Proceedings of the 7th IFAC Symposium on
Modelling and Control in Biomedical Systems
Aalborg, Denmark, August 12-14, 2009

Indistinguishability and identifiability of kinetic models for the Mur C reaction in peptidoglycan biosynthesis ⋆

J. Pérez-Velázquez * **J.G. Hattersley** * **M.J. Chappell** *
D. Bearup ** **D. Roper** ** **C. Dowson** ** **T. Bugg** ***
N.D. Evans *

* *School of Engineering, University of Warwick, Coventry, CV4 7AL, UK (e-mail: Neil.Evans@warwick.ac.uk).*
** *School of Biological Sciences, University of Warwick, Coventry, UK*
*** *School of Chemistry, University of Warwick, Coventry, UK*

Abstract: An important question in Systems Biology is the design of experiments to allow discrimination between two (or more) competing pathway models or biological mechanisms. In chemical kinetics a common assumption when studying reactions which release several products is to assume that they are all released in one step. A structural indistinguishability analysis is performed between two different models describing the kinetic mechanism of the Mur C reaction in the cytoplasmic phase of peptidoglycan biosynthesis. One model involves ordered substrate binding and ordered release of the three products; the competing model also assumes ordered substrate binding, but with fast release of the three products. The two versions are shown to be distinguishable both in the full version and under quasi-steady-state assumptions. A structural identifiability analysis is carried out for both models to ensure that the model output uniquely determines the unknown parameters. Similar analyses (indistinguishability and identifiability) are performed using other model simplifications (using conservation equations) and comparisons made with the results of the full model. The analysis forms an integrated step towards the modelling of the full pathway of the cytoplasmic phase of peptidoglycan biosynthesis.

Keywords: Identifiability, Indistinguishability, Experiment design, Mur C, Biomedical systems, Parameter identification

1. INTRODUCTION

In the biological sciences it is becoming increasingly common to collect data in high-throughput experiments on genomic, proteomic, and metabolic scales (Snoep and Westerhoff (2005); Sauer et al. (2007)). These data hold the promise of identifying the mechanisms of interactions that comprise large-scale regulatory biochemical networks. An important step in this identification is the design of experiments to allow discrimination between two (or more) competing pathway models or biological mechanisms. Mathematical models can be used to suggest ways to modify the way data are collected or registered, as they often involve parameters such as reaction rates, which have to be determined in accordance with measured data. Structural indistinguishability provides a formal approach to distinguish between competing model mechanisms (Kholodenko et al. (2005)).

In Systems Biology the mathematical/network models that are generated invariably include large numbers of state variables with numerous model parameters, many of which are unknown, or cannot be directly measured. With such highly complex systems there are often few direct measurements that can be made and limited access for inputs or perturbations. These limitations cause immense problems when investigating the existence of hidden pathways or attempting to estimate unknown parameters and these problems severely hinder model validation. Identifiability analysis provides a formal approach to determine what additional inputs and/or measurements are necessary in order to reduce, or remove, these limitations and permit the derivation of models that can be used for practical purposes with greater confidence (Snoep and Westerhoff (2005); Sauer et al. (2007); Kholodenko et al. (2005)).

Structural indistinguishability for systems models is concerned with determining the uniqueness between possible candidates for the model (or mechanism) structure (Evans et al. (2004)). The analysis is concerned with whether the underlying possibilities for the parametrised mathematical model can be distinguished using the inputs (perturbations or interventions) and observations (or measurements) available for the biological system under investigation. In chemical kinetics it is key to characterise reaction mechanisms, however often several different mechanisms are consistent with the available data, or may even give the same mathematical representation (see Érdi and Tóth (1989); Espenson (1995)), that is, it is not possible to prove that a reaction mechanism is correct. Mechanisms can only

⋆ EPSRC funded project *Indistinguishability analysis for model discrimination in Systems Biology: A Feasibility Study applied to Bacterial Peptidoglycan Biosynthesis* (EP/E057535/1).

be disproved by showing inconsistency with data, or with theoretical requirements for a model. Chemical kinetics is therefore an *ad hoc* field in which to apply indistinguishability analysis. In Schnell et al. (2006) the mechanism distinguishability problem in biochemical kinetics was considered via application of structural indistinguishability to classical models for a single-enzyme, single-substrate reaction. For linear systems the structural indistinguishability analysis is generally exhaustive with all competing mechanisms generated from a given one (see Godfrey and Chapman (1990)). For nonlinear systems approaches are generally only for pairs of candidate models, though in some cases a parametrised family of such candidates can be generated.

Structural identifiability arises in the inverse problem of inferring from the known, or assumed, properties of a biological system, estimates for the corresponding rate constants and other parameters. As such it can be considered as a special case of the structural indistinguishability problem. Structural identifiability analysis considers the uniqueness of the unknown model parameters from the input-output structure corresponding to proposed experiments to collect data for parameter estimation. This is an important, but often overlooked, prerequisite to experiment design, system identification and parameter estimation, since estimates for unidentifiable parameters are effectively meaningless. If parameter estimates are to be used to inform intervention or inhibition strategies, or other critical decisions, then it is essential that the parameters be uniquely identifiable. Numerous techniques for performing a structural identifiability analysis on linear parametric models exist (see Godfrey and DiStefano III (1987); Walter (1982)). In comparison, there are relatively few techniques available for nonlinear systems such as the Taylor series approach (Pohjanpalo (1978)), similarity transformation based approaches (Tunali and Tarn (1987); Vajda et al. (1989); Evans et al. (2002)) and differential algebra techniques (Ljung and Glad (1994); Saccomani et al. (2003)). In a structural identifiability analysis significant computational problems can arise even for relatively simple models. At present there has been relatively little work on techniques for large-scale, highly complex systems, which are typical in Systems Biology.

The purpose of this paper is to explore the possible effectiveness of using structural indistinguishability and identifiability techniques in model discrimination within systems biology networks, using Mur C as a case study; to this end, a structural indistinguishability analysis is performed between two different models describing the kinetic mechanism of the Mur C (Ter Ter) reaction in the cytoplasmic phase of bacterial peptidoglycan biosynthesis (BPB). One model assumes the kinetic mechanism proposed by Emanuele et al. (1997), which involves step-by-step release of the three products; the competing model assumes that the release of the products is simultaneous. It is shown that the two versions are distinguishable both in the full version and under quasi-steady-state assumptions. A structural identifiability analysis is also carried out for both models to ensure that the model output uniquely determines the unknown parameters. Similar analyses (indistinguishability and identifiability) are performed using other model simplifications (using conservation equations) and comparison made with the results of the full model. This forms an integrated step towards the modelling of the full pathway of the cytoplasmic phase of BPB.

2. STRUCTURAL INDISTINGUISHABILITY

Since an indistinguishability analysis can be seen as a generalisation of the identifiability problem, it can be studied by modifying existing approaches for identifiability. Here a modification of the Taylor series approach for identifiability is used.

Consider two uncontrolled systems of the form:

$$\Sigma(\boldsymbol{p}) \begin{cases} \dot{\boldsymbol{x}}(t,\boldsymbol{p}) = \boldsymbol{f}(\boldsymbol{x}(t,\boldsymbol{p}),\boldsymbol{p}), & \boldsymbol{x}(0,\boldsymbol{p}) = \boldsymbol{x}_0(\boldsymbol{p}) \\ \boldsymbol{y}(t,\boldsymbol{p}) = \boldsymbol{h}(\boldsymbol{x}(t,\boldsymbol{p}),\boldsymbol{p}) \end{cases} \quad (1)$$

$$\tilde{\Sigma}(\tilde{\boldsymbol{p}}) \begin{cases} \dot{\tilde{\boldsymbol{x}}}(t,\tilde{\boldsymbol{p}}) = \tilde{\boldsymbol{f}}(\tilde{\boldsymbol{x}}(t,\tilde{\boldsymbol{p}}),\tilde{\boldsymbol{p}}), & \tilde{\boldsymbol{x}}(0,\tilde{\boldsymbol{p}}) = \tilde{\boldsymbol{x}}_0(\tilde{\boldsymbol{p}}) \\ \tilde{\boldsymbol{y}}(t,\tilde{\boldsymbol{p}}) = \tilde{\boldsymbol{h}}(\tilde{\boldsymbol{x}}(t,\tilde{\boldsymbol{p}}),\tilde{\boldsymbol{p}}) \end{cases} \quad (2)$$

where $p \in \Omega \subseteq \mathbb{R}^q$ and $\tilde{\boldsymbol{p}} \in \tilde{\Omega} \subseteq \mathbb{R}^{\tilde{q}}$, both open subsets consisting of the admissible parameter vectors for the two systems; $\boldsymbol{f}$ and $\boldsymbol{h}$ are analytic on $M(\boldsymbol{p})$, an open and connected subset of $\mathbb{R}^n$ such that $\boldsymbol{x}_0(\boldsymbol{p}) \in M(\boldsymbol{p})$, and $\tilde{\boldsymbol{f}}$ and $\tilde{\boldsymbol{h}}$ are analytic on $M(\tilde{\boldsymbol{p}})$, an open and connected subset of $\mathbb{R}^{\tilde{n}}$ such that $\tilde{\boldsymbol{x}}_0 \in M(\tilde{\boldsymbol{p}})$; $\boldsymbol{p}$ and $\tilde{\boldsymbol{p}}$ are constant parameter vectors, Ω and $\tilde{\Omega}$ are the sets of admissible parameter vectors for the two models (1) and (2), respectively, $\boldsymbol{x}(t,\boldsymbol{p})$ and $\tilde{\boldsymbol{x}}(t,\tilde{\boldsymbol{p}})$ are the state variables for each model, which are the different species concentrations whose values are governed by the system of differential equations comprising the model, (1) and (2), respectively. These kinetics, and hence the solutions $\boldsymbol{x}(t,\boldsymbol{p})$ and $\tilde{\boldsymbol{x}}(t,\tilde{\boldsymbol{p}})$, are dependent on the particular parameter vectors $\boldsymbol{p} \in \Omega$ and $\tilde{\boldsymbol{p}} \in \tilde{\Omega}$ used in the models.

The indistinguishability problem arises because, in general, it is not possible to measure all reactants in a given chemical reaction. An experiment that is used to collect measurements of the process gives rise to an output structure for the model, the resulting output, or measurement vectors are $\boldsymbol{y}(t,\boldsymbol{p}) = (y_1(t,\boldsymbol{p}),\ldots,y_r(t,\boldsymbol{p}))^T$ and $\tilde{\boldsymbol{y}}(t,\tilde{\boldsymbol{p}}) = (y_1(t,\tilde{\boldsymbol{p}}),\ldots,y_r(t,\tilde{\boldsymbol{p}}))^T$, respectively, and it is these vectors that are compared with the collected experimental data during subsequent parameter estimation.

Suppose that there exists a $\boldsymbol{p} \in \Omega$ and a $\tilde{\boldsymbol{p}} \in \tilde{\Omega}$ such that $\boldsymbol{y}(t,\boldsymbol{p}) = \tilde{\boldsymbol{y}}(t,\tilde{\boldsymbol{p}})$ for all $t \geq 0$. Then it is not possible to distinguish between the model given by (1) with parameter vector $\boldsymbol{p}$ (i.e., $\Sigma(\boldsymbol{p})$) and the model given by (2) with parameter vector $\tilde{\boldsymbol{p}}$ (i.e., $\tilde{\Sigma}(\tilde{\boldsymbol{p}})$) from their outputs. Therefore, even with perfect data (continuous measurements that are noise-free and error-free) it is not possible to distinguish between the reaction schemes modelled by $\Sigma(\boldsymbol{p})$ and $\tilde{\Sigma}(\tilde{\boldsymbol{p}})$ from the proposed experiment. In this case the models $\Sigma(\boldsymbol{p})$ and $\tilde{\Sigma}(\tilde{\boldsymbol{p}})$ are said to be *indistinguishable*, written as $\Sigma(\boldsymbol{p}) \sim \tilde{\Sigma}(\tilde{\boldsymbol{p}})$.

The Taylor series approach of Pohjanpalo (1978) is used for experiments which can produce time series data. The basis of the Taylor series approach is that the components of the output or observation function $\boldsymbol{y}_i(t,\boldsymbol{p})$ and its successive time derivatives are evaluated at some known time point (usually an initial condition). These derivatives are thus expressed solely in terms of the system parameters $\boldsymbol{p}$ (and $\tilde{\boldsymbol{p}}$ for the indistinguishability case). Since the