recast as simple combinations of auxiliary parameters (enzyme state abundances  $\mathbf{e}$  and microscopic reversibilities  $\mathbf{R}$  – closely related to coefficients of equilibrium displacement in ORACLE) that are constrained by mass and energy conservation laws. These auxiliary parameters can be sampled uniformly and encompass all the allowable values for  $\tilde{\mathbf{k}}$ . Notably, the proposed parameterization does not require a fully defined reference point, but only needs definition of a reference flux distribution ( $\mathbf{v}^{\text{ref}}$ ) and thermodynamic driven forces for each reaction ( $\Delta G_r$ ). The behaviour of each ensemble model is then represented by a set of ODEs (Tran et al., 2008),

$$\frac{d\tilde{\mathbf{x}}}{dt} = \text{diag}(\mathbf{x}^{\text{ref}})^{-1} \cdot \mathbf{S} \cdot \mathbf{v}(\tilde{\mathbf{E}}; \tilde{\mathbf{x}}; \mathbf{e}; \mathbf{R}) \quad (28)$$

where  $\tilde{\mathbf{x}}$  and  $\tilde{\mathbf{E}}$  represent respectively normalized metabolite and enzyme concentrations, i.e.,  $\tilde{x}_i = x_i/x_i^{\text{ref}}$  and  $\tilde{E}_i = E_i/E_i^{\text{ref}}$ , and  $\text{diag}(\mathbf{x}^{\text{ref}})$  is a diagonal matrix with the reference metabolite concentrations. As steady-state data is typically used to train the ensemble,  $\mathbf{x}^{\text{ref}}$  does not need to be known exactly to determine the metabolic steady-state of the system. Also, as the EM approach relies on the ER-MA parameterization, any non-allosteric regulatory mechanism (i.e., substrate-level regulation) can be described. Once the model is defined and the experimental data collected, the EM approach generates an initial large ensemble of models that it is subsequently reduced as more data becomes available. The latter is achieved by removing models that do not agree with the observations within some defined error. The ultimate goal of this approach is to find a single model that accurately captures the system's behaviour (Tan et al., 2011; Tran et al., 2008). We note that this goal is counter to the original purpose of an ensemble. A complete overview of the EM approach is depicted in Fig. 7A.

**Fig. 7.** Overview of the Ensemble Modelling (EM) framework. **A** Ensemble Modelling workflow for kinetic model construction and selection. As more data is collected, the ensemble of models (herein represented by different parameterizations of the same model) is pruned and reduced until a single explanatory is left. **B** Applications of the EM paradigm have evolved and involve two new approaches for kinetic analysis with different scopes. The Ensemble Modelling with Optimization (EM-O) approach seeks to fit a large kinetic model to genome-scale networks for the prediction of metabolic states and yields using large-scale omics datasets from mutant strains. On the other hand, the Ensemble Modelling for Robustness Analysis (EM-RA) seeks to determine and define metabolic

targets and conditions that ensure robust metabolic performance when (over)producing valuable metabolites.

Application of the EM approach has yielded valuable insights into the kinetic behaviour of metabolic networks in model organisms such as *E. coli* (Contador et al., 2009; Rizk and Liao, 2009; Tan and Liao, 2012), human cells (Dean et al., 2010; Khazaei et al., 2012) as well as non-model organisms such as *Rhodobacter sphaeroides* (Rizk et al., 2011). More recently, an optimization-based strategy for parameter estimation was implemented to build large-scale models of the *E. coli* the central carbon metabolism (Khodayari and Maranas, 2016; Khodayari et al., 2014). By combining the EM approach for generating a population of feasible parameter candidates for each reaction and a genetic algorithm for identifying accurate combinations of parameters for different reactions, large-scale kinetic models can be constructed. We have designated this approach EM-O (Ensemble Modelling with Optimization, refer to Fig. 4). The last model reported by Khodayari and Maranas (2016) represents – to the best of our knowledge – the most comprehensive (457 reactions, 337 metabolites and 295 substrate-level regulatory interactions) and experimentally validated (25 mutant strains) kinetic representation of a microbial metabolism to date. Predictions of product yields for 320 engineered strains using this reconstruction are substantially superior than those derived from purely stoichiometric-based methods (Khodayari and Maranas, 2016) (Fig. 7B).

Another recent extension of the EM approach – known as Ensemble Modelling for Robustness Analysis (EM-RA) (Lee et al., 2014) – has been proposed for studying the robustness of metabolic systems upon enzymatic perturbations by combining EM with the continuation method for bifurcation analysis (Allgower and Georg, 2003). Briefly, EM-RA looks at the probability of failure in the ensemble upon manipulation of the activities of distinct enzymes to different extents (Fig. 7B). Application of this approach has been successfully performed for the dynamic analysis of non-native pathways (Lee et al., 2014) as well as cell-free systems (Theisen et al., 2016), highlighting in both cases the trade-off between robustness and performance in enzymatic systems. More recently, the EM-RA approach has been extended and applied to define the Kinetically Accessible Yield (KAY) for the production of exo-metabolites (specifically isobutanol in *E. coli* mutants) based on stability and robustness considerations (Rivera et al., 2017).

Main limitations of the EM approach for kinetic modelling relate to lack of 1) parameter uncertainty measures, and 2) careful consideration of allosteric interactions. In the former case, given the huge number of parameters involved in EM-derived kinetic models, representation of the system with a single accurate point estimate is insufficient to properly assess the uncertainty of the fitted. For instance, application of bootstrapping methods can be of help for estimating approximate confidence intervals or other relevant uncertainty measures. In the case of allosteric interactions, a previous EM study showed that inclusion of all known allosteric interactions in a model of the central metabolism of human cells severely hindered the convergence of the ensemble (Khazaei et al., 2012). Proper inclusion of allosteric interactions as described in Eq. 16 may help to overcome this limitation (Saa and Nielsen, 2015).

#### 4.5. Approximate Bayesian Computation and General Reaction and Assembly Platform (ABC-GRASP)

The most recent framework for kinetic modelling is GRASP. Initially formulated as a kinetic formalism for describing general reactions performed by oligomeric enzymes (Saa and Nielsen, 2015), GRASP evolved recently into statistical framework grounded on Bayesian principles for constructing feasible kinetic models of metabolic reactions (Saa and Nielsen, 2016c) and networks (Saa and Nielsen, 2016a). The original kinetic formalism relies on the decomposition of the reaction flux as the product of a regulatory and catalytic functions derived from the Generalized MWC model (for details refer to Section 2.2.2. Allosteric kinetics and the Monod-Wyman-Changeux model). This decomposition enables mechanistic description of the kinetics of any oligomeric enzyme, provided that the reaction mechanism and conformation transitions are defined. More importantly, the Generalized MWC model is fully compatible the ER-MA formalism, enabling ready implementation of the normalization around the reference state  $\mathbf{y}^{\text{ref}} = (\mathbf{v}^{\text{ref}}; \Delta\mathbf{G}_r)$  used in EM (Tran et al., 2008). Application of the latter normalization along with generalization of the ER-MA parameterization proposed in GRASP to both compulsory order and random reaction mechanisms, yields the following general expression for the calculation of the rate constants,

$$\tilde{\mathbf{k}} = \mathbf{P}_E(\mathbf{e}) \cdot \mathbf{\Gamma}_R(\mathbf{R}) \cdot \mathbf{r}_{\text{elem}} \quad (29)$$

where  $\mathbf{P}_E$  and  $\mathbf{\Gamma}_R$  represent respectively diagonal matrix functions of the enzyme state abundances and microscopic reversibilities, and  $\mathbf{r}_{\text{elem}}$  denotes a branching vector necessary to capture the split of elementary fluxes in random binding reaction mechanisms. For compulsory order mechanisms that latter factor is constant and given by  $\mathbf{r}_{\text{elem}} = \mathbf{1}$ . The mathematical forms of both  $\mathbf{P}_E$  and  $\mathbf{\Gamma}_R$  are determined by the catalytic mechanism of reaction. Using the general formula described in Eq. 29, the rate constants for the active (R) and tense (T) form of any oligomeric enzyme can be computed, hereby enabling mechanistic description of allosteric kinetics. Definition of Dirichlet distributions – i.e., a class of probability distribution defined on a simplex – for the enzyme states and log-reversibilities, and uniform distributions for  $\mathbf{r}_{\text{elem}}$ , enables exploration of the thermodynamically feasible kinetic space for any type of catalytic enzyme. Emergent properties relating to the behaviour of elasticities for different thermodynamic affinities and catalytic mechanisms, kinetic cooperativity and allosteric effects have been recently studied and revealed using GRASP (Saa and Nielsen, 2015).

Realization that the probability distributions defined in GRASP effectively act as thermodynamically feasible *priors* for the auxiliary parameters – which in turn determine uniquely the distributions for the kinetic parameters of metabolic reactions –, opened the door for the implementation of Bayesian strategies for parameter inference. The only element missing for performing full Bayesian inference is a likelihood function (refer to Eq. 25), which unfortunately is intractable for these applications. Indeed, due to the complex form of kinetic expressions and the unknown correlation between parameters, definition of a suitable likelihood function is hard. To overcome this issue, we turned to an Approximate Bayesian Computation (ABC) approach that bypasses the need for a likelihood using instead a comparison between simulations and data (Sunnåker et al., 2013a). By evaluating the discrepancy between the simulation and the experimental data, an approximate posterior sample can be generated and employed for parameter inference and predictions. We denote ABC-GRASP the combined approach using ABC for parameter inference, and GRASP for

prior distribution generation. A summarized overview of the latter framework is depicted in Fig. 8A.

Implementation of this approach has yielded novel insights into the consequences of kinetic constraints on the control structure of simple catalytic mechanisms, as well as for inferring likely kinetic features of complex mutant enzymes (e.g., P450 monooxygenases) from modest wild-type kinetic data (Saa and Nielsen, 2016c). A more recent application of the ABC-GRASP approach for studying the complex metabolic regulation in the mammalian methionine cycle, showed that the posterior samples do not only converge on the true parameter values, but also improve their prediction fidelity as more data is known (Saa and Nielsen, 2016a) (Fig. 8B). Furthermore, owing to its Bayesian character, advanced modelling tasks like model selection are also supported, which enabled partial identification of missing allosteric interactions in an incomplete model of the cycle (Fig. 8B).

**Fig. 8.** Overview of the Approximate Bayesian Computation and General Reaction Assembly and Sampling Platform (ABC-GRASP) framework. **A** General workflow of the ABC-GRASP framework. Of all the reviewed frameworks, this approach requires the highest amount of prior and experimental information for the construction of mechanistic kinetic models. The ABC-GRASP output corresponds to a distribution of *a posteriori* parameters that are both consistent with prior knowledge (mechanistic and thermodynamic information) and experimental observations. **B** ABC-GRASP has been able of identifying missing allosteric interactions, as well as predicting metabolic control and states with high accuracy in complex metabolic models (Saa and Nielsen, 2016a).

Altogether, the ABC-GRASP framework constitutes a complete modelling suite for Bayesian parameter inference in detailed kinetic models of metabolism. We note this framework holds additional modelling capabilities so far not exploited, e.g., (Bayesian) experimental design and prior priming using reported kinetic data from databases. Indeed, this framework can benefit from a range of statistical capabilities underpinned by the Bayesian paradigm. However, as in any other Bayesian approach, there are several challenges related to the Monte Carlo computation of the probabilistic quantities in high-dimensional spaces. Even efficient Sequential Monte Carlo methods struggles to converge on the posterior if the latter is too diffuse (Beskos et al., 2014). Considering the high-dimensionality of kinetic models generated using GRASP, proper representativeness of the parameter space can be difficult to attain in large kinetic models. As such, the scalability to large-scale kinetic models is the main limitation of the ABC-GRASP framework.

#### 4.6. Selection of modelling framework

Depending on the scope of the study and the availability of data and information, different frameworks will be more appropriate and effective for answering determined research questions. As shown earlier, each approach supports different modelling capabilities. While some frameworks are designed for the study of structural and dynamic properties of metabolic networks around a reference state (SKM, MASS and ORACLE), others have focused on the accurate prediction of metabolic states (EM and ABC-GRASP). Additional features such as scalability and data requirements are also important and must be considered when choosing a modelling framework for a specific metabolic network. Ultimately, the intended application is most critical when choosing a kinetic modelling framework. For

metabolic engineering purposes, ORACLE, EM and ABC-GRASP are the most suitable, with ORACLE and EM the most scalable. On the other hand, for exploratory studies with modest experimental data, the SKM and MASS provide convenient frameworks for interrogating large networks. Finally, for prospective studies where different model hypotheses are to be tested, ABC-GRASP offers the soundest mathematical framework for statistical inference and uncertainty analysis. A general chart summarizing the key features of each framework as well as their historical development is depicted in Fig. 9.

**Fig. 9.** Advanced modelling frameworks for kinetic modelling and analysis. The figure depicts the historical development of recent frameworks as well as the general supported capabilities (PRIMAL). The five considered frameworks are colour coded as follows: ORACLE (green), SKM (purple), MASS (red), EM (orange), and GRASP (blue). Abbreviations: MCA, Metabolic Control Analysis; SKM, Structural Kinetic Modelling; EM, Ensemble Modelling; ORACLE, Optimization and Risk Analysis of Complex Living Entities; EM-RA, Ensemble Modelling for Robustness Analysis; EM-O, Ensemble Modelling with Optimization; GRASP, General Reaction and Assembly Platform; ABC-GRASP, Approximate Bayesian Computation and GRASP; IMCA, Inverse Metabolic Control Analysis.

## 5. Future directions

Up to this point, we have presented a complete overview of the main modelling frameworks for formulating, constructing and analysing kinetic models of metabolism. In the following, we briefly discuss the advances and current limitations hindering the construction of kinetic models in the context of the literature reviewed. We finish with a brief comment on the outlook of the next generation of kinetic models for metabolic modelling.

### 5.1. Opportunities and advances

The greatest strength of kinetic models lies in their ability to integrate all the factors that determine reaction flux. This unique feature cannot easily be replicated, for example, by genome-scale stoichiometric models. Several strategies have leveraged and adapted the CBM formulation to incorporate kinetic information, yielding important improvements in the prediction fidelity of metabolic fluxes under different conditions (Chowdhury et al., 2014; Cotten and Reed, 2013; Yizhak et al., 2010). Even the most comprehensive metabolic mathematical representation of an organism captured in the so-called ME-model (Macromolecular Expression model), requires adequate parameterization of enzyme turnovers for accurately predicting by-products secretion (King et al., 2017). These kinetic features are naturally included in kinetic models and highlight the importance of kinetic information in the prediction of metabolic states. As the construction of kinetic representations becomes more frequent, we expect the predictive performance of stoichiometric models to increase by integrating the key kinetic features of the latter models.

More fundamentally, kinetic models can assist in understanding complex metabolic traits of biological systems. Complex behaviours such as activation, inhibition, stability, oscillations and ultra-sensitivity, are dynamic in nature, and hence, necessitate a kinetic representation. As these behaviours often arise from the interplay between different components of the system, their operation can be only grasped using globally interacting kinetic models. A remarkable example of the above is the model developed by Kotte et al. (2010), which describes an emergent property of metabolic adaptations in *E. coli* denominated 'distributed sensing'. By sensing specific flux-signalling metabolites, system-level adjustments of the network emerge within the metabolic network itself. Identification of such behaviour was achieved by the modular integration of a kinetic model of the central carbon metabolism with dynamic models of transcriptional and translational regulation. Another relevant application of detailed kinetic models in the field of systems medicine, identified metabolic targets and drugs capable of affecting parasites (*Trypanosoma brucei*) without collateral damage to the host (erythrocytes) (Haanstra et al., 2017). By employing a 'network-based' drug selectivity criterion based on MCA, tested metabolic targets showed excellent agreement in co-cultures experiments of both cell types. These examples highlight the potential of kinetic models for unravelling nontrivial metabolic behaviours as well as the mechanisms underpinning the latter.

Probably the most intuitive application of kinetic models is for the optimization of cell factories in biotechnological processes (Almquist et al., 2014). As the availability of platform or 'chassis' strains increases (Nielsen and Keasling, 2016), i.e., strains that already produce high amounts of various precursor intermediates, the requirement of detailed kinetic models capturing the fine regulation of the production pathway should also increase. Thanks to detailed information embedded in kinetic models, sophisticated metabolic engineering strategies can be identified and implemented for improving cell and process performance.



The reviewed modelling frameworks offer different alternatives for arriving to such rational strategies.

## 5.2.Challenges and limitations

Despite recent advances in kinetic modelling frameworks, kinetic model development remains an ongoing challenge in the field. Main limitations hindering further progress are the disparity and high-dimensionality of rate laws, incomplete knowledge of regulatory interactions, paucity of broad datasets for kinetic model construction, and proper description of complex regulatory events. The above limitations render construction of detailed kinetic models a challenging task and explains the limited use of kinetic modelling compared to the more established stoichiometric models for metabolic modelling.

A key challenge for kinetic modelling is the structural and quantitative uncertainty surrounding kinetic representations. Our knowledge about regulatory interactions is quite limited and often consists of putatively assigned regulatory functions. The latter complicates fitting of kinetic models as all relevant regulatory interactions must be included *a priori* to parameterize the model. In this regards, modelling frameworks capable of sampling the space of possible regulatory interactions can be of great assistance (Saa and Nielsen, 2016a; Sunnåker et al., 2014). A promising novel approach called SIMMER based on Bayesian statistics using Monte Carlo sampling and combining a comprehensive omics dataset – proteomic, fluxomic, and metabolomic data – from yeast cultivation under different conditions not only recapitulated known metabolic regulators, but also yielded novel ones (e.g., inhibition of pyruvate kinase flux by citrate/isocitrate) (Hackett et al., 2016). However, these approaches may still fail to identify key interactions given the huge size of the interaction space and large amounts of data needed to discern between similar structures. Furthermore, even if all interactions are known, the complex nonlinear nature of mechanistic expressions makes it difficult to fit their parameters using conventional fitting approaches, especially when trying to construct large-scale kinetic models (Heijnen and Verheijen, 2013). The use of sampling approaches and – to a lesser extent due to their lack of kinetic detail – approximate kinetic laws can help to address these limitations; however, improving the quality and quantity of data is ultimately what is required to overcome this stumbling block. Considering that *in vivo* measurements are subject to homeostatic control and they are often limited to few metabolites and proteins, generation of the necessary dataset for the construction of genome-scale kinetic models still represents a great obstacle. Even though the recent large-scale kinetic reconstruction of the *E.coli* central metabolism invites us to think otherwise, we highlight that the achievement of the training dataset required integration of multiple wide datasets and substantial data pre-processing and normalization (Khodayari and Maranas, 2016).

From a theoretical standpoint, mathematical integration into a unified framework of the several layers of regulation at the transcription, translation, and post-translation levels is probably one of the toughest challenges (Chowdhury et al., 2015). At the post-translational level, cooperative and allosteric regulation has long been known to play a critical role in the behaviour of metabolic pathways (Gerhart and Pardee, 1962). Recent studies have particularly highlighted the role of this type of regulation during rapid metabolic adaptations (Link et al., 2013; Xu et al., 2012). Description of these regulations can be achieved using mechanistic (e.g., Generalized MWC) or approximate (e.g., reversible Hill equation) kinetic formalisms. The use of the former formalisms is advantageous as it enables discovery of the

underlying mechanisms of action and it can be supported by recent sampling frameworks (Saa and Nielsen, 2015). Signal transduction (i.e., protein de-/phosphorylation), translation (i.e., protein synthesis and degradation), and gene-expression regulation (i.e., transcription factor binding, synthesis and degradation) remain open problems from a fundamental, kinetic modelling point of view. Despite the difficulty of bringing together these different layers of regulation in a unified mechanistic formalism, definition of an operational framework grounded in metabolic and hierarchical control analysis and borrowing concepts from control engineering, may be the way forward for the design and analysis of such integrated systems (He et al., 2016).

### 5.3.Outlook

Kinetic models should realize the long-sought dream of predicting how metabolic phenotypes are shaped dynamically by the genetic content and environmental conditions. As such, we envision a future where kinetic models are not merely symbolic abstractions of cell metabolism, but instrumental to understand the complex regulation of metabolic networks. Furthermore, these mathematical representations will be the cornerstone for driving discoveries in the field of biomedicine and advancing biotechnological applications. In addition, we foresee an increase in the scope and degree of integration of additional layers of metabolic regulation supported by recent developments in modelling frameworks.

Key aspects that will be critical in the advancement of large-scale kinetic models are 1) how they can be brought to larger scales, 2) how uncertainty is captured and described, and 3) how models are effectively communicated. The first point relies not only on the continuous development and advancement of modelling frameworks, but also on the generation of the required datasets for constructing such models. Although there are few comprehensive multi-omics datasets available for such purpose (Buescher et al., 2012; Hackett et al., 2016; Ishii et al., 2007), we highlight that these are still limited and an increased effort must be devoted to the generation of high quality experimental data. The second and third points are related and they address how kinetic models are constructed and shared in the community. Conventional fitting methods fail to capture parameter uncertainty in large-scale kinetic models, and hence, Monte Carlo-based approaches are increasingly used to overcome this limitation. The latter implies that the kinetic parameterization is not unique, but instead described by a large ensemble or population of parameter sets. So far, communication of models represented in this way has not been properly addressed by the community. If kinetic models of this kind are to move forward and drive biological discovery, this pending task must be resolved.

### Acknowledgements

PAS acknowledges the financial support of the Becas-Chile and Australian Government Research Training Program Scholarships.

## References

1. Achcar, F., Kerkhoven, E.J., Bakker, B.M., Barrett, M.P., Breitling, R., Consortium, S., 2012. Dynamic modelling under uncertainty: the case of *Trypanosoma brucei* energy metabolism. *Plos Comput Biol* 8, e1002352.
2. Adair, G.S., 1925. The hemoglobin system. VI. The oxygen dissociation curve of hemoglobin. *J Biol Chem* 63(2), 529-545.
3. Akaike, H., 1973. Information theory and the maximum likelihood principle, in: Petrov, B.N., Csaki, F. (Eds.), *Second International Symposium on Information Theory*. Akademiai Kiado, Budapest, pp. 610-624.
4. Alberty, R.A., 1953. The Relationship between Michaelis Constants, Maximum Velocities and the Equilibrium Constant for an Enzyme-Catalyzed Reaction. *J Am Chem Soc* 75(8), 1928-1932.
5. Allgower, E.L., Georg, K., 2003. *Introduction to Numerical Continuation Methods*. Society for Industrial and Applied Mathematics, Philadelphia.
6. Almaas, E., Kovács, B., Vicsek, T., Oltvai, Z.N., Barabási, a.-L., 2004. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* 427, 839-843.
7. Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., Jirstrand, M., 2014. Kinetic models in industrial biotechnology - Improving cell factory performance. *Metab Eng* 24, 38-60.
8. Alvarez-Vasquez, F., Sims, K.J., Cowart, L.A., Okamoto, Y., Voit, E.O., Hannun, Y.A., 2005. Simulation and validation of modelled sphingolipid metabolism in *Saccharomyces cerevisiae*. *Nature* 433(7024), 425-430.
9. Alves, R., Savageau, M.A., 2000. Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways. *Bioinformatics* 16, 534-547.
10. Andreatti, S., Chakrabarti, A., Soh, K.C., Burgard, A., Yang, T.H., Van Dien, S., Miskovic, L., Hatzimanikatis, V., 2016a. Identification of metabolic engineering targets for the enhancement of 1,4-butanediol production in recombinant *E. coli* using large-scale kinetic models. *Metab Eng* 35, 148-159.
11. Andreatti, S., Miskovic, L., Hatzimanikatis, V., 2016b. iSCHRUNK - In silico approach to characterization and reduction of uncertainty in the kinetic models of genome-scale metabolic networks. *Metab Eng* 33, 158-168.
12. Ashyraliyev, M., Fomekong-Nanfack, Y., Kaandorp, J.A., Blom, J.G., 2009. Systems biology: parameter estimation for biochemical models. *Febs J* 276(4), 886-902.
13. Atkinson, M.R., Savageau, M.A., Myers, J.T., Ninfa, A.J., 2003. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* 113(5), 597-607.
14. Bailey, J.E., Ollis, D.F., 1977. *Biochemical Engineering Fundamentals*. McGraw-Hill.
15. Battogtokh, D., Asch, D.K., Case, M.E., Arnold, J., Schuttler, H.B., 2002. An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of *Neurospora crassa*. *P Natl Acad Sci USA* 99(26), 16904-16909.
16. Bernshtein, V., Oref, I., 2004. Termolecular collisions: Comparison between analytical expression and trajectory calculations. *J Phys Chem A* 108(39), 8131-8133.
17. Beskos, A., Crisan, D.O., Jasra, A., Whiteley, N., 2014. Error bounds and normalising constants for sequential Monte Carlo samplers in high dimensions. *Advances in Applied Probability* 46(1), 279-306.
18. Bock, H.G., Carraro, T., Jäger, W., Körkel, S., Rannacher, R., Schlöder, J., 2013. *Model Based Parameter Estimation: Theory and Applications*. Springer Science & Business Media.

19. Bodenstein, M., 1913. Eine Theorie der photochemischen Reaktionsgeschwindigkeiten. *Zeitschrift für Physikalische Chemie* 85, 329-397.
20. Bohr, C., 1904. Die Sauerstoffaufnahme des genuinen Blutfarbstoffes und des aus dem Blute dargestellten Hämoglobins. *Zentralblatt für Physiologie* 23, 688-690.
21. Bordbar, A., McCloskey, D., Zielinski, D.C., Sonnenschein, N., Jamshidi, N., Palsson, B., 2015. Personalized whole-cell kinetic models of metabolism for discovery in genomics and pharmacodynamics. *Cell Systems* 1, 283-292.
22. Briggs, G.E., Haldane, J.B.S., 1925. A note on the kinetics of enzyme action. *Biochem J* 19(2), 338-339.
23. Buescher, J.M., Liebermeister, W., Jules, M., Uhr, M., Muntel, J., Botella, E., Hessling, B., Kleijn, R.J., Le Chat, L., Lecoite, F., Mader, U., Nicolas, P., Piersma, S., Rugheimer, F., Becher, D., Bessieres, P., Bidnenko, E., Denham, E.L., Dervyn, E., Devine, K.M., Doherty, G., Drulhe, S., Felicori, L., Fogg, M.J., Goelzer, A., Hansen, A., Harwood, C.R., Hecker, M., Hubner, S., Hultschig, C., Jarmer, H., Klipp, E., Leduc, A., Lewis, P., Molina, F., Noirot, P., Peres, S., Pigeonneau, N., Pohl, S., Rasmussen, S., Rimm, B., Schaffer, M., Schnidder, J., Schwikowski, B., Van Dijl, J.M., Veiga, P., Walsh, S., Wilkinson, A.J., Stelling, J., Aymerich, S., Sauer, U., 2012. Global network reorganization during dynamic adaptations of *Bacillus subtilis* metabolism. *Science* 335(6072), 1099-1103.
24. Burgard, A.P., Pharkya, P., Maranas, C.D., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* 84, 647-657.
25. Busetto, A.G., Hauser, A., Krummenacher, G., Sunnaker, M., Dimopoulos, S., Ong, C.S., Stelling, J., Buhmann, J.M., 2013. Near-optimal experimental design for model selection in systems biology. *Bioinformatics* 29(20), 2625-2632.
26. Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C.A., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D.S., Karp, P.D., 2016. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 44(D1), D471-D480.
27. Cauchy, A., 1847. Méthodes générale pour la résolution des systèmes d'équations simultanées. *C.R. Acad. Sci. Par.* 25, 536-538.
28. Cedersund, G., Roll, J., 2009. Systems biology: model based evaluation and comparison of potential explanations for given biological data. *Febs J* 276(4), 903-922.
29. Chakrabarti, A., Miskovic, L., Soh, K.C., Hatzimanikatis, V., 2013. Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotechnol J* 8, 1043-1057.
30. Changeux, J.-P., 2012. Allostery and the Monod-Wyman-Changeux model after 50 years. *Annu Rev Biophys* 41, 103-133.
31. Chassagnole, C., Noisommit-Rizzi, N., Schmid, J.W., Mauch, K., Reuss, M., 2002. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol Bioeng* 79(1), 53-73.
32. Chelliah, V., Juty, N., Ajmera, I., Ali, R., Dumousseau, M., Glont, M., Hucka, M., Jalowicki, G., Keating, S., Knight-Schrijver, V., Lloret-Villas, A., Natarajan, K.N., Pettit, J.B., Rodriguez, N., Schubert, M., Wimalaratne, S.M., Zhao, Y.Y., Hermjakob, H., Le Novère, N., Laibe, C., 2015. BioModels: ten-year anniversary. *Nucleic Acids Res* 43(D1), D542-D548.
33. Chen, M., Shao, Q., Ibrahim, J., 2000. Monte Carlo Methods in Bayesian Computation. Springer.

34. Childs, D., Grimbs, S., Selbig, J., 2015. Refined elasticity sampling for Monte Carlo-based identification of stabilizing network patterns. *Bioinformatics* 31(12), 214-220.
35. Chou, I.C., Martens, H., Voit, E.O., 2006. Parameter estimation in biochemical systems models with alternating regression. *Theor Biol Med Model* 3, 25.
36. Chou, I.C., Voit, E.O., 2009. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math Biosci* 219(2), 57-83.
37. Chowdhury, A., Khodayari, A., Maranas, C.D., 2015. Improving prediction fidelity of cellular metabolism with kinetic descriptions. *Curr Opin Biotech* 36, 57-64.
38. Chowdhury, A., Zomorodi, A.R., Maranas, C.D., 2014. k-OptForce: integrating kinetics with flux balance analysis for strain design. *Plos Comput Biol* 10, e1003487.
39. Cintolesi, A., Clomburg, J.M., Rigou, V., Zygourakis, K., Gonzalez, R., 2012. Quantitative analysis of the fermentative metabolism of glycerol in *Escherichia coli*. *Biotechnol Bioeng* 109(1), 187-198.
40. Cleland, W.W., 1963. Kinetics of enzyme-catalyzed reactions with 2 or more substrates or products. 1. Nomenclature and rate equations. *Biochim Biophys Acta* 67(1), 104-187.
41. Congdon, P., 2006. Bayesian Statistical Modelling. John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester.
42. Contador, C.a., Rizk, M.L., Asenjo, J.a., Liao, J.C., 2009. Ensemble modeling for strain development of L-lysine-producing *Escherichia coli*. *Metab Eng* 11, 221-233.
43. Corne, D., Dorigo, M., Glover, F., 1999. *New Ideas in Optimization*. McGraw-Hill.
44. Cornish-Bowden, A., 2012. *Fundamentals of Enzyme Kinetics*. Wiley-Blackwell.
45. Cornish-Bowden, A., 2014. Understanding allosteric and cooperative interactions in enzymes. *Febs J* 281, 621-632.
46. Cornish-Bowden, A., 2015. One hundred years of Michaelis–Menten kinetics. *Perspectives in Science* 4, 3-9.
47. Costa, R.S., Verissimo, A., Vinga, S., 2014. KiMoSys: a web-based repository of experimental data for KInetic MOdels of biological SYStems. *Bmc Syst Biol* 8, 85.
48. Cotten, C., Reed, J.L., 2013. Mechanistic analysis of multi-omics datasets to generate kinetic parameters for constraint-based metabolic models. *Bmc Bioinformatics* 14, 32.
49. Cramér, H., 1946. *Mathematical methods of statistics*. Princeton University Press.
50. Curien, G., Bastien, O., Robert-Genthon, M., Cornish-Bowden, A., Cardenas, M.L., Dumas, R., 2009. Understanding the regulation of aspartate metabolism using a model based on measured kinetic parameters. *Mol Syst Biol* 5, 271.
51. Curto, R., Voit, E.O., Sorribas, A., Cascante, M., 1997. Validation and steady-state analysis of a power-law model of purine metabolism in man. *Biochem J* 324, 761-775.
52. Dagher, R., Peng, S., Gioria, S., Feve, M., Zeniou, M., Zimmermann, M., Pigault, C., Haiech, J., Kilhoffer, M.C., 2011. A general strategy to characterize calmodulin-calcium complexes involved in CaM-target recognition: DAPK and EGFR calmodulin binding domains interact with different calmodulin-calcium complexes. *Bba-Mol Cell Res* 1813(5), 1059-1067.
53. Daniels, B.C., Chen, Y.J., Sethna, J.P., Gutenkunst, R.N., Myers, C.R., 2008. Sloppiness, robustness, and evolvability in systems biology. *Curr Opin Biotech* 19(4), 389-395.
54. Dean, J.T., Rizk, M.L., Tan, Y., Dipple, K.M., Liao, J.C., 2010. Ensemble modeling of hepatic fatty acid metabolism with a synthetic glyoxylate shunt. *Biophys J* 98, 1385-1395.
55. Degenring, D., Froemel, C., Dikta, G., Takors, R., 2004. Sensitivity analysis for the reduction of complex metabolism models. *J Process Contr* 14(7), 729-745.
56. Diccio, T.J., Romano, J.P., 1988. A review of bootstrap confidence-intervals. *J Roy Stat Soc B Met* 50(3), 338-354.

57. Du, B., Zielinski, D.C., Kavvas, E.S., Drager, A., Tan, J., Zhang, Z., Ruggiero, K.E., Arzumanyan, G.A., Palsson, B.O., 2016. Evaluation of rate law approximations in bottom-up kinetic models of metabolism. *Bmc Syst Biol* 10, 40.
58. Ederer, M., Gilles, E.D., 2007. Thermodynamically feasible kinetic models of reaction networks. *Biophys J* 92(6), 1846-1857.
59. Ederer, M., Gilles, E.D., 2008. Thermodynamic constraints in kinetic modeling: Thermodynamic-Kinetic Modeling in comparison to other approaches. *Eng Life Sci* 8(5), 467-476.
60. Edwards, J.S., Ibarra, R.U., Palsson, B.O., 2001. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nat Biotechnol* 19(2), 125-130.
61. Efron, B., 1979. Bootstrap Methods - Another Look at the Jackknife. *Ann Stat* 7(1), 1-26.
62. Efron, B., 1981. Nonparametric estimates of standard error - the Jackknife, the bootstrap and other methods. *Biometrika* 68(3), 589-599.
63. Eydgahi, H., Chen, W.W., Muhlich, J.L., Vitkup, D., Tsitsiklis, J.N., Sorger, P.K., 2013. Properties of cell death models calibrated and compared using Bayesian approaches. *Mol Syst Biol* 9, 644.
64. Franceschini, G., Macchietto, S., 2008. Model-based design of experiments for parameter precision: State of the art. *Chem Eng Sci* 63, 4846-4872.
65. Fröhlich, F., Theis, F.J., Hasenauer, J., 2014. Uncertainty analysis for non-identifiable dynamical systems: Profile likelihoods, bootstrapping and more, in: Mendes, P., Dada, J.O., Smallbone, K. (Eds.), *Computational Methods in Systems Biology: 12th International Conference*. Springer International Publishing, Manchester, UK, pp. 61-72.
66. Gerhart, J.C., Pardee, A.B., 1962. Enzymology of control by feedback inhibition. *J Biol Chem* 237(3), 891-896.
67. Gerlach, J., 1994. Accelerated convergence in newton method. *Siam Rev* 36(2), 272-276.
68. Gibbs, J.W., 1902. *Elementary Principles in Statistical Mechanics*. Charles Scribner's Sons, New York.
69. Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc.
70. Grimbs, S., Selbig, J., Bulik, S., Holzhütter, H.-G., Steuer, R., 2007. The stability and robustness of metabolic states: identifying stabilizing sites in metabolic networks. *Mol Syst Biol* 3, 146.
71. Guldberg, C., Waage, P., 1864. Studier i affiniteten. (Translation: Studies on affinities.). *Forhandlinger i Videnskabs-Selskabet i Christiania* 35.
72. Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R., Sethna, J.P., 2007. Universally sloppy parameter sensitivities in systems biology models. *Plos Comput Biol* 3(10), 1871-1878.
73. Guus, C., Boender, E., Romeijn, H.E., 1995. *Stochastic Methods*. Kluwer Academic Publishers, Dordrecht.
74. Haanstra, J.R., Gerding, A., Dolga, A.M., Sorgdrager, F.J.H., Buist-Homan, M., du Toit, F., Faber, K.N., Holzhütter, H.-G., Szöör, B., Matthews, K.R., Snoep, J.L., Westerhoff, H.V., Bakker, B.M., 2017. Targeting pathogen metabolism without collateral damage to the host. *Sci Rep-Uk* 7, 40406.
75. Hackett, S.R., Zanolli, V.R.T., Xu, W.X., Goya, J., Park, J.O., Perlman, D.H., Gibney, P.A., Botstein, D., Storey, J.D., Rabinowitz, J.D., 2016. Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* 354(6311), aaf2786.
76. Hadlich, F., Noack, S., Wiechert, W., 2009. Translating biochemical network models between different kinetic formats. *Metab Eng* 11(2), 87-100.
77. Haldane, J.B.S., 1930. *Enzymes*. Longmans, London.

78. Hatzimanikatis, V., Bailey, J.E., 1997. Effects of spatiotemporal variations on metabolic control: approximate analysis using (log)linear kinetic models. *Biotechnol Bioeng* 54, 91-104.
79. Hatzimanikatis, V., Floudas, C.A., Bailey, J.E., 1996. Analysis and design of metabolic reaction networks via mixed-integer linear optimization. *Aiche J* 42(5), 1277-1292.
80. He, F., Murabito, E., Westerhoff, H.V., 2016. Synthetic biology and regulatory networks: where metabolic systems biology meets control engineering. *J R Soc Interface* 13(117), 20151046.
81. Heijnen, J.J., 2005. Approximative kinetic formats used in metabolic network modeling. *Biotechnol Bioeng* 91(5), 534-545.
82. Heijnen, J.J., Verheijen, P.J.T., 2013. Parameter identification of in vivo kinetic models: Limitations and challenges. *Biotechnol J* 8(7), 768-775.
83. Heinrich, R., Rapoport, T., 1974. A linear steady-state treatment of enzymatic chains. *Eur J Biochem* 105, 97-105.
84. Hendrix, E.M.T., Tóth, B.G., 2010. *Introduction to Nonlinear and Global Optimization*. Springer-Verlag, New York.
85. Hill, A.V., 1910. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *Journal of Physiology* 40, iv-vii.
86. Hill, C.M., Waight, R.D., Bardsley, W.G., 1977. Does any enzyme follow Michaelis-Menten equation? *Mol Cell Biochem* 15(3), 173-178.
87. Hofmeyr, J., Cornish-Bowden, A., 1997. The reversible Hill equation: how to incorporate cooperative enzymes into metabolic models. *Computer applications in the biosciences: CABIOS* 13, 377-385.
88. Hooke, R., Jeeves, T.A., 1961. Direct search solution of numerical and statistical problems. *J Acm* 8(2), 212-229.
89. Horn, F., Jackson, R., 1972. General mass action kinetics. *Arch Ration Mech An* 47(2), 81-116.
90. Hug, S., Raue, A., Hasenauer, J., Bachmann, J., Klingmuller, U., Timmer, J., Theis, F.J., 2013. High-dimensional Bayesian parameter estimation: Case study for a model of JAK2/STAT5 signaling. *Math Biosci* 246(2), 293-304.
91. Ibarra, R.U., Edwards, J.S., Palsson, B.O., 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 420(6912), 186-189.
92. Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., Ho, P.Y., Kakazu, Y., Sugawara, K., Igarashi, S., Harada, S., Masuda, T., Sugiyama, N., Togashi, T., Hasegawa, M., Takai, Y., Yugi, K., Arakawa, K., Iwata, N., Toya, Y., Nakayama, Y., Nishioka, T., Shimizu, K., Mori, H., Tomita, M., 2007. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* 316, 593-597.
93. Jahan, N., Maeda, K., Matsuoka, Y., Sugimoto, Y., Kurata, H., 2016. Development of an accurate kinetic model for the central carbon metabolism of *Escherichia coli*. *Microb Cell Fact* 15, 112.
94. Jamshidi, N., Palsson, B.O., 2008. Formulating genome-scale kinetic models in the post-genome era. *Mol Syst Biol* 4, 171.
95. Jamshidi, N., Palsson, B.O., 2010. Mass action stoichiometric simulation models: Incorporating kinetics and regulation into stoichiometric models. *Biophys J* 98(2), 175-185.
96. Jankowski, M.D., Henry, C.S., Broadbelt, L.J., Hatzimanikatis, V., 2008. Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys J* 95(3), 1487-1499.

97. Jaqaman, K., Danuser, G., 2006. Linking data to models: data regression. *Nature reviews. Molecular cell biology* 7, 813-819.
98. Joshi, M., Seidel-Morgenstern, A., Kremling, A., 2006. Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metab Eng* 8(5), 447-455.
99. Kacser, H., Burns, J.A., 1973. The control of flux. *Symp. Soc. Exp. Biol.* 27, 65-104.
100. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44(D1), D457-D462.
101. Keleti, T., 1986. 2 Rules of Enzyme-Kinetics for Reversible Michaelis-Menten Mechanisms. *Febs Lett* 208(1), 109-112.
102. Kerkhoven, E.J., Achcar, F., Alibu, V.P., Burchmore, R.J., Gilbert, I.H., Trybilo, M., Driessen, N.N., Gilbert, D., Breitling, R., Bakker, B.M., Barrett, M.P., 2013. Handling uncertainty in dynamic Models: The pentose phosphate pathway in *Trypanosoma brucei*. *Plos Comput Biol* 9(12), e1003371.
103. Khazaei, T., Mcguigan, A., Mahadevan, R., 2012. Ensemble modeling of cancer metabolism. *Frontiers in physiology* 3, 135.
104. Khodayari, A., Maranas, C.D., 2016. A genome-scale *Escherichia coli* kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. *Nat Commun* 7, 13806.
105. Khodayari, A., Zomorodi, A.R., Liao, J.C., Maranas, C.D., 2014. A kinetic model of *Escherichia coli* core metabolism satisfying multiple sets of mutant flux data. *Metab Eng* 25, 50-62.
106. King, Z.A., O'Brien, E.J., Feist, A.M., Palsson, B.O., 2017. Literature mining supports a next-generation modeling approach to predict cellular byproduct secretion. *Metab Eng* 39, 220-227.
107. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. *Science* 220(4598), 671-680.
108. Klotz, I.M., 1946. The application of the law of mass action to binding by proteins - Interactions with calcium. *Arch Biochem* 9(1), 109-117.
109. Koshland, D.E., Némethy, G., Filmer, D., 1966. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry-US* 5, 365-385.
110. Kotte, O., Zaugg, J.B., Heinemann, M., 2010. Bacterial adaptation through distributed sensing of metabolic fluxes. *Mol Syst Biol* 6.
111. Koutroumpas, K., Ballarini, P., Votsi, I., Cournède, P.-H., 2016. Bayesian parameter estimation for the Wnt pathway: an infinite mixture models approach. *Bioinformatics* 32(17), i781-i789.
112. Kresnowati, M.T., van Winden, W.A., Heijnen, J.J., 2005. Determination of elasticities, concentration and flux control coefficients from transient metabolite data using linlog kinetics. *Metab Eng* 7(2), 142-153.
113. Kreutz, C., Raue, A., Kaschek, D., Timmer, J., 2013. Profile likelihood in systems biology. *Febs J* 280(11), 2564-2571.
114. Kuepfer, L., Peter, M., Sauer, U., Stelling, J., 2007. Ensemble modeling for analysis of cell signaling dynamics. *Nat Biotechnol* 25(9), 1001-1006.
115. Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E., 1998. Convergence properties of the Nelder-Mead simplex method in low dimensions. *Siam J Optimiz* 9(1), 112-147.
116. Lee, S.Y., Kim, H.U., 2015. Systems strategies for developing industrial microbial strains. *Nat Biotechnol* 33(10), 1061-1072.
117. Lee, Y., Lafontaine Rivera, J.G., Liao, J.C., 2014. Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways. *Metab Eng* 25, 63-71.



118. Lehmann, E., Romano, J.P., 2005. *Testing Statistical Hypothesis*. Springer-Verlag New York.
119. Leskovac, V., 2003. *Comprehensive Enzyme Kinetics*. Springer Science & Business Media.
120. Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. *Q Appl Math* 2(2), 164-168.
121. Lewis, F., Butler, A., Gilbert, L., 2011. A unified approach to model selection using the likelihood ratio test. *Methods Ecol Evol* 2(2), 155-162.
122. Lewis, N.E., Nagarajan, H., Palsson, B.O., 2012. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol* 10(4), 291-305.
123. Liebermeister, W., Klipp, E., 2005. Biochemical networks with uncertain parameters. *IEE Proceedings - Systems Biology* 152(3), 97.
124. Liebermeister, W., Klipp, E., 2006a. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor Biol Med Model* 3, 41.
125. Liebermeister, W., Klipp, E., 2006b. Bringing metabolic networks to life: integration of kinetic, metabolic, and proteomic data. *Theor Biol Med Model* 3, 42.
126. Liebermeister, W., Uhlenendorf, J., Klipp, E., 2010. Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics* 26, 1528-1534.
127. Liepe, J., Holzhutter, H.G., Bellavista, E., Kloetzel, P.M., Stumpf, M.P.H., Mishto, M., 2015. Quantitative time-resolved analysis reveals intricate, differential regulation of standard- and immuno-proteasomes. *Elife* 4, e07545.
128. Link, H., Christodoulou, D., Sauer, U., 2014. Advancing metabolic models with kinetic information. *Curr Opin Biotech* 29C, 8-14.
129. Link, H., Kochanowski, K., Sauer, U., 2013. Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. *Nat Biotechnol* 31(4), 357-361.
130. Mackay, D.J.C., 1998. Introduction to Monte Carlo methods, in: Jordan, M. (Ed.) *Nato Adv Sci I D-Beh*. Springer, Netherlands, pp. 175-204.
131. Mahadevan, R., Schilling, C.H., 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* 5, 264-276.
132. Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math* 11(2), 431-441.
133. Michaelis, L., Menten, M.L., 1913. Die Kinetik der Invertinwirkung. *Biochem Z* 49, 333-369.
134. Mišković, L., Hatzimanikatis, V., 2010. Production of biofuels and biochemicals: in need of an ORACLE. *Trends in biotechnology* 28, 391-397.
135. Mišković, L., Hatzimanikatis, V., 2011. Modeling of uncertainties in biochemical reactions. *Biotechnol Bioeng* 108(2), 413-423.
136. Moles, C.G., Mendes, P., Banga, J.R., 2003. Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Res* 13(11), 2467-2474.
137. Monod, J., Wyman, J., Changeux, J.P., 1965. On the nature of allosteric transitions: a plausible model. *J Mol Biol* 12, 88-118.
138. Mukherjee, S., Speed, T.P., 2008. Network inference using informative priors. *P Natl Acad Sci USA* 105(38), 14313-14318.
139. Muller, T.G., Faller, D., Timmer, J., Swameye, I., Sandra, O., Klingmuller, U., 2004. Tests for cycling in a signalling pathway. *J R Stat Soc C-Appl* 53, 557-568.

140. Murabito, E., Smallbone, K., Swinton, J., Westerhoff, H.V., Steuer, R., 2011. A probabilistic approach to identify putative drug targets in biochemical networks. *J R Soc Interface* 8(59), 880-895.
141. Murabito, E., Verma, M., Bekker, M., Bellomo, D., Westerhoff, H.V., Teusink, B., Steuer, R., 2014. Monte-Carlo modeling of the central carbon metabolism of *Lactococcus lactis*: Insights into metabolic regulation. *Plos One* 9(9), e106453.
142. Nelder, J.A., Mead, R., 1965. A simplex-method for function minimization. *Comput J* 7(4), 308-313.
143. Neves, A.R., Ventura, R., Mansour, N., Shearman, C., Gasson, M.J., Maycock, C., Ramos, A., Santos, H., 2002. Is the glycolytic flux in *Lactococcus lactis* primarily controlled by the redox charge? Kinetics of NAD(+) and NADH pools determined in vivo by C-13 NMR. *J Biol Chem* 277(31), 28088-28098.
144. Nielsen, J., 1997. Metabolic control analysis of biochemical pathways based on a thermokinetic description of reaction rates. *Biochem J* 138, 133-138.
145. Nielsen, J., Jørgensen, H.S., 1995. Metabolic control analysis of the penicillin biosynthetic pathway in a high-yielding strain of *Penicillium chrysogenum*. *Biotechnol Progr* 11, 299-305.
146. Nielsen, J., Keasling, J.D., 2016. *Engineering Cellular Metabolism*. *Cell* 164(6), 1185-1197.
147. Noor, E., Haraldsdottir, H.S., Milo, R., Fleming, R.M., 2013. Consistent estimation of Gibbs energy using component contributions. *Plos Comput Biol* 9(7), e1003098.
148. Olivier, B.G., Rohwer, J.M., Snoep, J.L., Hofmeyr, J.H.S., 2006. Comparing the regulatory behaviour of two cooperative, reversible enzyme mechanisms. *Iee P Syst Biol* 153(5), 335-337.
149. Olivier, B.G., Snoep, J.L., 2004. Web-based kinetic modelling using JWS Online. *Bioinformatics* 20(13), 2143-2144.
150. Pauling, L., 1935. The oxygen equilibrium of hemoglobin and its structural interpretation. *P Natl Acad Sci USA* 21, 186-191.
151. Peskov, K., Goryanin, I., Demin, O., 2008. Kinetic model of phosphofructokinase-1 from *Escherichia coli*. *Journal of Bioinformatics and Computational Biology* 6(4), 843-867.
152. Peskov, K., Mogilevskaya, E., Demin, O., 2012. Kinetic modelling of central carbon metabolism in *Escherichia coli*. *Febs J* 279(18), 3374-3385.
153. Placzek, S., Schomburg, I., Chang, A., Jeske, L., Ulbrich, M., Tillack, J., Schomburg, D., 2016. BRENDA in 2017: new perspectives and new tools in BRENDA. *Nucleic Acids Res* 45, D380-D388.
154. Poolman, M.G., Fell, D.A., Thomas, S., 2000. Modelling photosynthesis and its control. *Journal of Experimental Botany* 51(suppl\_1), 319-328.
155. Popova, S.V., Sel'kov, E.E., 1975. Generalization of the model by Monod, Wyman and Changeux for the case of reversible monosubstrate reaction. *Febs Lett* 53, 269-273.
156. Popova, S.V., Sel'kov, E.E., 1976. Generalization of the Monod-Wyman-Changeux model for the case of multisubstrate reactions. *Molekuliarnaia Biologiia* 10, 1116-1126.
157. Rao, C.R., 1945. Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society* 37, 81-89.
158. Raphson, J., 1690. *Analysis aequationum universalis*. Apud Abelem Swalle, London.
159. Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmuller, U., Timmer, J., 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25(15), 1923-1929.

160. Rivera, J.G.L., Theisen, M.K., Chen, P.W., Liao, J.C., 2017. Kinetically accessible yield (KAY) for redirection of metabolism to produce exo-metabolites. *Metab Eng* 41, 144-151.
161. Rizk, M.L., Laguna, R., Smith, K.M., Tabita, F.R., Liao, J.C., 2011. Redox homeostasis phenotypes in RubisCO-deficient *Rhodobacter sphaeroides* via ensemble modeling. *Biotechnol Progr* 27(1), 15-22.
162. Rizk, M.L., Liao, J.C., 2009. Ensemble modeling for aromatic production in *Escherichia coli*. *Plos One* 4, e6903.
163. Rohwer, J.M., Hanekom, A.J., Hofmeyr, J.S., 2006. A universal rate equation for systems biology, ESCEC. Rüdeshelm/Rhein, Germany.
164. Rossignol, R., Letellier, T., Malgat, M., Rocher, C., Mazat, J.P., 2000. Tissue variation in the control of oxidative phosphorylation: implication for mitochondrial diseases. *Biochem J* 347, 45-53.
165. Rottenberg, H., 1973. The Thermodynamic Description of Enzyme-Catalyzed Reactions: The Linear Relation between the Reaction Rate and the Affinity. *Biophys J* 13(6), 503-511.
166. Rubinstein, R.Y., Kroese, D.P., 2011. Simulation and the Monte Carlo method, 2nd edition ed. John Wiley & Sons.
167. Saa, P.A., Nielsen, L.K., 2015. A general framework for thermodynamically consistent parameterization and efficient sampling of enzymatic reactions. *Plos Comput Biol* 11(4), e1004195.
168. Saa, P.A., Nielsen, L.K., 2016a. Construction of feasible and accurate kinetic models of metabolism: A Bayesian approach. *Sci Rep-Uk* 6, 29635.
169. Saa, P.A., Nielsen, L.K., 2016b. ll-ACHRB: a scalable algorithm for sampling the feasible solution space of metabolic networks. *Bioinformatics* 32(15), 2330-2337.
170. Saa, P.A., Nielsen, L.K., 2016c. A probabilistic framework for the exploration of enzymatic capabilities based on feasible kinetics and control analysis. *Biochimica et Biophysica Acta-General Subjects* 1860(3), 576-587.
171. Sacher, J., Saa, P., Carcamo, M., Lopez, J., Gelmi, C.A., Perez-Correa, R., 2011. Improved calibration of a solid substrate fermentation model. *Electron J Biotechnol* 14, 5.
172. Sánchez, B.J., Soto, D.C., Jorquera, H., Gelmi, C.A., Pérez-Correa, J.R., 2014. HIPPO: An iterative reparametrization method for identification and calibration of dynamic bioreactor models of complex processes. *Ind Eng Chem Res* 53(48), 18514-18525.
173. Savageau, M.A., 1969. Biochemical Systems Analysis. 1. Some mathematical properties of rate law for component enzymatic reactions. *J Theor Biol* 25(3), 365-369.
174. Savageau, M.A., 1988. Introduction to S-Systems and the underlying power-law formalism. *Math Comput Model* 11, 546-551.
175. Savageau, M.A., Coelho, P.M.B.M., Fasani, R.A., Tolla, D.A., Salvador, A., 2009. Phenotypes and tolerances in the design space of biochemical systems. *P Natl Acad Sci USA* 106(16), 6435-6440.
176. Savoglidis, G., dos Santos, A.X.D., Riezman, I., Angelino, P., Riezman, H., Hatzimanikatis, V., 2016. A method for analysis and design of metabolism using metabolomics data and kinetic models: Application on lipidomics using a novel kinetic model of sphingolipid metabolism. *Metab Eng* 37, 46-62.
177. Schaber, J., Flottmann, M., Li, J., Tiger, C.F., Hohmann, S., Klipp, E., 2011. Automated ensemble modeling with modelMaGe: Analyzing feedback mechanisms in the Sho1 branch of the HOG pathway. *Plos One* 6(3), e14791.
178. Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., Sauer, U., 2012. Multidimensional optimality of microbial metabolism. *Science* 336, 601-604.

179. Schwarz, G., 1978. Estimating the dimension of a model. *Ann Stat* 6(2), 461-464.
180. Smallbone, K., Messiha, H.L., Carroll, K.M., Winder, C.L., Malys, N., Dunn, W.B., Murabito, E., Swainston, N., Dada, J.O., Khan, F., Pir, P., Simeonidis, E., Spasic, I., Wishart, J., Weichart, D., Hayes, N.W., Jameson, D., Broomhead, D.S., Oliver, S.G., Gaskell, S.J., McCarthy, J.E., Paton, N.W., Westerhoff, H.V., Kell, D.B., Mendes, P., 2013. A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes. *Febs Lett* 587(17), 2832-2841.
181. Soh, K.C., Miskovic, L., Hatzimanikatis, V., 2012. From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks. *FEMS yeast research* 12, 129-143.
182. Sorensen, D.C., 1982. Newton method with a model trust region modification. *Siam J Numer Anal* 19(2), 409-426.
183. Srinivasan, S., Cluett, W.R., Mahadevan, R., 2015. Constructing kinetic models of metabolism at genome-scales: A review. *Biotechnol J* 10(9), 1345-1359.
184. Stanford, N.J., Lubitz, T., Smallbone, K., Klipp, E., Mendes, P., Liebermeister, W., 2013. Systematic construction of kinetic models from genome-scale metabolic networks. *Plos One* 8, e79195.
185. Steuer, R., Gross, T., Selbig, J., Blasius, B., 2006. Structural kinetic modeling of metabolic networks. *P Natl Acad Sci USA* 103, 11868-11873.
186. Steuer, R., Nesi, A.N., Fernie, A.R., Gross, T., Blasius, B., Selbig, J., 2007. From structure to dynamics of metabolic pathways: application to the plant mitochondrial TCA cycle. *Bioinformatics* 23(11), 1378-1385.
187. Sunnåker, M., Busetto, A.G., Numminen, E., Corander, J., Foll, M., Dessimoz, C., 2013a. Approximate Bayesian computation. *Plos Comput Biol* 9(1), e1002803.
188. Sunnåker, M., Zamora-Sillero, E., Dechant, R., Ludwig, C., Busetto, A.G., Wagner, A., Stelling, J., 2013b. Automatic generation of predictive dynamic models reveals nuclear phosphorylation as the key Msn2 control mechanism. *Sci Signal* 6, ra41.
189. Sunnåker, M., Zamora-Sillero, E., López García de Lomana, A., Rudroff, F., Sauer, U., Stelling, J., Wagner, A., 2014. Topological augmentation to infer hidden processes in biological systems. *Bioinformatics* 30, 221-227.
190. Tan, Y., Liao, J.C., 2012. Metabolic ensemble modeling for strain engineers. *Biotechnol J* 7, 343-353.
191. Tan, Y., Rivera, J.G.L., Contador, C., Asenjo, J., Liao, J.C., 2011. Reducing the allowable kinetic space by constructing ensemble of dynamic models with the same steady-state flux. *Metab Eng* 13, 60-75.
192. Teusink, B., Passarge, J., Reijenga, C.A., Esgalhado, E., van der Weijden, C.C., Schepper, M., Walsh, M.C., Bakker, B.M., van Dam, K., Westerhoff, H.V., Snoep, J.L., 2000. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *Eur J Biochem* 267(17), 5313-5329.
193. Theisen, M.K., Lafontaine Rivera, J.G., Liao, J.C., 2016. Stability of ensemble models predicts productivity of enzymatic systems. *Plos Comput Biol* 12(3), e1004800.
194. Tipton, K.F., Armstrong, R.N., Bakker, B.M., Bairoch, A., Cornish-Bowden, A., Halling, P.J., Hofmeyr, J.-H., Leyh, T.S., Kettner, C., Raushel, F.M., Rohwer, J., Schomburg, D., Steinbeck, C., 2014. Standards for reporting enzyme data: The STRENDA consortium: What it aims to do and why it should be helpful. *Perspectives in Science* 1(1-6), 131-137.
195. Tolman, R.C., 1925. The principle of microscopic reversibility. *P Natl Acad Sci USA* 11, 436-439.
196. Tolman, R.C., 1938. *The Principles of Statistical Mechanics*. Oxford at the Clarendon Press.

197. Tran, L.M., Rizk, M.L., Liao, J.C., 2008. Ensemble modeling of metabolic networks. *Biophys J* 95, 5606-5617.
198. Turanyi, T., Tomlin, A.S., 2014. Analysis of Kinetic Reaction Mechanisms, illustrated ed. Springer.
199. van der Meer, R., Westerhoff, H.V., van Dam, K., 1980. Linear Relation between Rate and Thermodynamic Force in Enzyme-Catalyzed Reactions. *Biochim Biophys Acta* 591(2), 488-493.
200. van Eunen, K., Bouwman, J., Daran-Lapujade, P., Postmus, J., Canelas, A.B., Mensonides, F.I.C., Orij, R., Tuzun, I., van den Brink, J., Smits, G.J., van Gulik, W.M., Brul, S., Heijnen, J.J., de Winde, J.H., de Mattos, M.J.T., Kettner, C., Nielsen, J., Westerhoff, H.V., Bakker, B.M., 2010. Measuring enzyme activities under standardized in vivo-like conditions for systems biology. *Febs J* 277(3), 749-760.
201. van Eunen, K., Kiewiet, J.A., Westerhoff, H.V., Bakker, B.M., 2012. Testing biochemistry revisited: how in vivo metabolism can be understood from in vitro enzyme kinetics. *Plos Comput Biol* 8(4), e1002483.
202. Vanlier, J., Tiemann, C.A., Hilbers, P.A.J., van Riel, N.A.W., 2013. Parameter uncertainty in biochemical models described by ordinary differential equations. *Math Biosci* 246(2), 305-314.
203. Vasilakou, E., Machado, D., Theorell, A., Rocha, I., Noh, K., Oldiges, M., Wahl, S.A., 2016. Current state and challenges for dynamic metabolic modeling. *Curr Opin Microbiol* 33, 97-104.
204. Vera, J., Balsa-Canto, E., Wellstead, P., Banga, J.R., Wolkenhauer, O., 2007. Power-law models of signal transduction pathways. *Cell Signal* 19(7), 1531-1541.
205. Visser, D., Heijnen, J.J., 2003. Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metab Eng* 5(3), 164-176.
206. Visser, D., Schmid, J.W., Mauch, K., Reuss, M., Heijnen, J.J., 2004. Optimal re-design of primary metabolism in *Escherichia coli* using linlog kinetics. *Metab Eng* 6, 378-390.
207. Visser, D., Van der Heijden, R., Mauch, K., Reuss, M., Heijnen, S., 2000. Tendency modeling: A new approach to obtain simplified kinetic models of metabolism applied to *Saccharomyces cerevisiae*. *Metab Eng* 2(3), 252-275.
208. Voit, E.O., 2000. Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists. Cambridge University Press.
209. Voit, E.O., Martens, H.A., Omholt, S.W., 2015. 150 years of the mass action law. *Plos Comput Biol* 11(1), e1004012.
210. Walter, C., 1974. Validity of using a quasi-steady state approximation for reversible Michaelis-Menten mechanism of enzyme action. *J Theor Biol* 44(1), 1-5.
211. Wang, F.S., Ko, C.L., Voit, E.O., 2007. Kinetic modeling using S-systems and lin-log approaches. *Biochem Eng J* 33(3), 238-247.
212. Wang, L., Birol, I., Hatzimanikatis, V., 2004. Metabolic control analysis under uncertainty: framework development and case studies. *Biophys J* 87, 3750-3763.
213. Wang, L., Hatzimanikatis, V., 2006. Metabolic engineering under uncertainty--II: analysis of yeast metabolism. *Metab Eng* 8(2), 142-159.
214. Wegscheider, R., 1901. Über simultane Gleichgewichte und die Beziehungen zwischen Thermodynamik und Reaktionskinetik homogener Systeme. *Monatshefte für Chemie* 22(8), 849-906.
215. Westerhoff, H.V., Van Dam, J., 1987. Thermodynamics and Control of Biological Free-energy Transduction. Elsevier, New York.
216. Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., Alga, E., Weidemann, A., Sauer-Danzwith, H., Mir, S., Krebs, O., Bittkowski, M., Wetsch, E.,

- Rojas, I., Muller, W., 2012. SABIO-RK-database for biochemical reaction kinetics. *Nucleic Acids Res* 40(D1), D790-D796.
217. Wu, L., Mashego, M.R., Proell, A.M., Vinke, J.L., Ras, C., van Dam, J., van Winden, W.A., van Gulik, W.M., Heijnen, J.J., 2006. In vivo kinetics of primary metabolism in *Saccharomyces cerevisiae* studied through prolonged chemostat cultivation. *Metab Eng* 8(2), 160-171.
218. Wu, L., Wang, W.M., van Winden, W.A., van Gulik, W.M., Heijnen, J.J., 2004. A new framework for the estimation of control parameters in metabolic pathways using lin-log kinetics. *Eur J Biochem* 271(16), 3348-3359.
219. Xu, Y.-F., Amador-Noguez, D., Reaves, M., Feng, X.-J., Rabinowitz, J., 2012. Ultrasensitive regulation of anapleurosis via allosteric activation of PEP carboxylase. *Nat Chem Biol* 8, 562-568.
220. Yang, J., Bruno, W.J., Hlavacek, W.S., Pearson, J.E., 2006. On imposing detailed balance in complex reaction mechanisms. *Biophys J* 91(3), 1136-1141.
221. Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J.D., Osterhout, R.E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H.B., Andrae, S., Yang, T.H., Lee, S.Y., Burk, M.J., Van Dien, S., 2011. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat Chem Biol* 7(7), 445-452.
222. Yizhak, K., Benyamini, T., Liebermeister, W., Ruppin, E., Shlomi, T., 2010. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics* 26, i255-i260.
223. Yonetani, T., Park, S., Tsuneshige, A., Imai, K., Kanaori, K., 2002. Global allosteric model of hemoglobin - Modulation of O<sub>2</sub> affinity, cooperativity, and Bohr effect by heterotropic allosteric effectors. *J Biol Chem* 277(37), 34508-34520.
224. Zamora-Sillero, E., Hafner, M., Ibig, A., Stelling, J., Wagner, A., 2011. Efficient characterization of high-dimensional parameter spaces for systems biology. *Bmc Syst Biol* 5, 142.

## Figure captions

**Fig. 1.** Model-centric workflow for metabolic networks reconstruction and analysis. **A** Network reconstruction starts with the annotation of the genome sequence with the encoded metabolic enzymes. Relationships between gene, proteins and reactions are stored in GPR (Gene-Protein-Reaction) associations, enabling rational representation of the biological information flow. Discrepancies and missing information are resolved (where possible) with the support of data from the literature and comprehensive databases (e.g., BRENDA (Placzek et al., 2016), MetaCyc (Caspi et al., 2016), KEGG (Kanehisa et al., 2016)) in a thorough process of manual curation. **B** The set of curated enzymatic reactions determines the spectrum of metabolic capabilities of the cell. This spectrum is mathematically described by the stoichiometry of the biochemical reactions and defines the topology of the reconstructed network. **C** Integration of diverse ‘omics’ data with the metabolic reconstruction enables construction of a metabolic model amenable for rational interrogation and biological discovery. **D** Structural analysis of the metabolic model is readily achieved using constrained-based modelling methods. Stoichiometric, thermodynamic and kinetic (capacity) constraints defines the space of possible network states which can be readily explored using state-of-the-art optimization and/or sampling methods. This structural analysis is however static, and, it does not quantitatively explain how fluxes are achieved. **E** Inclusion of kinetic descriptions for the all the enzymes in the model enables prediction of metabolic states as well as dynamic interrogation of the system. The resulting kinetic model can reconcile higher amounts of data; however, it requires substantially more information for its construction.

**Fig. 2.** Historical development and key features of popular mathematical formalisms for describing enzyme kinetics. **A** While during the late 40s to the early 70s considerable efforts were devoted to formulating more general mechanistic-based formalisms to explain fundamental kinetic behaviours, since the late 70s efforts have been focused on developing more compact approximate formulations to ease the construction of useful kinetic models. **B** Depending on the degree of detail, kinetics supported and modelling assumptions, different formalisms can be more appropriate for a given task provided a certain amount of data for fitting a multitude of parameters. The validity of the selected formalism depends largely on the intended use of the modeller. Abbreviations: MWC, Monod-Wyman-Changeux; GMA, Generalized Mass Action; KFN, Koshland-Némethy-Filmer; TKM, Thermodynamic-Kinetic Modelling.

**Fig. 3.** Workflows for kinetic model development and statistical approaches for parameter estimation. **A** Depending on the availability of experimental data and kinetic information, kinetic models can be developed in a bottom-up or top-down fashion. The key difference between these approaches lies on whether the model is constructed reaction-by-reaction using *in vitro* and data from different sources, or whether it is built globally from *in vivo* measurements. **B** Parameter estimation of kinetic models can be achieved using either frequentist or Bayesian inference approaches. While both approaches employ the likelihood function to weight how likely a parameter is, the use of this function differs greatly in each approach. Frequentist approaches compute a ‘best’ point estimate of the parameter known as Maximum Likelihood Estimator (MLE) using optimization, and then, use different approaches to quantify the uncertainty in this estimate. In contrast, Bayesian approaches treat parameters as truly random variables defined by probability density functions known as

*priors*. Combination of prior distribution (subjective) with the information contained in the data through the likelihood, yields the parameter's *posterior* distribution which represents the uncertainty in the parameter. Computation of this *posterior* is typically achieved using sampling techniques.

**Fig. 4.** Overview of the Structural Kinetic Modelling (SKM) framework. **A** Starting from the network structure and with minimal experimental and kinetic information, the SKM workflow generates a population of Jacobians that enable exploration of the dynamic properties of a metabolic. **B** Main applications of the SKM framework include the identification of key regulatory (allosteric) checkpoints responsible for an increased stability of the central carbon metabolism of the RBC (Grimbs et al., 2007), as well as identification and determination of metabolite levels required to attain a more robust operation of the TCA cycle in plants; in this case, low relative levels of pyruvate and oxaloacetate (Steuer et al., 2007).

**Fig. 5.** Overview of the MASS framework. **A** The MASS framework is likely the most scalable framework for large-scale kinetic model construction. By using the network topology and thermodynamics combined with available omics data at a reference state, this framework enables appraisal of the kinetic ( $k_{\text{PERC}}$ ) and dynamic (Jacobians) properties of large metabolic systems. **B** Application of the MASS framework for the construction of personalized whole-cell models of RBCs from different individuals revealed kinetic differences and network susceptibilities to metabolic drugs, as well as organization of the network in reaction modules operating at distinct time scales (Bordbar et al., 2015).

**Fig. 6.** Overview of the ORACLE framework. **A** Formulation of the network structure integrated with fluxomic data supported by directionalities based on thermodynamic and metabolomic data, enables definition of a reference state for the ORACLE application. This framework generates a population of control coefficients, Jacobians and elasticities that together fully characterize the dynamic state of the system in the neighbourhood of the reference state. **B** Recent application of the ORACLE to a BDO-overproducing *E. coli* strain showed excellent agreement with reported data for the identification of non-intuitive pathways and reactions controlling BDO production and yield in a large-scale kinetic model (Andreozzi et al., 2016a).

**Fig. 7.** Overview of the Ensemble Modelling (EM) framework. **A** Ensemble Modelling workflow for kinetic model construction and selection. As more data is collected, the ensemble of models (herein represented by different parameterizations of the same model) is pruned and reduced until a single explanatory is left. **B** Applications of the EM paradigm have evolved and involve two new approaches for kinetic analysis with different scopes. The Ensemble Modelling with Optimization (EM-O) approach seeks to fit a large kinetic model to genome-scale networks for the prediction of metabolic states and yields using large-scale omics datasets from mutant strains. On the other hand, the Ensemble Modelling for Robustness Analysis (EM-RA) seeks to determine and define metabolic targets and



conditions that ensure robust metabolic performance when (over)producing valuable metabolites.

**Fig. 8.** Overview of the Approximate Bayesian Computation and General Reaction Assembly and Sampling Platform (ABC-GRASP) framework. **A** General workflow of the ABC-GRASP framework. Of all the reviewed frameworks, this approach requires the highest amount of prior and experimental information for the construction of mechanistic kinetic models. The ABC-GRASP output corresponds to a distribution of *a posteriori* parameters that are both consistent with prior knowledge (mechanistic and thermodynamic information) and experimental observations. **B** ABC-GRASP has been able of identifying missing allosteric interactions, as well as predicting metabolic control and states with high accuracy in complex metabolic models (Saa and Nielsen, 2016a).

**Fig. 9.** Advanced modelling frameworks for kinetic modelling and analysis. The figure depicts the historical development of recent frameworks as well as the general supported capabilities (PRIMAL). The five considered frameworks are coloured coded as follows: ORACLE (green), SKM (purple), MASS (red), EM (orange), and GRASP (blue). Abbreviations: MCA, Metabolic Control Analysis; SKM, Structural Kinetic Modelling; EM, Ensemble Modelling; ORACLE, Optimization and Risk Analysis of Complex Living Entities; EM-RA, Ensemble Modelling for Robustness Analysis; EM-O, Ensemble Modelling with Optimization; GRASP, General Reaction and Assembly Platform; ABC-GRASP, Approximate Bayesian Computation and GRASP; IMCA, Inverse Metabolic Control Analysis.

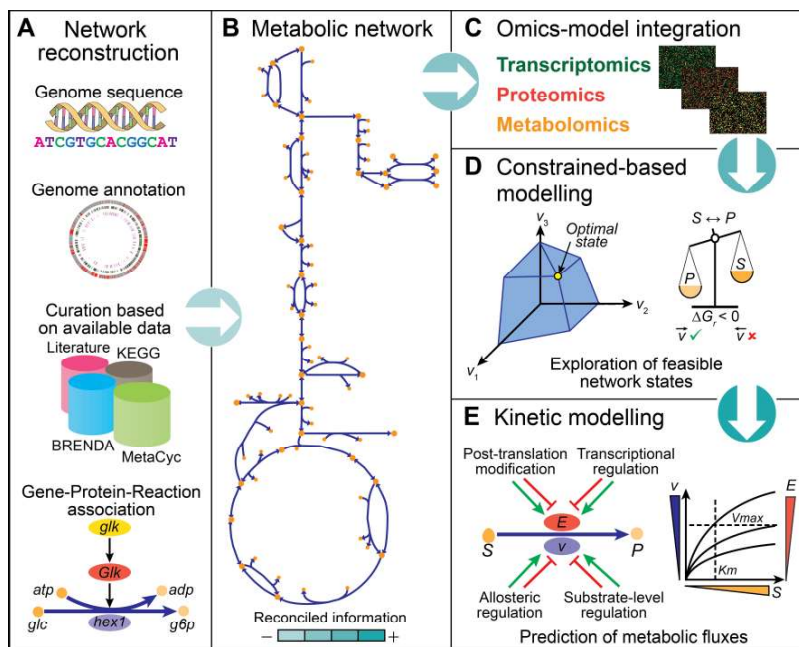


Figure 1

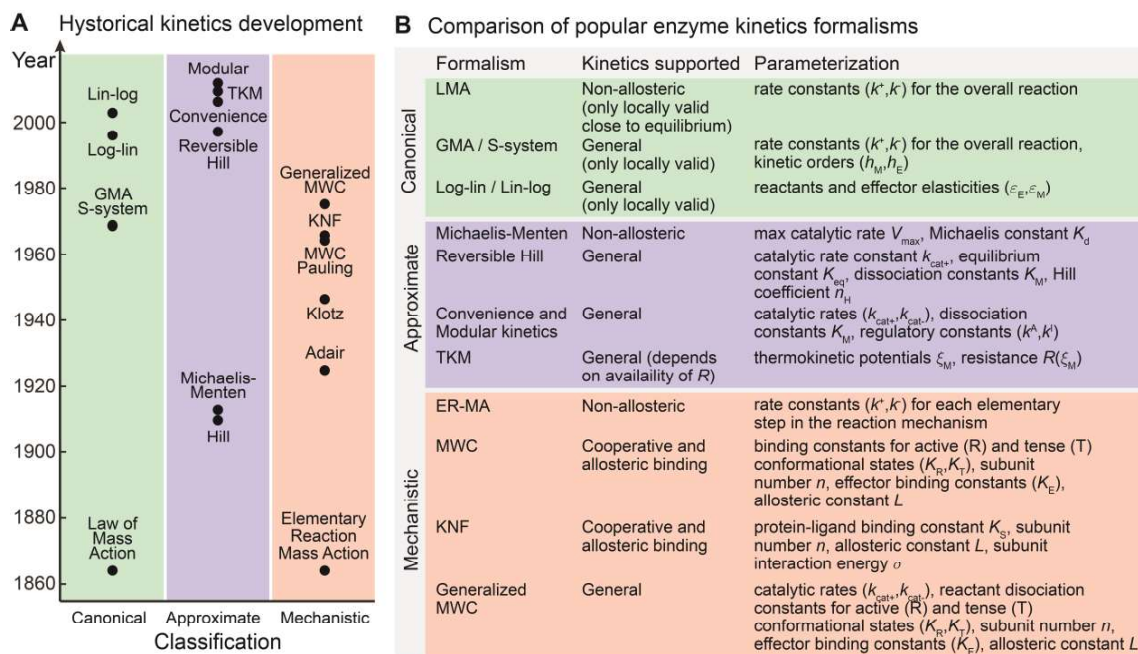


Figure 2

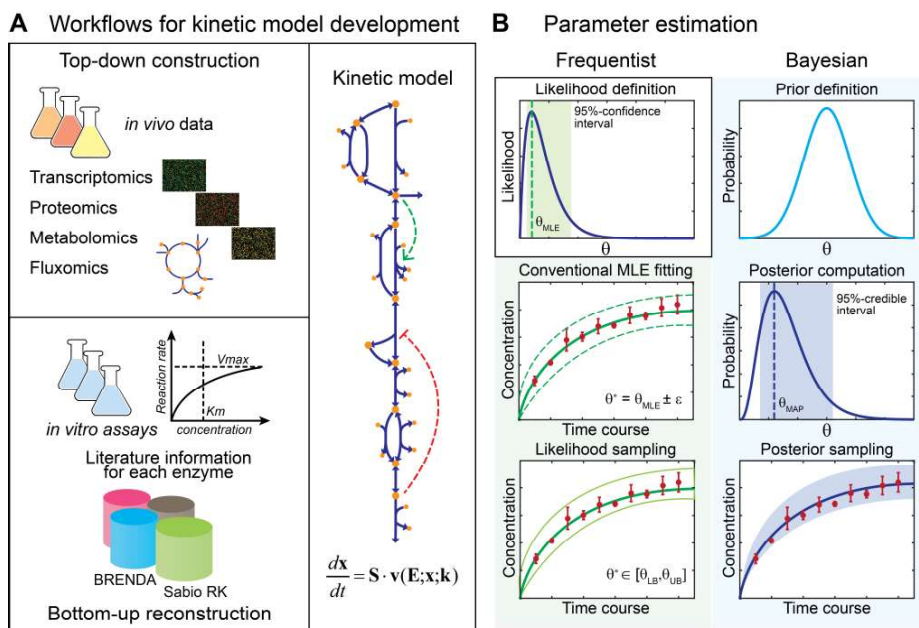


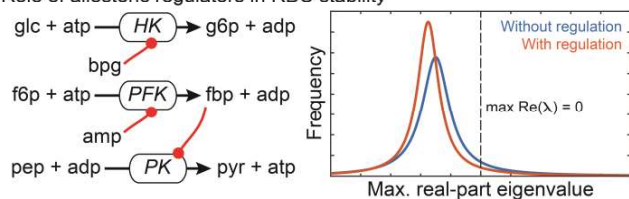
Figure 3

**A SKM workflow**

	Structure	Data	Kinetics
INPUTS	Network structure ( $\mathbf{S}$ )	Fluxomics ( $\mathbf{v}^{\text{ref}}$ )	Parameterize each reaction in terms of saturation degrees $\theta_x^v$
	Thermodynamic constraints ( $v_i \geq 0$ )	Metabolomics ( $\mathbf{x}^{\text{ref}}$ )	
COMPUTE	Build structural matrix $\Lambda$ using reference data ( $\mathbf{v}^{\text{ref}}, \mathbf{x}^{\text{ref}}$ )		
	Sample uniformly degree of saturation and build elasticity matrix $\theta_x^v$		
	Construct Jacobian of the system $\mathbf{J}_x^{\text{ref}} = \Lambda \cdot \theta_x^v$		
OUTPUT	Evaluate Jacobian eigenvalues $\lambda$		
	Ensemble of Jacobians and eigenvalues for stability and robustness analysis		

**B SKM application**

Role of allosteric regulators in RBC stability



Stabilizing metabolite levels in plant TCA

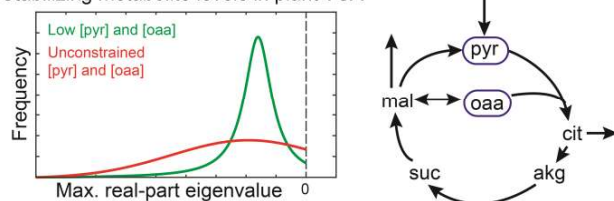


Figure 4

**A** MASS workflow

	Structure	Data	Kinetics
<b>INPUTS</b>	Network structure ( <b>S</b> ) Thermodynamic constraints ( $v_i \geq 0$ )	Thermodynamics ( $K^{eq}$ ) Metabolomics ( $\mathbf{x}^{ref}$ )	LMA for each reaction using ( $k_{PERC}, K^{eq}$ )
<b>COMPUTE</b>	Construct thermodynamic matrix $\Gamma(K^{eq}, \mathbf{x}^{ref})$ Sample uniformly flux distribution at reference state $\mathbf{v}^{ref}$ Compute $\mathbf{k}_{PERC}$ consistent with thermodynamics ( $K^{eq}$ ), metabolite concentrations ( $\mathbf{x}^{ref}$ ) and sampled flux distribution ( $\mathbf{v}^{ref}$ ) Construct Jacobian of the system $\mathbf{J}_x^{ref} = \mathbf{S} \cdot \text{diag}(\mathbf{k}_{PERC}) \cdot \Gamma$		
<b>OUTPUT</b>	Population of Jacobians and $\mathbf{k}_{PERC}$ for the study of kinetic variations and time-scale separation		

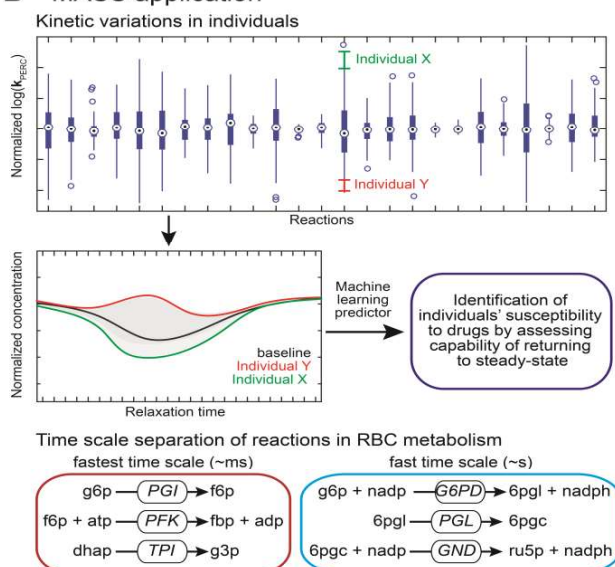
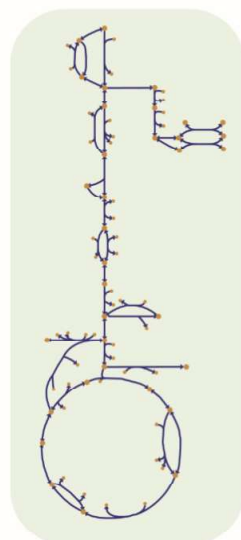
**B** MASS application

Figure 5

## A ORACLE workflow

	Structure	Data	Kinetics
INPUTS	Network structure ( $\mathbf{S}$ )	Fluxomics ( $\mathbf{v}^{\text{ref}}$ )	Various approximate and mechanistic formalisms
	Thermodynamic constraints ( $v_i \geq 0$ )	Metabolomics ( $\mathbf{x}^{\text{ref}}$ ) Thermodynamics ( $\mathbf{K}^{\text{eq}}$ )	
COMPUTE	If not fully determined, sample/define representative ( $\mathbf{v}^{\text{ref}}, \mathbf{x}^{\text{ref}}$ )		
	For enzymes with known kinetics, sample uniformly equilibrium coefficients ( $\gamma^{\text{ref}}$ ) and enzyme state abundances ( $\mathbf{e}^{\text{ref}}$ ) and compute elasticities ( $\epsilon^{\text{ref}}$ ). Otherwise, sample elasticities directly		
OUTPUT	Construct Jacobian of the system $\mathbf{J}_x^{\text{ref}}$ as function of $\epsilon^{\text{ref}}$		
	If $\mathbf{J}_x^{\text{ref}}$ stable, compute control coefficients. Otherwise reject.		
Population of elasticities, Jacobians and control coefficients for dynamic analysis of the system			

## B ORACLE application



Identification of metabolic engineering targets for improved BDO production

### *E. coli* core

175 internal reactions  
106 metabolites  
1 biomass reaction

### Data

Metabolomics and fluxomics for production strain

Distribution of control involved in BDO production  
central glycolysis (*PGI*, *PFK*)

lower TCA (*PPC*)

BDO production route (*4HBDH2*)

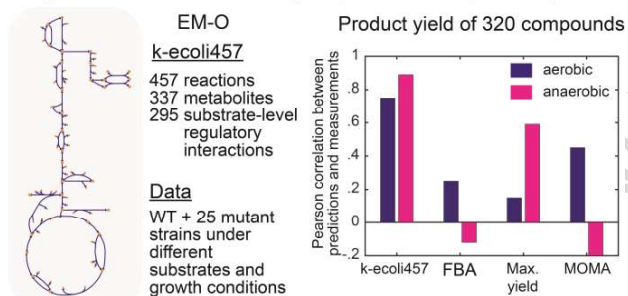
Figure 6

## A EM workflow

	Structure	Data	Kinetics
INPUTS	Network structure ( $\mathbf{S}$ )	Fluxomics ( $\mathbf{v}^{\text{ref}}, \mathbf{v}^{\text{exp}}$ ) Proteomics ( $\mathbf{E}^{\text{exp}}$ )	ER-MA for all the reactions ( $\mathbf{k}$ )
	Thermodynamic constraints ( $v_i \geq 0$ )	Metabolomics ( $\mathbf{x}^{\text{ref}}, \mathbf{x}^{\text{exp}}$ ) Thermodynamics ( $\mathbf{K}^{\text{eq}}$ )	
COMPUTE	If not fully determined, define/sample representative ( $\mathbf{v}^{\text{ref}}, \Delta \mathbf{G}^{\text{ref}}$ )		
	Sample uniformly reversibilities ( $\mathbf{R}^{\text{ref}}$ ) and enzyme state abundances ( $\mathbf{e}^{\text{ref}}$ )		
	Compute normalized rate parameter vector $\tilde{\mathbf{k}}(\mathbf{R}^{\text{ref}}, \mathbf{e}^{\text{ref}}, \mathbf{v}^{\text{ref}}, \Delta \mathbf{G}^{\text{ref}})$		
OUTPUT	Simulate steady-states ( $\mathbf{v}^{\text{sim}}, \mathbf{x}^{\text{sim}}$ ) using $\tilde{\mathbf{k}}$ for different experimental perturbations ( $\mathbf{E}^{\text{exp}}, \mathbf{x}^{\text{exp}}$ ) and prune based on $\mathbf{v}^{\text{exp}}$		
	Kinetic model(s) for prediction of metabolic states and system robustness		

## B EM application

Large-scale kinetic modelling for prediction of product yields



## Robustness of enzymatic systems (EM-RA)

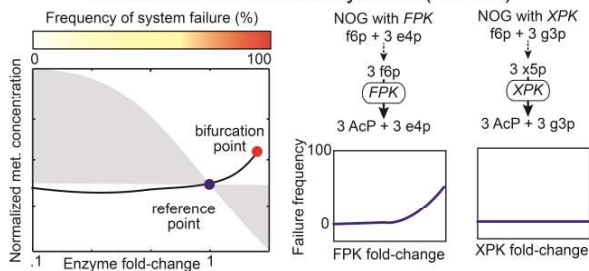


Figure 7

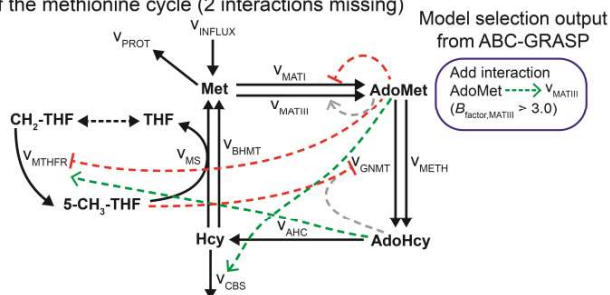


## A ABC-GRASP workflow

	Structure	Data	Kinetics
INPUTS	Network structure ( $\mathbf{S}$ )	Fluxomics ( $\mathbf{v}^{\text{ref}}, \mathbf{v}^{\text{exp}}$ )	ER-MA and MWC
	Thermodynamic constraints ( $v_i \geq 0$ )	Metabolomics ( $\mathbf{x}^{\text{ref}}, \mathbf{x}^{\text{exp}}$ )	
	Prot. structural info.	Thermodynamics ( $\mathbf{K}^{\text{eq}}$ )	
		Proteomics ( $\mathbf{E}^{\text{exp}}$ )	$(L, \tilde{\mathbf{k}}_R, \tilde{\mathbf{k}}_T, \tilde{\mathbf{k}}_E)$
COMPUTE	Sample feasible state ( $\mathbf{v}^{\text{ref}}, \Delta \mathbf{G}^{\text{ref}}$ )		
	Sample uniformly reversibilities ( $\mathbf{R}^{\text{ref}}$ ), enzyme state abundances ( $\mathbf{e}^{\text{ref}}$ ) and allosteric parameters (if necessary)		
	Compute kinetic parameters $\tilde{\mathbf{k}}(\mathbf{R}^{\text{ref}}, \mathbf{e}^{\text{ref}}, \mathbf{v}^{\text{ref}}, \Delta \mathbf{G}^{\text{ref}})$ and assemble reaction rates $\mathbf{v}(L, \tilde{\mathbf{k}}_R, \tilde{\mathbf{k}}_T, \tilde{\mathbf{k}}_E)$		
	Simulate from the model $\mathbf{v}^{\text{sim}} \leftarrow \mathbf{S} \cdot \mathbf{v}(\mathbf{E}^{\text{exp}}, \mathbf{x}^{\text{exp}}) = \mathbf{0}$ , compare against $\mathbf{v}^{\text{exp}}$ , and keep if they are close enough		
OUTPUT	Sample from ABC-posterior composed of a population of feasible kinetic models for prediction of metabolic states and identification of key regulatory interactions		

## B ABC-GRASP application

Identification of key regulations in an incomplete model of the methionine cycle (2 interactions missing)



Prediction of metabolic states and control in the methionine cycle

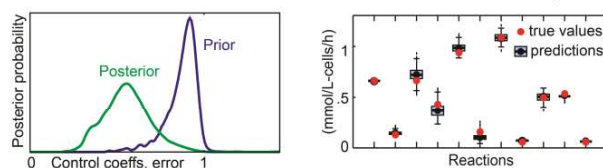


Figure 8

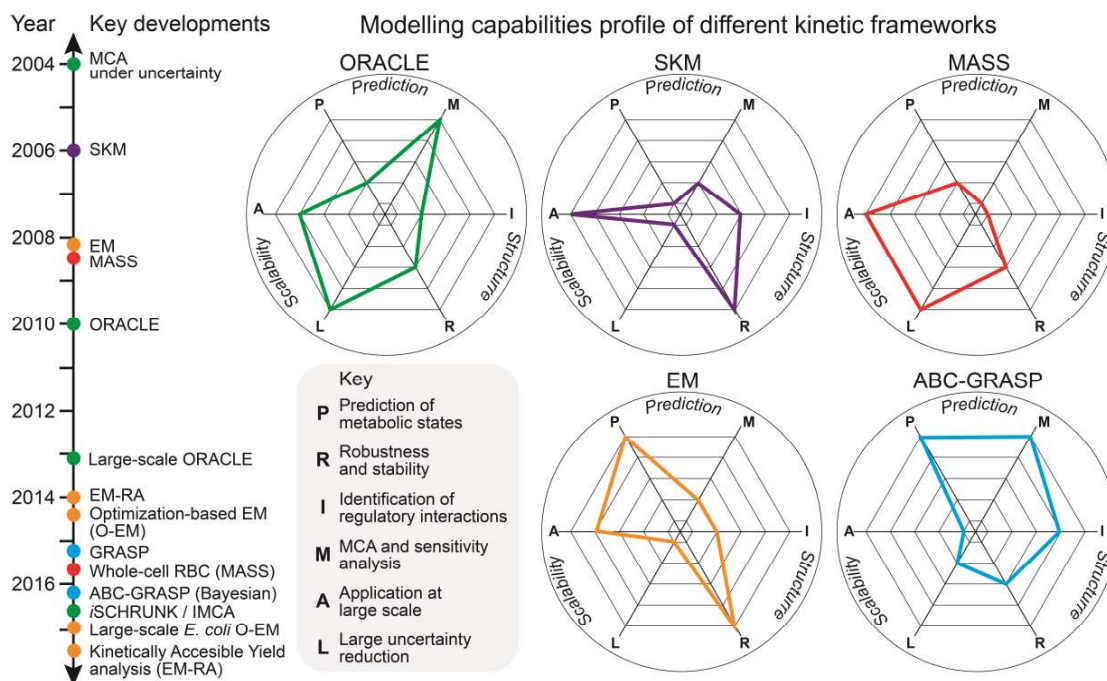


Figure 9

### Highlights

- Kinetic models can accurately capture the complexity of metabolic systems.
- A comprehensive revision of kinetic modelling frameworks is presented.
- Kinetic formalisms, construction and analysis methods are reviewed.
- Opportunities and challenges of large-scale kinetic modelling are discussed.

ACCEPTED MANUSCRIPT