Current state and challenges for dynamic metabolic modeling
Eleni Vasilakou¹, Daniel Machado², Axel Theorell³, Isabel Rocha²,⁴, Katharina Nöh³, Marco Oldiges³,⁵ and S Aljoscha Wahl¹

While the stoichiometry of metabolism is probably the best studied cellular level, the dynamics in metabolism can still not be well described, predicted and, thus, engineered. Unknowns in the metabolic flux behavior arise from kinetic interactions, especially allosteric control mechanisms. While the stoichiometry of enzymes is preserved in vitro, their activity and kinetic behavior differs from the in vivo situation. Next to this challenge, it is infeasible to test the interaction of each enzyme with each intracellular metabolite in vitro exhaustively. As a consequence, the whole interacting metabolome has to be studied in vivo to identify the relevant enzymes properties. In this review we discuss current approaches for in vivo perturbation experiments, that is, stimulus response experiments using different setups and quantitative analytical approaches, including dynamic carbon tracing. Next to reliable and informative data, advanced modeling approaches and computational tools are required to identify kinetic mechanisms and their parameters.

Addresses
¹ Department of Biotechnology, Delft University of Technology, van der Maasweg 9, 2629 HZ Delft, The Netherlands
² Institute for Biotechnology and Bioengineering, University of Minho, Braga, Portugal
³ Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, 52425 Jülich, Germany
⁴ SilicoLife, Braga, Portugal
⁵ Institute of Biotechnology, RWTH Aachen University, Worringer Weg 1, D-52056 Aachen, Germany

Corresponding author: Wahl, S Aljoscha (a.wahl@tudelft.nl)

http://dx.doi.org/10.1016/j.mib.2016.07.008
1369-5274/© 2016 Published by Elsevier Ltd.

Introduction
Modeling of microbial systems has two major aims: (1) to provide a systemic understanding of cellular behavior and (2) to guide the design of microbial host, to optimize, for example, the production of chemicals. Metabolic network analysis has guided the genetic engineering of cells, leading to significantly improved production hosts [1,2]. Especially, steady-state analysis has delivered insights to metabolic fluxes in many different microorganisms [3**]. This includes the discovery of unknown pathways and activities including unusual routes in carbohydrate metabolism in pathogenic hosts [4], amino acid degradation pathways [5] or uncommon shunts in cyanobacteria [6].

However, most current models fail to predict cellular operation [7]. The metabolic flux not only depends on the enzyme concentration, but a variety of cellular functions and mechanisms, like transcription, translation, post-translational modifications and allosteric control. For each level, techniques have been developed to monitor changes in vivo, but the integration of data and its interpretation remain highly challenging. Experimental data sets for modeling are often derived from well-defined and controlled environmental conditions, whereas cells in production processes are faced with sub-optimal conditions, for example, limited oxygen, switching substrate availability or product inhibition. Such environmental factors are one source leading to a limited accuracy of model predictions for dynamic process conditions.

Without doubt, metabolism is the best studied cellular level. For most common hosts like Escherichia coli, Saccharomyces cerevisiae, Bacillus subtilis, Corynebacterium glutamicum and more, the metabolic network stoichiometry is arguably completely described [8,9]. Unknowns in metabolic activity arise from kinetic interactions, especially allosteric control mechanisms. While the stoichiometry of enzymes is preserved in vitro, its activity and behavior differs from the in vivo situation [10]. As a consequence, the whole interacting metabolome has to be studied to identify the enzymatic properties in vivo [11]. Experiments and modeling of enzyme kinetic mechanisms have been pioneered by Reuss et al. [12,13] using stimulus–response experiments (SRE). While crucial new insights have been generated, these approaches only partly succeeded to identify enzyme mechanisms (structural) or kinetic (quantitative) parameters [7].

There are different aspects that lead to non-identifiability (i.e., the inability of the data to sufficiently determine the
model's structure and its quantitative parameters: (1) Carbon effluxes from central carbon metabolism cannot be quantified with sufficient accuracy during the short term of the experiment. (2) Parallel reaction rates and reaction cycles cannot be distinguished. (3) Parameter estimation quality remains low because of high correlations of the model parameter and limited regulatory information content of intracellular concentration measurements [14].

The review focuses on approaches to overcome named challenges, especially approaches that (1) increase the information content by addition of isotopic tracers, like $^{13}$C, (2) combinatorial approaches that allow for inference of different enzyme kinetic mechanisms, (3) novel developments in parameter estimation.

**Coupling experimental observations with modeling approaches**

Identification of *in vivo* kinetic mechanisms is challenging as the system can only be perturbed by extracellular stimuli and/or genetic modifications (Figure 1). The experiments have therefore to be designed with the modeling and the required model resolution and accuracy in mind. In particular, the experimental data must show precise quantitative properties to distinguish between the different hypotheses and deliver sufficient accuracy and coverage for the parameter identification. These criteria, coming from the study aim and the modeling approach, define the measurements and approaches needed, that is, to decide whether additional, quantitative metabolite measurements need to be developed or complementary observables, like carbon labeling [15], are required.

**Experimental approaches**

The aim to reach predictive kinetic models requires sufficient informative experimental data for parameter identification. In this context, the term ‘informative’ means accurate, robust and quantitative data gathered for relevant conditions. Commonly, metabolic flux is observed under steady-state conditions, while dynamic flux estimation is more challenging in several experimental and computational aspects. The aim of this review article is not a complete description of all variants of experimental approaches, but to emphasize how they contribute to the construction of kinetic metabolic models. All these experimental approaches have in common that they must be conducted under well-controlled, reproducible conditions.

To identify kinetic parameters from steady-state experiments, the analysis of a series of different steady-states is required [16–18]. An obvious challenge in such a series of experiments is to keep the cellular properties comparable. To this end, continuous cultivation in chemostat with different dilution rates has been employed.

**Figure 1**

![Diagram showing the coupling of experimental and modeling approaches](current-opinion-in-microbiology-2016-33-97-104.png)

Modeling and the experimental approach are determined by the biological question, that is, the approaches need to be fine-tuned to identify the relevant parameters. The biological system needs to be perturbed by modification of the metabolic network (using genetic modifications) and/or the extracellular conditions (substrate pulse, temperature, among others). The response of the system is monitored using (advanced) analytical methods including $^{13}$C tracing to provide the researcher with quantitative *in vivo* data. The data is then used to calibrate metabolic models which need to be chosen based on the biological question and available data. Modeling and parameter estimation delivers information on the intracellular kinetics including kinetic features of the reaction steps and allows for new biological insights.
To keep the enzymatic properties constant while gathering sufficient information on the kinetic mechanisms, the so-called stimulus–response experiment was proposed by Theobald et al. [12] and became a widely-used, yet very challenging approach. More specifically, the cells are exposed to strong and abrupt perturbations in substrate supply in a short timeframe, that is, much shorter than protein turnover times. Pioneering work has been performed in yeast and bacteria by substrate pulses [12,18–23]. An experimental challenge in SREs is the rapid monitoring of intracellular metabolites, that is, rapid sampling, quenching and analysis of the low concentrated intracellular metabolites by quantitative analytical techniques. The available setups range from fast manual sampling [13] to automated sampling devices coupled to conventional bioreactors [24,25] or plug-flow bioreactor units like the BioScope [26,27].

Besides precise analytical determination of metabolite concentrations, the quantification at intracellular levels is influenced by imperfect quenching procedures that have to be considered [28,29], that is, aspects of metabolite leakage or significant presence of metabolites already in culture supernatant. However, procedures like the differential method with total broth extraction [30] or metabolite balancing including error propagation with all three types of samples (i.e. cell extract, quenching and culture supernatant) [31] have been developed to overcome this. Nevertheless, such methods need to be validated for each novel microbial species.

SREs generate a comprehensive time course of intracellular metabolite concentrations in time, that can be used to identify reaction kinetic parameters [32] and putative regulatory mechanisms [33]. For example, Chassagnole et al. [19] designed a dynamic model accounting for the phosphotransferase system (PTS), glycolysis and the pentose-phosphate pathway in E. coli. Using the data of intracellular metabolite concentrations after the disturbance of steady-state with a glucose pulse, it was shown that the PTS adjusts in sub-seCONDS to the new condition and exhibits a major flux control in E. coli metabolism.

The SRE approach has also been applied to other microorganisms with the aim to highlight the importance of compartmentation for the regulation of glycolysis in yeast [12], to shed light on the valine/leucine pathway kinetics in C. glutamicum [20], or to study the dependency of penicillin-G production on the mechanisms of transport of phenylacetic acid and the product over the cell membrane in Penicillium chrysogenum [18,23].

While SREs with single pulse are highly informative to obtain insights into microbial kinetics and metabolic responses, it is not yet clear if this type of perturbation mimics well the ‘non-laboratory’ biotechnological conditions experienced by cells in large-scale bioreactors, especially when the network has been conditioned to the substrate limited steady-state before the perturbation. There is evidence from literature that the metabolic response of the first substrate pulse differs from a series of perturbations in E. coli [34].

To study such ‘training’ phenomena where metabolic networks are ‘trained’ under periodically changing conditions, a series of scale-down approaches have been applied. Block-wise feeding regimes have been used in scale-down experiments, generating a repetitive dynamic environment. One of the first studies applying block-wise feeding investigated the impact of dynamics on the energy metabolism in yeast strains [35], especially evaluating the yield of biomass and products in comparison to steady-state conditions. Later, this type of feast/famine experiments was used to study metabolism in vivo, with focus on storage metabolism in P. chrysogenum [36] and S. cerevisiae [33].

Suarez-Mendez et al. [33] also showed that this kind of experimental regime not only simulates the cell transition from substrate excess to starvation conditions, but also facilitates the reproducibility of metabolic response measurements. Especially, several (identical) cycles can be sampled allowing for higher time-resolution and replicate measurements compared to the single-pulse experiment.

Continuous dynamic perturbations can also be generated in two-compartment bioreactors that mimic large-scale conditions. This efficient scale-down approach can simulate inhomogeneity inside large-scale bioreactors, by circulating cells between either two stirred-tank reactors (STR-STR) or from one STR to a plug flow reactor (PFR) [37,38].

While all these experimental setups can generate frequent observations and high coverage of metabolic concentration profiles, the relevant information for the identification of kinetic parameters might still be limited, especially for branch-point metabolic nodes [39]. In recent years, these limitations have been overcome with the use of 13C tracer experiments, a powerful method that enables the quantification of intracellular fluxes and provides reliable information on parallel or bidirectional reactions [40,41]. In 13C-based metabolic flux analysis (MFA), 13C-labeled substrates are fed and the labeling enrichment is traced through the metabolic network by either mass spectrometry-based techniques or nuclear magnetic resonance spectroscopy (NMR) [42]. In the traditional isotopic steady-state method only the labeling data of the metabolites is required to inform about the particular flux distribution, whereas under isotopic dynamic conditions, both the labeling and concentrations of metabolites need to be measured [14]. Link et al. [43**]
used $^{13}$C isotopic labeling to identify allosteric metabolite–protein interactions (allosteric mechanisms) that have an impact on the switch between gluconeogenesis to glycolysis in *E. coli*. The cells were cultured on filter material allowing for a very fast exchange of the cultivation medium, for example switching from glucose to pyruvate. The authors measured the metabolic response to such shifts and applied a modeling approach, using a large set of different kinetic hypothesis to identify the most relevant allosteric mechanisms.

**High-throughput opportunities and developments**

The experimental approaches discussed can only generate results for one strain under one perturbation condition. In recent years, high-throughput experimental approaches have been developed to miniaturize the experiments and study more strains and conditions in parallel. A first characterization of metabolic phenotypes can be obtained from the analysis of the extracellular space (metabolic footprinting) [44].

Fuhret *et al.* [45] screened the intracellular metabolome of several *E. coli* mutants, using a microtiter cultivation system coupled to flow-injection mass-spectrometry. This system allows for up to 1400 sample measurements per day. Hollinshead *et al.* [46] have applied metabolic fingerprinting together with $^{13}$C tracing using a series of different tracer substrates, allowing to identify key metabolic flux phenotypes of less common microorganisms.

While the classical milliliter scale cultivation can only be performed in batch mode, novel systems combine automated liquid-handling and optical sensors to control small scale cultivations [47]. For example, the Biolector system can handle 48 parallel cultivation wells [48]. Heux *et al.* [49*] developed a robotic flux profiling system from isotopic fingerprints that enables the generation of 20 flux profiles per day though.

**Analytical techniques**

To obtain as much information as possible about the $^{13}$C patterns of metabolites, advanced analytical techniques are of major importance. Mass spectrometry and tandem mass spectrometry are the most common devices. With the ambition of kinetic modeling in mind, the focus in this review is on quantitative approaches, while untargeted approaches are only briefly touched.

The ambition of quantitative intracellular measurements not only requires highly sensitive instruments to detect the low concentrated metabolites, but also a careful sample preparation. Continuous improvements and validation of protocols for new organisms are crucial to ensure good data quality. Especially, the cellular matrix is challenging, as ionization is sensitive to varying backgrounds. Standard addition or introduction of internal standards is required to correct for matrix effects. In 2005, Mashego *et al.* [50,51] introduced an internal standard for each metabolite, by the addition of U-$^{13}$C labeled cell extract, which is, since then, frequently applied in current quantitative metabolomics. This internal standard can be added at an early stage of the sample processing and enables to correct for losses during the processing [31,52].

For measuring isotopic labeling, precisely the mass isotopomer distribution of intracellular metabolites, mass spectrometry, coupled to gas-chromatography or liquid chromatography, has shown significant advances in recent years. Tandem mass spectrometry has proven to enhance the sensitivity and additionally increase the resolution, with respect to the labeling composition by MS/MS [53]. Therefore, the metabolic flux estimation can be improved, compared to single MS or NMR based techniques [14,54*,55].

Next to these targeted, quantitative approaches, untargeted approaches are necessary for the determination of novel metabolites and pathways. Since they provide broader coverage, untargeted metabolomics data is extremely complex and software tools are indispensable. Examples are the XCMS platform [56] for traditional metabolomics and X$^{13}$CMS [57*], and DynaMet [58**], MathDAMP [59], or MID Max [60] for identification of isotopic labeling enrichments in detected metabolites.

**Modeling approaches**

The parameterized kinetic model should be able to (1) reproduce the experimental observations, (2) allow for the prediction of genetic or environmental perturbation. With predictive models at hand, optimization of the host and the process conditions will deliver more efficient bioprocesses. The advances in technology have enabled the construction of detailed mechanistic models that link metabolite concentrations with enzyme activities. Major limitations of practical applicability are the sheer amount of model parameters lacking identifiability, the size of the network or the accuracy of the kinetic expressions [61]. Here it is important to recognize that for predictive models not necessarily all parameters are required to be well determined [62]. This perception unlocks the use of sampling approaches, where average model predictions over a range of parameters are investigated. Approximative kinetic formats are a suitable alternative, as they are represented by canonical equations and usually contain fewer parameters. Some of the earliest approaches include power–law formats (GMA, S-Systems) and linearized formats (log-lin, lin-log). However, these formats can lead to inconsistent thermodynamic states, a problem that is addressed by recent formats such as modular rate laws and convenience kinetics [61,63].

Although kinetic parameters can often be found in the literature, they are determined using *in vitro* experiments.
that can differ significantly from \textit{in vivo} conditions. Hence, the final step to obtain a working model is to calibrate its parameters using \textit{in vivo} data. The quality of calibration will depend on the model complexity and amount of available data. True estimates of some parameters may not be possible due to structural or practical identifiability problems [64].

Ensemble modeling approach is a powerful approach to tackle these problems [65–69]. It consists on building an ensemble of alternative models that complies with experimental observations. In especial, models with different complexity are generated and compared with respect to their ability to reproduce key features of the data. To overcome data scarcity and inaccuracies (noise), sampling based approaches have become popular to yield surrogates for missing knowledge in parameter values. Sampling of metabolite concentrations, kinetic parameters, enzyme levels and fluxes have been used to identify average properties on a system level, even when the available data is insufficient for actual parameter inference [70*,71,72*,73]. Having fast simulators and smart stochastic sampling schemes at hand, Bayesian approaches could emerge as the ‘swiss army knife’ that unlocks the consistent incorporation of all prior knowledge.

Irrespective of the biological question, modeling includes several common elements. In particular, fast and accurate numerical integrators, robust parameter fitting and advanced statistical tools are required, capable to deal with the non-linear and often ill-posed dynamic problems. Particularly, badly determined or non-identifiable parameters, often non-intuitively correlated pose distinct numerical challenges to model calibration. Parameter uncertainty is addressed by the calculation of confidence intervals, often using the Fisher information matrix, bootstrapping or profile likelihoods. For addressing uncertainty in potentially non-identifiable parameters, profile likelihoods have proven the most reliable [74]. With a dynamic model at hand, analysis for the rate limiting and controlling steps can be performed. One frequently used approach is Metabolic Control Analysis, a sensitivity analysis framework [75–77]. MCA computes the effects of small parameter perturbations resulting in flux control coefficients which describe the effect of a change in the activity of an enzyme on all network fluxes.

Conclusions and outlook
With predictive kinetic models at hand, the design and understanding of microbial cell factories could receive a boost in development. The construction of valid metabolic models is highly challenging and requires further developments, in both experimental and computational approaches:

- Design experimental systems that generate sufficient perturbations, while still being representative for natural and industrial environments and allow for accurate monitoring of the cellular dynamics.
- Develop these platforms for high-throughput analysis, to study a series of external and internal conditions.
- Rigorous dynamical systems theory and systems analysis to elucidate mathematical structures that can be beneficially exploited [78].
- New computational tools for parameter exploration and identification in high-dimensional (>100) spaces.
- Enhancement of model building frameworks (like KiMoSys [79] for kinetic modeling) by various features to assist modelers with the complex tasks of gathering and integrating the available information.
- Establish comprehensive model databases (like BioModels [80] for kinetic modeling). To this end, standards, structured repositories for the experimental omics data and associated protocols (meta-data) are needed [81].

Ultimately, predictive metabolic models could then integrate into whole-cell models, which also include transcription, translation and post-translational mechanisms [82**]. Next to cell-focused models, the integration of the extracellular environment with spatial inhomogeneity due to transport limitation (mixing) are relevant for the development of industrial bioprocesses [83,84].

Acknowledgements
The authors EV, AT, KN, IR, MO, DM and AW are part of the ERA-IB funded consortium DYNAMICS (ERA-IB-IB-14-081, NWO 053.80.724).

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


35. van Kleef BH, Kuenen JG, Heijnen JJ: Heat flux measurements for the fast monitoring of dynamic responses to glucose additions by yeasts that were subjected to different feeding regimes in continuous culture. Biotechnol Prog 1996, 12:510-518.


This paper describes an approach to extract information on relevant metabolic activity outside the typical main scope of central carbon metabolism, that is, detecting labeling enrichments in untargeted metabolomics.


